基于改进 YOLOv8 的小目标检测方法研究

韦 伟[1] 王翔翔[1] 陶亚明[1]
WEI Wei WANG Xiangxiang TAO Yaming

摘 要 针对 YOLOv8 算法在小目标检测上效果不佳、难度大的问题，提出一种改进 YOLOv8 网络结构的小目标识别方法。为解决小目标检测中由于检测点感受野不足而导致的问题，在骨干网络中引入了可变形卷积模块。通过调整卷积核的形变，使其能够更好地适应小目标的特征，从而提高检测的准确性，减少漏检和误检情况；借鉴 Wise-IoU 的思想，提出一种名为 HIoUd1 全新的损失函数计算方法，随时调整不同部分在损失函数中的权重，以在训练的不同阶段更好地处理小目标。通过这种调整，模型能够更准确地检测小目标并将其回归到真实标注框，以此显著提升了对小目标的检测性能。实验结果表明，优化后的模型能够更有效地捕捉和识别小目标，相对于 YOLOv8 在小目标检测方面表现出了显著的提升。

关键词 深度学习；小目标检测；YOLOv8 算法

doi：10.3969/j.issn.1672-9528.2024.10.021

0 引言

随着 YOLO 系列算法的不断演进，越来越多的目标检测方法采用了无锚框的检测方式。相较于基于锚框的检测方法，无锚框目标检测算法在泛化性能和计算效率上表现更为出色，其通过直接预测目标的位置和大小，避免了传统锚框方法中需要大量的候选框生成和调整的步骤，简化了整个检测流程，减少了冗余计算，提升了检测的准确性和速度 [1]。但小目标往往具有较低的信噪比和少量的可见特征，容易被其他背景干扰或者遮挡，因此很难准确地被检测和定位。虽然无锚框目标检测算法在许多方面已经取得了显著的进展，但对于小目标的检测仍然存在一定的挑战 [2]。

为解决小目标检测难度大的问题，本文对 YOLOv8 的损失函数计算方法和骨干网络进行针对优化。为了改进 YOLOv8 的性能，对其骨干网络添加可变形卷积模块来解决由于感受野不足而导致的漏检或误检问题。提出名为 HIoU 的一种全新 IoU 损失函数计算方法，来随时调整不同部分在损失函数中的权重，以在训练的不同阶段更好地处理小目标 [3]。最后通过实验验证，优化后的模型相对于原模型对小目标的检测表现有提高。

1 YOLOv8 算法

1.1 YOLOv8 网络结构

YOLOv8 的网络结构主要由以下三个大部分组成。

（1）Backbone：它采用了一系列卷积和反卷积层来提取特征，同时也使用了残差连接和瓶颈结构来减小网络的大小和提高性能。该部分采用了 C2f 模块作为基本构成单元。与 YOLOv5 的 C3 模块相比，C2f 模块具有更少的参数量和更优秀的特征提取能力 [4]。

（2）Neck：它采用了多尺度特征融合技术，对来自 Backbone 的不同阶段的特征图进行融合，以增强特征表示能力。具体来说，YOLOv8 的 Neck 部分包括一个 SPPF 模块、一个 PAA 模块和两个 PAN 模块。

（3）Head：它负责最终的目标检测和分类任务，包括一个检测头和一个分类头。检测头包含一系列卷积层和反卷积层，用于生成检测结果；分类头则采用全局平均池化来对每个特征图进行分类。

1.2 YOLOv8 算法原理

（1）数据预处理：主要采用翻转、旋转、缩放、裁剪、高斯噪声、亮度调节等常规方式，以及基于图像混叠的 Mosaic 增强、Mixup 增强的数据增强方式和空间扰动、颜色扰动两个增强手段 [5]。

（2）骨干网络结构：通过使用步长为 2 的 3*3 卷积层进行降采样，有效地减少特征图的尺寸，同时保留重要的特征信息。C2f 模块则起到了进一步优化特征表达的作用，该模块的核心参数为“3/6/9/3”，这意味着它包含了不同深度的卷积核来捕捉不同尺度的特征，使得模型更好地适应不同大小和形状的目标，提高了模型的泛化能力。

（3）FPN-PAN 结构。采用 FPN+PAN 结构来构建 YOLO 的特征金字塔，使多尺度信息之间进行充分的融合。

1. 安徽工业大学管理科学与工程学院 安徽马鞍山 243032

（4）Detection head 结构。YOLOv8 引入了解耦头的结构，通过将目标检测任务拆分为类别识别和位置定位两个并行的任务，使得模型在学习类别信息和位置信息时更加独立和有效，分别使用一层 1*1 卷积层降低特征图的维度，保留重要特征信息的同时完成分类和定位的任务 [6]。

2 YOLOv8 算法改进

2.1 基于可变形卷积的目标检测改进

可变形卷积是一种卷积操作，其在输入特征图的每个像素点上学习并应用一组偏移量（通常是 x 和 y 方向的偏移）。这意味着可变形卷积在每个位置上都可以自适应地调整其感受野，以更好地适应目标的形状和位置变化。通过学习这些偏移量，网络可以有效地捕获目标的空间变化和形状变化，从而提高了模型对目标的感知能力。

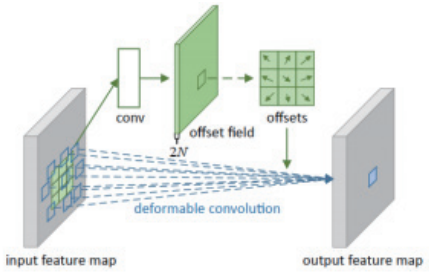

图 1 DCN 模块结构

通过引入可变形卷积到 YOLO 中可以为目标检测任务带来显著的改进。传统的卷积操作在处理规则矩形感受野时效果良好，但在处理不规则形状的目标时表现不佳。而可变形卷积通过学习每个像素点的偏移量，使得卷积操作可以自适应地调整感受野，更好地适应目标的形状和位置变化。此外，可变形卷积还能够拓展模型的感受野，使其能够有效地捕获更广泛的上下文信息。通过在卷积操作中引入可变形卷积，YOLO 可以在不增加额外计算成本的情况下，提高对目标的感知能力，进而提升整体检测性能。

在 YOLOv8 中，骨干网络负责从输入图像中提取不同层次的特征。这些特征提取有助于人们更深入地理解和描述图像。通过结合这些不同层次的特征，可以获得更全面、更准确的图像特征表示，从而提升模型的性能和泛化能力。在 YOLOv8 的骨干网络中引入可变形卷积模块，这样做可以增强图像特征的提取能力，使得模型能够捕捉更丰富的细节信息。

2.2 基于改进 IoU 的目标检测

在目标检测任务中，交并比（intersection over union，IoU）是一项至关重要的指标，用于评估模型的预测结果与真实标注之间的重叠程度。其计算方法是通过计算模型预测框和真实标注框的交集面积与并集面积之比来衡量两者的相似程度。IoU 的取值范围在 0 ～ 1 之间，其中 1 表示完全重合，0 表示没有重叠。通常情况下，如果模型的预测框与真实框之间的 IoU 值大于预先设定的阈值（如 0.5 或 0.75），则可以认为该预测框与真实框匹配成功，从而被视为一个正确的检测结果。

记锚框为 B=[x y w h]，目标框为 B_{gt}=[x_{gt} y_{gt} w_{gt} h_{gt}]；DIoU（distance-IoU）将惩罚项定义为中心点连接的归一化长度：

$$R=\frac{(x-x_{gt})^2+(y-y_{gt})^2}{W_g^2+H_g^2} \tag{1}$$

Wise-IoU 采用动态方法计算类别预测损失中的 IoU 损失，定义为：

$$L_{WIoU}=R_{WIoU}L_{IoU} \tag{2}$$

$$R=\exp\left(\frac{(x-x_{gt})^2+(y-y_{gt})^2}{(W_g^2+H_g^2)^*}\right) \tag{3}$$

由于在训练数据中不可避免地存在低质量示例，传统的几何度量可能在评估目标检测模型时施加较大的惩罚，导致模型的泛化能力下降。一个良好的损失函数应该在模型的预测框与目标框重叠良好时，减少对几何度量的惩罚，从而避免模型在训练数据上过度拟合。

通过 WIoU 的思想，提出一种名为 HIoU 的新的 IoU 损失函数计算准则，定义为：

$$L_{HIoU}=R_{WIoU}\times L_{IoU}+\frac{1}{2}\left(\frac{(x-x_{gt})^2+(y-y_{gt})^2}{W_g^2+H_g^2}+\alpha v\right) \tag{4}$$

在训练初期，模型通常面临着预测候选框与真实物体标注框之间的较低 IoU 的情况。在这种情况下，模型的优先重点应该是提高预测候选框与真实物体标注框之间的 IoU，以便更好地匹配目标。当训练进入后期时，预测候选框与真实物体标注框之间的 IoU 会因为模型在训练过程中逐渐学习到了目标的特征和位置，从而更准确地预测出目标框而保持在较高的水平，并且不会有太大的变化。在这种情况下，新的损失函数中的 IoU 惩罚项的位置调整项（L_{IoU}）可以逐渐减小，甚至衰减到一个非常小的值 [7]。此时，模型会自动将重点放在中心点和横纵比的回归上，通过进一步优化中心点和横纵比，以进一步提高预测候选框与真实物体标注框的匹配度，从而改善模型的性能。

HIoU 是一种新颖的方法，旨在改进传统 IoU 的计算方式，以更全面地考虑预测框和真实框之间的多个因素，从而提高目标检测的准确性。传统的 IoU 主要关注预测框与真实框之间的重叠度，但它忽略了其他重要因素。HIoU 通过在调整边界框回归损失的同时降低对距离和纵横比等几何度量的惩罚水平，来改善传统 IoU 的不足。它通过减半了损失函数中的第二和第三项，使模型对于这些几何度量的优化更为温和。这种温和的方式有助于模型在训练过程中更平稳地学习，同时提高了其在未见过的数据上的泛化能力。

3 实验与分析

3.1 数据集

本文选用的数据集是视觉领域中常用的数据集之一。数据集中涵盖了18个常见物体类别，如人脸、汽车、交通标志等。数据集总共有2200张图像，其中有700张图像被用于测试，剩下的1500张图像则被用于训练和验证。每张图像都配有物体的注释信息，这些注释信息为模型训练和评估提供了标准化的参考，保证了对模型泛化能力的客观评估。

3.2 实验环境

本文的实验环境配置如下：Intel(R)Core(TM)i7-8750H CPU@2.20 GHz 2.21 GHz处理器，CPU（NVIDIA GEFORCE GTX 1050 Ti）。对改进的算法模型进行训练，生成的P-R曲线对比如图2所示。曲线下的面积，即平均精度，反映了模型的分类或检测能力。而mAP则是通过计算所有类别的平均精度得出的均值平均精度[8]。P-R曲线下的面积越大，代表模型在不同类别上的检测效果越好，从而对应着更高的mAP值。

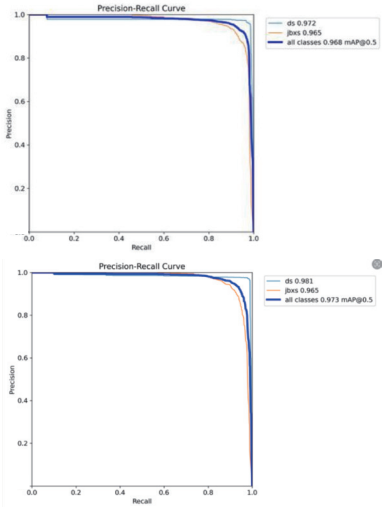

图2 优化后的模型与原模型的对比结果图

实验结果表明，改进后的YOLOv8模型在每一类目标检测效果都表现较好，即其在整体上的表现也是最好的。调用训练生成的权重文件对图片进行测试，结果如图3所示。

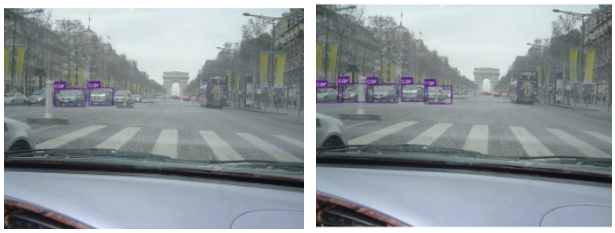

图3 测试结果对比

3.3 算法测试性能比对

通过使用同一训练集进行对比实验，分别使用YOLOv5算法、YOLOv8算法和改进后的YOLOv8三种网络，每次实验都使用了各自训练过程中得到的最终权重文件，对同一张图片进行目标检测。检测结果如表1所示。

表1 不同网络结构测试性能对比

	YOLOv5	YOLOv8	改进后的YOLOv8
recall	0.896	0.907	0.967
precision	0.924	0.937	0.985
mAP_0.5	0.924	0.928	0.992
mAP_0.5:0.95	0.794	0.857	0.977
模型大小	46	62	126
FPS	79.6	91.2	52.9

由表1可知，改进后的YOLOv8相较改进前，召回率、精确率和mAP都有显著提升，提高了模型的检测表现。

4 结论

本文提出一种基于改进YOLOv8的小目标检测方法。在骨干网络中引入可变形卷积模块来改善特征图上检测点的感受野，并且为了网络能够在不同阶段可以随时调整不同部分在损失函数中的权重来设计IoU准则。通过实验表明，优化后的YOLOv8，mAP提高了2.4%，在一定程度上提高了小目标检测的准确性，并在训练及推理速度不变的条件下，增强了对小目标的检测效果。

参考文献：

[1] 常飞，王奔，张小旭，等．基于改进YOLOv5的小目标检测方法研究[J]. 智慧轨道交通，2024,61(2):7-13.
[2] 陈成琳，鲍春，曹杰，等．基于改进YOLOv3的遥感小目标检测网络[J]. 计算机仿真，2023,40(8):30-35.
[3] 王莹．基于深度学习的小目标检测方法研究[D]. 南京：南京信息工程大学，2023.
[4] 刘易宸．基于改进YOLOv8的输电线路多目标检测算法研究[D]. 阜新：辽宁工程技术大学，2023.
[5] 王艺成，张国良，张自杰．基于改进YOLOv5的小目标检测方法[J]. 计算机与现代化，2023(5):100-105.
[6] 韩强．面向小目标检测的改进YOLOv8算法研究[D]. 长春：吉林大学，2023.
[7] 王继千，刘唤唤，廖涛，等．基于改进YOLOv3的目标检测方法研究[J]. 现代信息科技，2022,6(16):71-74.
[8] 解尧婷．基于深度学习的输电线路小目标识别算法研究[D]. 太原：中北大学，2021.

【作者简介】

韦伟（1970—），男，安徽马鞍山人，本科，副教授，硕士生导师，研究方向：人工智能、设备管理信息化建设。

王翔翔（2000—），男，安徽宿州人，硕士研究生，研究方向：质量管理与工程。

（收稿日期：2024-07-16）