

基于优化YOLOv8的唐卡图像目标检测方法

程维龙, 僧冰枫, 刘晓静

(青海大学计算机技术与应用学院, 青海 西宁 810016)

摘要: 唐卡中的元素绘制复杂且存在的多尺度情况会影响目标检测技术任务的准确度。为此, 提出一种优化YOLOv8模型的唐卡元素目标检测方法。首先, 采用级联融合网络提取图像特征, 并将特征提取参数用于后续的特征融合, 以有效增加参数利用效率; 其次, 借鉴双向特征金字塔网络的思想, 在同一层的特征信息传递层中增加一条额外的路径实现跨尺度连接, 以提升模型特征融合能力; 最后, 在检测头的回归损失函数中引入EIOU-Loss和CIOU-Loss, 考虑边界框回归的多种因素, 结合宽高和宽高比参数来提升模型目标定位效率和准确率。实验表明, 优化后的YOLOv8模型相较于原有模型参数量、计算量分别下降7.21%、7.23%, mAP50、mAP50-95分别提升3.72%、4.55%; 相较于其他目标检测算法优势明显; 消融实验也验证了不同改进模块对模型的积极作用。

关键词: 唐卡图像目标检测; YOLOv8模型; 级联融合网络; 跨尺度连接; 回归损失函数

DOI: 10.11907/rjtk.241837

开放科学(资源服务)标识码(OSID):

中图分类号: TP391

文献标识码: A

文章编号: 1672-7800(2025)001-0102-07



An Optimized Thangka Image Object Detection Method Based on YOLOv8

CHENG Weilong, SENG Bingfeng, LIU Xiaojing

(School of Computer Technology and Application, Qinghai University, Xining 810016, China)

Abstract: Complex and multi-scale rendering of elements in thangka can affect the accuracy of object detection technology tasks. Therefore, a Thangka element object detection method is proposed to optimize the YOLOv8 model. Firstly, a cascaded fusion network is used to extract image features, and the feature extraction parameters are used for subsequent feature fusion to effectively increase parameter utilization efficiency; Secondly, drawing on the idea of bidirectional feature pyramid network, an additional path is added in the feature information transmission layer of the same layer to achieve cross scale connection, in order to enhance the model's feature fusion capability; Finally, EIoU Loss and CIOU Loss are introduced into the regression loss function of the detection head, taking into account various factors of bounding box regression, and combining width to height and aspect ratio parameters to improve the efficiency and accuracy of model target localization. The experiment showed that the optimized YOLOv8 model reduced the parameter and computational complexity by 7.21% and 7.23% respectively compared to the original model, while mAP50 and mAP50-95 increased by 3.72% and 4.55% respectively; Compared to other object detection algorithms, it has significant advantages; The ablation experiment also verified the positive effects of different improvement modules on the model.

Key Words: Thangka image object detection; YOLOv8 model; cascaded fusion network; cross-scale connection; regression loss function

0 引言

唐卡是松赞干布时期兴起的一种独特绘画艺术, 具有鲜明的民族特点, 被称为藏族百科全书, 于2006年列入第

一批国家级名录^[1]。唐卡图像内容涉及较多的艺术元素, 对提升中国的文化软实力具有重要作用, 但因其年代久远、构成元素复杂, 通过人工检测唐卡图像的难度很大且耗时长, 因此目标检测技术在唐卡领域的研究具有一定实际价值^[2]。

收稿日期: 2024-09-30

扫描二维码阅读全文:



基金项目: 国家自然科学基金项目(61862053); 青海省自然科学基金青年项目(2023-ZJ-951Q)

作者简介: 程维龙(2000-), 男, 青海大学计算机技术与应用学院硕士研究生, 研究方向为目标检测、图像处理; 僧冰枫(1992-), 男, 青海大学计算机技术与应用学院讲师, 研究方向为图像处理、地理信息系统; 刘晓静(1978-), 女, 青海大学计算机技术与应用学院教授, 研究方向为图像处理与三维可视化。

早期, 目标检测主要依赖于手工设计的特征进行提取和传统机器学习方法, 通常使用边缘检测、颜色直方图、纹理等低级特征表示物体, 并基于支持向量机或决策树等分类器识别物体。然而, 传统方法在处理复杂场景、多尺度物体和遮挡时表现不佳, 原因为上述方法对特征选择和物体的表示高度依赖人工经验。

随着深度学习技术兴起, 目标检测领域经历了革命性变革。卷积神经网络等深度学习模型的出现使目标检测任务更准确和高效, 既能自动学习图像中的特征表示, 又无需要手动设计特征。例如, 薛盼盼^[3]针对唐卡主尊头饰与台座提出一种基于 ResNet 和可变形卷积的小样本检测方法。张效娟等^[4]在检测方法中增加旋转检测框来克服漏检情况。何国源^[5]研究出一种改进的 YOLOX 目标检测算法, 通过 TransFPN 得到更丰富的特征与语义信息。

此外, 深度学习还能处理更复杂的场景, 具备更强的泛化能力, 目前已出现了许多经典的目标检测模型, 主要分成两阶段和一阶段算法。例如, Girshick^[6]提出的 R-CNN 系列模型, 通过候选框生成、特征提取两个阶段识别物体, 精确度较高但计算成本较大。后续, Fast R-CNN^[7]、Faster R-CNN^[8]通过引入候选区域生成网络等手段来优化检测过程。Terven 等^[9]提出单阶段算法 YOLO, 采用单一

神经网络将目标检测任务建模为回归问题, 并同时预测目标的类别和边界框, 因此具有更高的检测速度。类似的, YOLOv2 引入锚框来处理多尺度物体, 并采用 Darknet-19 作为基准网络提升模型的特征提取能力。后续改进的 YOLO 算法通过优化主干网络、损失函数等手段进一步提升模型检测精度, 现已推出 YOLOv8 版本, 在同期的主流目标检测算法中具有较高的检测精度与速度^[10]。

针对唐卡图像场景下存在的复杂元素, 本文提出了一种改进 YOLOv8 的算法。首先, 在主干网络中引入级联融合模块丰富特征融合参数信息; 其次, 在融合模块的中层输出检测层中增加跳跃连接来增强特征融合能力; 再次, 在回归损失函数中加入高宽、高宽比等几何因素, 确保目标框所在位置包含了完整的物体; 最后, 通过网络爬虫等手段自建唐卡数据集并进行 YOLO 格式转换。

1 YOLOv8 结构

YOLOv8 是 YOLO 系列近年提出的算法, 在网络结构方面相较于前者改进较大, 网络结构如图 1 所示。本文主要介绍 YOLOv8 的网络结构组成(主干网络、融合模块、检测头等)及其相较于 YOLOv5 的改进部分^[11]。

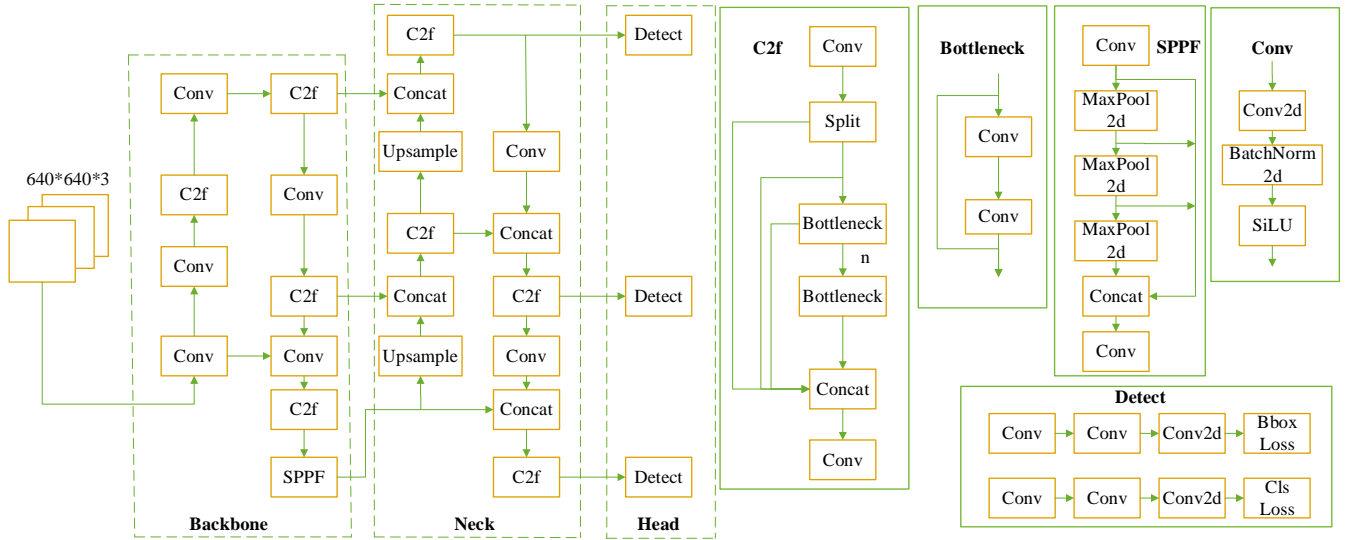


Fig. 1 YOLOv8 network structure

图1 YOLOv8网络结构

同时, 针对 YOLOv8 网络结构的特点, 主要介绍 YOLOv8 的网络结构的具体优化工作(主干网络、融合模块、损失函数等), 优化后的 YOLOv8 网络结构如图 2 所示。

由图 1、图 2 可见, 骨干网络将输入图像转换为多尺度信息的特征图, 通过多次卷积与池化操作为后续特征融合提供语义信息。YOLOv8 的主干特征提取网络与 YOLOv5 相似, 借鉴了 CSPDarkNet 结构, 为了保留全局和局部信息, 将主干骨干特征分为两部分: 一部分与辅助骨干特征融合; 另一部分直接传递给后续层^[12]。YOLOv8 相较于 YOLOv5 而言, 主要将 Backbone 的 C3 模块替换为 C2f 模块,

C2f 为一种改进的卷积层, 结合了 C3 模块与 YOLOv7 中 ELAN 的思想, 多次利用卷积核大小为 1x1、3x3 的卷积操作降低特征图维度, 实现特征提取和压缩^[13]。同时, 并行更多梯度流分支(Bottleneck 模块)获得更丰富的梯度信息, 在保持较高准确度的同时降低了模型复杂度。

Neck 模块结合不同层级的特征图生成具有多尺度信息的特征图, 以提升目标检测准确度。在 YOLOv8 中, 通过上采样、下采样等操作构建特征金字塔, 并利用横向连接融合不同分辨率的特征图后传递给目标检测头部, 用于目标定位和类别预测。YOLOv8 相较于 YOLOv5 而言, 删

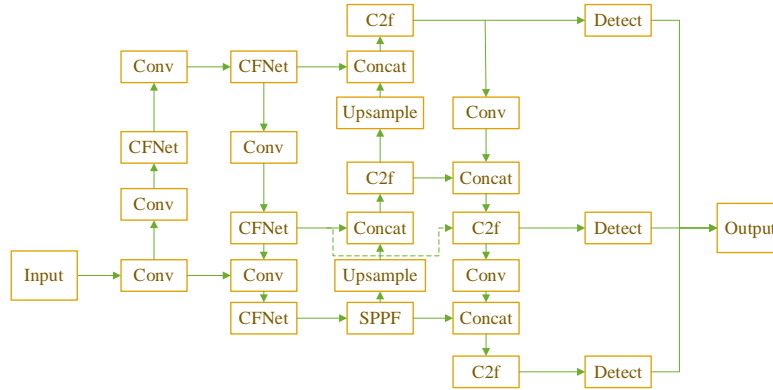


Fig. 2 Optimized YOLOv8 network structure

图2 优化后的YOLOv8网络结构

除了上采样阶段中1×1卷积结构,将Backbone不同阶段输出的特征直接送入上采样操作来提升计算效率,并在特征融合过程将C3模块替换为C2f模块。如图1所示,YOLOv8特征融合模块包含双向通路网络,首先自下向上的路径使底层信息更容易传递到顶层;其次经过多次下采样操作将输入分辨率为640×640的图片转换为3种不同分辨率的图片;最后将各图片输入PANet中,经过上采样通道融合操作将PANet的3个输出分支送入Detect head中进行Loss等数据计算。

Head模块根据Backbone提的特征来预测目标位置和类别。在YOLOv8的模型中,通过卷积块进行通道数降维和特征图缩放,融合不同层级的特征图得到更丰富的特征信息,从而匹配实际目标位置并将信息映射到类别分数等信息中。YOLOv8相较于YOLOv5而言,将耦合头更换为解耦头,并在回归分支使用Distribution Focal Loss中提出的积分形式表示法,以交叉熵形式优化与标签最接近的两个位置的概率,让网络更快聚焦到目标位置邻近区域的分布^[14];将回归坐标视为一个确定性单值,在正负样本匹配上由原来的静态策略更换为TaskAlignedAssigner,通过控制分类得分与IoU优化来引导网络动态关注并对齐高质量的Anchor,以准确预测物体位置和大小^[15]。

2 本文方法

2.1 特征提取网络优化

唐卡图像背景较为复杂,可能同时包含较小的法器细

节和较大的佛像画面,具有显著的尺度变化和纹理细节信息,因此单一尺度特征提取方法难以兼顾不同尺度目标的检测需求。目前,大多数方法首先通过分类主干提取多尺度特征,其次基于轻量级模块进行特征融合。然而,传统模型参数数量无法充分利用不同层级特征的信息,在检测部分复杂图像时效果较差。

为此,Zhang等^[16]提出一种新颖的多尺度融合方法——级联融合网络(CFNet),每个阶段包括一个用于特征提取的子主干和一个用于特征集成的极其轻量级的过渡块,通过多个级联阶段逐层融合不同尺度的特征,以逐步增强特征表示能力。图3为级联融合网络总体结构,主干由两个步长为2的3×3卷积组成,每个卷积后都有一个LayerNorm层和一个GELU单元。首先将输入的RGB图像分成一个主干和N个连续的块,用来提取空间尺寸为H/4×W/4的高分辨率特征;其次将提取的高分辨率特征馈送到M个级联之前,通过步长为2的2×2卷积进行下采样,而后所有级联共享相同结构,但具有不同的尺寸,即每个阶段包含一个特征提取子网络和一个特征融合子网络,以提取、集成不同尺度的特征;最后在每个阶段的最后块组中应用焦点块,通过扩展深度卷积与两个跳跃连接来合并细粒度局部特征和粗粒度全局特征,每一级以8、16、32的步幅输出特征P3、P4、P5,但只有P3被馈送到后一级,最后阶段输出P3、P4和P5的融合特征。实验表明,这种层级间信息传递机制能有效增强特征表达能力,使高层特征充分利用低层特征信息。

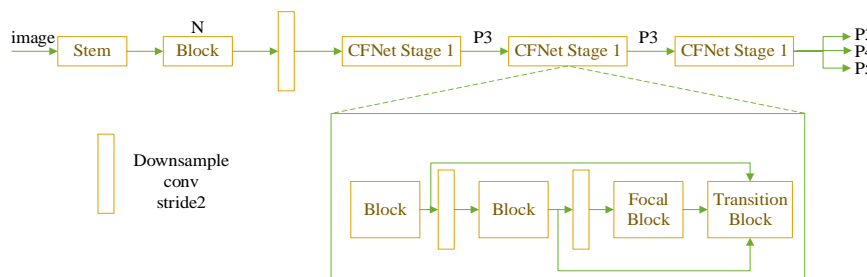


Fig. 3 Cascade fusion network structure

图3 级联融合网络结构

2.2 特征融合网络优化

融合网络中传统的特征金字塔网络 (Feature Pyramid Networks, FPN) 通过自上而下的路径实现特征融合, 仅限于将高层特征传递到低层特征中, 这种单向信息流动方式可能导致低层特征信息在高层融合过程中丢失, 从而影响多尺度目标的检测性能。为此, Tan 等^[17]提出一种优化多尺度融合进行的方法 BiFPN, 目的在于解决目标检测中的性能与效率之间的平衡问题, 以在不同硬件和计算资源上进行高效的目标检测。

本文借鉴了 BiFPN 的思想, 在同层特征信息传递中增加一条额外路径, 在中等大小的特征输出层中由第 6 层进行卷积上采样等操作后输出第 12 层特征, 并与 16 层特征进行融合得到最后输出, 即在第 6 层与第 16 层中添加一条额外的路径进行跳跃连接 (如图 4 弧线) 来融合更多特征。

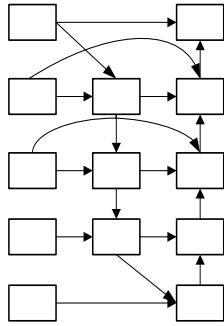


Fig. 4 Bidirectional pyramid network structure

图 4 双向金字塔网络结构

2.3 损失函数优化

目标检测一般可分为定位和检测两部分, 定位准确率主要由回归损失函数决定。YOLOv8 采用 CIoU 损失函数综合考虑了重叠面积、中心点距离、高宽比等因素对目标进行定位, 但其公式表示的是高宽比的差异, 而非高宽分别与其置信度的真实差异, 因此在唐卡场景下检测形状多变的元素时存在一定的困难^[18]。为此, Zhang 等^[19]在 CIoU 的基础上将高宽比拆开, 提出一种对边界框损失函数的优化函数 EIou Loss, 包含 IoU 损失、距离损失、高宽损失 (重叠面积、中心点距离、高宽比) 3 个部分。其中, 高宽损失直接最小化预测目标边界框与真实边界框的高宽差异, 使其具有更快的收敛速度和更好的定位效果, 具体公式如式 (1) 所示。

$$EIou_{Loss} = 1 - IOU + \frac{\rho^2(b, b^{gt})}{(w^c)^2 + (h^c)^2} + \frac{\rho^2(w, w^{gt})}{(w^c)^2} + \frac{\rho^2(h, h^{gt})}{(h^c)^2} \quad (1)$$

式中: b^{gt} 、 b 分别为预测框和目标框中心点的坐标; $\rho^2(b, b^{gt})$ 表示两者的欧氏距离; w^c 、 h^c 为预测框与目标框最小外界矩形的宽高。

在此基础上本文对 EIou 进行优化, 并应用到 YOLOv8 中。首先通过 CIoU 将预测帧的宽高收敛到一个合适的范

围; 其次通过 EIou 将每条边收敛到正确值, 具体计算公式为:

$$EIou_{Loss} = 1 - IOU + \alpha v + \frac{\rho^2(b, b^{gt})}{(w^c)^2 + (h^c)^2} + \frac{\rho^2(w, w^{gt})}{(w^c)^2} + \frac{\rho^2(h, h^{gt})}{(h^c)^2} \quad (2)$$

$$\alpha = \frac{v}{(1 - IOU) + v} \quad (3)$$

$$v = \frac{4}{\pi^2} \left(\arctan \frac{w^{gt}}{h^{gt}} - \arctan \frac{w}{h} \right)^2 \quad (4)$$

式中: α 为权重函数; v 用来度量宽高比的一致性。

3 实验结果与分析

3.1 数据集与实验环境

实验采用 Linux 操作系统, 编程语言为 Python 3.8, 搭载 Intel(R) Xeon(R) CPU E5-2603 v4 @ 1.70GHz 处理器, GPU 为 GTX 1080Ti 12 GB, 深度学习框架为 Pytorch 1.12.1, CUDA 11.8。由于网络上没有公开的唐卡数据集, 本文主要通过学校图书馆已有数据及网络爬虫获取相关唐卡图片, 在了解唐卡知识后将唐卡元素分为佛像 (如释迦牟尼)、法器 (如金轮)、头饰 (如僧帽) 3 类, 共涉及 57 个标签。如图 5 所示, 图中正中央为释迦牟尼佛像且手持法器钵均为检测任务的识别对象。

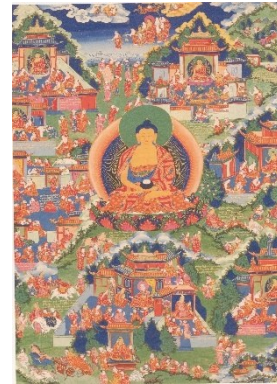


Fig. 5 Original image of dataset

图 5 数据集原图

本文首先通过在线网站进行标注, 生成包含 4 000 余张图片的数据集; 其次按照 YOLOv8 官网推荐的在线数据集网站将图片标注转换成 YOLO 格式; 最后按照 7:2:1 的比例划分训练集、验证集、测试集。为了提升模型训练速率, 训练过程中的参数批处理大小设置为 16, 优化函数采用 SGD, 模型初始学习速率设置为 0.01。对于超参数文件的选择, 由于本文训练的是一个小型的自定义数据集, 为了避免数据扰动, 使用数据增强较少的配置文件。

3.2 评价指标

本文使用召回率、mAP、FLOPs (计算量)、Params (参数量) 和 Box_loss (边界框损失) 评价模型。其中, 精准率 P 表

示所有预测为正样本的结果中预测正确的比率;召回率 R 表示所有预测为正样本的结果中为正样本的概率。具体计算公式如式(5)、式(6)所示。

$$P = \frac{TP}{(TP + FP)} \quad (5)$$

$$R = \frac{TP}{(TP + FN)} \quad (6)$$

式中: TP 为模型预测为正样本且与真实标注匹配的检测框数量; FP 为模型预测为正样本且与真实标注不匹配的检测框数量; FN 为模型未检测到的真实标注数量。

通过 P 、 R 可计算 mAP , 即单个类别的平均准确率。其中, $mAP50$ 表示在 IOU 阈值为 0.5 的情况下, 各个类别预测的平均准确率; $mAP50-95$ 表示在区间 $[0.5, 0.95]$ 内, 以 0.05 为步长取 10 个 IOU 阈值, 分别计算这 10 个 IOU 阈值下的 mAP 再取平均值。FLOPs 表示处理一张图像所需要的浮点运算次数, 用来衡量算法复杂度, 参数量主要表示模型的大小, 单位通常为 MB。Box_loss 用来衡量模型预测的边界框与真实边界框之间的差异, 较小的 Box_loss 值表示模型边界框预测更准确。

3.3 比较实验

为了评估所提算法相较于现有算法的性能, 将改进后的 YOLOv8 算法与 YOLOv5、YOLOv10n^[20]、YOLOv8n、YOLOv7-Tiny、Faster-RCNN、SSD^[21]、RetinaNet^[22] 进行比较, 评估指标采用 $mAP50$ 、召回率、精确率、回归损失及分类损失, 实验结果如表 1 所示。

Table 1 Experiment results of each model

表 1 各模型实验结果

方法	mAP50	mAP50-95	Recall	FLOPs/GB	Params/MB
SSD	0.841	0.548	0.662	98.629	36.038
RetinaNet	0.893	0.664	0.732	259	56.961
Faster-RCNN	0.892	0.664	0.736	175	41.753
YOLOv5n	0.836	0.643	0.763	7.8	2.655
YOLOv7-Tiny	0.82	0.609	0.741	13.7	6.167
YOLOv8n	0.834	0.659	0.754	8.3	3.022
YOLOv10n	0.826	0.643	0.804	8.3	2.591
YOLOv8n 优化	0.865	0.689	0.822	7.7	2.804

注: 加粗数据为最优值

由表 1 可知, 传统 YOLOv8 算法在计算量和参数量远小于 SSD、RetinaNet、Faster-RCNN, 且平均准确率和召回率与上述算法差距不大; 优化后的 YOLOv8 算法在进一步减少参数量和计算量的同时提升了算法平均准确度。同时, 优化后的 YOLOv8 在 $mAP50-95$ 指标上略高于 RetinalNet、Faster-RCNN 算法, 但前者的 $mAP50$ 略低于后者, 原因可能为 RetinalNet、Faster-RCNN 网络模型的大小及复杂度较高、目标位置参数较多, 因此能够精确定位, 对不同数据集具有一定的普适性, 从而在较高的 IOU 阈值下可保持较高的准确率。此外, 相较于其他版本的 YOLO 算法, 优化后的算法将计算量、参数量压缩至与 YOLOv5 相近时, 准确率还有所提升, 原因可能为先前版本算法并未充分利

用特征融合参数。YOLOv8n 优化相较于最新版本的 YOLOv10 而言除了参数量略高之外, 其余指标均最优。

为了了解损失函数对模型的优化效果, 将改进后的损失函数 ECIOU 与 EIOU、CIOU、SIOU、GIOU、WIOU、DIOU 进行比较, 评估指标采用 $mAP50$ 、 $mAP50-95$ 及边界框损失, 实验结果如表 2 所示。由此可知, 优化后的损失函数在精度和目标框的预测效果方面均为最优。

Table 2 Comparison test results of loss function

表 2 损失函数比较实验结果

方法	mAP50	mAP50-95	Box_loss
EIOU	0.842	0.644	0.708
CIOU	0.835	0.649	0.801
SIOU	0.842	0.653	0.683
GIOU	0.852	0.655	0.681
WIOU	0.848	0.665	0.676
DIOU	0.837	0.647	0.675
ECIOU 优化	0.853	0.686	0.371

注: 加粗数据为最优值

3.4 消融实验

为了评估所提优化算法相较于原有算法的优势, 在消融试验中对每种优化方法单独进行测试, 涉及 CFNet 模块、特征融合添加融合路线、回归损失函数, 实验结果如表 3 所示。其中, YOLOv8n-C 表示仅在原模型的骨干网络增加 CFNet 模块的优化算法; YOLOv8n-N 表示仅在原模型增加跳跃连接后的优化算法; YOLOv8n-L 表示仅优化边界框的损失函数的模型。

Table 3 Results of ablation experiments

表 3 消融实验结果

方法	CFNet	Neck	Loss	mAP50	mAP50-95	FLOPs/GB	Params/MB
YOLOv8n	×	×	×	0.834	0.649	8.3	3.022
YOLOv8n-C	√	×	×	0.855	0.663	7.7	2.788
YOLOv8n-N	×	√	×	0.844	0.659	8.3	3.038
YOLOv8n-L	×	×	√	0.853	0.658	8.3	3.022
YOLOv8n 优化	√	√	√	0.865	0.689	7.7	2.804

注: 加粗数据为最优值

由表 3 可知, YOLOv8n-C 相较于原模型在 $mAP50$ 、 $mAP50-95$ 的指标上均提升 2%, 计算量和参数量更少, 证明了 CFNet 模块在特征融合时利用了子主干和特征块, 能通过更多参数生成多尺度特征且相较于原 C2f 模块更轻量。YOLOv8n-N 通过增加额外连接生成多尺度特征参数, 以提升模型特征融合能力, 只以少量参数的代价在 $mAP50$ 、 $mAP50-95$ 上均提升了 1%。YOLOv8n-L 既提升了模型预测目标位置的精确度, 又提升了 IOU 在高阈值下的准确度。优化后的 YOLOv8n 相较于原有模型参数量、计算量分别下降 7.21%、7.23%, $mAP50$ 、 $mAP50-95$ 分别提升 3.72%、4.55%。

3.5 结果分析

为进一步分析实验结果, 在数据集中将优化的 YO-

LOv8n、YOLOv8n、YOLOv5n 和 SSD 算法进行比较, 具体如图 6 所示。从左至右依次为 YOLOv8n 优化图像、YOLOv8n 图像、YOLOv5n 图像、SSD 图像。由此可见, 对于第一列图中发髻、头冠、长寿瓶等小目标, 后 3 种算法均无法识别, 而优化的 YOLOv8n 能准确识别; 对于第二列图像(手绘的唐卡图像)而言, SSD 将马头金刚佛像错检为金刚善护法,

而 YOLOv8n、YOLOv5n 均能准确识别; 第三列图像存在不完整的佛像元素, SSD 无法识别, 而 YOLO 系列模型均能准确识别; 针对第四列图像中的三股叉元素, 优化后的检测框相较于原算法定位更准确, 主要原因为改进后的损失函数更优秀。综上, 本文算法在唐卡数据集上的检测效果明显优于其他几种算法。

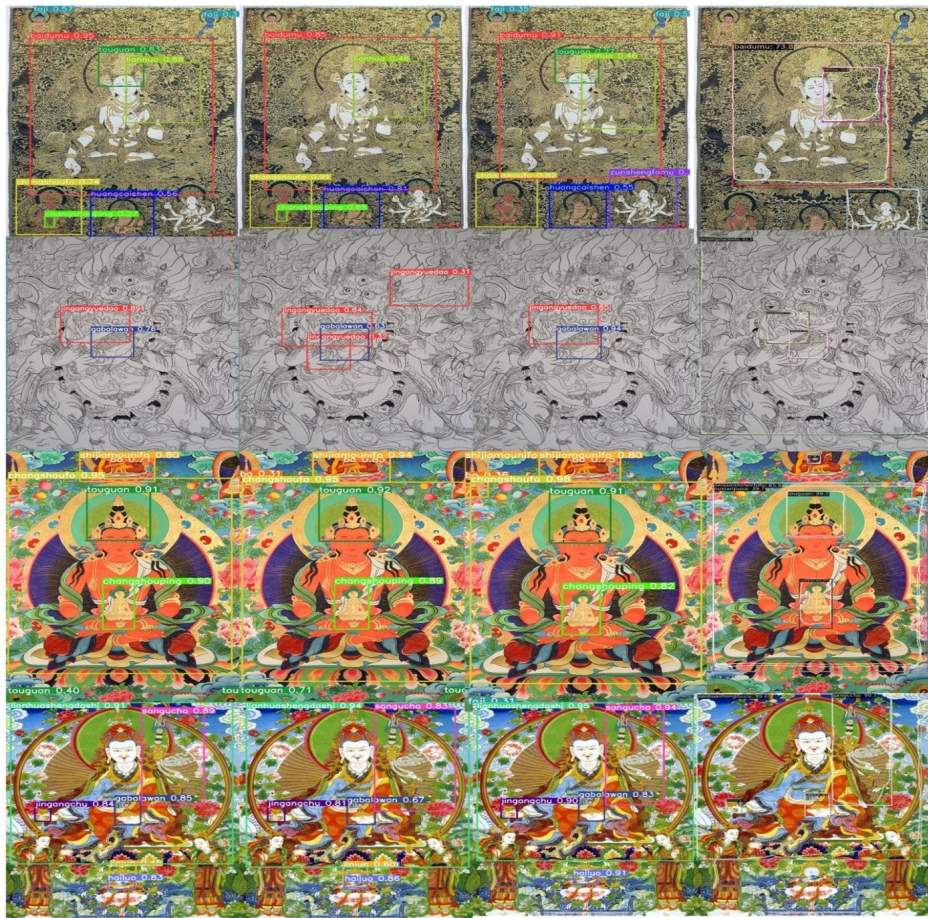


Fig. 6 Experimental results of Thangka image

图 6 唐卡图像检测结果

4 结语

本文针对唐卡中元素绘制复杂且存在的多尺度情况, 提出一种基于 YOLOv8n 的改进算法提升目标检测技术的准确度。首先引入 CFNet 模块并借鉴 BiFPN 思想, 在原有模型中增加特征融合参数比例及跳跃连接来增强多尺度特征融合能力, 能在一定程度上提升小目标检测准确率; 其次在损失函数中引入高宽等影响因素来提升模型检测框的精度, 并在注重提取有效信息的同时提升了原有 YOLOv8 算法识别复杂图像的能力。

实验表明, 所提算法在唐卡图像的目标检测领域相较于原算法准确度更高, 各优化部分对提升模型精度具有积极作用。然而, YOLO 系列作为目标检测的代表算法还有很多问题需要解决, 例如根据数据集的长尾分布特点来针

对性的提升模型精度; 目前已收集的唐卡图片数量较少。为此, 在后续研究中将继续优化相关算法, 并与文化领域相结合, 以推进非物质文化遗产的保护进程。

参考文献:

- [1] QIU Y Z. Analysis of the digital innovation protection and inheritance strategy of Thangka[J]. Exhibition Economy, 2023, 73(3):38-40. 邱意之. 唐卡数字化创新保护传承策略探析[J]. 商展经济, 2023, 73(3):38-40.
- [2] ZHAO Y H. Research and application of automatic detection of Thangka elements based on deep learning[D]. Xining: Qinghai Normal University, 2022. 赵元豪. 基于深度学习的唐卡元素自动检测研究与应用[D]. 西宁: 青海师范大学, 2022.
- [3] XUE P P. Research on the classification of Thangka masters and small sample target detection algorithm based on deep learning [D]. Lanzhou: Northwest University for Nationalities, 2022.

- 薛盼盼. 基于深度学习的唐卡主尊分类及小样本目标检测算法研究[D]. 兰州:西北民族大学, 2022.
- [4] ZHANG X J, ZHAO Y H, ZHAO Y. Thangka element automatic detection algorithm based on circular smoothing YOLOv5-Ghost[J]. Journal of Shanxi University (Natural Science Edition), 2023, 46(2):342-351.
张效娟, 赵元豪, 赵洋. 基于环形平滑 YOLOv5-Ghost 的唐卡元素自动检测算法[J]. 山西大学学报(自然科学版), 2023, 46(2):342-351.
- [5] HE G Y. Research on Thangka image target detection algorithm based on deep learning [D]. Lanzhou: Northwest University for Nationalities, 2022.
何国源. 基于深度学习的唐卡图像目标检测算法研究[D]. 兰州:西北民族大学, 2022.
- [6] GIRSHICK R. Fast R-CNN [C]// Proceedings of the IEEE International Conference on Computer Vision, 2015: 1440-1448.
- [7] XIE X, CHENG G, WANG J, et al. Oriented R-CNN for object detection [C]// Proceedings of the IEEE International Conference on Computer Vision, 2021: 3520-3529.
- [8] REN S, HE K, GIRSHICK R, et al. Faster R-CNN: towards real-time object detection with region proposal networks [DB/OL]. <https://arxiv.org/abs/1506.01497>.
- [9] JIANG P, ERGU D, LIU F, et al. A review of YOLO algorithm developments [J]. Procedia Computer Science, 2022, 199: 1066-1073.
- [10] TERVEN J, CORDOVA-ESPARZA D. A comprehensive review of YOLO: from YOLOv1 to YOLOv8 and beyond [DB/OL]. <https://arxiv.org/pdf/2304.00501v1>.
- [11] ZHU X, LYU S, WANG X, et al. TPH-YOLOv5: improved YOLOv5 based on transformer prediction head for object detection on drone-captured scenarios [C]// Proceedings of the IEEE International Conference on Computer Vision, 2021: 2778-2788.
- [12] ZHANG Z, LU X, CAO G, et al. ViT-YOLO: transformer-based YOLO for object detection [C]// Proceedings of the IEEE International Conference on Computer Vision, 2021: 2799-2808.
- [13] WANG C Y, BOCHKOVSKIY A, LIAO H Y M. YOLOv7: trainable bag-of-freebies sets new state-of-the-art for real-time object detectors [C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2023: 7464-7475.
- [14] LI X, WANG W, WU L, et al. Generalized focal loss: learning qualified and distributed bounding boxes for dense object detection [J]. Advances in Neural Information Processing Systems, 2020, 33: 21002-21012.
- [15] WANG X, GAO H, JIA Z, et al. BL-YOLOv8: an improved road defect detection model based on YOLOv8 [J]. Sensors, 2023, 23(20): 8361.
- [16] ZHANG G, LI Z, LI J, et al. CFNet: cascade fusion network for dense prediction [DB/OL]. <https://arxiv.org/abs/2302.06052v1>.
- [17] TAN M, PANG R, LE Q V. Efficientdet: scalable and efficient object detection [C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2020: 10781-10790.
- [18] ZHENG Z, WANG P, LIU W, et al. Distance-IoU loss: faster and better learning for bounding box regression [C]// Proceedings of the AAAI Conference on Artificial Intelligence, 2020: 12993-13000.
- [19] ZHANG Y F, REN W, ZHANG Z, et al. Focal and efficient IOU loss for accurate bounding box regression [J]. Neurocomputing, 2022, 506: 146-157.
- [20] WANG A, CHEN H, LIU L, et al. YOLOv10: real-time end-to-end object detection [DB/OL]. <https://arxiv.org/abs/2405.14458>.
- [21] LIU W, ANGUELOV D, ERHAN D, et al. SSD: single shot multibox detector [C]// 14th European Conference on Computer Vision, 2016: 21-37.
- [22] WANG Y, WANG C, ZHANG H, et al. Automatic ship detection based on RetinaNet using multi-resolution Gaofen-3 imagery [J]. Remote Sensing, 2019, 11(5): 531.

(责任编辑:刘嘉文)