

基于改进 Yolov5s 的增强现实文物识别方法

张元^{1,2,3}, 关瑜^{1,2,3}, 熊风光^{1,2,3}, 庞敏^{1,2,3}, 况立群^{1,2,3}

- 中北大学 计算机科学与技术学院, 山西 太原 030051;
- 机器视觉与虚拟现实山西省重点实验室, 山西 太原 030051;
- 山西省视觉信息处理及智能机器人工程研究中心, 山西 太原 030051)

摘要:将增强现实技术应用于数字博物馆的文物展览,有助于拉近参观者与文物的距离,使展览更具趣味性。针对增强现实技术在文物展览应用场景中,增强现实设备采集的文物目标图像背景复杂、文物形状纹理丰富而导致的误检、识别准确率低的问题,提出一种基于改进的 Yolov5s 的文物识别方法。在 Yolov5s 网络结构中的骨干网络与颈部网络之间引入了 CBAM 注意力机制,并在骨干网络中的 Bottleneck 模块中,使用多头注意力机制替换普通卷积,有效捕获局部信息,降低了无用信息的干扰。为了提高识别网络对于目标文物的边界框定位精度,采用 DIoU-NMS 方法挑选最优的目标识别框作为最终的预测框。实验结果表明,该方法提高了文物的平均识别精度,比原模型更适用于 AR 应用文物的目标识别。

关键词:增强现实;文物识别;Yolov5s;注意力机制;多头自注意力机制;DIoU-NMS

中图分类号:TP311

文献标识码:A

文章编号:1673-629X(2024)07-0017-07

doi:10.20165/j.cnki.ISSN1673-629X.2024.0118

An Augmented Reality Cultural Relics Recognition Method Based on Improved Yolov5s

ZHANG Yuan^{1,2,3}, GUAN Yu^{1,2,3}, XIONG Feng-guang^{1,2,3}, PANG Min^{1,2,3}, KUANG Li-qun^{1,2,3}

- School of Computer Science and Technology, North University of China, Taiyuan 030051, China;
- Shanxi Provincial Key Laboratory of Machine Vision and Virtual Reality, Taiyuan 030051, China;
- Shanxi Province Visual Information Processing and Intelligent Robot Engineering Research Center, Taiyuan 030051, China)

Abstract: Applying augmented reality technology to the cultural relics exhibition of the digital museum helps to shorten the distance between the visitors and the cultural relics, and make the exhibition more interesting. Aiming at the problems of false detection and low recognition accuracy caused by complex background and rich shape and texture of cultural relics collected by augmented reality equipment in the application scene of cultural relics exhibition, an improved cultural relics recognition method based on Yolov5s is proposed. The CBAM attention mechanism is introduced between the backbone network and the neck network in the Yolov5s network structure. In addition, a multi-head attention mechanism is used to replace common convolution in the Bottleneck module of the backbone network, which effectively captures local information and reduces the interference of unnecessary information. In order to improve the location accuracy of the boundary frame of the target cultural relics, the DIoU-NMS method is used to select the optimal target identification frame as the final prediction frame. Experimental results show that the proposed algorithm improves the average detection accuracy of the cultural relics, which is more suitable for the target detection of cultural relics than the original algorithm in AR application.

Key words: augmented reality; cultural relic identification; Yolov5s; attention mechanism; multi-head self-attention; DIoU-NMS

0 引言

目前,文物展览主要以博物馆展示的形式存在,文物被安放在展览柜中或以图文并茂的形式呈现给参观者。然而,这种展览形式仅给予参观者远观的机会,文

物的细节信息无法清晰地展现出来,参观者无法真正地身临其境,深入感受文物的历史、文化和艺术价值^[1]。为了突破传统博物馆文物展览存在的限制,增强现实技术逐渐被应用到了博物馆中,从而使文物展

收稿日期:2023-10-10

修回日期:2024-02-21

基金项目:山西省科技重大专项计划“揭榜挂帅”项目(202201150401021)

作者简介:张元(1976-),女,硕士,副教授,CCF 专业会员(48948M),通信作者,研究方向为人工智能与计算机视觉;关瑜(1996-),女,硕士研究生,研究方向为计算机视觉。

览的形式变得更加丰富,给博物馆的数字化新的契机。例如,Hou Weiting^[2]开发的基于 HoloLens 的博物馆展览应用系统,以其清晰的界面和高完成度得到了用户的认可;Alana A. U. Kennedy 等^[3]学者在拉布雷亚(LaBreak)沥青坑博物馆(LaBreak Hotel)中以增强现实技术为理论依据,进行了一项以情感和兴趣为基础的古物陈列实验;Jiang T 等^[4]学者设计的基于混合现实技术的新型智能数字博物馆系统,通过人机交互方式让人感受到更加真实和多样化的文物信息。

文物识别算法是数字博物馆的核心技术之一,其准确率直接影响到基于增强现实的文物展览效果^[5-8]。基于增强现实设备的文物展览对文物进行识别分类,并根据分类结果检索相应文物模型,精准的文物目标识别对后续的文物展览系统提供正确的文物模型。Philipp-Foliguet 等^[9]制作出专用于古代文物及艺术品的三维模型检索与分类的系统。Desai 等^[10]使用基于内容的图像检索技术支持考古学家对考古发现进行评估和分类。他们使用 CBIR 技术对考古遗迹进行自动分类,使用视觉特征形状和纹理来研究艺术形式并从参考收藏中检索相似图像。Guyot 等^[11]使用深度学习卷积神经网络方法自动识别、描绘和描述异常地形。

目前,深度学习主流的目标识别模型依据有无候选框而被分为两大类,一类是基于候选框的目标识别模型,包含 RCNN^[12]、Fast RCNN^[13]、Mask RCNN^[14]等模型,此类模型虽然识别精度较高,但识别速度偏低,无法达到增强现实对速度的需求。第二类是以 YOLO 和 SSD^[15]为代表的基于端到端的目标识别模型,此类模型在目标识别过程中不需要使用传统的候选框,而是对被识别目标直接进行回归与分类操作,在保持识别精度的同时提高了识别速度,适用于实时目标识别。YOLOv5^[16]模型是一种具有高准确率和快速识别能力的目标识别模型,其识别效果优于 Faster RCNN 模型和 SSD 模型。因此,该文选择 YOLOv5 模型来进行文物检测识别。

为了采集文物数据,使用了增强现实设备 HoloLens 眼镜。在文物识别部分,该文共包含了 4 个大类和 7 个小类的文物。这些小类文物具有各自独特的形状和纹理特征,且这些特征非常复杂。在实际应用中,使用 HoloLens 设备采集的目标图像中往往包含大量的背景信息,导致文物检测精度降低,出现漏检和误检的问题。由于传统 YOLOv5 模型难以解决这些问题,该文对 YOLOv5 算法进行了改进,以更好地应对具有复杂形状特征的文物识别任务。这样的优化方法能够提高文物的检测识别精度,为实际应用中的文物识别提供了更可靠的解决方案。

1 基于改进 Yolov5s 的增强现实文物识别方法

Yolov5s 是 YOLOv5 系列中的一个模型,采用单阶段目标检测算法,具有高检测精度和快速检测速度。它由输入端、网络头部、骨干网络、颈部网络和输出层组成。骨干网络采用 Darknet-53 进行图像特征提取。同时,引入了特征金字塔机制实现多尺度目标识别,确保对不同尺寸的目标准确检测。在分类任务中采用逻辑回归替代 softmax,提供了更高效和准确的分类概率处理多分类问题。为了增强网络对文物等细节丰富的目标的识别能力,该文在 Yolov5s 引入 CBAM 注意力机制^[17],能够使网络更加关注重要的特征,并抑制无关信息,从而提升目标检测的精度。另外,在骨干网络的 Bottleneck 模块中,在 Yolov5s 中采用 MHSA (Multi-Head Self-Attention) 机制^[18]替换了普通卷积。MHSA 能够更好地捕捉目标间的长程依赖关系,有效地降低了文物纹理和复杂背景对目标识别精度的影响。最后,在目标框回归阶段,为了提升被检目标的识别框回归的稳健性,采用 DIou-NMS 方法取代传统 Yolov5s 中的非极大抑制方法,通过计算目标框之间的交并比以及目标框的距离,能够更加准确地选择最优的检测结果,从而提高了目标检测的准确性。

1.1 改进后的 Yolov5s 网络结构

针对文物漏检、误检的问题,对 Yolov5s 算法进行了优化,具体优化方法如下:优化后 Yolov5s 的网络结构如图 1 所示。优化的具体方法如下:

首先,考虑到文物具有丰富的纹理细节特征,为了更好地获取文物的局部信息并提高识别精度,在 Yolov5s 网络结构的骨干网络与颈部网络之间引入了 CBAM 注意力机制。注意力机制可以自动地关注图像中的重要区域,从而更好地捕捉文物的关键特征。

其次,为了降低文物纹理和复杂背景对目标识别精度的影响,在骨干网络中的 Bottleneck 模块使用 MHSA 机制替换了普通卷积。MHSA 机制可以更好地建模局部和全局上下文信息,从而更好地提取文物的特征表示。

最后,为了提高被检测目标的识别框回归的稳健性,采用了 DIou-NMS 方法替代了 Yolov5s 算法中的传统非极大抑制方法。DIou-NMS 方法考虑了检测框与真实框之间的重叠程度和距离,从而更准确地确定目标的位置和大小。

1.2 CBAM 注意力机制

在文物检测识别任务中,CBAM 模块的优势在于关注重要特征、增强特征提取能力、减轻过拟合和提高计算效率。针对文物形状纹理细节特征丰富的特点,为了更好地获取文物的局部信息,提升文物识别精度,

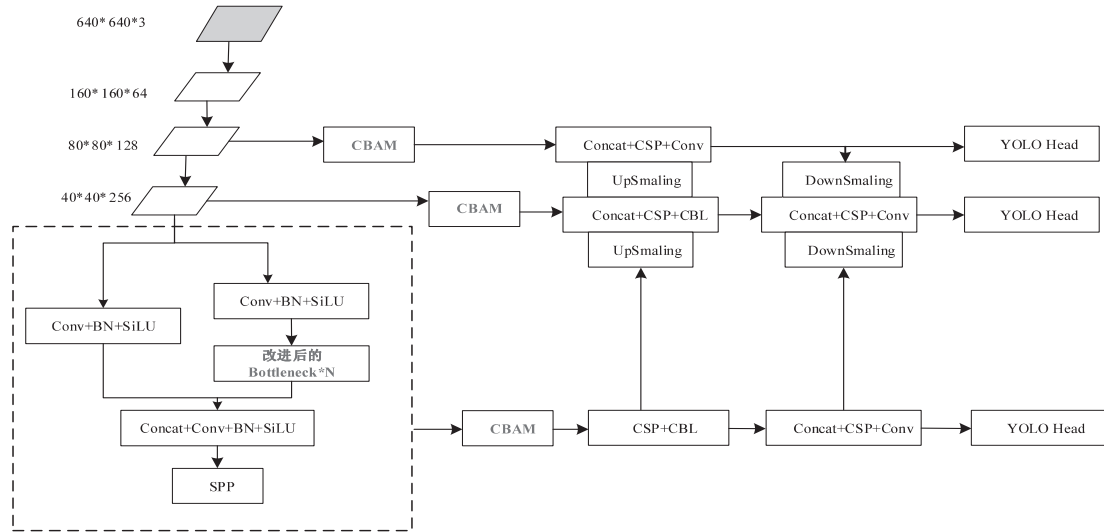


图 1 改进后的 Yolov5s 网络结构

降低模型的复杂度,在 Yolov5s 网络结构中的骨干网络与颈部网络之间引入了 CBAM 注意力机制^[15]。注意力机制的原理是通过提高对感兴趣的局部区域信息的关注,从而抑制图像中无用信息的干扰。CBAM (Convolutional Block Attention Module) 注意力机制主要分为两个模块:通道注意力模块、空间注意力模块。CBAM 模块结构如图 2 所示。

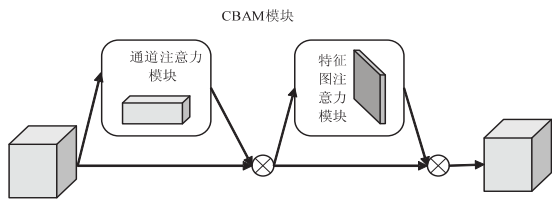


图 2 CBAM 模块结构

将大小为 $H * W * C$ 的特征图 F 输入到 CBAM 的通道注意力机制后,分别进行 MaxPool 和 AvgPool 操作,输出两个大小为 $1 * 1 * C$ 的特征图,并将其分别作为输入特征图经过一个两层共享的神经网络,先通过一个全连接层改变通道数量,再通过另一个全连接层恢复通道数。然后,将神经网络输出的特征进行加权和操作,经过激活函数操作最终形成 $MC(F)$ 特征图。通道注意力模块如图 3 所示。

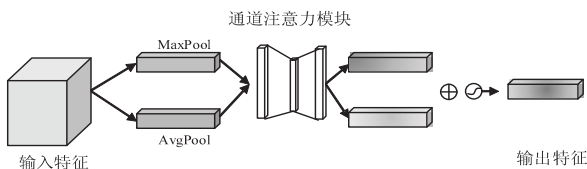


图 3 通道注意力机制模块

将通道注意力模块输出的 $MC(F)$ 特征图与特征图 F 进行内积运算,得到的特征图作为输入空间注意力机制特征图,分别进行 MaxPool 和 AvgPool 操作,输出两个大小为 $H * W * 1$ 的特征图。融合输出的两个特征图,通过卷积操作与激活函数操作,最终输出 $MS(F)$ 特征图。将 $MS(F)$ 特征图与特征图 F 进行内积运算,输出的特征图是整个 CBAM 模块的输出。空间注意力模块如图 4 所示。

(F) 特征图。将 $MS(F)$ 特征图与特征图 F 进行内积运算,输出的特征图是整个 CBAM 模块的输出。空间注意力模块如图 4 所示。

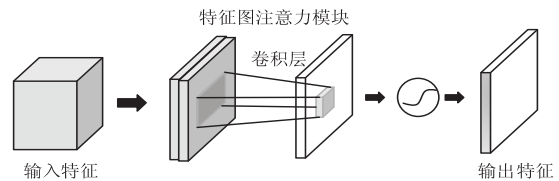


图 4 空间注意力机制模块

1.3 改进 Bottleneck 模块

文物识别部分共包含 7 个不同种类的文物,此 7 类文物拥有的形状特征各不相同且纹理特征复杂,需要进行识别的文物目标图像通过增强现实设备采集,而增强现实设备采集的图像分辨率大小为 $1408 * 792$,因而采集到的目标图像背景部分较多且复杂。在实验过程中发现,丰富的文物纹理和复杂的背景会影响文物特征提取的效果,从而降低文物识别的准确性,导致文物漏检和误检的问题。Yolov5s 的 Backbone 部分采用 CSPdarknet-53 网络结构进行特征提取,将输入的大小为 $640 * 640$ 的图像转化为 $20 * 20$ 的特征图。其中,该网络结构中包含了多个 CSP 模块,该模块将原始输入特征图分别进行两种不同的操作,分别为卷积操作和卷积操作与 N 个 Bottleneck 模块相结合的操作,并将操作后的特征图进行融合。CSP 模块如图 5 所示。

针对上述问题,对 CSP 模块中的 Bottleneck 模块进行改进。改进后的 Bottleneck 模块使用多头注意力机制 (Multi-Head Self-Attention, MHSA)^[18] 替换 $3 * 3$ 的普通卷积。普通卷积可以有效捕获局部信息,但在目标识别中需要对长距离相关性建模,自注意力机制看到上下文的同时聚焦与它相关性更高的内容,避免来自对外部信息的干扰。注意力机制有 3 个输入,分

别为:键、查询和值。注意力机制的计算公式如式 1 所示。

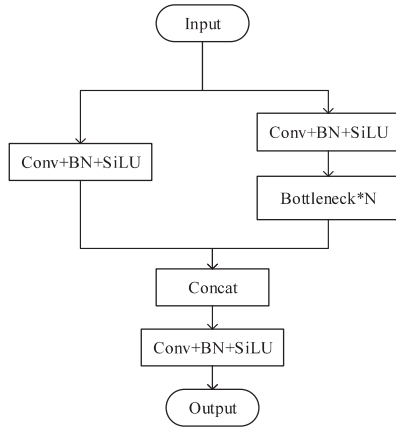


图 5 CSP 模块

$$\text{Attention}(\mathbf{Q}, \mathbf{K}, \mathbf{V}) = \text{softmax}\left(\frac{\mathbf{Q}\mathbf{K}^T}{\sqrt{d_k}}\right)\mathbf{V} \quad (1)$$

其中, $\mathbf{Q}, \mathbf{K}, \mathbf{V}$ 分别代表查询、键和值的矩阵, d_k 表示键的维数。与单头注意力机制不同, 多头注意力机制可以使模型同时关注多个重点区域, 并为模型的注意力层提供了若干表示子空间, 多头自注意力中将会有多个学习 $\mathbf{Q}, \mathbf{K}, \mathbf{V}$ 的权重矩阵, 将矩阵均独立地进行随机初始化操作, 并在此基础上将输入向量与若干不同的子空间进行映射, 进而丰富了信息的特征表达。为了使模型可以从注意力层中的若干表示子空间中获取更多学习信息, 提出 MHSA 机制。将自注意力头数设置为 8, 通过独立学习将 MHSA 机制输出的 8 组不同线性投影变换矩阵对键、查询和值进行变换, 并将其融入注意力中。然后, 将融入注意力中的输出进行拼接操作, 再次进行线性投影变换, 得到最终的输出公式如式 2、式 3 所示。

$$\text{MultiHead}(\mathbf{Q}, \mathbf{K}, \mathbf{V}) = \text{Concat}(\text{head}_1, \text{head}_2, \dots, \text{head}_h)\mathbf{W}^o \quad (2)$$

$$\text{head}_i = \text{Attention}(\mathbf{Q}\mathbf{W}_i^q, \mathbf{K}\mathbf{W}_i^k, \mathbf{V}\mathbf{W}_i^v) \quad (3)$$

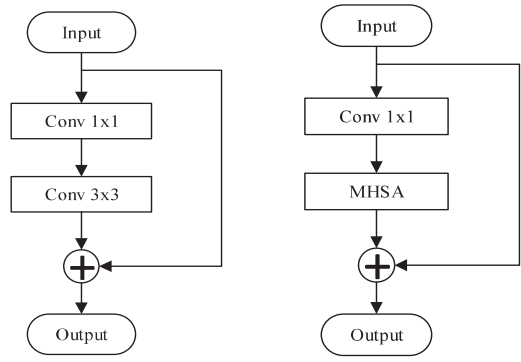
其中, $\mathbf{W}^o, \mathbf{W}_i^q, \mathbf{W}_i^k, \mathbf{W}_i^v$ 是参数矩阵。CNN 具有局部连接性, 对局部信息的处理更有优势, 但对全局信息的注意力较差。该文在 Bottleneck 模块引用 MHSA, 增强对局部信息的注意, 但在文物识别中, 采集的文物图像尺寸较大, 导致训练和推理的开销增大。因此, 该文只在低分辨率的特征图中引用 MHSA 处理和汇总卷积捕获的特征映射中包含的信息。Yolov5s 骨干网络中包含多个 CSP 模块, 在最后一个 CSP 模块的 Bottleneck 中使用 MHSA 替换 3×3 普通卷积, 骨干网络架构如表 1 所示。

捕获全局信息可以有效地增强前景和背景之间的可分辨性, 有利于提高复杂背景下的文物识别率, 降低文物漏检的概率。改进前后的 Bottleneck 模块如图 6

所示。其中, 图 6(a) 为改进前的 Bottleneck 模块, 图 6(b) 为改进后的 Bottleneck 模块。

表 1 骨干网络架构

Stage	Output	YOLOv5	文中算法
Conv+BN+SiLU	$160 \times 160 \times 64$		
CSP1	$160 \times 160 \times 64$	$\begin{bmatrix} 1 & * & 1 \\ 3 & * & 3 \end{bmatrix} \times 1$	$\begin{bmatrix} 1 & * & 1 \\ 3 & * & 3 \end{bmatrix} \times 1$
CSP2	$80 \times 80 \times 128$	$\begin{bmatrix} 1 & * & 1 \\ 3 & * & 3 \end{bmatrix} \times 3$	$\begin{bmatrix} 1 & * & 1 \\ 3 & * & 3 \end{bmatrix} \times 3$
CSP3	$40 \times 40 \times 256$	$\begin{bmatrix} 1 & * & 1 \\ 3 & * & 3 \end{bmatrix} \times 3$	$\begin{bmatrix} 1 & * & 1 \\ 3 & * & 3 \end{bmatrix} \times 3$
CSP4	$20 \times 20 \times 512$	$\begin{bmatrix} 1 & * & 1 \\ 3 & * & 3 \end{bmatrix} \times 1$	$\begin{bmatrix} 1 & * & 1 \\ 3 & * & 3 \end{bmatrix} \times 1$ [MHSA]



(a) Bottleneck 模块

(b) 改进后的 Bottleneck 模块

图 6 Bottleneck 模块结构

1.4 优化的非极大抑制方法

目标识别推理过程中, 一个目标会产生多个识别框, 但最终被识别目标仅需要一个识别框。非极大抑制 (NMS) 的作用是除去冗余的识别框。从众多识别框中选择得分高的识别框, 并降低其他识别框的分数, 挑选最优的目标识别框作为最终的预测框。

在文物识别过程中, 部分文物的纹理相似, 当此类文物距离相近或出现遮挡时, 可能导致文物的漏检、误检问题。该文采用了 DIoU-NMS 方法代替 Yolov5s 模型中传统的 NMS 方法, DIoU-NMS 方法使用 DIoU 作为阈值代替普通的 IoU, 其目的是在计算过程中考虑目标与锚框间的距离、尺度以及重叠率, 使被识别目标的识别框回归更具稳健性。DIoU 计算公式如式 4 所示。

$$\text{DIoU} = \text{IoU} - \frac{\rho^2(b, b^{gt})}{c^2} \quad (4)$$

其中, $\rho^2(b, b^{gt})$ 表示两个识别框与中心点距离的平方, c^2 表示两个识别框最小外接矩形的对角线长度的平方。DIoU-NMS 计算公式如式 5 所示。

$$s_i = \begin{cases} s_i, \text{IoU} - R_{\text{DIoU}}(M, B_i) < \varepsilon, \\ 0, \text{IoU} - R_{\text{DIoU}}(M, B_i) \geq \varepsilon, \end{cases} \quad R_{\text{DIoU}} = \frac{\rho^2(b, b^{gt})}{c^2} \quad (5)$$

2 实验结果分析

2.1 数据集介绍

2.1.1 数据采集

该文使用的数据集共包含礼器类、陶器类、瓷器类和玉器首饰类共4个大类7个小类不同的文物,通过互联网途径和现实拍摄收集上述文物图像样本。为提高系统对于复杂环境下的文物识别适应能力,现实拍摄中在不同角度、环境下,充分采集文物在顺光与逆光等不同光线下3000张图像样本。并通过互联网收集此7类文物在不同场景中共3000张图像样本。经过整理、筛选后,共得到4000张图像样本。代表性文物示例如图7所示。

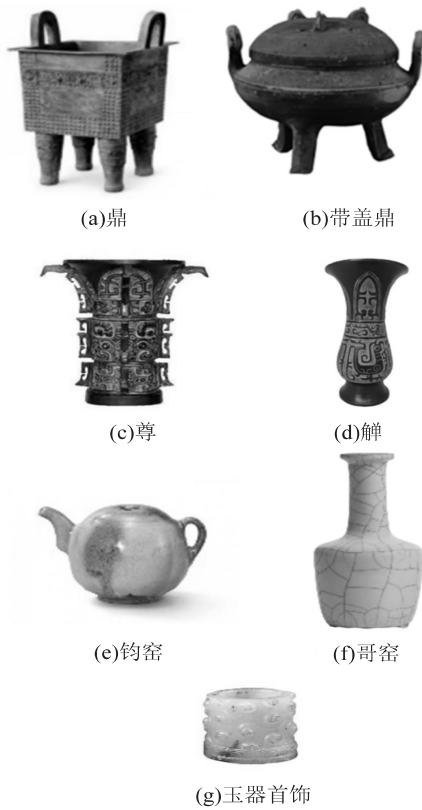


图7 文物数据

2.1.2 数据集准备

数据集的标注采用 LabelImage 标注工具,通过手动拉取矩形框标出整个目标文物并对其进行相应的标记,标注完成后每一张图片得到一张 txt 文件。由于文物的纹理较为复杂,同一种类的文物形状可能不同,模型训练过程中,为防止过度拟合,缓解各类样本由于数据量过少和数据分布不均的问题,对图片样本进行数据扩充。通过对图像样本数据依据其数量比例依次进行垂直镜像、水平镜像以及旋转不同角度等操作进行扩充。样本扩充如图8所示。经上述扩充操作后的文物数据集的总数量达到了11507张,其中包括训练集8057张,测试集3450张,最终作为用于识别文物的数据集。样本种类及数量分布如表2所示。

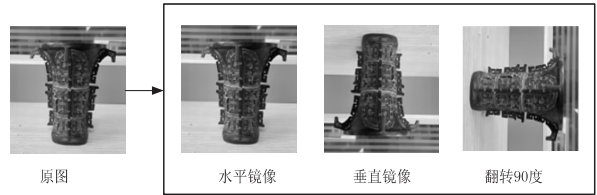


图8 样本扩充

表2 样本种类及数量分布

文物大类	文物小类	数量
青铜器	鼎	1 650
	尊	1 200
陶瓷	带盖鼎	1 600
	觥	1 800
瓷器	钧窑	1 100
	哥窑	1 600
玉器	首饰	2 207

在数据集扩充旋转操作中,不仅对图像进行整数倍旋转,也对其进行30°、60°等角度的旋转,图像在进行非整数倍角度的旋转操作后会产生大片纯白的区域,该区域与原图背景的像素差异较大,且大量的背景缺失容易影响文物识别训练效果。因此,该文使用均值填充方法解决旋转后样本背景缺失的问题。均值填充是图像预处理中的算法之一,其主要原理是使用图像像素的平均值代替缺失的空白区域中的像素,使数值缺失的区域的颜色与图像背景的颜色保持一致。旋转30°的样本数据与填充效果图如图9所示。

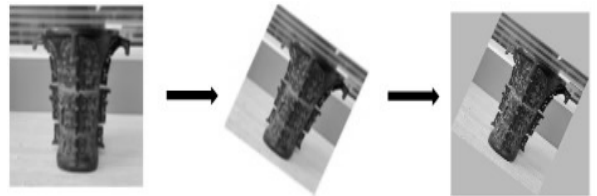


图9 旋转30°样本数据

2.2 实验环境及评价指标

实验采用 Windows10 操作系统,显卡为英伟达 RTX3080;开发环境为 PyTorch 1.0.2, OpenCV 4.1.1, Python 3.8。原 YOLOv5s 算法和优化后的 YOLOv5s 算法都是在 PyTorch 环境下运行,以8057张文物图片做训练集,3450张做测试集。设置的学习率为0.001,训练迭代次数为3000。最终的 Loss 值稳定在0.5左右。

评价指标不仅可以对算法性能的优劣进行科学的评估,还可以辅助分析算法中存在的缺陷,从而对算法进行优化。评价文中算法性能的指标分别有:准确率、召回率、每张图片运行时间以及平均识别精度均值。目标识别中, Precision 为被识别为文物的识别精度,如式6所示; Recall 为被正确识别为文物的概率,如式7所示; mAP 为文物平均识别精度,如式8所示; 识别速度如式9所示。

$$\text{Precision} = \frac{\text{TP}}{\text{FP} + \text{TP}} \quad (6)$$

$$\text{Recall} = \frac{\text{TP}}{\text{FN} + \text{TP}} \quad (7)$$

$$\text{mAP} = \frac{1}{n} \sum_{i=0}^N \text{AP}_i \quad (8)$$

$$\text{FPS} = \frac{\text{Frames}}{\text{Time}} \quad (9)$$

在文物识别中, TP (True Positives)、FP (False Positives) 分别代表文物被正确识别的数量与错误识别的数量, FN (False Negatives) 代表文物被误检的数量, AP (Average-Precision) 为单个类别的目标识别精度, N 为所有类别的数量, Frames 为文物图像数量, Time 为文物所需时间。

2.3 结果分析

针对文物识别中文物纹理特征丰富、采集目标图像背景复杂的问题, 提出了优化骨干网络中的 Bottleneck 模块, 为更好地提取文物特征信息加入 CBAM 注意力机制模块, 为满足一个文物可以多标签标注并缓解因遮挡而产生的文物漏检的问题引入了 DIoU-NMS。数据集共有 7 小类文物, 分别为鼎、带盖鼎、尊、钧窑、觚、哥窑、玉器首饰。该文选取了玉器首饰、鼎、尊、觚、哥窑的识别结果进行对比分析, 对应的标签分别分 yuqi、ding、zun、zhi、geyao。原 YOLOv5 算法和优化后的 YOLOv5 算法的文物检测识别对比如图 10 所示。

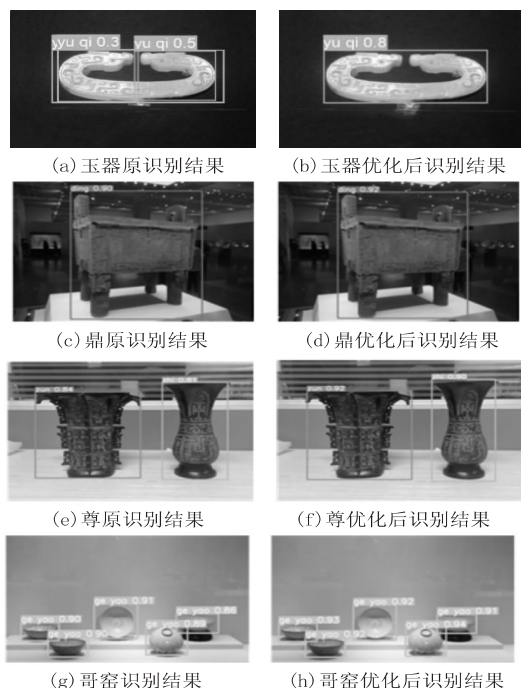


图 10 文物识别结果对比

由图 10 可见, a 组中包含单个有缺口的玉器, 在原 YOLOv5 算法中存在对该玉器识别不准确的问题, 将其错误地识别为两个玉器, 而改进后的 YOLOv5 算

法不仅成功解决了这一问题, 还显著提高了识别精度。在 b 组, 含有单个尊, 由于尊的形状和纹理相对单一, 因此原 YOLOv5s 算法和改进后的算法在识别精度上相差不大。在 c 组中, 左侧为尊, 右侧为觚。d 组包含多个哥窑, 相对于原 YOLOv5s 算法, 改进后的 YOLOv5s 算法在多目标文物的场景中识别精度都有所提升, 其提升幅度在 0 到 0.1 之间。综合来看, 改进后的算法在单个文物和多目标文物场景中都表现出较好的性能, 为特定场景下的文物识别精度提升提供了显著的改进。

为了更好地验证提出的目标识别算法的性能, 使用自制的文物数据集设计了训练实验, 并使用测试集进行测试。对提出的优化算法进行消融实验, 分析并对比不同算法的召回率、平均识别精度 (mAP) 和每秒帧率 (FPS)。算法性能对比如表 3 所示。在原 YOLOv5s 算法基础上, 分别改进骨干网络中 Bottleneck 模块、引入 CBAM 注意力机制以及采用非极大抑制 (DIoU-NMS) 方法, 实验结果表示, 每一步算法的 mAP 都有所提升, 原 YOLOv5s 模型的召回率为 95.5%, 优化后的 YOLOv5s 模型召回率为 97.3%, 比原模型提高了 1.8 个百分点。原 YOLOv5s 模型平均检测精度 (mAP) 为 84.6%, 改进后的 YOLOv5s 模型平均检测精度 (mAP) 为 86.7%, 比原模型提高了 2.1 个百分点。虽然文中模型的每秒帧率 (FPS) 相较于原模型有所降低, 但文中模型对于文物的召回率、准确率与平均检测精度均有所提升, 而文中系统的完整度依赖于目标检测的精度, 因此改进后的模型更适用于文中系统的文物检测。

表 3 消融实验

模型	召回率/%	mAP/%	FPS/(f · s ⁻¹)
Yolov5s	95.5	84.6	31
Yolov5s_DIoU NMS	97.1	85.2	33
Yolov5s_Bottleneck	96.2	86.0	28
Yolov5s_CBAM	97.0	85.7	30
改进后的 YOLOv5s	97.3	86.7	27

针对增强现实技术在文物展览应用场景中, 由增强现实设备采集的文物目标图像背景复杂、文物纹理丰富而导致的误检、识别准确率低的问题, 将不同目标识别模型与优化后的 YOLOv5s 模型的各项性能做了进一步对比, 如表 4 所示。由表 4 可知, Faster RCNN 模型、Mask RCNN 模型和 SSD 模型的平均识别精度 (mAP) 分别为 83.3%, 82.1%, 78.0%, 原 YOLOv5s 模型平均识别精度 (mAP) 为 84.6%, 优化后的 YOLOv5s 模型的平均识别精度 (mAP) 为 86.7%, 比原模型提高了 2.1 个百分点。由对比实验可得, 每一类文物的平均识别精度 (mAP) 都有所提升。综上所述, 对 YOLOv5 模型改进后, 更适合文物的识别。

表4 模型性能对比

模型	鼎	带盖鼎	尊	钧窑	觚	哥窑	玉器首饰	mAP/%
Faster RCNN ^[13]	85.4	82.3	78.4	82.1	87.6	83.7	83.3	83.3
Mask RCNN ^[14]	80.8	84.6	81.2	83.3	83.1	79.2	82.6	82.1
SSD ^[15]	75.1	71.9	76.2	83.4	82.6	71.6	85.2	78.0
YOLOv5S ^[16]	84.3	85.1	79.6	83.5	87.9	85.0	86.8	84.6
改进后的 YOLOv5s	85.9	88.7	81.3	85.9	86.4	90.1	88.7	86.7

为展示文物识别应用效果,使用增强现实设备 HoloLens 眼镜对目标文物进行信息采集,使用文中模型对其进行识别通过关键字检索相应模型进行展示, HoloLens 文物展示如图 11 所示。

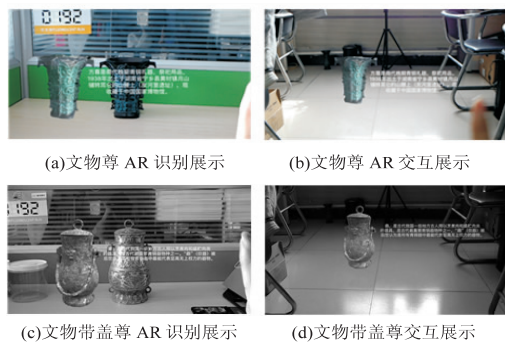


图 11 文物 AR 展示

3 结束语

针对增强现实技术在文物展览应用场景中存在的问题,提出了一种优化后的基于 YOLOv5s 模型的文物检测方法。通过实验结果的分析表明,该模型在文物识别方面有较好的性能表现,具有更高的精度和鲁棒性,相比于 Faster RCNN 模型、Mask RCNN 模型、SSD 模型和原 YOLOv5 模型,具有更好的适应性和实用性。未来,将进一步探索和研究该方法的实际应用,提高其在文物展览和保护方面的实用性和应用效果,为数字博物馆的发展和文物保护做出贡献。

参考文献:

- [1] 赵婧. 数字博物馆 2.0—智能化时代线上博物馆的未来[J]. 博物院, 2023(3):55-62.
- [2] HOU W. Augmented reality museum visiting application based on the Microsoft HoloLens[J]. Journal of Physics: Conference Series, 2019, 1237(5):052018.
- [3] KENNEDY A A U, THACKER I, NYE B D, et al. Promoting interest, positive emotions, and knowledge using augmented reality in a museum setting[J]. International Journal of Science Education, Part B, 2021, 11(3):242-258.
- [4] JIANG T, GAN X, LIANG Z, et al. AIDM: artificial intelligent for digital museum autonomous system with mixed reality and software-driven data collection and analysis[J]. Automated Software Engineering, 2022, 29(1):22.
- [5] 朱叶. 基于增强现实的博物馆动态展示应用——以桂林博物馆为例[J]. 中国民族博览, 2022(11):197-200.
- [6] 蔚小清. 移动端数字化文物展示的交互体验设计研究[D]. 沈阳:沈阳航空航天大学, 2018.
- [7] 侯伟婷. 基于 HoloLens 的博物馆沉浸式导览系统[D]. 天津:天津大学, 2019.
- [8] 韩玉, 韩晓雯, 卢颖, 等. 虚拟现实与增强现实技术在中医药博物馆科学普及中的应用分析[J]. 中国医药科学, 2020, 10(20):228-233.
- [9] PHILIPP-FOLIGUET S, JORDAN M, NAJMAN L, et al. Artwork 3D model database indexing and classification[J]. Pattern Recognition, 2011, 44(3):588-597.
- [10] DESAI P, PUJARI J, AYACHIT N H, et al. Classification of archaeological monuments for different art forms with an application to CBIR[C]//Proceedings of 2013 international conference on advances in computing, communications and informatics (ICACCI). Mysore:IEEE, 2013:1108-1112.
- [11] GUYOT A, LENNON M, LORHO T, et al. Combined detection and segmentation of archeological structures from LiDAR data using a deep learning approach[J]. Journal of Computer Applications in Archaeology, 2021, 4(1):1.
- [12] 王婉婷, 姜国龙, 褚云飞, 等. 从 RCNN 到 YOLO 系列的物体检测系统综述[J]. 齐鲁工业大学学报, 2021, 35(5):9-16.
- [13] 杜芸彦. 基于 Faster R-CNN 改进的少样本目标识别算法研究[D]. 北京:中国科学院大学(中国科学院光电技术研究所), 2022.
- [14] 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.
- [15] LIU W, ANGUELOV D, ERHAN D, et al. SSD: single shot MultiBox detector[C]//Proceedings of the 14th European conference on computer vision (ECCV). Amsterdam:Springer, 2016:21-37.
- [16] JOCHER G. YOLOv5 [EB/OL]. [2020-08-10]. <https://github.com/ultralytics/yolov5>.
- [17] WOO S, PARK J, LEE J Y, et al. CBAM: convolutional block attention module[C]//Proceedings of the 15th European conference on computer vision (ECCV). Munich:Springer, 2018:3-19.
- [18] XIAO X, ZHANG D, HU G, et al. CNN - MHSA: a convolutional neural network and multi-head self-attention combined approach for detecting phishing websites[J]. Neural Networks, 2020, 125:303-312.