

多尺度特征融合与锚框自适应的目标检测算法

张润梅^{1,2}, 毕利君¹, 汪方斌^{1,2,3}, 袁彬^{1,2*}, 罗谷安¹, 姜怀震¹

¹安徽建筑大学机械与电气工程学院, 安徽 合肥 230601;

²工程机械智能制造重点实验室, 安徽 合肥 230601;

³安徽建筑大学建筑机械故障诊断与预警重点实验室, 安徽 合肥 230601

摘要 针对 Faster R-CNN 算法存在特征提取不充分、检测框定位不准确导致检测精度不高的问题, 提出一种多尺度特征融合和锚框自适应相结合的目标检测算法。首先, 通过双向融合方法充分提取相邻层级间的深层特征和浅层特征; 然后, 均衡化处理多尺度特征, 使集成的特征能获得来自不同分辨率下等量的语义信息和细节信息, 提高目标的识别能力; 最后, 在区域提议网络(RPN)中利用目标的特征信息, 通过自适应预测锚框的位置和形状来生成锚框。基于 VOC 数据集对算法的实验结果表明: 与基于 ResNet50 的 Faster R-CNN 算法相比, 所提算法中的多尺度特征融合策略加强了算法对不同尺度目标的检测能力, 自适应锚框机制能够提高定位精度并避免小目标的漏检, 算法整体的检测结果具有较好表现, 平均检测精度提升了 3.20 个百分点。

关键词 机器视觉; 目标检测; Faster R-CNN 算法; 特征融合; 锚框自适应

中图分类号 TP391.4

文献标志码 A

DOI: 10.3788/LOP202259.1215019

Multiscale Feature Fusion and Anchor Adaptive Object Detection Algorithm

Zhang Runmei^{1,2}, Bi Lijun¹, Wang Fangbin^{1,2,3}, Yuan Bin^{1,2*}, Luo Gu'an¹, Jiang Huaizhen¹

¹School of Mechanical and Electrical Engineering, Anhui Jianzhu University, Hefei 230601, Anhui, China;

²Key Laboratory of Intelligent Manufacturing of Construction Machinery, Hefei 230601, Anhui, China;

³Key Laboratory of Construction Machinery Fault Diagnosis and Early Warning Technology, Anhui Jianzhu University, Hefei 230601, Anhui, China

Abstract Aiming at the problems of low detection accuracy resulting from insufficient feature extraction and inaccurate detection box positioning in the Faster R-CNN algorithm, an object detection algorithm based on multiscale feature fusion and anchor adaptation is proposed. First, the high- and low-level features between adjacent levels were fully extracted using the two-way fusion method; then, the multiscale features were balanced so that the integrated features could obtain the same amount of semantic information and detailed information with different resolutions, improving the object recognition ability. Finally, the anchor was generated by adaptively predicting the position and shape of the anchor using the characteristic information of the object in the region proposals network(RPN). The experimental results of the algorithm based on VOC dataset show that compared with the Faster R-CNN algorithm based on ResNet50, the multiscale feature fusion strategy in the proposed algorithm strengthens the detection ability for objects with different scales. The adaptive anchor mechanism can improve the positioning accuracy and avoid missed detection of small objects, and the overall detection results of the proposed algorithm have good performances. The proposed algorithm improves the average detection accuracy by approximately 3.20 percentage points.

Key words machine vision; object detection; Faster R-CNN algorithm; feature fusion; anchor adaptation

收稿日期: 2021-11-16; 修回日期: 2021-12-02; 录用日期: 2021-12-21

通信作者: *yuanbinwork@163.com

1 引言

目标检测作为一类基础性机器视觉技术,主要完成目标的识别和定位任务,要求计算机不仅能判断出物体类别,更要得到它在图片中的具体位置。通过计算机和视觉模块^[1-2]的结合,目标检测有效降低了对人力资源的消耗,因此被广泛应用于无人驾驶^[3]、医疗服务^[4]和军事^[5]等领域,成为近几年理论研究和应用的热点。

传统目标检测大多以图像识别为基础,通过穷举法选出所有物体可能出现的区域,然后利用其特征和模式识别方法进行分类,最后对成功分类的区域进行非极大值抑制,进而输出结果。随着计算机软硬件设备的进步和机器学习技术的发展,基于深度学习的目标检测方法逐渐成为主流。2014年, Girshick 等摒弃传统滑动窗口和人工选取特征的方法,将选择性搜索(SS)算法^[6]和卷积神经网络(CNN)^[7-8]结合,通过边框回归模型对检测框的位置进行修正,提出二阶段 R-CNN 算法^[9],开启了深度学习应用于目标检测的先河,在 PASCAL VOC 数据集^[10]上的测试结果相对传统检测算法提高了 30%。但此方法要求输入图像尺寸固定,并且在训练过程中需要进行裁切或变形缩放等操作,导致一定程度的图像信息丢失或变形,限制了检测精度的提高。

通过分析 R-CNN 架构发现,在网络实现过程中,可生成任意大小特征图的卷积层,并不需要输入固定尺寸的图像,仅在全连接层需要对图像尺寸进行固定。为此, He 等^[11]在 2015 年引入 spatial pyramid pooling network(SPP-Net),将卷积层输出的任意尺寸特征图池化为统一规格的特征向量并传递给全连接层,使整个网络可以输入任意尺寸的图像,解决了 R-CNN 算法在候选区域缩放时存在的畸变和提取特征时因重复计算导致的模型性能与速度瓶颈问题。然而,该方法仍需单独训练 CNN 和支持向量机(SVM)^[12]分类器,难以实现并行训练,占用巨大运行存储空间,并且训练过程复杂,速度较慢。

2015 年 Girshick^[13]在前述工作的基础上,将整幅图像作为输入,使用卷积和池化操作一次性处理图像,生成对应的特征映射,并通过增加感兴趣区域池化(ROI Pooling)模块实现 CNN 和边框回归的同时训练,达到简化模型的目的,提出的 Fast R-

CNN 算法有效解决了模型训练速度慢、所占空间大的问题,但此方法在实现过程中过分依赖 SS^[6],产生候选域,时间损失较大。基于此, Ren 等^[14]提出 Faster R-CNN 算法,设计区域提议网络(RPN)取代 SS 算法来提取候选区域,使 RPN 和 Fast R-CNN 能共享卷积特征,实现了模型端对端的训练,综合性能得到显著提升。

尽管 Faster R-CNN 算法已经取得了不错的检测效果,但在进行多目标检测时,物体所含特征信息量不一致,特别是尺寸较小的目标,使用单一尺度进行特征提取的能力不足;此外,通过滑窗法生成固定大小和长宽比的锚框缺乏针对性,难以适应不同尺度的目标,模型整体检测精度的提高受到限制。基于此,本文提出一种多尺度特征融合和锚框自适应相结合的目标检测算法。首先,改进特征金字塔结构,均衡融合各个尺度的语义特征和细节特征,使集成的特征更具辨别力,从而提高模型的检测精度;然后,在 RPN 中,让模型根据物体特征自适应预测合适的候选区域,生成更高质量的锚框,提高检测框的定位精度。

2 相关工作

2.1 Faster R-CNN 模型

Faster R-CNN 模型如图 1 所示,主要包括骨干网络(backbone)、RPN、ROI Pooling、分类与回归 4 个步骤。首先将图像输入骨干网络进行特征提取,获得特征图;然后将特征图输入区域提议网络,经滑动窗口在每个位置产生固定大小和长宽比的锚框,并对其进行分类和回归来产生候选框;再将特征图和候选框同时送入感兴趣区域池化,以固定候选框的尺寸;最后通过 Softmax 分类函数计算候选框中包含目标的概率,利用边框回归对候选框进行微调并最终输出结果。

2.2 特征融合

特征融合通常是对从输入图像提取的各尺度特征图进行深层语义和浅层细节特征融合的过程,通过金字塔结构表征不同尺度的目标,是目前提高检测性能的一种重要方式。Faster R-CNN 模型常采用特征金字塔网络(FPN)^[15-16],如图 2 所示,通过由上到下的融合路径和横向连接方式,不同尺度特征图都具有较强的上下文信息。此方法虽然能改善对多尺度目标的表征能力,但也应注意到算法仅加强了对低层特征的语义表达,没有考虑卷积和池

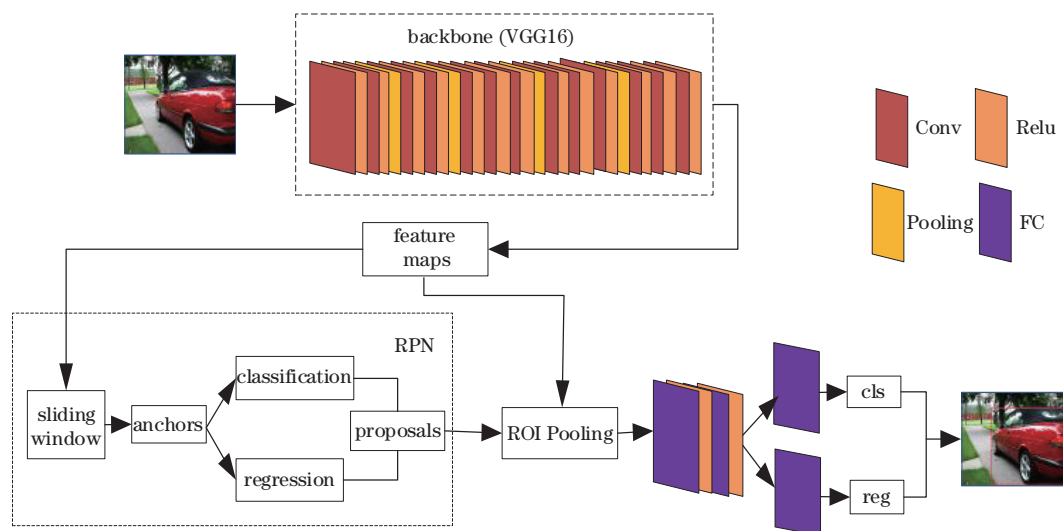


图1 Faster R-CNN 模型

Fig. 1 Faster R-CNN model

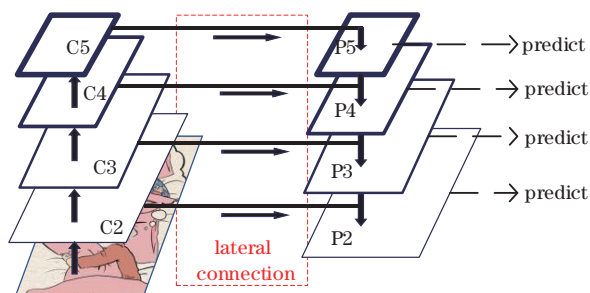


图2 FPN 结构

Fig. 2 FPN construction

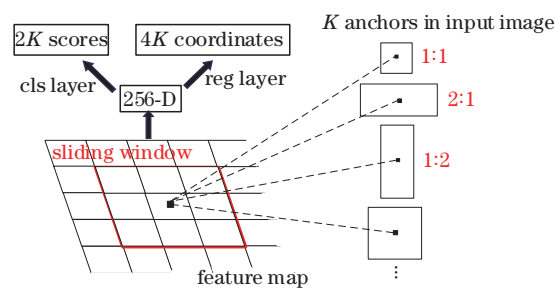


图3 RPN 检测图

Fig. 3 RPN detection diagram

化操作可能造成细节信息丢失的情况,因而对于小尺度物体的检测存在局限性;另一方面,在融合过程中,特征更多关注相邻层级,非相邻层级的特征信息被削弱,这也不可避免地会造成信息量的丢失,导致检测精度的下降。

2.3 区域提议网络

RPN作为Faster R-CNN算法的核心,主要用来提取候选框。它可以任意大小的特征图作为输入,通过滑动窗口对其进行卷积操作,获得可能包含目标的候选区域。检测过程如图3所示,窗口每滑动一次,预定义尺度和长宽比的检测头会在每个位置生成K个区域提议(即锚框),然后二分类Softmax函数获得2K个分数用来判断框中是否包含目标,边框回归得到4K个坐标,即锚框的宽高和中心点坐标^[17]。此方法获得锚框的过程简单但大小需要预先定义,参数的选取在多目标检测上具有差异;此外,目标在图像中的位置分散,基于滑动窗口生成锚框的方式没有针对性,导致大量锚框分布在背景域内,

模型训练效果欠佳,影响检测框的定位精度。

3 所提方法内容

3.1 改进的Faster R-CNN模型

图4为所提改进的Faster R-CNN目标检测算法,在改进模型中,选用ResNet50作为骨干网络,在不改变其结构的前提下,分别提取四组残差块的最后一层特征图进行B-PAFPN特征融合。首先通过path aggregation feature pyramid network(PA-FPN)加强对底层特征的利用和提取,改善对小尺度目标的检测能力;然后进行balanced feature pyramid network(B-FPN)均衡化处理,使集成的特征能获得来自各尺度特征图中等量的语义信息和细节信息,增强目标的表征能力,并通过金字塔结构解决目标检测的多尺度问题;此外,将融合后的多尺度特征图输入改进的RPN,让模型根据目标特征自适应去预测锚框的位置和形状,克服传统RPN中由滑窗法带来形状固定、定位不准确的问题。

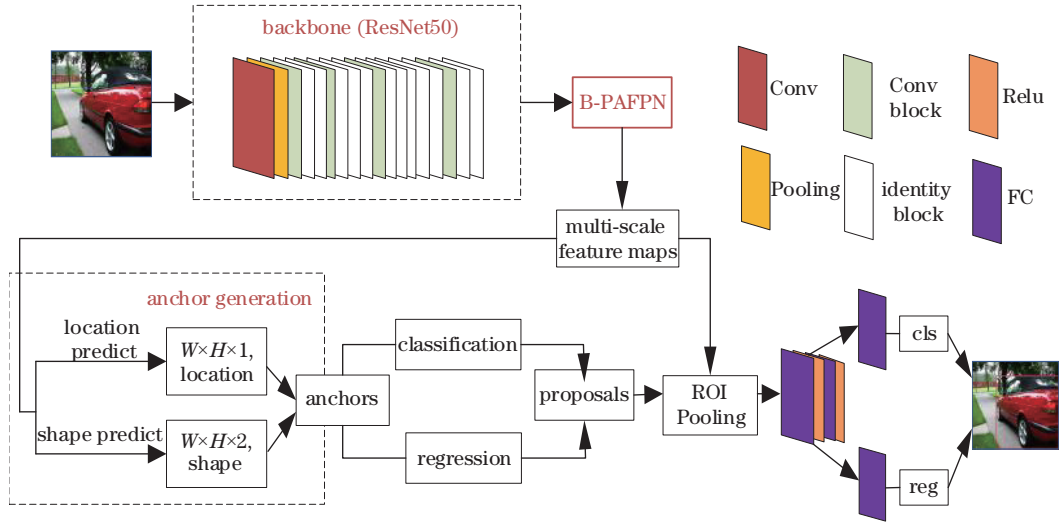


图 4 改进的 Faster R-CNN 模型
Fig. 4 Improved Faster R-CNN model

3.2 多尺度特征融合策略

针对 FPN 存在的不足, PA-FPN 方法^[18]在此基础上添加一路自下向上的路径分支, 加强对底层信息的表征能力, 将传统单向融合转为双向融合, 实现语义特征和细节特征的高效融合; Libra R-CNN^[19]综合金字塔中各尺度进行特征集成, 通过平衡信息流, 特征兼具各分辨率层级信息, 使用 B-FPN 方式达到特征融合的目的。本文结合两者特点, 在 PA-FPN 基础上引入均衡化特征融合策略, 使用相同深度集成的平衡语义信息加强对特征的表征能力。

PA-FPN 结构如图 5 所示, 包括 FPN 中自上向下的 a 路径和添加的自下向上融合路径 b 两部分。由图 5 可知, 补充的 b 路径能解决底层特征在经过多次卷积运算和池化操作后细节信息丢失严重的问题, 通过横向连接融合深层特征和浅层特征以实现信息的双向传递。显然, 这种方式仍旧只使用了相邻层级特征, 没有考虑非相邻层级间的关系, 在信息传递过程中, 特征每融合一次, 非相邻层级的

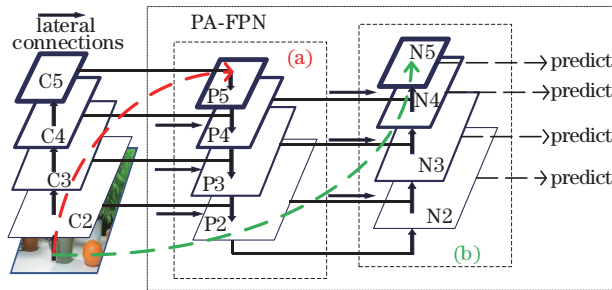


图 5 PA-FPN 结构
Fig. 5 PA-FPN construction

语义信息就会被稀释一次, 造成特征融合的低效率, 影响目标检测精度。

因此, 基于 PA-FPN 又引入 B-FPN 方式, 让金字塔中的各尺度特征图都能获得来自其他不同尺度特征图中等量的特征信息, 通过平衡语义和细节特征提高对目标的表征能力。首先调整 N_l 中各尺度特征图为 N_4 大小^[20], 通过池化减小 N_2 和 N_3 尺寸, 对 N_5 进行上采样, 通过双线性插值增大图像尺寸, 并分别进行 1×1 的卷积操作, 使各尺度特征都具有相同的 256 深度。然后进行特征均衡化:

$$N = \frac{1}{L} \sum_{l_{\min}}^{l_{\max}} N_l, \quad (1)$$

式中: L 表示多尺度特征图数量; 使用的最高和最低分辨率特征图索引分别为 l_{\max} 和 l_{\min} ; 分辨率级别为 l 的特征表示为 N_l 。由式(1)可知, 通过均衡化操作可获得来自各尺度特征图中等量的语义信息和细节信息。此外, 考虑到卷积核固有的局部性特点, 为提高特征表达能力, 对集成特征使用嵌入式高斯函数非局部注意力 (embedded Gaussian nonlocal attention) 模块^[21]进行提炼:

$$y_i = \frac{1}{C(\mathbf{x})} \sum_{v_j} f(\mathbf{x}_i, \mathbf{x}_j) \mathbf{g}(\mathbf{x}_j), \quad (2)$$

式中: \mathbf{x} 为输入特征; \mathbf{y} 为输出特征; $C(\mathbf{x})$ 是归一化因子; $f(\mathbf{x}_i, \mathbf{x}_j)$ 为嵌入式高斯函数, 表示输入特征不同位置间的关联程度; $\mathbf{g}(\mathbf{x}_j)$ 为特征在 j 位置处的表示。

$$f(\mathbf{x}_i, \mathbf{x}_j) = e^{\theta(\mathbf{x}_i)^T \phi(\mathbf{x}_j)}, \quad (3)$$

$$\mathbf{g}(\mathbf{x}_j) = W_g \mathbf{x}_j, \quad (4)$$

式中： $\theta(x_i) = W_o x_i$, $\phi(x_j) = W_\phi x_j$, W 是可学习权重矩阵。最后，将提炼后的特征重新缩放，再通过图 6 中的双箭头连接线与路径 b 的各尺度特征图 N_i

进行反向融合，进一步加强原始 N_i 特征，并通过特征金字塔结构解决目标检测中的多尺度问题，具体流程如图 6 所示。

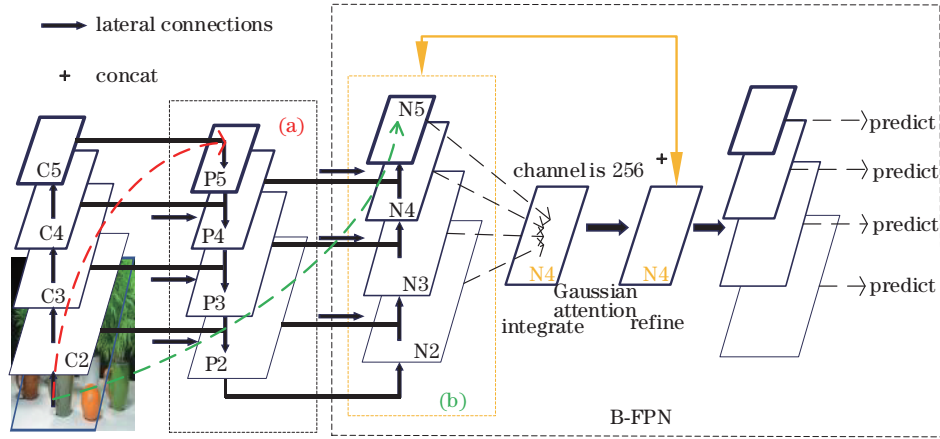


图 6 改进后的特征融合方式

Fig. 6 Improved feature fusion method

3.3 自适应锚框生成机制

基于滑窗生成锚框的方法虽简单可行，但对于尺度范围特殊的目标，固定尺寸和长宽比下获得的锚框具有局限性和预定义参数选取的不确定性。故而，引用自适应锚框生成(AAG)算法^[22]，通过位置预测和形状预测实现锚框的生成过程，基本流程如图 7 所示。

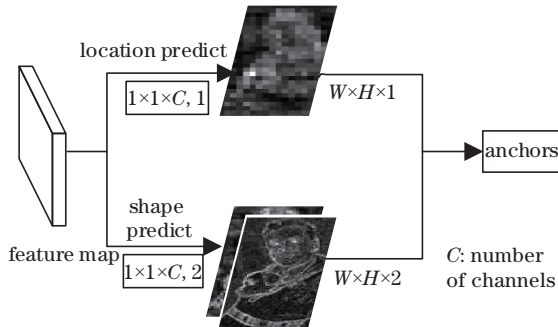


图 7 锚框自适应流程

Fig. 7 Adaptive flow chart of the anchors

当输入图像为 I 时，锚框模型可定义为

$$P(x, y, w, h) = P(x, y|I)P(w, h|x, y, I), \quad (5)$$

式中： (x, y) 表示锚框中心点坐标； (w, h) 为锚框的宽高； $P(x, y|I)$ 是位置概率分布函数，用来对锚框进行位置预测； $P(w, h|x, y, I)$ 是锚框形状的概率分布，由位置信息实现锚框的形状预测。

位置预测时，首先提取特征融合后的多尺度特征图，通过 1×1 卷积将所有通道的特征信息整合为一个通道，获得与特征图尺寸相同的得分图；然后通过 Sigmoid 激活函数将得分转为概率值，得到特征图像中每个像素位置对应原图某区域内存在目标的可能性大小；最后通过特定阈值判断物体在图像中的可能位置。

形状预测时，第一步将位置预测得到的中心坐标作为输入，通过两个 1×1 的卷积预测锚框的宽和高，生成一个包含宽度信息和高度信息的二通道特征图；第二步将传统锚框机制转化为对锚框形状的直接预测过程^[23]，直接预测宽高带来的不稳定性；第三步通过计算真实框和锚框的交并比 (IOU) 来衡量两者之间的重叠程度，预测出和最邻近目标真实框 IOU 最大的形状，最后通过损失函数确定锚框的预测结果。解决宽高带来的不稳定性的公式和损失函数为

$$\begin{cases} w = \sigma \times s \times e^{d_w} \\ h = \sigma \times s \times e^{d_h} \\ L_{\text{shape}} = L \left[1 - \min \left(\frac{P_w}{G_{tw}}, \frac{G_{tw}}{P_w} \right) \right] + L \left[1 - \min \left(\frac{P_h}{G_{th}}, \frac{G_{th}}{P_h} \right) \right] \end{cases}, \quad (6)$$

式中： s 为步幅； σ 为经验尺度因子，设为 8； L 是平滑 L1 损失函数； G_{tw} 和 G_{th} 是真实框的宽度和高度； P_w

和 P_h 表示生成候选框的宽和高。

4 实验部分

4.1 实验平台与数据集

实验基于 Linux 系统,模型的环境配置如表 1 所示。实验选用公开 PASCAL VOC2007 数据集,共包含 5011 张训练图片和 4952 张测试图片,总计 20 类。输入图像尺寸设为 1000×600 ,模型初始学习率设为 0.0025,训练迭代 12 个批次并从第 8 个批次开始将学习率降为 0.00025,使用随机梯度下降 (SGD) 方法作为优化器进行训练,动量因子和权重衰减因子分别设为 0.9 和 0.001,防止模型出现过拟合。

表 1 实验环境配置

Table 1 Experimental environment configuration

Hardware	Software
CPU: AMDR5-4600H	Operating System
GeForce GTX1650Ti-4G	Ubuntu 20.0
GPU: NVIDIA RTX1650	Frame: mmdetection 2.0
512 GB SSD	Language: python

4.2 评价指标

对于多目标检测,首先使用平均精度 (AP) 评估

每一类目标的检测结果,包括精度 (P) 和召回率 (R)。精度表示模型检测正确的目标占模型所有检测目标的比例,用来判断模型的误检,召回率表示模型检测正确的目标占所有待检测目标的比例,用来判断模型的漏检。最后通过计算所有目标类别 AP 的平均值 (mAP) 衡量整个模型的性能。精度和召回率的计算公式^[24]分别为

$$P = \frac{N_{TP}}{N_{TP} + N_{FP}}, \quad (7)$$

$$R = \frac{N_{TP}}{N_{TP} + N_{FN}}, \quad (8)$$

式中: N_{TP} 是标签为正样本且预测为正样本的边框; N_{FP} 是标签为负样本但预测为正样本的边框; N_{FN} 是标签为正样本但预测为负样本的边框。

4.3 实验过程

4.3.1 不同特征融合对 20 类目标的检测结果对比

为证明改进特征融合策略的有效性,实验采用基于 ResNet50 的 Faster R-CNN 模型,分别对 VOC07 数据集中的 20 个类别进行检测,将 FPN、PA-FPN、B-FPN 和所提融合方法进行对比。实验结果如表 2 所示,其中 Variation1、Variation2 是所提方法分别相

表 2 不同特征融合方式对 20 个目标类的 AP

Table 2 AP of different feature fusion methods for twenty target categories

unit: %

Category	FPN	PA-FPN	B-FPN	Proposed method	Variation1 /percent	Variation2 /percent
Aero plane	75.9	79.0	79.0	77.8	-1.2	-1.2
Bicycle	77.2	80.0	79.8	79.7	-0.3	-0.1
Bird	69.1	71.1	70.8	70.2	-0.9	-0.6
Boat	57.9	60.1	59.2	60.9	+0.8	+1.7
Bottle	56.4	58.9	57.8	59.2	+0.3	+1.4
Bus	75.2	82.9	78.5	78.0	-4.9	-0.5
Car	79.3	80.5	80.3	80.7	+0.2	+0.4
Cat	81.7	86.3	83.9	85.2	-1.1	+1.3
Chair	52.2	54.1	54.4	54.3	+0.2	-0.1
Cow	75.8	77.4	81.2	77.3	-0.1	-3.9
Dining table	63.4	65.4	66.7	65.7	+0.3	-1.0
Dog	82.9	85.3	83.6	85.2	-0.1	+1.6
Horse	78.8	80.7	81.1	80.7	0	-0.4
Motorbike	76.2	78.9	79.2	80.2	+1.3	+1.0
Person	77.9	79.3	79.2	79.6	+0.3	+0.4
Potted plant	42.2	44.1	43.2	45.2	+1.1	+2.0
Sheep	71.1	73.4	72.3	75.4	+2.0	+3.1
Sofa	68.2	69.9	69.6	73.0	+3.1	+3.1
Train	77.6	78.8	78.3	79.9	+1.1	+1.6
Tv monitor	67.0	68.8	69.2	68.8	0	-0.4
mAP	70.3	72.745	72.365	72.85	+0.105	+0.485

对 PA-FPN 和 B-FPN 的变化情况,“+”表示改进算法相对原算法的增长百分比,“-”为改进算法相对原算法的下降百分比。

由表 2 可得:改进的特征融合相较于传统 FPN 方法检测精度得到提升,双向融合 PA-FPN 在重合度较高或存在遮挡的目标类中有良好表现,如对公交车、猫目标的平均精度比原始算法分别提高了 7.7 个百分点、4.6 个百分点;均衡融合方法在特征明显、重合度比较低的目标类中能保持较高精度,如对奶牛、餐桌等目标的精度提升较大;所提方法综合两种方法,先双向融合相邻层级的语义和细节特征,再从整体上提炼所有尺度特征,通过对 20 个类别进行检测,结果显示相比于单一 PA-FPN 算法,提高了 60% 目标类的平均检测精度,对部分所占面积小、受环境影响大的目标类检测,如对船、沙发等的检测精度得到了有效提升;而相比 B-FPN 方

式,所提方法也得到了近 70% 目标类的精度提升,其中对于边缘比较明显的物体,例如对狗、山羊和盆栽分别得到了 1.6 个百分点、3.1 个百分点和 2.0 个百分点的精度提升。

4.3.2 锚框生成结果对比

为了验证锚框自适应生成机制的有效性,实验随机选取数据集内的图像,对比原始 RPN 和所提改进方法的锚框生成效果,结果如图 8 所示。图 8(a)、(b)为滑窗法生成的固定大小、长宽比的 1000、100 个锚框,图 8(c)、(d)是基于目标特征自适应预测位置和形状得到的 1000、100 个锚框。由图 8 可得,当提取较多数量候选框时,改进方法能更有针对性地聚焦到目标,降低背景域内的锚框数量;而当候选框数量较少时,改进方法也能避免部分小尺度目标的漏检,提高检测器性能。

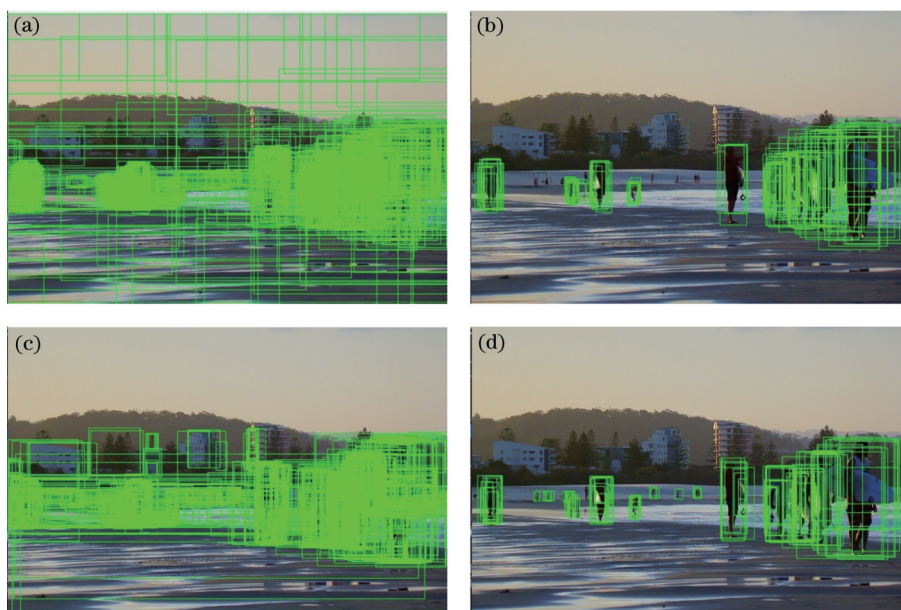


图 8 锚框结果对比

Fig. 8 Comparison results of anchors

4.3.3 消融实验

消融实验作为深度学习领域常用的检验方法,主要用来分析不同分支对整个模型性能的影响。为进一步验证所提方法对 Faster R-CNN 算法的影响,将实验分成 4 组进行训练。第一组为未改动的原始 Faster R-CNN 算法;第二组将 backbone 替换为 ResNet50,特征融合使用经典 FPN 方法;第三组在第二组的基础上将 FPN 改为所提的均衡化路径聚合特征金字塔网络(B-PAFPN);第四组基于第三组实验,改变 RPN 中的锚框机制,引入自适应锚框生

成方法。四组消融实验结果如表 3 所示,其中“√”表示包含该结构,“×”表示不含该结构。

表 3 消融实验结果

Table 3 Ablation experiment results

Group	ResNet50	B-PAFPN	Adaptive-anchor	mAP / %
G1	×	×	×	69.9
G2	√	×	×	70.3
G3	√	√	×	72.85
G4	√	√	√	73.5

由表 3 可知:基于 VGG16 的 Faster R-CNN 在 VOC07 数据集上的 mAP 为 69.9%;将 backbone 替换为轻量化的 ResNet50 之后,G2 检测精度提升了 0.4 个百分点,原因是神经网络层数增加,提取特征的能力提高但受限于特征融合的不充分性,精度提升有限;而第三组实验改进了 G2 中的 FPN 结构,应用改进的多尺度特征融合策略,整体 mAP 提升了 2.55 个百分点,表明改进的特征融合在提取特征任务上发挥了重要作用,加强了对目标特征的判别能力,提升了模型的检测效果;最后 G4 实验改变 RPN 中的锚框机制,通过自适应预测框的位置和形状,提高对建议框的定位准确性,mAP 基于 G3 提升了 0.65 个百分点,改善了检测器性能。

4.4 算法对比实验

为了进一步验证改进算法的有效性,将所提算法和部分经典目标检测算法进行对比实验,结果如表 4 所示。

表 4 不同算法的 mAP 比较

Table 4 mAP comparison of different algorithms

Algorithm	Backbone	mAP / %
Faster R-CNN	VGG16	69.9
Faster R-CNN ^[25]	Res2Net	74.4
Libra R-CNN ^[19]	ResNet50	72.6
YOLO ^[26]	Dark-Net	63.4
SSD ^[27]	VGG16	68.0
Proposed algorithm	ResNet50	73.5

对比表 4 不同检测算法得到的 mAP 结果,所提多尺度特征融合与锚框自适应相结合的改进 Faster R-CNN 算法在检测精度上得到了一定提升。相较于经典 Faster R-CNN 算法,所提算法精度提高了 3.6 个百分点;而 Faster R-CNN 和 Libra R-CNN 是基于原算法的改进方法,所提方法与它们相比具有一定竞争力;此外,当与传统单阶段检测算法^[26-27]进行比较时,所提算法精度分别提高了 10.1 个百分点和 5.5 个百分点。

4.5 目标检测结果

目标检测的任务是识别出图像中感兴趣的物体并确定它们的类别和位置。为了更直观地体现改进算法的性能,实验随机选取部分图像,根据最终检测框对目标的分类结果和得分,对比所提方法和原 Faster R-CNN 算法的检测效果,证明改进算法的有效性,结果如图 9 所示。

由图 9 可得:原 Faster R-CNN 算法存在目标识

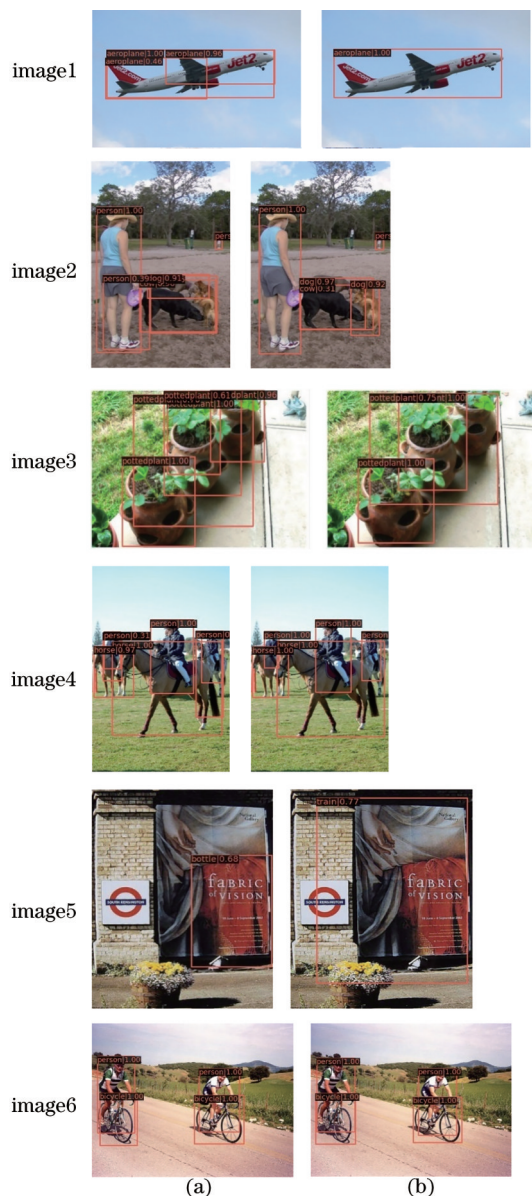


图 9 图片检测比较。(a)原算法检测结果;(b)改进算法检测结果

Fig. 9 Image detection comparison. (a) Detection results of the original algorithm; (b) detection results of the improved algorithm

别准确度不高、检测框冗余的问题;所提改进算法通过 B-PAFPN 特征融合和 AAG 方法,加强了特征的提取和判别能力,改善对不同目标物体的检测效果,同时减少了检测框的数量。检测框可以更高效地框选目标,提高定位精度。在 Image1、Image3 和 Image6 中,通过采用高质量的区域提议能获得更有优势的检测框,提高检测框的定位精度;Image2 和 Image4 中,通过改进的特征融合策略增强了多目标检测的表征能力,提高目标的检测准确度。虽然所提方法仍存在问题,如 Image5 中盆栽的漏检情况,

但整体而言,所提改进算法的精度相较于原算法仍取得了一定提高。

5 结 论

针对 Faster R-CNN 算法检测精度不高的问题,使用均衡化双向融合策略提取特征不充分的问题,先双向融合深层浅层特征,将各尺度特征调整为相同深度,再通过相加求平均的方式实现特征均衡化;同时在 RPN 中让模型基于物体特征自适应地预测锚框位置和形状,使用高质量的区域提议来获取候选框,从而提高检测框的定位精度。基于 VOC 数据集的实验结果表明,所提方法相较于经典 Faster R-CNN 算法,平均检测精度有较好表现。而下一步工作将是针对小目标检测准确性不高、易出现误检和漏检的情况进行更深入的研究。

参 考 文 献

- [1] Zhuo L Q, Fan P P, Zhang S, et al. High-performance fiber-integrated multifunctional graphene-optoelectronic device with photoelectric detection and optic-phase modulation[J]. *Photonics Research*, 2020, 8(12): 1949-1957.
- [2] Wang X Y, Jin X, Li J Q. Blind position detection for large field-of-view scattering imaging[J]. *Photonics Research*, 2020, 8(6): 920-928.
- [3] 王科俊, 赵彦东, 邢向磊. 深度学习在无人驾驶汽车领域应用的研究进展[J]. *智能系统学报*, 2018, 13(1): 55-69.
Wang K J, Zhao Y D, Xing X L. Deep learning in driverless vehicles[J]. *CAAI Transactions on Intelligent Systems*, 2018, 13(1): 55-69.
- [4] 龙坤. 医疗图像小目标检测与分割研究[D]. 成都: 电子科技大学, 2021.
Long K. Research on small objects detection and segmentation in medical images[D]. Chengdu: University of Electronic Science and Technology of China, 2021.
- [5] 邓雪亚, 肖秦琨. 基于优化候选区域的战场目标检测研究[J]. *计算机仿真*, 2021, 38(10): 9-14, 42.
Deng X Y, Xiao Q K. Research on battlefield target detection based on optimize candidate regions[J]. *Computer Simulation*, 2021, 38(10): 9-14, 42.
- [6] Uijlings J R R, van de Sande K, Gevers T, et al. Selective search for object recognition[J]. *International Journal of Computer Vision*, 2013, 104(2): 154-171.
- [7] Li J X, Mengu D, Luo Y, et al. Class-specific differential detection in diffractive optical neural networks improves inference accuracy[J]. *Advanced Photonics*, 2019, 1: 046001.
- [8] 丁文锐, 刘春蕾, 李越, 等. 二值卷积神经网络综述[J]. *航空学报*, 2021, 42(6): 024504.
Ding W R, Liu C L, Li Y, et al. Binary convolutional neural network: review[J]. *Acta Aeronautica et Astronautica Sinica*, 2021, 42(6): 024504.
- [9] Girshick R, Donahue J, Darrell T, et al. Rich feature hierarchies for accurate object detection and semantic segmentation[C]//2014 IEEE Conference on Computer Vision and Pattern Recognition, June 23-28, 2014, Columbus, OH, USA. New York: IEEE Press, 2014: 580-587.
- [10] Everingham M, Gool L, Williams C K I, et al. The pascal visual object classes (VOC) challenge[J]. *International Journal of Computer Vision*, 2010, 88(2): 303-338.
- [11] He K M, Zhang X Y, Ren S Q, et al. Spatial pyramid pooling in deep convolutional networks for visual recognition[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2015, 37(9): 1904-1916.
- [12] Cristianini N, Shawe-Taylor J. An introduction to support vector machines and other kernel-based learning methods[M]. Cambridge: Cambridge University Press, 2000.
- [13] Girshick R. Fast R-CNN[C]//2015 IEEE International Conference on Computer Vision, December 7-13, 2015, Santiago, Chile. New York: IEEE Press, 2015: 1440-1448.
- [14] Ren S Q, He K M, Girshick R, et al. Faster R-CNN: towards real-time object detection with region proposal networks[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2017, 6(1): 1137-1149.
- [15] 陈睿龙, 罗磊, 蔡志平, 等. 基于深度学习的实时吸烟检测算法[J]. *计算机科学与探索*, 2021, 15(2): 327-337.
Chen R L, Luo L, Cai Z P, et al. Algorithm for real-time smoking detection based on deep learning[J]. *Journal of Frontiers of Computer Science and Technology*, 2021, 15(2): 327-337.
- [16] 韩松臣, 张比浩, 李炜, 等. 基于改进 Faster-RCNN 的机场场面小目标物体检测算法[J]. *南京航空航天大学学报*, 2019, 51(6): 735-741.
Han S C, Zhang B H, Li W, et al. Small target detection in airport scene via modified Faster-RCNN

- [J]. Journal of Nanjing University of Aeronautics & Astronautics, 2019, 51(6): 735-741.
- [17] 周兵, 李润鑫, 尚振宏, 等. 基于改进的 Faster R-CNN 目标检测算法[J]. 激光与光电子学进展, 2020, 57(10): 101009.
Zhou B, Li R X, Shang Z H, et al. Object detection algorithm based on improved Faster R-CNN[J]. Laser & Optoelectronics Progress, 2020, 57(10): 101009.
- [18] Liu S, Qi L, Qin H F, et al. Path aggregation network for instance segmentation[C]//2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, June 18-23, 2018, Salt Lake City, UT, USA. New York: IEEE Press, 2018: 8759-8768.
- [19] Pang J M, Chen K, Shi J P, et al. Libra R-CNN: towards balanced learning for object detection[C]//2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), June 15-20, 2019, Long Beach, CA, USA. New York: IEEE Press, 2019: 821-830.
- [20] 贺宇哲, 何宁, 张人, 等. 面向深度学习目标检测模型训练不平衡研究[J/OL]. 计算机工程与应用: 1-11 [2021-10-18]. <http://kns.cnki.net/kcms/detail/11.2127.TP.20201231.1451.023.html>.
He Y Z, He N, Zhang R, et al. Research on imbalanced training of deep learning target detection mode[J/OL]. Computer Engineering and Applications: 1-11 [2021-10-18]. <http://kns.cnki.net/kcms/detail/11.2127.TP.20201231.1451.023.html>.
- [21] Wang X L, Girshick R, Gupta A, et al. Non-local neural networks[C]//2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, June 18-23, 2018, Salt Lake City, UT, USA. New York: IEEE Press, 2018: 7794-7803.
- [22] Wang J Q, Chen K, Yang S, et al. Region proposal by guided anchoring[C]//2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), June 15-20, 2019, Long Beach, CA, USA. New York: IEEE Press, 2019: 2960-2969.
- [23] 陈运忠. 基于特征融合和自适应锚框的目标检测算法研究[D]. 开封: 河南大学, 2020.
Chen Y Z. Research on object detection algorithm based on feature fusion and adaptive anchor framework[D]. Kaifeng: Henan University, 2020.
- [24] 伍锡如, 凌星雨. 基于改进的 Faster RCNN 面部表情检测算法[J]. 智能系统学报, 2021, 16(2): 210-217.
Wu X R, Ling X Y. Facial expression recognition based on improved Faster RCNN[J]. CAAI Transactions on Intelligent Systems, 2021, 16(2): 210-217.
- [25] Gao S H, Cheng M M, Zhao K, et al. Res2Net: a new multi-scale backbone architecture[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2021, 43(2): 652-662.
- [26] 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, June 27-30, 2016, Las Vegas, NV, USA. New York: IEEE Press, 2016: 779-788.
- [27] Liu W, Anguelov D, Erhan D, et al. SSD: single shot multi-box detector[M]//Leibe B, Matas J, Sebe N, et al. Computer vision-ECCV 2016. Lecture notes in computer science. Cham: Springer, 2016, 9905: 21-37.