



面向甲骨文目标检测的大规模数据集生成技术

杨富勇¹, 李华飙^{1,2}, 孟睿伟¹

(1. 中国国家博物馆, 北京 100006; 2. 馆藏资源活化技术文化和旅游部重点实验室, 北京 100006)

摘要 甲骨文目标检测是甲骨文数字化研究中重要一环, 主要依靠深度学习模型实现对甲骨文图像中位置信息和分类信息的识别。为了避免模型过拟合, 深度学习模型的训练一般需要依赖大规模的数据集, 而在甲骨文目标检测领域, 目前可用于深度学习的大规模数据集较少, 很多研究所用的数据集均依靠专家人工标注和整理, 这使得甲骨文目标检测数据集存在整理成本较高、数据量较小、数据质量不高、类别间均衡性差等问题。提出了动态两阶段 Mosaic 算法及甲骨文大规模数据集生成技术, 解决传统 Mosaic 算法在处理甲骨文图像中存在的拼接图数量有限、图像的多样性和差异性不足、空白背景较大、信息缺失等问题, 并设计了完整的数据集生成流程, 实现了从甲骨文单字符图片到大规模数据集生成的流程化、智能化处理, 从根本上解决了甲骨文目标检测领域的数据困境。通过此研究方法, 生成了标注位置信息和类别信息且规模庞大的甲骨文数据集, 共生成 57 万张甲骨文图像和 57 万份对应的标注文件, 包含甲骨文类别 416 类, 样本数量最少的类别包含了 516 个甲骨文字符, 且数据集规模和各类别样本数量可动态调整以避免类别间样本不均衡。采用 YOLOv8 模型对生成后的大规模数据集进行训练, 在经过 200 批次训练后, 模型精度 (Precision) 达到 96.45%, mAP50 值为 97.75%, mAP50-95 值为 96.96%, 从模型训练曲线看, 训练过程表现出较好的稳定性和高效性, 模型训练结果表明, 研究的数据集生成技术可应用于甲骨文目标检测。

关键词 甲骨文; 深度学习; 目标检测; 数据集; YOLOv8 算法

中图分类号: TP391.4 **DOI:** 10.16152/j.cnki.xdxbr.2025-01-003

Large-scale dataset generation technology for oracle object detection

YANG Fuyong¹, LI Huabiao^{1,2}, MENG Ruiwei¹

(1. National Museum of China, Beijing 100006, China;

2. Key Laboratory of Collection Resources Revitalising Technology, Ministry of Culture and Tourism, Beijing 100006, China)

Abstract Oracle bone inscription object detection is an important part of oracle bone inscription digitization research. This work mainly relies on deep learning models to realize the recognition of position information and classification information in oracle bone inscription images. In order to avoid model overfitting, deep learning models need to rely on large-scale datasets. In the field of oracle bone inscription object detection, there are currently few large-scale data sets available for deep learning. Many research datasets rely on experts to manually annotate and organize, which makes oracle bone inscription object detection datasets face problems such

收稿日期: 2024-10-15

基金项目: 古文字与中华文明传承发展工程规划项目 (G1811); 国家语委科研项目 (WT145-9)。

第一作者: 杨富勇, 男, 博士后, 工程师, 从事文物大数据、人工智能、机器学习、图像处理研究, yfy317@163.com。

通信作者: 李华飙, 男, 正高级工程师, 从事图像处理、文物大数据、智慧博物馆建设研究, lihuabiao@chnmuseum.cn。

as high cost, small data volume, low data quality, and poor balance between categories. This study proposes a dynamic two-stage Mosaic algorithm and oracle bone inscription large-scale dataset generation technology to solve the problems of limited number of mosaic images, insufficient image diversity and difference, large blank background, and missing information in the traditional Mosaic algorithm in processing oracle bone images. A complete dataset generation process is designed to realize the process-based and intelligent processing from oracle bone inscription single character images to dataset generation, which fundamentally solves the data dilemma in the field of oracle bone inscription object detection. Using the method in this study, a large-scale oracle bone inscription dataset with labeled position information and category information was generated. A total of 570 000 oracle bone inscription images and 570 000 corresponding annotation files were generated, including 416 oracle bone inscription categories, and the minimum category contained 516 oracle bone inscription characters. The dataset size and the number of samples in each category can be adjusted dynamically to avoid the problem of sample imbalance between categories. This research uses the YOLOv8 model to train the generated large-scale dataset. After 200 batches of training, the model precision reached 96.45%, the mAP50 value was 97.75%, and the mAP50-95 value was 96.96%. From the model training curve, the training process showed good stability and efficiency. The model training results show that the dataset generation technology in this paper can be applied to oracle bone inscription target detection research.

Keywords oracle; deep learning; object detection; dataset; YOLOv8 algorithm

文化遗产是中华五千年悠久历史的重要载体,是传承文化的特殊资源,也是连接过去现在与未来的重要途径^[1]。甲骨文作为重要的文化遗产之一,被认为是已知最早的中文书写形式和世界上最著名的文字体系之一^[2-5],蕴含了商代人生活的丰富信息,记录了当时文明发展情况。开展甲骨文研究,对于推进汉语词源学研究以及了解商代、古代中国乃至世界文化和历史均具有重要意义^[3]。甲骨文被发现后,专家学者一直试图解读甲骨文内涵^[6],随着人工智能和深度学习在实际中的应用,一些学者开始利用计算机辅助技术来开展甲骨文研究。这些研究工作中,甲骨文目标检测是字符破译的基础,是计算机视觉领域一个重要研究方向^[7]。早期甲骨文字识别方法主要通过分析甲骨文字的拓扑结构^[8],利用支持向量机^[9]和分形几何^[10]等方法进行分类,这些方法不仅精度低,而且泛化能力较差。近年来,深度学习技术在图像分类和识别等方面取得了突出成果,许多学者尝试将深度卷积神经网络应用在拓片甲骨文字的识别工作当中。一些网络模型在图像领域取得了显著效果,如 AlexNet^[11]、VGG^[12]、ResNet^[13]、ResNeSt 网络^[14]、Inception-v4^[15]、YOLO 模型^[7]等都被应用到拓片甲骨文字识别任务中。

开展甲骨文目标检测研究需要依赖大量的甲骨文图像数据,中国各地已出土超过 15 万块甲骨

文和龟甲碎片^[16],整理这些原始材料的图像数据一般需要具备较高专业知识的甲骨文研究专家手工辅助完成^[3]。目前大约有 4 500 种不同甲骨文,已解读甲骨文字只有 2 200 种左右,在出土的甲骨化石中寻找和整理甲骨文字需要大量的劳动力成本^[6]。现有可直接用于目标检测模型训练的甲骨文大规模数据集较少。有些是由现代人手写得到的,如 Li 等整理的数据集^[17],有些只有甲骨文分类信息,而缺乏对甲骨文位置的标注信息,如 Huang 等整理的数据集^[3],有些只有位置信息而缺乏分类信息,如殷墟文渊整理的甲骨文字检测数据集^[18]。此外,甲骨文出现频率严重失衡,也给数据集的整理工作带来了难度,陈婷珠发现,在包含 1 425 个单字的 5.6 万个甲骨字样本中,常用字 366 个,次常用字 500 个,罕见字 559 个^[19]。学者在开展甲骨文目标检测研究时,所用数据集多数是通过人工标注^[15,20-22],不仅数量有限,耗时耗力,而且样本间很难做到均衡。因此,整理和构建可应用于深度学习,从而实现甲骨文目标检测任务的甲骨文大规模数据集,可从根源上解决甲骨文目标检测任务中面临的数据匮乏问题。

本文主要做了以下工作和贡献。①提出了动态两阶段 Mosaic 算法,解决传统 Mosaic 算法在处理甲骨文图像中存在的拼接图数量有限、图像多样性和差异性不足、空白背景较大、信息缺失等问

题。②构建了甲骨文大规模数据集生成流程和技术,通过本文方法,可生成同时具备甲骨文位置信息和类别信息,且数量和类别均可根据需要动态调整,适用于目标检测领域模型训练的甲骨文大规模数据集。③运用本文方法生成了甲骨文大规模数据集,数据量 57 万张,类别 416 类。④采用 YOLOv8 模型,验证了本文生成的大规模数据集的可用性。

1 相关工作

1.1 甲骨文数据集整理情况

目前甲骨文图像数据集主要有以下几类。

1) 无标注信息的高清图像集或拓片图像。高清图片集如中国国家博物馆馆藏甲骨文图像(见图 1),这些图像来源于真实甲骨藏品的拍摄,但由于数据集未经标注,无法直接用于甲骨文深度学习和目标检测研究。拓片图像数据库如台湾历史语言研究所收集整理的甲骨文拓片,共计 4 万余件^[23],均为无标注信息的图像。



图 1 中国国家博物馆馆藏甲骨文图片

Fig. 1 Pictures of oracle bones in the National Museum of China

2) 标注位置信息但未做分类的图像集。典型的例子如殷墟文渊甲骨文检测数据集(见图 2),该数据集来源于“殷契文渊”网站(<https://jgw.aynu.edu.cn/home/index.html>)。该数据集图像来源于《甲骨文字合集》《甲骨文字合集补编》与《怀特氏等收藏甲骨文集》。通过高分辨率扫描仪转为电子书后手工截选并标注。该数据集共有甲骨文拓片 9 823 个,黑白色拓片,图片标注有位置信息(json 格式),但并未标注分类信息。由于该数据集只标注了甲骨文的位置信息,而未对每个甲骨文字符进行分类,因此在目标检测研究时,只能用于识别甲骨文位置,而无法对甲骨文

进行更详细的分类。

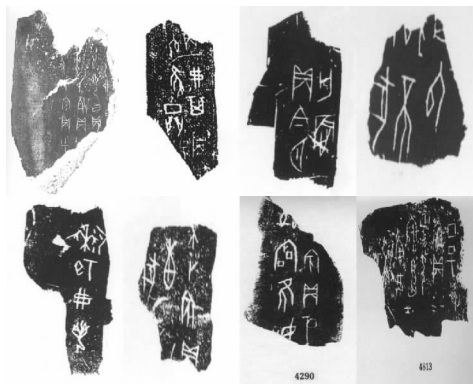


图 2 殷墟文渊甲骨文检测数据集示例

Fig. 2 Yinxu Wen Yuan oracle bone script detection dataset example

3) 已做分类的甲骨文单字符图像集。殷墟文渊 OBC306 数据库^[3](见图 3),是拓片甲骨文字形类别的数据库,其数据来自于以下 8 种甲骨文出版物:《甲骨文字合集》《小屯南地甲骨》《英国所藏甲骨集》《苏德美日所见甲骨集》《甲骨文字合集补编》《怀特士所藏甲骨文集》《东京大学东洋文化研究所藏甲骨文字》《天理大学附属天理参考馆藏甲骨文字》。该数据集共有 309 551 个样本,涉及 306 类,每类对应一个甲骨文文字,单字符图片来源于真实拓片。但该数据集没有现代汉字与甲骨文的对应关系,也没有图片标注的位置信息,由于缺乏位置信息,较难直接用于甲骨文目标检测的深度学习模型中。



图 3 殷墟文渊 OBC306 数据集示例

Fig. 3 Example of Yinxu Wen Yuan OBC306 dataset

4) 通过计算机模拟或手工撰写的“甲骨文”。这些甲骨文图片通过计算机模拟或者是请当代甲骨文专家手写产生,虽然可能与甲骨文字形上相似性较高,但并非来自真实的甲骨藏品或拓片。以殷墟文渊识别训练手写(HWOBC)数据集^[17]为代表(见图 4),该数据集是手写甲骨文字数据集,由 22 位来自不同专业的甲骨学研究者,通过手写甲

骨字搜集软件比照甲骨字标准字形书写并整理而成。数据集共搜集 83 245 张样本图片,按照字库编码分为 3 881 类。

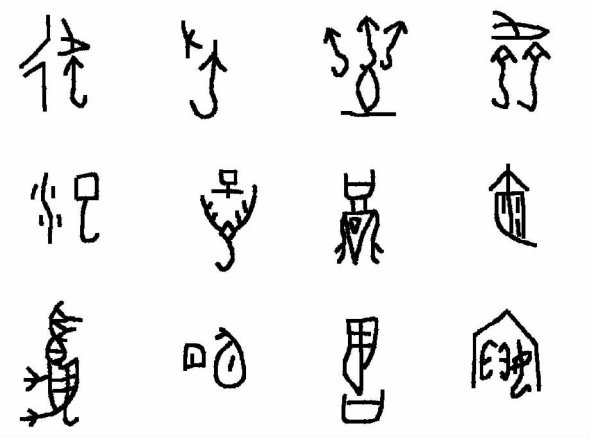


图 4 殷墟文渊识别训练手写 (HWOCB) 数据集示例
Fig. 4 Yin Xu Wen Yuan handwriting recognition training (HWOCB) dataset example

综上,当前很多甲骨文数据集有些缺少标注信息或标注信息不完备,有些并非来自于真实甲骨藏品、拓片或数据量不能满足研究需要。本文提出动态两阶段 Mosaic 算法和大规模数据集生成技术,用单字符的甲骨文图像集生成包含位置、分类等标注信息的甲骨文大规模图像集,以解决当前甲骨文数据困境。本文生成的大规模数据集有 4 个特点。①图像中的甲骨文文字来源于真实的甲骨藏品(甲骨文拓片或甲骨文照片)。②具备标注信息,包括标注框(位置信息)和分类信息。③图像接近真实的甲骨藏品照片,即每张图像上有多个甲骨文文字,且文字在图像上的位置具备一定的随机性。④图像数量足够庞大,且各类别中样本可根据需要动态调整。

1.2 数据增强方法

当前,深度卷积神经网络已成功应用于计算机视觉任务,如图像分类、目标检测和图像分割^[24],其主要目的是使用大数据学习训练样本中的模式和表达。深度学习在视觉上的成功可以归因于具有高能力的模型、计算能力增强以及大规模标记数据的可用性^[25],它是一种数据驱动技术,为了避免模型过拟合,这些模型在很大程度上需要依赖于大数据。2012 年以来,模型表示能力和 GPU 计算能力方面取得了重大进展,随着高性能计算技术的不断发展,文化遗产保护也将迎来新的发展阶段^[26],但数据集对甲骨文等特定领域的模型训练仍是最大制约因素之一。

Sun 等研究表明,深度学习模型的性能随着训练数据量的增加而呈对数增长^[25]。这意味着训练样本的数量越大,得到的模型泛化能力越好,性能也越好。普遍认为,当训练样本相对较小时,在实际应用过程中容易出现过拟合^[27-28],而多样化的训练样本则可以防止模型过拟合^[29-30]。但训练样本的收集和产生有时候需要很高成本,因此低成本和简单的数据增强方法已成为防止模型过拟合较为常见的选择。图像的数据增强主要基于基本图像操作的数据增强,常见图像增强方法有:裁剪(cropping)、翻转变换(flipping)、旋转/反射变换(rotation/reflection)、颜色变换(color space)、几何变换(geometric transformations)、噪声注入(noise injection)、移动(translation)、随机擦除(random erasing)、内核过滤器(kernel filters)、混合图像(mix)、缩放变换(zoom)、特征空间增强(feature space augmentation)、对抗生成(adversarial training)、基于 GAN 的数据增强(GAN-based data augmentation)、神经风格转换(neural style transfer)和元学习数据增强(meta learning data augmentations)等^[24,27,31-34]。这些技术已被证明对特定数据集有效,例如,随机裁剪和水平翻转技术对 CIFAR 数据集的识别任务有帮助^[35]。

然而,随着卷积神经网络的发展,更多新的网络架构被提出,如 VGG-16^[36]、ResNet^[37]、DenseNet^[38]等。这些架构具有更深网络层、更复杂结构和更多参数,因此过拟合风险也在增加^[39]。仅使用传统图片转换方法难以有效抑制过拟合的发生^[40]。因此,更先进的数据增强方法被提出。例如,DeVries 和 Taylor 提出了 Cutout 算法^[41],即随机裁剪图像中一个正方形补丁,并将其替换为“0”像素,通过掩蔽将噪声引入图像中,使卷积神经网络模型对噪声图像更具稳健性,然而,这类方法可能导致图像像素信息丢失^[42]。Zhang 等提出了 Mixup 算法^[43],通过按照一定比例对两幅图像执行凸线性插值^[27],然后将它们融合到新的训练样本中,提高了模型的泛化能力。Yun 等提出了 Cutmix 算法^[44],将另一个输入图像的裁剪区域叠加到补丁区域上,解决了裁剪算法中像素信息丢失的问题。

Bochkovskiy 等在 Cutmix 算法基础上提出了传统 Mosaic 算法^[45],该算法与 Cutmix 算法在理论上有一定相似性。Cutmix 算法是裁剪和拼接数据集中 2 个图像,而 Mosaic 算法是将 4 个图像

拼接成一个新图像。使用多幅图像拼接主要有两个目的,一是增加图像内容复杂度,二是增加图像中目标物体数量^[31]。这些算法都可以提升训练模型检测性能和泛化能力^[42]。此外,由于传统 Mosaic 算法将 4 个图像拼接为一个图像,因此在批量归一化计算激活统计^[46]操作期间,每一层都可以处理 4 个图像数据,从而降低对训练设备的性能要求。

传统 Mosaic 算法主要包括以下 5 个图像处理步骤(见图 5)。①从原始图片集中随机选取 4 幅原始图片作为拼接图像;②对原始图片应用随机变换以获得变换后图像,例如做图片缩放、翻转等处理;③使用随机函数在创建的 Mosaic 图像上获得中心点拼接坐标(x_c, y_c);④围绕中心点拼接坐标,按照左上、右上、左下、右下的顺序,将第二步变换后的图像放入 Mosaic 图像中;⑤裁剪 Mosaic 图像的周边,得到最终的 Mosaic 图像。

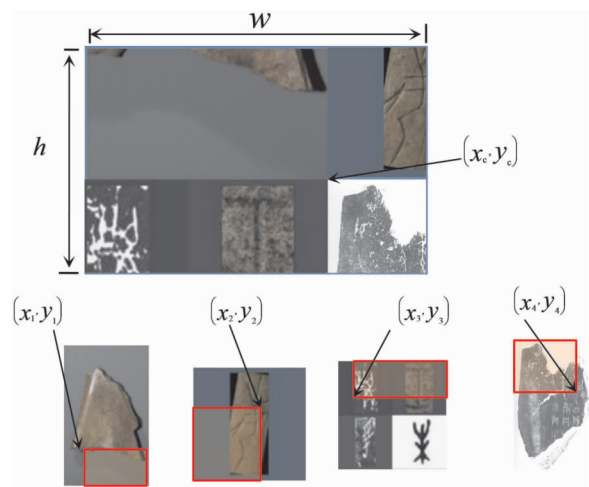


图 5 传统 Mosaic 算法示例

Fig. 5 Traditional Mosaic algorithm

然而,针对甲骨文目标检测领域,传统 Mosaic 算法在处理甲骨文图像中也存在一些不足,主要表现在以下几个方面。

1) 拼接图数量有限。每张 Mosaic 图像只包含 4 张拼接图,如果用于处理单字符甲骨文图片集,每张 Mosaic 图像只包含 4 个甲骨文字符。

2) Mosaic 图像多样性和差异性不足。传统 Mosaic 算法中,确定拼接点坐标后,4 个拼接图的位置也随即固定,这在一定程度上影响了图片的多样性。

3) 存在较大面积空白背景。空白背景是在 Mosaic 图像初始化时创建的,传统 Mosaic 算法处理后的图像由 4 幅拼接图组成,但受到中心点拼

接位置和拼接图像本身尺寸影响当中心点拼接位置靠近 Mosaic 图像边界或拼接图像本身尺寸较小时,拼接图像无法覆盖整个区域,生成的 Mosaic 图像中容易出现大面积的空白背景。空白背景中不存在有价值的目标物体,当 Mosaic 图像中出现大面积空白背景时,将减少 Mosaic 图像中包含的信息量^[42]。

4) 很多超出边界的图像被裁剪掉。由于甲骨文图像的特殊性标注框内的甲骨文应尽量在 Mosaic 图像中较为完整展示,而传统 Mosaic 算法对于拼接后超出 Mosaic 图像的内容将被裁剪掉,这可能导致信息丢失并影响模型结果。

2 动态两阶段 Mosaic 算法

本文提出了动态两阶段 Mosaic 算法,重点针对传统 Mosaic 算法^[45]在处理甲骨文图像中存在的不足做出以下几点改进。

1) 针对“拼接图数量有限”的不足,利用单字符甲骨文图片,通过两次动态 Mosaic 算法处理,增加拼接图像数量,实现了一张 Mosaic 图像包含 16 个单字符甲骨文,从而增加 Mosaic 图像内容复杂性和信息量,使其更适用于甲骨文目标检测任务需要。

2) 针对“Mosaic 图像多样性和差异性不足”,在第一阶段动态 Mosaic 处理时,除沿用传统 Mosaic 算法对拼接点坐标进行随机调整外,本文还加入了随机变量对 4 张拼接图的摆放位置进行了随机动态调整,确保即便拼接点坐标确定后,4 个拼接图的摆放位置仍然是随机的,从而增加了 Mosaic 图像的多样性。

传统 Mosaic 算法在拼接点坐标确定以后,4 张拼图是在拼接点坐标处进行拼接的。而本文在第一阶段动态 Mosaic 处理时,当拼接点坐标确定后,对 4 张拼接图摆放位置不再是在拼接点位置,而是进一步做了动态调整,让 4 个拼接图在 Mosaic 图像中随机摆放,并且让 4 张拼接图均至少有 3/4 的区域在 Mosaic 图像中显示。

如图 6 所示,要确定 Mosaic 图像左上拼接图的摆放位置,只需计算拼接图左边线和上边线分别与 Mosaic 图像的左边线和上边线的距离(dx_{il} 和 dy_{il})。其中 dx_{il} 和 dy_{il} 由随机函数计算得到,从而实现拼接图的动态随机摆放,公式为

$$\begin{cases} dx_{i1} = \text{rand}(-\frac{1}{4} \times nw_{i1}, w \times x_{ci} - \frac{3}{4} \times nw_{i1}) \\ dy_{i1} = \text{rand}(-\frac{1}{4} \times nh_{i1}, h \times y_{ci} - \frac{3}{4} \times nh_{i1}) \end{cases} \quad (1)$$

其中,函数 $\text{rand}(a, b)$ 是随机函数, a 是随机函数的下限(最小值), b 是随机函数的上限(最大值)。 w 和 h 分别为 Mosaic 图像的宽和高,是一个确定值。 (x_{ci}, y_{ci}) 为第 i 张 Mosaic 图像拼接点坐标相对于 Mosaic 图像宽和高的比例,这两个数据由随机函数产生,其中 $x_{ci} = \text{rand}(0.3, 0.7)$, $y_{ci} = \text{rand}(0.3, 0.7)$ 。 nw_{i1} 和 nh_{i1} 分别为第 i 张 Mosaic 图像中第 1 个拼接图(左上)的宽和高,由于拼接图原来的宽和高是确定值,当拼接图缩放比例确定后, nw_{i1} 和 nh_{i1} 的值也随之确定。从式(1)可看出, dx_{i1} 和 dy_{i1} 是一个动态随机数。

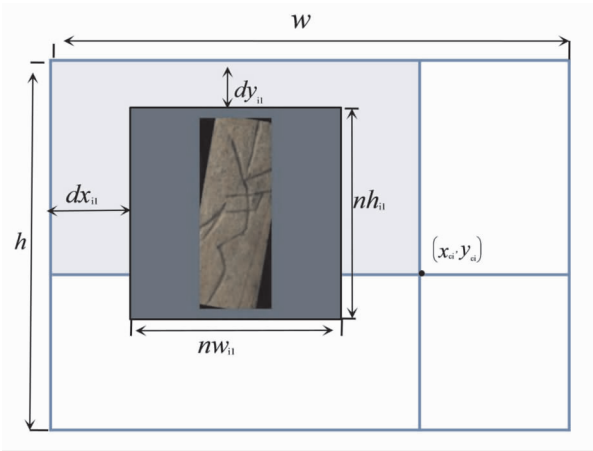


图 6 动态两阶段 Mosaic 算法(第一阶段)

Fig. 6 Dynamic two-stage Mosaic algorithm (first stage)

同理,右上拼接图左边线和上边线分别与 Mosaic 图像的左边线和上边线的距离(dx_{i2} 和 dy_{i2}) 计算公式为

$$\begin{cases} dx_{i2} = \text{rand}(w \times x_{ci} - \frac{1}{4} \times nw_{i2}, w - \frac{3}{4} \times nw_{i2}) \\ dy_{i2} = \text{rand}(-\frac{1}{4} \times nh_{i2}, h \times y_{ci} - \frac{3}{4} \times nh_{i2}) \end{cases}$$

左下拼接图左边线和上边线分别与 Mosaic 图像的左边线和上边线的距离(dx_{i3} 和 dy_{i3}) 计算公式为

$$\begin{cases} dx_{i3} = \text{rand}(-\frac{1}{4} \times nw_{i3}, w \times x_{ci} - \frac{3}{4} \times nw_{i3}) \\ dy_{i3} = \text{rand}(h \times y_{ci} - \frac{1}{4} \times nh_{i3}, h - \frac{3}{4} \times nh_{i3}) \end{cases}$$

右下拼接图左边线和上边线分别与 Mosaic 图像的左边线和上边线的距离(dx_{i4} 和 dy_{i4}) 计算

公式为

$$\begin{cases} dx_{i4} = \text{rand}(w \times x_{ci} - \frac{1}{4} \times nw_{i4}, w - \frac{3}{4} \times nw_{i4}) \\ dy_{i4} = \text{rand}(h \times y_{ci} - \frac{1}{4} \times nh_{i4}, h - \frac{3}{4} \times nh_{i4}) \end{cases}$$

3) 针对“存在较大面积空白背景”的不足,在第二阶段动态 Mosaic 处理时,对 4 张拼接图大小进行了动态调整,通过动态调整拼接图像尺寸来填补 Mosaic 图像中空白背景,使得 4 个拼图能覆盖 Mosaic 图像,避免出现过多空白背景,增强了 Mosaic 图片的信息量。

动态两阶段 Mosaic 算法的第二阶段主要是动态调整拼接图的图片大小,减少 Mosaic 图像中无价值区域的比例,增加图像内容复杂度。经过动态第一阶段 Mosaic 处理后,可得到如图 7 的 Mosaic 图像。这些图像已标注了位置信息和类别信息。但一张 Mosaic 图像只有 4 个标注框,对于甲骨文目标检测任务而言,图像的丰富度不够,因此进一步对传统 Mosaic 算法进行改进,进行动态两阶段 Mosaic 处理。

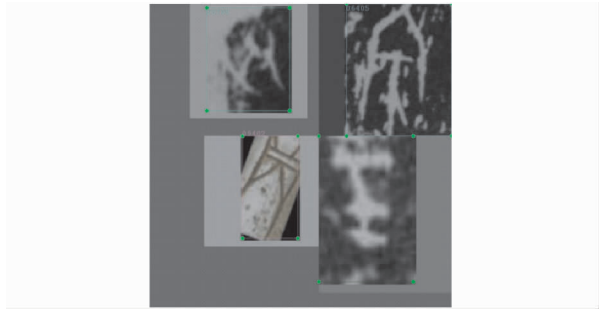


图 7 动态两阶段 Mosaic 图像示例(第一阶段)

Fig. 7 Dynamic two-stage mosaic image example (first stage)

在动态两阶段 Mosaic 算法的第二阶段中,拼接图缩放比例不是固定值,而是计算机通过判断拼接点坐标,按照拼接点坐标位置对拼接图进行动态缩放处理,从而避免过多空白背景产生。

动态两阶段 Mosaic 算法第二阶段中,4 张拼接图的缩放公式如下。

左上拼接图:

$$\begin{cases} nw_{i1} = w \times x_{ci} \\ nh_{i1} = h \times y_{ci} \end{cases}$$

左下拼接图:

$$\begin{cases} nw_{i2} = w \times x_{ci} \\ nh_{i2} = h \times (1 - y_{ci}) \end{cases}$$

右下拼接图:

$$\begin{cases} nw_{i3} = w \times (1 - x_{ci}) \\ nh_{i3} = h \times (1 - y_{ci}) \end{cases}$$

右上拼接图:

$$\begin{cases} nw_{i4} = w \times (1 - x_{ci}) \\ nh_{i4} = h \times y_{ci} \end{cases}$$

其中, nw_{i1} 、 nh_{i1} 分别表示左上拼接图缩放后的宽和高(其他拼图类推)。 w 、 h 表示 Mosaic 图像的宽和高。 (x_{ci}, y_{ci}) 为第 i 张 Mosaic 图像拼接点坐标, 相对于 Mosaic 图像宽和高的比例, 这两个数据由随机函数产生, 其中 $x_{ci} = \text{rand}(0.3, 0.7)$, $y_{ci} = \text{rand}(0.3, 0.7)$ 。

通过该阶段处理后的 Mosaic 图像如图 8 所示, 从图中可以看出, 经过动态两阶段 Mosaic 算法处理后, Mosaic 图像背景的空白区域较小, 图片中的甲骨文排列相对无序, 单字符甲骨文大小也是随机的, 所有图像均包含了位置和分类信息的标注框, 且标注框至少约 3/4 的区域在 Mosaic 拼图中展示。



图 8 动态两阶段 Mosaic 处理后的图像

Fig. 8 Image after dynamic two-stage Mosaic processing

4) 针对“很多超出边界的图像被裁剪掉”的不足, 对传统 Mosaic 算法的参数进行调整, 保留了传统 Mosaic 算法关于图片缩放的操作, 但将拼接图的缩放比例调整为只允许缩小而不做放大(因为放大后, 可能出现拼图超出 Mosaic 图像而被裁掉的情况), 同时设置拼接图随意摆放位置的上下限, 确保 4 个拼图均至少有 3/4 的区域在 Mosaic 图像中显示, 从而确保拼接图大部分区域在 Mosaic 图像中显示。

3 甲骨文大规模数据集生成方法

制作甲骨文目标检测大规模数据集, 传统工作流程是搜集大量甲骨文图像, 用标注软件(如 Labelimg 软件)逐字符进行标注, 除标注位置信息

外, 还需对每个甲骨文进行正确分类, 这需要大量工作量, 并且要求一定的甲骨文专业基础。而本文甲骨文大规模数据集生成方法, 从单字符甲骨文图像入手, 流程化、智能化生成大规模可应用于甲骨文目标检测任务的数据集。工作流程大致分为 4 个阶段: 甲骨文单字符图片集制作; 单字符图片集处理(数据增强和智能标注); 大规模数据集生成(运用动态两阶段 Mosaic 算法); 数据集动态调整(见图 9)。

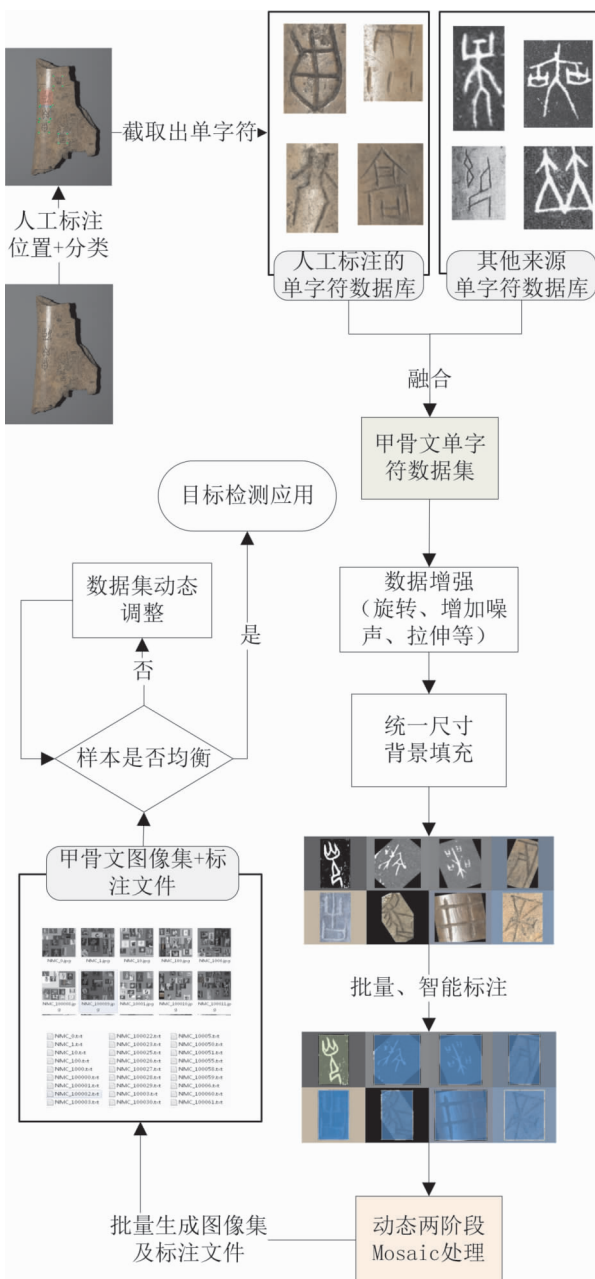


图 9 甲骨文大规模数据集生成流程

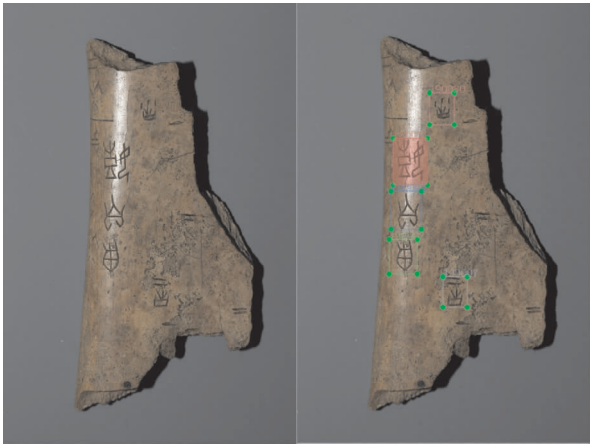
Fig. 9 Oracle large-scale dataset generation process

3.1 甲骨文单字符图片集制作

本文采用的原始图片主要来源于中国国家博物馆的馆藏甲骨文图片和殷墟文渊 OBC306 数据集。

通过人工标注了 118 张中国国家博物馆馆藏甲骨文图片,截取出其中单字符甲骨文图片,每张单字符甲骨文图片对应一个甲骨文字符,共计 1 000 余张甲骨文单字符图片。采用传统数据增强方法将这些图片的数量增强到了近 2 万张。进一步将这近 2 万张图片与殷墟文渊 OBC306 数据集进行融合,最终得到了 416 类、共计 36 万多张甲骨文单字符图片,最后将这些图片统一为同一大小,最终形成甲骨文单字符图片集。具体流程如下。

1)对中国国家博物馆馆藏甲骨文图片进行数据标注。选取中国国家博物馆馆藏 118 张甲骨文图片,人工进行标注和分类,从而得到了 118 张甲骨文图片的位置和分类信息,涉及 164 个字符,标注后的图片如图 10(b)所示。



(a)标注前 (b)标注后
图 10 人工标注前后的甲骨文图片

Fig. 10 Oracle bone inscriptions picture before and after manual annotation

2)对标注后的馆藏甲骨文图片截取出单字符甲骨文。对标注后的馆藏图片,先分割出单个字符,并对字符进行分类存放,最终从 118 张图片中截取出 1 000 余张共计 164 类甲骨文字符。

3)对单字符甲骨文图片进行数据增强。对这些单字符图片进行旋转、左右翻转、上下翻转、增加噪声(椒盐噪声、高斯噪声)、变亮变暗、拉伸、随机透视、颜色翻转等数据增强操作,将图片数量加到了近 2 万张。

4)将数据增强后的单字符图片与殷墟文渊

OBC306 图片集融合。将数据增强后的单字符图片集与殷墟文渊 OBC306 图片集融合,最终形成甲骨文融合图片集,该融合图片集包含了 32 万余张单字符图片,共 416 类。但这些图片像素不一致,图片像素分布如图 11 所示。

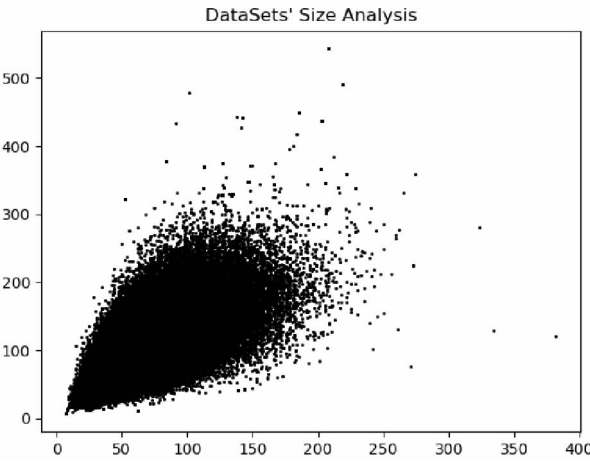


图 11 融合后的单字符甲骨文图片集像素分布情况
Fig. 11 Pixel distribution of the fused single-character oracle bone inscription image set

5)对融合图片集进行数据增强。对第 4 步得到的融合数据集,选择不高于 10% 的比例进行图片翻转,翻转的角度在($-45^{\circ} \sim 45^{\circ}$)之间。数据增强后,新增了 3 万多张单字符图片,此时共有 36 万多张单字符的甲骨文融合图片。

6)将融合图片统一为同一大小,并进行背景填充。为了便于后续的 Mosaic 处理,在不扭曲单字符图片的情况下将图片统一为同一像素(本文统一像素为 224×224)。即制作一个像素为 224×224 的背景图,将所有单字符图片等比例缩放或放大,并放置在背景图中间,为了图片的多样性,背景图空白区域的像素采用随机生成方式,即随机从单字符图片中选择一个点的像素作为背景图填充像素值。最终处理后产生了如图 12 的甲骨文单字符图片。

3.2 甲骨文单字符图片集智能标注

甲骨文单字符图片没有标注信息,如果对 36 万多张图片进行人工标注,工作量和难度极其巨大。本文通过计算机批量处理对单字符图片进行智能标注。

首先,用 Labelimg 软件人工标注一张图片,生成标注文件,批量复制该文件,使标注文件数量与单字符图片数量相同。其次,批量修改每个标注文件的内容。将标注文件的文件名修改成与图片文件名一致,标注文件路径相应修改;将标注文

件中的 size 参数修改为与图片大小一致,为 224×224 ;标注文件的类别名(name)修改为图片的类别名;将标注文件的标注框(bdbbox)参数分别修改为与单字符图片中甲骨文字符所在位置一致。通过复制和智能修改标注文件,最终实现对所有单字符的图片进行智能标注。

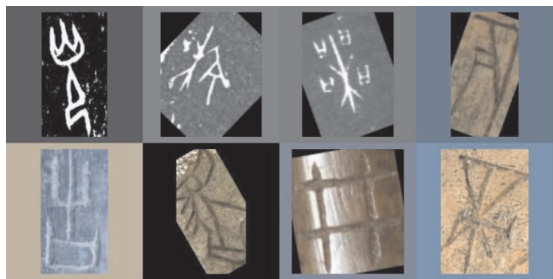


图 12 处理后的甲骨文字单字符图片

Fig. 12 Processed single-character image of oracle bone inscription

其中,标注框的坐标信息公式为

$$\begin{cases} x_{\min_i} = \frac{w}{2} - \frac{\gamma_i}{2} + \alpha \\ x_{\max_i} = \frac{w}{2} - \frac{\gamma_i}{2} - \alpha \\ y_{\min_i} = \frac{h}{2} - \frac{\beta_i}{2} + \alpha \\ y_{\max_i} = \frac{h}{2} - \frac{\beta_i}{2} - \alpha \end{cases}$$

其中, $x_{\min_i}$ 、 $x_{\max_i}$ 、 $y_{\min_i}$ 、 $y_{\max_i}$ 分别表示第 i 张图片标注框的坐标点信息,也就是标注文件中需要修改的标注框坐标信息。 w 、 h 分别表示 Mosaic 图像的宽和高,本文分别为 224 和 224。 γ_i 、 β_i 分别表示单字符缩放后的宽和高。 α 为较小的常数,是为避免边框与单字符图片大小完全重合而设置的,本文设置为 3。标注后的单字符图片如图 13 所示。

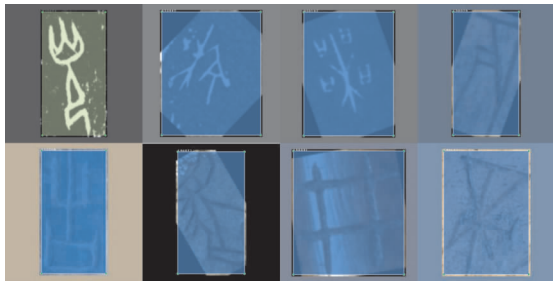


图 13 智能标注后的甲骨文字单字符图片

Fig. 13 Oracle single character image after intelligent annotation

3.3 大规模数据集生成

标注后的甲骨文字单字符图片,具备了位置信息和分类信息,但每张图片均为一个甲骨文字符。运用本文提出的动态两阶段 Mosaic 算法对甲骨文字单字符图片集进行处理,可生成大规模甲骨文目标检测数据集。

首先,进行动态一阶段 Mosaic 处理。设置生成图片的数量为 50 万张,生成图片大小统一为 224×224 。处理后的图片如图 7 所示。

其次,进行动态二阶段 Mosaic 处理。对动态一阶段 Mosaic 处理后的图片,进一步做动态 Mosaic 处理,设置生成图片的数量同样为 50 万张,生成图片的大小统一为 448×448 ,这 50 万张甲骨文图像中共涉及 416 类共计 800 万个甲骨文字符。处理后的图片示例如图 8 所示。

3.4 数据集动态调整

由于甲骨文字单字符图片类别间分布不均衡,加上 Mosaic 算法是从原始图片集中随机抽取图片进行处理的,经过两阶段 Mosaic 处理后的甲骨文图像集可能存在类别间样本分布不均衡的情况,因此需要对数据集进行动态调整,具体方法如下。

将各个类别(416 类)的样本量从小到大排序,选择低于第一四分位数的类别,分别对这些类别的单字符甲骨文进行数据增强,并采用动态两阶段 Mosaic 算法生成甲骨文图像,生成的图片数量设定为第二分位数的值,生成后加入之前的图像集中。同样的方法做第二次动态调整,可根据需要多次做数据集动态调整,从而解决样本分布不均衡问题。

经过两次数据集动态调整后,共生成 57 万张甲骨文图像和 57 万份标注文件,每个标注文件对应一张图像,并标注了每张图像中每个甲骨文字符所在位置和甲骨文对应的分类,数据集共涉及甲骨文字符 416 类,每个类别中样本量最少的一类包含 516 个字符。从样本分布情况看,多数样本分布相对均衡(见图 14),Sun 等^[25]认为,随着预训练数据量的扩大,性能呈对数增长,但发现类别的大小对模型的表现影响不大,Shahinfar 等^[47]认为,在资源有限的情况下,每类 150 ~ 500 张图像足以实现特定模型的合理分类精度,由此看出,本文生成的甲骨文大规模数据集基本符合甲骨文目标检测任务的需要。

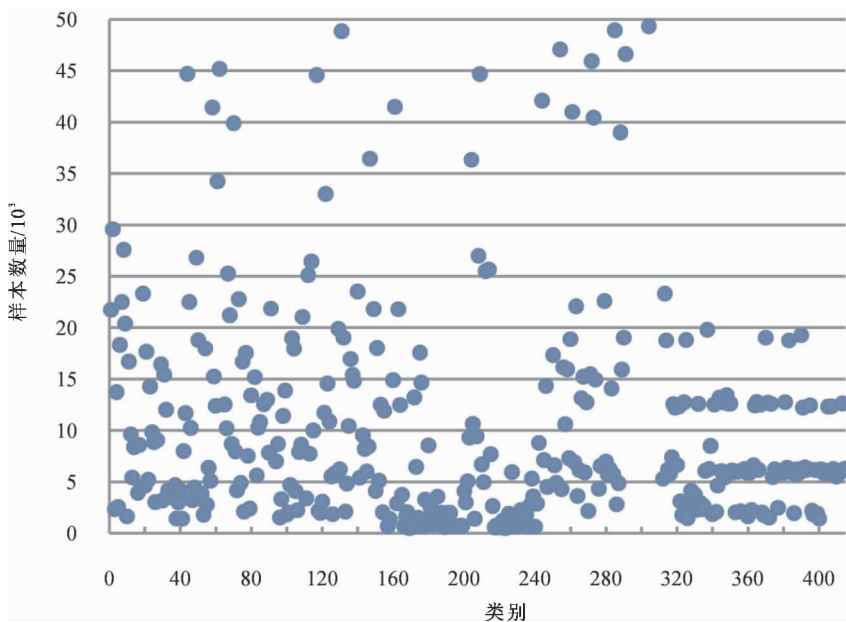


图14 甲骨文数据集中各类别样本量分布情况

Fig. 14 Distribution of sample size of each category in the oracle dataset

4 实验

本文采用 YOLOv8 模型来检验数据集的有效性。YOLO^[48] 系列网络模型在目标检测领域具有较好表现,其中 YOLOv8 网络^[49] 是比较先进的目标检测网络之一,2023 年由 Ultralytics 公司发布。模型训练所用数据集为本文方法生成的 57 万张甲骨文图像和 57 万份标注文件,类别数为 416 类。本文数据集开源地址: <https://github.com/windbrown/Dynamic-Mosaic-Oracle-Bone-Dataset>。

4.1 实验环境及参数配置

本文训练模型进行了 200 次迭代,每批次输入图片数量大小为 16,采用随机梯度 (SGD) 下降优化器,采用现行学习率调度策略,初始学习率为 0.01,最终学习率为 0.000 1。warm epochs 为 3.0,动量参数为 0.8,用于加速训练过程,权重衰减系数为 0.000 5。其中,环境配置为操作系统为 windows10,64 位;CPU 型号为 Intel(R) Xeon(R) Platinum 8362 CPU @ 2.80 GHz;GPU 型号为 RTX 3090(24 GiB);运行内存为 45 GiB;编程语言为 Python3.8.0;采用 PyTorch2.0.2 深度学习框架,GPU 加速库为 CUDA11.8。

4.2 模型评价指标

精确率 (Precision) 计算公式如式(2)所示,精确率的分子为真正例 (true positive, TP),即正确预测为正例的数量,分母则是 TP 加上假正例

(false positive, FP),即错误预测为正例的数量。

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \quad (2)$$

召回率 (Recall) 计算公式如式(3)所示,随着召回率的增加,假负例 (false negative, FN) 的数目降低,表明把正例错判为负例的情况变少,意味着实际正例被成功识别的比例更高。因此,较高的召回率反映了更低的漏检率。

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \quad (3)$$

将每个类别的准确率 Precision 和召回率 Recall 绘制成 P-R 曲线,对该曲线进行积分以后,就得到了平均精度 (average precision, AP),如式(4)所示。

$$\text{AP} = \int_0^1 P(r) dr \quad (4)$$

平均精度均值 (mean average precision, mAP) 计算公式如式(5)所示,是将所有类别的平均精度进行汇总取平均后得到的指标。AP 衡量的是单个类别预测准确性的标准,而 mAP (mean average precision) 综合了所有类别的 AP,从而评估模型整体的预测性能。mAP (式中用 m_{AP} 表示) 值的提升意味着精确率召回率 (PR) 曲线与坐标轴之间的区域面积增加。

$$m_{\text{AP}} = \frac{1}{N} \sum_{i=1}^N \text{AP}_i \quad (5)$$

其中, AP_i 表示第 i 个类别的 AP 值, N 表示类别数量。

4.3 模型训练结果分析

首先,从模型评价指标看,经过 200 批次训练后,模型精度(Precision)达到 96.45%,mAP50 值为 97.75%,mAP50-95 值为 96.96%,表明模型训练效果较好。

其次,分析训练曲线的损失函数和评估指标趋势。损失函数是机器学习中评估模型预测准确性的关键指标,它表明了模型预测结果与真实值之间的偏差。图 15 为 YOLOv8 模型在训练过程中损失和性能指标的变化情况,图中可看出,训练集和验证集上的边界框损失(box_loss)、分类损失(cls_loss)和目标定位损失(dfl_loss)都随着训练的进行而快速降低并趋于稳定,表明模型在学习过程中持续改进,对目标的检测和分类能力在

持续增强。损失函数曲线的平滑下降和收敛表明训练过程是稳定的,没有出现过拟合或欠拟合的情况,这对于模型泛化能力是一个积极的信号。从精确度(Precision)和召回率(Recall)看,二者都随训练进行而逐步提高,说明模型在识别真正的目标方面越来越精准,同时减少了漏检的情况。

最后,分析模型的平均精度(mAP)指标,特别是 mAP50 和 mAP50-95 指标。mAP50 度量了 IoU(交并比)阈值为 50% 时模型的平均精度,而 mAP50-95 则计算了在 IoU 从 50% 到 95% 这一范围内不同阈值的平均精度。从图 16 可以看出,这两个指标稳步上升,表明该模型在各种阈值下都表现出色。

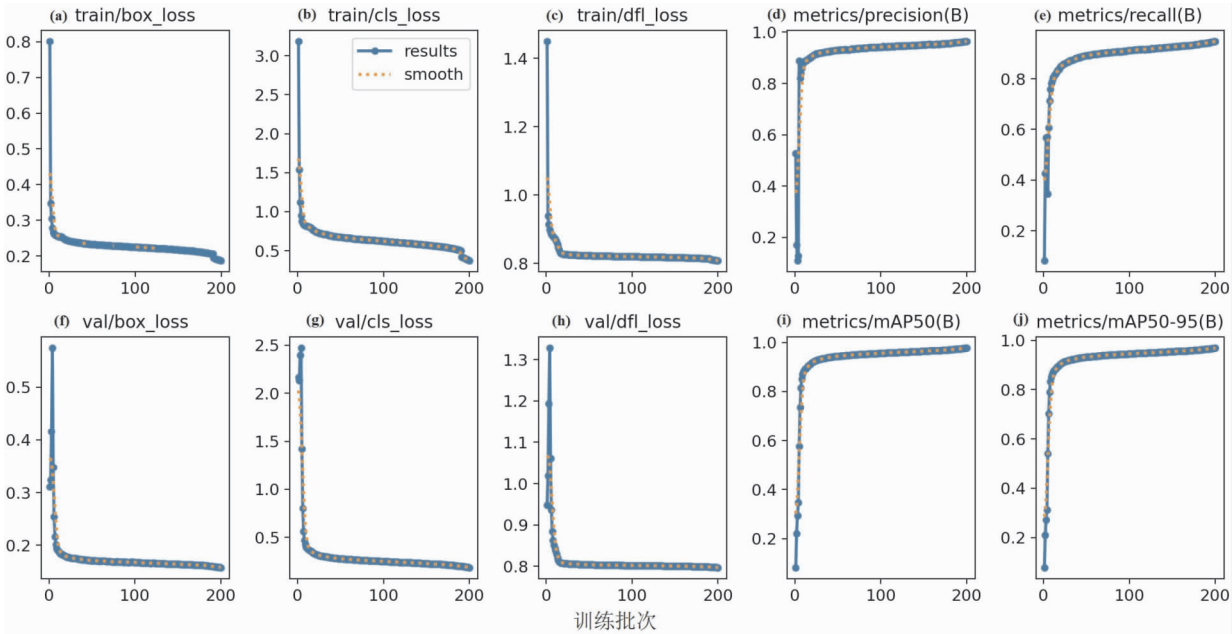


图 15 YOLOv8 模型训练曲线
Fig. 15 YOLOv8 model training curve

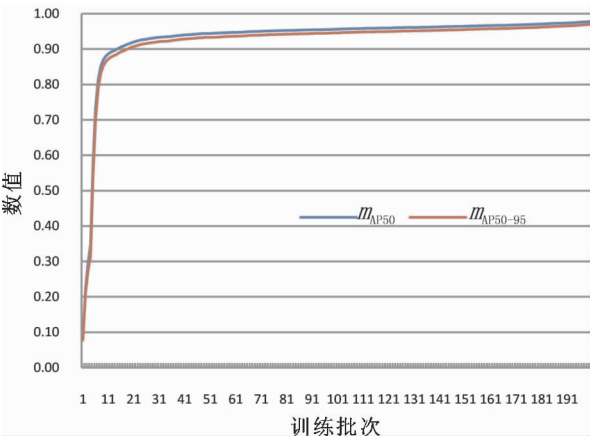


图 16 mAP50 和 mAP50-95 变化趋势图
Fig. 16 mAP50 and mAP50-95 change trend chart

整体而言,利用 YOLOv8 模型对本文生成的数据集进行训练,模型精度较高,模型训练过程表现出较好的稳定性和高效性。损失函数下降趋势以及精确度和召回率的提升,都表明模型的性能在持续提高。此外,mAP 指标的提高也反映了模型在不同 IoU 阈值下都保持了良好的检测精度。可见,运用本文技术和流程生成的甲骨文大规模数据集,在目标检测领域的应用相对成功。

5 结论

针对当前甲骨文目标检测研究任务中用于深

度学习的大规模数据集较为匮乏等问题,本文提出了动态两阶段 Mosaic 算法,综合使用了中国国家博物馆馆藏甲骨文图像和殷墟文渊 OBC306 数据库数据,设计了一套较为科学有效的大规模数据集生成流程,通过本文提出的大规模数据集生成方法和流程,可生成标注了位置和分类信息的大规模甲骨文图像集。从生成的甲骨文大规模图像数据集来看,一是生成的图像信息完整,每张甲骨文图像均标注了深度学习所需的甲骨文字符的位置信息和分类信息,并且一定程度上解决了传统 Mosaic 算法在甲骨文图像处理中存在的问题。二是生成的图像数量理论上没有上限,从而满足了甲骨文深度学习的数据量要求。三是在平衡类别间样本数量时,可以结合数据增强方法进行动态调整,从而解决了样本分布不均衡问题。运用本文的流程和方法,生成了包含 57 万张甲骨文图像的目标检测数据集,为验证技术的可行性,本文采用 YOLOv8 模型,训练本文生成的大规模数据集,从训练结果看,模型的精度较高(Precision 值为 96.45%,mAP50 值为 97.75%,mAP50-95 值为 96.96%)。从模型训练曲线看,训练过程表现出较好稳定性和高效性。模型训练结果表明,本文数据集生成技术可应用于甲骨文目标检测研究。

参考文献

- [1] 耿国华,冯龙,李康,等. 秦陵文物数字化及虚拟复原研究综述[J]. 西北大学学报(自然科学版), 2021, 51(5): 710-721.
GENG G H, FENG L, LI K, et al. A literature review on the digitization and virtual restoration of cultural relics in the Mausoleum of Emperor Qinshihuang [J]. Journal of Northwest University (Natural Science Edition), 2021, 51(5): 710-721.
- [2] FLAD R K. Divination and power: A multiregional view of the development of oracle bone divination in early China [J]. Current Anthropology, 2008, 49(3): 403-437.
- [3] HUANG S P, WANG H B, LIU Y G, et al. OBC306: A large-scale oracle bone character recognition dataset [C]//2019 International Conference on Document Analysis and Recognition (ICDAR). September 20-25, 2019. Sydney, Australia. IEEE, 2019: 681-688.
- [4] CHINASAGE. Early Chinese writing on oracle bones [R/OL]. (2021-02-10) [2024-07-07]. <https://www.chinasage.info/oracle-bones.htm>.
- [5] ZHANG C S, ZONG R X, CAO S, et al. AI-powered oracle bone inscriptions recognition and fragments re-joining[C]//Proceedings of the Twenty-Ninth International Joint Conference on Artificial Intelligence. January 7-15, 2021. Yokohama, Japan: ACM, 2021: 5309-5311.
- [6] GAO F, ZHANG J P, LIU Y G, et al. Image translation for oracle bone character interpretation[J]. Symmetry, 2022, 14(4): 743.
- [7] ZHEN Q Q, WU L, LIU G Y. An oracle bone inscriptions detection algorithm based on improved YOLOv8 [J]. Algorithms, 2024, 17(5): 174.
- [8] 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.
GU S T. Identification of oracle-bone script fonts based on topological registration[J]. Computer & Digital Engineering, 2016, 44(10): 2001-2006.
- [9] 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.
LIU Y G, LIU G Y. Oracle bone inscription recognition based on SVM[J]. Journal of Anyang Normal University, 2017(2): 54-56.
- [10] 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.
GU S T. Identification of oracle-bone script fonts based on fractal geometry[J]. Journal of Chinese Information Processing, 2018, 32(10): 138-142.
- [11] 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: 郑州大学, 2020.
- [12] 王琦琦. 基于深度卷积神经网络的甲骨文精确识别[D]. 南昌: 江西科技师范大学, 2021.
- [13] 高旭. 基于卷积神经网络的甲骨文识别研究与应用[D]. 长春: 吉林大学, 2021.
- [14] 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.
MAO Y F, BI X J. Rubbing oracle bone character recognition based on improved ResNeSt network [J]. CAAI Transactions on Intelligent Systems, 2023, 18(3): 450-458.
- [15] 王浩彬. 基于深度学习的甲骨文检测与识别研究[D]. 广州: 华南理工大学, 2019.
- [16] CHEUNG C. The Chinese history that is written in bone [R/OL]. (2018-01-23) [2024-07-10]. <https://www.sapiens.org/archaeology/chinese-oracle-bones-history/>.
- [17] LI B, DAI Q W, GAO F, et al. HWOBC-a handwriting oracle bone character recognition database [J].

- Journal of Physics: Conference Series, 2020, 1651 (1): 012050.
- [18] 殷契文渊. 甲骨文字检测数据集[DB/OL]. (2020-09-23) [2023-07-14]. <http://jgw.aynu.edu.cn/home/down/detail/index.html?sysid=3>.
- [19] 陈婷珠. 殷商甲骨文字形系统再研究[D]. 上海: 华东师范大学, 2007.
- [20] 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.
LIU F, LI H B, MA J, et al. Automatic detection and recognition of oracle rubbings based on Mask R-CNN [J]. Data Analysis and Knowledge Discovery, 2021, 5(12): 88-97.
- [21] GUO J, WANG C H, 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.
- [22] IZUMI T, MENG L. A combined recognition system for oracle bone inscriptions[J]. International Journal of Advanced Mechatronic Systems, 2017, 7(4): 235.
- [23] 台湾历史语言研究所. 甲骨文數位典藏資料庫[DB/OL]. (2020-07-22) [2023-08-22]. <https://rub.ihp.sinica.edu.tw/~oracle/main4.htm>.
- [24] SHORTEN C, KHOSHGOFTAAR T M. A survey on image data augmentation for deep learning[J]. Journal of Big Data, 2019, 6(1): 60.
- [25] SUN C, SHRIVASTAVA A, SINGH S, et al. Revisiting unreasonable effectiveness of data in deep learning era[C]//2017 IEEE International Conference on Computer Vision (ICCV). October 22-29, 2017. Venice, Italy: IEEE, 2017: 843-852.
- [26] 孟宪佳, 傅利平, 刘栋, 等. 高性能计算发展现状及其在文化遗产保护中的应用展望[J]. 西北大学学报(自然科学版), 2021, 51(5): 807-815.
MENG X J, FU L P, LIU D, et al. Development status of high performance computing and its application prospect in cultural heritage protection[J]. Journal of Northwest University (Natural Science Edition), 2021, 51(5): 80.
- [27] TAKAHASHI R, MATSUBARA T, UEHARA K. Data augmentation using random image cropping and patching for deep CNNs[J]. IEEE Transactions on Circuits and Systems for Video Technology, 2020, 30(9): 2917-2931.
- [28] ZHANG C Y, BENGIO S, HARDT M, et al. Understanding deep learning (still) requires rethinking generalization[J]. Communications of the ACM, 2021, 64(3): 107-115.
- [29] SCHMIDT L, SANTURKAR S, TSIPRAS D, et al. Adversarially robust generalization requires more data [C]//Proceedings of the 32nd International Conference on Neural Information Processing Systems. December 3-8, 2018. Montréal, Canada: ACM, 2018: 5019-5031.
- [30] HESTNESS J, NARANG S R, ARDALANI N, et al. Deep learning scaling is predictable, empirically[EB/OL] (2017-12-01) [2024-08-08]. <http://arxiv.org/abs/1712.00409>.
- [31] SUMMERS C, DINNEEN M J. Improved mixed-example data augmentation[C]//2019 IEEE winter conference on applications of computer vision (WACV). January 7-11, 2019. Hawaii, United States: IEEE, 2019: 1262-1270.
- [32] ZHONG Z, ZHENG L, KANG G L, et al. Random erasing data augmentation[J]. Proceedings of the AAAI Conference on Artificial Intelligence, 2020, 34(7): 13001-13008.
- [33] NIU J S, CHEN Y F, YU X H, et al. Data augmentation on defect detection of sanitary ceramics[C]//IECON 2020 The 46th Annual Conference of the IEEE Industrial Electronics Society. October 18-21, 2020. Singapore, Singapore: IEEE, 2020: 5317-5322.
- [34] INOUE H. Data augmentation by pairing samples for images classification[EB/OL]. (2018-04-11) [2024-08-08]. <http://arxiv.org/abs/1801.02929>.
- [35] KRIZHEVSKY A, SUTSKEVER I, HINTON G E. ImageNet classification with deep convolutional neural networks[J]. Communications of the ACM, 2017, 60(6): 84-90.
- [36] SIMONYAN K, ZISSERMAN A. Very deep convolutional networks for large-scale image recognition[EB/OL]. (2015-04-10) [2024-08-08]. <http://arxiv.org/abs/1409.1556>.
- [37] HE K M, ZHANG X Y, REN S Q, et al. Deep residual learning for image recognition[EB/OL]. (2015-12-10) [2024-08-08]. <http://arxiv.org/abs/1512.03385>.
- [38] HUANG G, LIU Z, VAN DER MAATEN L, et al. Densely connected convolutional networks[C]//2017. IEEE Conference on Computer Vision and Pattern Recognition (CVPR). July 21-26, 2017. Honolulu, HI, USA: IEEE, 2017: 2261-2269.
- [39] XIE S N, GIRSHICK R, DOLLÁR P, et al. Aggregated residual transformations for deep neural networks

[C] //2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). July 21-26, 2017. Honolulu, HI, USA; IEEE, 2017: 5987-5995.

[40] MORENO-BAREA F J, STRAZZERA F, JEREZ J M, et al. Forward noise adjustment scheme for data augmentation[C] //2018 IEEE symposium series on computational intelligence (SSCI). November 18-21, 2018. Bangalore, India; IEEE, 2018: 728-734.

[41] DEVRIES T, TAYLOR G W, ASSIRI Y. Improved regularization of convolutional neural networks with cutout[EB/OL]. (2017-11-29) [2024-08-08]. <http://arxiv.org/abs/1708.04552>.

[42] LI Y H, CHENG R, ZHANG C Y, et al. Dynamic Mosaic algorithm for data augmentation[J]. Mathematical Biosciences and Engineering, 2023, 20 (4): 7193-7216.

[43] ZHANG H Y, CISSE M, DAUPHIN Y N, et al. Mix-up: Beyond empirical risk minimization [EB/OL]. (2017-11-29) [2024-08-08]. <http://arxiv.org/abs/1708.04552>, 2017.

[44] YUN S, HAN D, OLF S J, et al. CutMix: Regularization strategy to train strong classifiers with localizable features[EB/OL]. (2019-08-07) [2024-08-08]. <http://arxiv.org/abs/1905.04899>.

[45] BOCHKOVSKIY A, WANG C Y, LIAO H M. YOLOv4: Optimal speed and accuracy of object detection [EB/OL]. (2020-04-23) [2024-08-08]. <http://arxiv.org/abs/2004.10934>.

[46] IOFFE S, SZEGEDY C, PARANHOS L, et al. Batch normalization: Accelerating deep network training by reducing internal covariate shift [EB/OL]. (2015-03-02) [2024-08-08]. <http://arxiv.org/abs/1502.03167>.

[47] SHAHINFAR S, MEEK P, FAIZON G. "How many images do I need?" Understanding how sample size per class affects deep learning model performance metrics for balanced designs in autonomous wildlife monitoring [J]. Ecological Informatics, 2020, 57: 101085.

[48] REDMON J, DIVVALA S, GIRSHICK R, et al. You only look once: Unified, real-time object detection[C] //2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). June 27-30, 2016. Las Vegas, NV, USA; IEEE, 2016: 779-788.

[49] JOCKER G, CHAURASIA A, QIU J. Ultralytics YOLO (Version 8.0.0) [CP/OL]. (2023-07-20) [2024-03-17]. <https://github.com/ultralytics/ultralytics>.

(编辑 邵煜)