

基于改进 YOLOv8 的细长物体检测方法

李佳兴, 文峰

(沈阳理工大学 信息科学与工程学院, 沈阳 110159)

摘要: 在计算机视觉领域, 无锚框的检测算法能够较好地检测并定位任意几何形状的物体, 有利于检测细长物体。从更好保留细长物体信息、防止信息的混淆以及更好表达非局部信息之间关联的角度出发, 提出改进 YOLOv8 的细长物体检测方法。在 YOLOv8 的骨干和头部网络中加入有助于处理细长物体的 SPD-Conv 构建块, SPD-Conv 充分利用空间分割与非跨步卷积处理技术, 有效减少细长物体特征的丢失; 改进 YOLOv8 骨干网络中的卷积运算, 解决参数共享的问题, 对图像不同位置 and 不同通道的特征赋予不同的重要性和含义; 为更好标定细长目标, 使用非局部 non_local 注意力机制理解图像中的长距离依赖关系和相关性。实验结果表明: 与 YOLOv8 原型相比, 改进算法平均精度提升了 11.35%, 推理速度略有降低, 为每秒 20 帧, 基本达到了实时性要求。

关键词: YOLOv8; 目标检测; 注意力机制; 细长物体识别

中图分类号: TP391.41 **文献标志码:** A **DOI:** 10.3969/j.issn.1003-1251.2024.04.003

Method of Detecting Slender Object Based on Improved YOLOv8

LI Jiaying, WEN Feng

(Shenyang Ligong University, Shenyang 110159, China)

Abstract: In the field of computer vision, the detection algorithm without anchor frame can be used for detecting slender objects and better locate objects with any geometric shape. From the perspective of better retaining the information of elongated objects, preventing the confusion of information and better expressing the correlation between non-local information, an improved elongated object detection method of YOLOv8 is proposed. In the backbone and head network of YOLOv8, SPD-Conv building blocks are added to help process elongated objects, and SPD-Conv fully uses spatial segmentation and non-step-step convolution processing techniques to effectively reduce the loss of features of elongated objects. The convolution operation in the YOLOv8 backbone network is improved, the problem of parameter sharing is solved, and different importance and meaning are attached to the features of different positions and channels of the image. In order to better calibrate slender targets, attention mechanism non_local is used to learn long-distance dependencies and correlations in images. Experimental results show that the average accuracy of the improved algorithm is increased by 11.35% compared with the YOLOv8 prototype, and although the inference speed is slightly reduced, the inference speed of 20 frame per second is maintained, which basically meets the real-time requirements.

Key words: YOLOv8; object detection; attention mechanism; slender object recognition

在计算机视觉领域,目标检测算法通过对输入图像或视频进行扫描,识别出图像或视频中感兴趣的目标,给出目标的坐标、大小和类别等信息。常见的一阶段目标检测算法有 YOLO 系列^[1-3]、SSD^[4-5]、RCNN、RetinaNet^[6-7]、FCOS^[8-10]算法等,一阶段算法通常在处理速度上具有一定的优势。常见的二阶段目标检测算法有 Faster R-CNN^[11-12]、Mask R-CNN^[13]、SPP-Net^[14]等,二阶段算法的优势是具有更高的检测精度,对于复杂背景和多目标情景具有较好的鲁棒性,但检测速度较慢。

上述算法大多为基于锚框的目标检测算法,由于锚框通常根据特定的尺度和长宽比设置,不适用于处理尺度变化较大的物体。当目标物体的尺度与锚框的尺度不匹配时,锚框的回归和分类结果可能会受到影响,从而导致对物体的检测结果不准确。

细长物体主要指图像中呈现细长形状的目标,如牙刷、勺子、棒球棒等,在日常生活中随处可见,但由于细长物体具有长宽比大、低对比度、容易与背景混合等特点,在目标检测领域检测细长物体仍然具有一定挑战性。文献[9]对 FCOS 算法进行了改进,提升了检测精度。文献[15]设计了一个通用的特征自适应模块,用以改进对细长目标的检测。尽管对该方面的研究取得了一定进展,依然难以达到满意的检测效果。

YOLOv8 是基于无锚框的目标检测算法,无锚框方法不需要生成大量的锚框,能够有效降低计算复杂度,并且提升检测速度和效率。为有效提高细长物体的检测精度,增强提取细长物体重要特征的能力,减少高级语义信息的丢失,本文提出改进 YOLOv8 的细长物体检测方法,采用网格化的方式预测整体图像用以代替传统检测方法的候选框选取操作,具有较快的检测速度,能够满足实时性需求。本文改进内容包括如下几点:

1) 设计了细长目标图像细节保留机制,通过使用 SPD-Conv^[16] 构建块,利用切片技术以及非跨步卷积操作能够保留细长物体关键的像素信息,解决了跨步卷积操作导致跳过大量像素信息的问题;

2) 采用新型卷积方法^[17],通过池化、普通卷积以及全连接操作的结合计算动态权重,并对特征图进行加权操作以提升对细长物体的检测精度,解决利用卷积核进行卷积操作时的参数共享问题;

3) 使用非局部(non_local)注意力机制^[18],可以实现对图像中长距离依赖的建模,而无需受限于两个相邻位置之间的交互情况,有效地提高模型性能。

1 YOLOv8 检测算法

YOLOv8 的核心思想是将目标检测任务视为一个端到端的回归问题,通过一个卷积神经网络同时预测目标的类别和位置。相比于传统的锚框方法,YOLOv8 采用无锚框的方式进行目标检测,适应各种形状和尺寸的物体。本文将 YOLOv8 算法进行改进以期获得良好的检测效果。

YOLOv8 的网络结构如图 1 所示。图中 Backbone 表示 YOLOv8 网络的骨干部分,Head 表示网络的头部,Conv、C2f、Concat、Upsample、SPPF、Detect 是 YOLOv8 网络的主要组成模块。Conv 是卷积模块,该模块用于提取图像信息;C2f 模块更好地捕捉到图像中的复杂特征,有助于在目标检测任务中取得更好的效果;Concat 是拼接模块,用于将两组特征图在通道维度上进行拼接;Upsample 是上采样模块,用于将特征图的长宽成倍地放大;SPPF 模块用于自适应融合不同尺度的

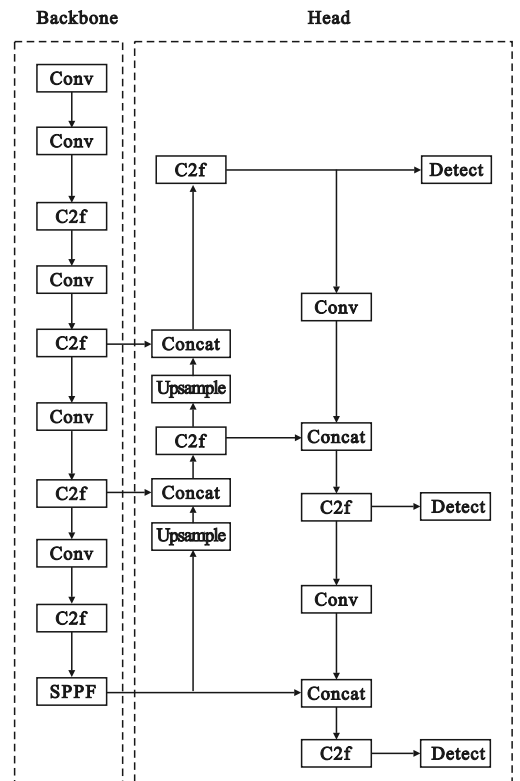


图 1 YOLOv8 网络结构

Fig. 1 YOLOv8 network structure

信息;Detect 模块用于输出检测结果。这些模块内部主要由卷积、池化等方法组成。

2 改进 YOLOv8 检测算法

对 YOLOv8 改进后的结构如图 2 所示。改进算法采用了 SPD-Conv 模块、新型卷积以及注意力机制 non_local,通过参数对齐的方法将上述模块无缝嵌入到网络结构中并充分发挥作用。图 2 中 Conv、SPD-Conv、新型卷积、non_local 注意力机制、C2f、Concat、Upsample、SPPF、Detect 是改进 YOLOv8 网络的主要组成模块。SPD-Conv 模块的最终效果是对输入的特征图进行 2 倍下采样;与普通的 Conv 操作相比,新型卷积能够更好地保留图像信息,避免由卷积核参数共享所带来的信息混淆问题;non_local 模块能够更好地关注图像非相邻点之间的关联,从而提升检测精度。

2.1 SPD-Conv 构建块

卷积神经网络(CNN)使用步长大于 1 的卷积和池化层,会导致关键特征的丢失且难以对非常规特征进行学习。为此,采用一种新的 CNN 构建块 SPD-Conv 取代每个步长大于 1 的卷积层和每个池化层。SPD-Conv 是一种数据处理技术,用于将空间分割与非跨步卷积技术相结合,能够更好地支持多种 CNN 架构。SPD 层对特征图进行下采样,但保留通道维度中的所有信息,因此不存在信息丢失。图 3 展示了 SPD 卷积操作的过程,图中假设下采样倍率为 2,对于长宽都为 S 、通道数为 C_1 的特征图,SPD-Conv 会通过切片操作将

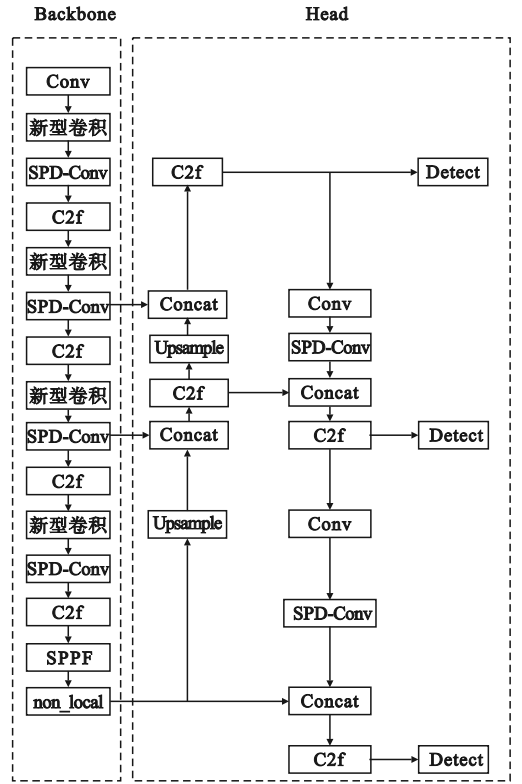


图 2 改进的 YOLOv8 网络结构

Fig. 2 Improved YOLOv8 network structure

输入的特征图切成四个子特征图,每个子特征图的长宽都为 $S/2$,通道数依然为 C_1 ,然后再将四个子特征图在通道维度上进行拼接,得到通道数为 $4 \times C_1$ 的特征图,对其进行步长为 1 的普通卷积操作改变通道数,最终得到的特征图长宽均变为 $S/2$,通道数变为 C_2 ,从而达到了和下采样倍率为 2 的普通卷积一样的效果,并保留了图像的一些细节信息。

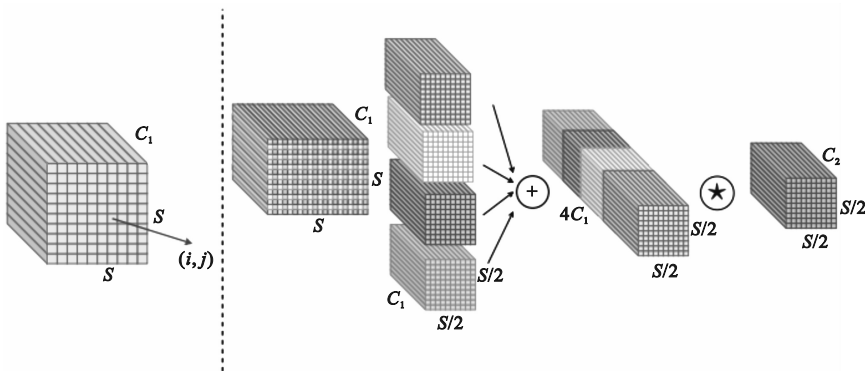


图 3 SPD-Conv 操作

Fig. 3 SPD-Conv operations

2.2 新型卷积方法

CNN 存在的另一个主要问题是参数共享。参数共享是指卷积核在提取特征时使用相同的参

数值,相较于全连接操作可以节省大量的参数和计算量,但参数共享会造成信息的丢失与混淆,因为不同位置和不同通道的特征可能有不同的重要

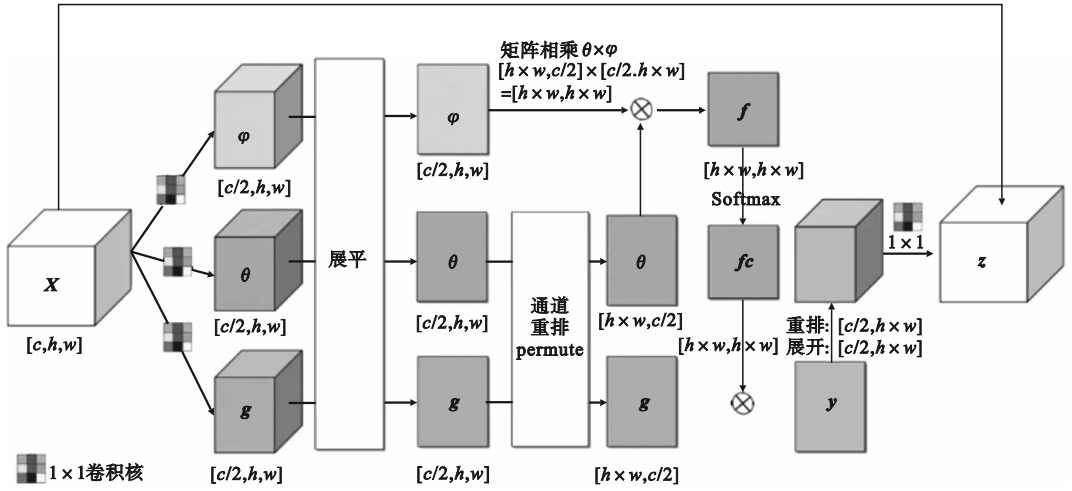


图 6 non_local 注意力机制计算流程图

Fig. 6 non_local attention mechanism calculation diagram

3 实验与分析

3.1 数据集与评估指标

本文使用从 COCO 数据集提取的数据,充分考虑到细长物体长宽比大的结构特点,从数据集中选取了包括书本、勺子、叉子、牙刷、棒球棒等长宽比大的细长物体。通过将数据集划分为训练集和验证集对模型进行训练和测试,其中训练集包含 25 424 张图片,验证集包含 1 077 张图片。本文采用的评估指标主要为均值平均精度 (mAP_{0.5})、准确率 (Precision) 和召回率 (Recall)。

3.2 实验环境与实验设置

实验采用 Ubuntu16.04 操作系统,在 GPU 为 NVIDIA Tesla V100 16 GB 显卡上进行模型训练和测试,使用的编程框架是 PyTorch 1.9.0,编程语言为 Python。

本文采用随机梯度下降的学习方法训练 300 轮,动量设置为 0.937,热身阶段轮数为 3 轮,权重衰减设置为 0.000 5,热身阶段初始动量为 0.8,热身阶段的初始偏置学习率为 0.1,初始学习率为 0.01,边界框损失权重为 7.5,类别损失权重为 0.5。

3.3 消融实验

本文以 YOLOv8 网络为基础在 COCO 数据集上进行消融实验,研究每个模块对模型检测精度的影响,结果如表 1 所示。通过在基础网络 YOLOv8 中分别添加新型卷积方法、SPD-Conv 以及 non_local 注意力机制,精度值 mAP_{0.5}较原型均有所上升。通过在骨干网络中将部分普通卷积替

换成新型卷积方法,在不同 IoU 阈值下的精度值均有所提升,mAP_{0.5}值提升了 1.17%,从而验证了新型卷积的有效性;添加 SPD-Conv 后,mAP_{0.5}提升到 55.8%,经过分析认为使用 SPD 卷积操作能够避免细长物体信息的大量损失,并且也通过消融实验证明了使用 SPD 卷积操作能够有效提升对细长物体的检测精度;添加 non_local 后,mAP_{0.5}提升到 52.0%。

消融实验的结果表明改进算法能够对细长物体的检测起到积极的影响,体现出改进算法的实用性。

表 1 消融实验效果

Table 1		Results of ablation experiment		%		
YOLOv8	SPD-Conv	新型卷积	non_local	Precision	Recall	mAP _{0.5}
✓				61.8	47.9	51.1
✓	✓			60.5	53.0	55.8
✓		✓		59.0	50.0	51.7
✓			✓	63.6	47.2	52.0
✓	✓	✓	✓	65.5	52.7	56.9

本文对改进的 YOLOv8 进行了比较实验,表 2 展示了改进后的 YOLOv8 检测算法与基于锚框的二阶段算法 Faster R-CNN、一阶段算法 FCOS、YOLOv8 原型算法以及基于 transformer 的 DETR 算法在 COCO 数据集中检测性能的对比结果。由表 2 可见:在细长目标数据集中,由于 Faster R-CNN 是基于锚框的算法,不适于处理尺度变化较大的物体,对细长物体的检测精度较低,其 mAP_{0.5} 值仅为 35.1%;一阶段 FCOS 算法的精度值达到了 38.6%;基于 transformer 的 DETR 算法则相较

于 FCOS 算法精度提高了 27.7%, DETR 通过 Transformer 编码器-解码器结构中的自注意力机制能够充分利用全局上下文信息,推断每个目标的位置和类别;基于无锚框的 YOLOv8 不依赖于预先定义好的锚框,能够预测出目标的真实尺度,在处理尺度变化较大的物体时更加稳健和准确,更能适应细长目标;改进后的 YOLOv8 有效提升了 mAP 值, $mAP_{0.5}$ 从 51.1% 提升到了 56.9%, 提升了 11.35%, 在细长目标的特征提取和语义信息的完整性方面给予了充分的考虑。对比实验结果表明改进后的 YOLOv8 算法相较于其他算法有效提升了检测精度,并保持了一定的推理速度,体现了该细长目标检测算法的实用性较强。

表 2 COCO 数据集中各模型检测性能

Table 2 Representation of each model in the COCO dataset

算法	$mAP_{0.5}/\%$	检测速度/ s^{-1}
Faster R-CNN	35.1	17
FCOS	38.6	22
DETR	49.3	27
YOLOv8	51.1	41
改进的 YOLOv8	56.9	20

图 7 展示了利用改进 YOLOv8 算法进行细长物体检测的效果。可见,改进模型准确地检测出叉子的类别和位置,且其置信度为 0.78。



图 7 改进的 YOLOv8 算法检测效果

Fig. 7 Detection result of improved YOLOv8 algorithm

4 结论

本文设计了一种基于改进 YOLOv8 的细长目标检测算法。通过在骨干网络中将卷积替换为新型卷积,使其能够解决参数共享的问题,提高卷积操作的效率和灵敏度,加入新型卷积后模型的精度提高了 1.17%;通过采用 SPD-Conv 卷积操作能够有效地避免细长物体目标信息的损失,进

而提高对细长物体检测的精度,通过实验得出的数据发现其精度增长了 9.20%,有效地提高了模型对细长物体的检测能力;通过采用 non_local 注意力机制使得算法能够更好地理解图像中的长距离依赖关系和相关性,提高模型的鲁棒性和泛化能力,实验表明,加入了 non_local 机制后使模型精度提升了 1.76%。本文改进算法的平均精度值提升了 11.35%,验证了本文算法对细长物体检测的有效性。

参考文献 (References):

- [1] 王梦妮,顾奇南,王化佳,等. 基于改进 YOLOv5s 模型的茶叶嫩芽识别方法[J]. 农业工程学报,2023,39(12):150-157.
WANG M N, GU J N, WANG H J, et al. Method for identifying tea buds based on improved YOLOv5s model[J]. Transactions of the Chinese Society of Agricultural Engineering, 2023,39(12):150-157. (in Chinese)
- [2] 苗荣慧,李志伟,武锦龙. 基于改进 YOLO v7 的轻量化樱桃番茄成熟度检测方法[J]. 农业机械学报,2023,54(10):225-233.
MIAO R H, LI Z W, WU J L. Lightweight maturity detection of cherry tomato based on improved YOLO v7[J]. Transactions of the Chinese Society for Agricultural Machinery, 2023,54(10):225-233. (in Chinese)
- [3] 崔海彬,浦东兵,陆云凤,等. 基于 CA-YOLO 的安全帽佩戴检测[J]. 东北师大学报(自然科学版),2023,55(3):94-100.
CUI H B, PU D B, LU Y F, et al. Helmet wearing detection based on CA-YOLO[J]. Journal of Northeast Normal University(Natural Science Edition), 2023,55(3):94-100. (in Chinese)
- [4] 曹梧汉,贺波涛. 基于改进 SSD 的交通标志检测算法[J]. 计算机与数字工程,2023,51(8):1771-1775.
CAO W H, HE B T. Traffic sign detection algorithm based on improved SSD[J]. Computer & Digital Engineering, 2023,51(8):1771-1775. (in Chinese)
- [5] 胡凯,骆润玫,刘泽乾,等. 基于改进 SSD 的广佛手病虫害检测方法[J]. 南京农业大学学报,2023,46(4):813-821.
HU K, LUO R M, LIU Z Q, et al. Detection of bergamot diseases and pests based on improved SSD[J]. Journal of Nanjing Agricultural University, 2023,46(4):813-821. (in Chinese)
- [6] GAO S, LIU H L. RetinaNet-based compact polarization SAR ship detection[J]. IEEE Journal on Miniaturization for Air and Space Systems, 2022,3(3):146-152.
- [7] MIAO T, ZENG H C, YANG W, et al. An improved lightweight RetinaNet for ship detection in SAR images[J]. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 2022,15:4667-4679.
- [8] 王梅,胡晓杰. 基于改进 FCOS 的细长物体检测算法[J]. 沈阳理工大学学报,2022,41(4):8-13,19.
WANG M, HU X J. Algorithm for slender object detection based on improved FCOS[J]. Journal of Shenyang Ligong University, 2022,41(4):8-13,19. (in Chinese)

(下转第 26 页)