

基于结构重参数化的目标检测模型

吕昌¹ 尹和² 邵叶秦¹

(1.南通大学交通与土木工程学院 南通 226019; 2.南通大学信息科学技术学院 南通 226019)

摘要: 虽然多尺度感受野特征融合能显著提升目标检测模型的精度,但同时也大大增加了模型的运算成本。针对这一问题,本文提出了基于结构重参数化的目标检测模型。首先,使用深度卷积代替 SPP 中的最大池化,并利用结构重参数化降低模块运算量,提出了新的感受野特征融合模块 CspRepSPP。接着,基于结构重参数化技术,提出了新的特征提取模块 RepBottleNeck。实验结果表明,在 VOC 2012 数据集上,相比原模型 YOLOv5s,本文模型在 mAP0.5:0.95 指标上提升了 3.22%,单张图片的推理速度提升了 0.5 ms,GFLOPs 降低了 1.0。与其他 YOLOv5s 改进算法相比,本文算法检测精度更高,推理速度更快,参数量更低。

关键词: 结构重参数化;多尺度感受野;目标检测;卷积神经网络

中图分类号: TP391.4 **文献标识码:** A **国家标准学科分类代码:** 520.20

Object detection model based on structural re-parameterization

Lyu Chang¹ Yin He² Shao Yeqin¹

(1. School of Transportation and Civil Engineering, Nantong University, Nantong 226019, China;

2. School of Information Science and Technology, Nantong University, Nantong 226019, China)

Abstract: The fusion of multi-scale receptive field feature can remarkably improve the detection accuracy of models, but it also greatly increases the computational cost of models at the same time. To address this issue, we propose the object detection model based on structural re-parameterization. Firstly, max pooling in SPP is substituted by depthwise convolution, while structural re-parameterization is utilized to reduce computational complexity of module simultaneously. Based on this, we propose a new multi-scale receptive field feature fusion module, called CspRepSPP. Additionally, a new feature extraction module, named RepBottleNeck, is proposed according to structural re-parameterization. Experimental results show that, compared with the original YOLOv5s model, the mAP0.5:0.95 of our model is improved by 3.22 percentage points, the detection speed of single image is improved by 0.5 ms, and the GFLOPs is reduced by 1.0. Compared with other improved methods based on YOLOv5s, our method shares higher detection accuracy, faster inference speed, and lower number of parameters.

Keywords: structural re-parameterization; multi-scale receptive fields; object detection; convolutional neural network

0 引言

目标检测是计算机视觉领域中十分重要的技术,其被广泛应用于公共交通、医学图像、遥感检测等多个场景中。目前基于卷积神经网络(convolutional neural network, CNN)的目标检测算法通常分为两类:1)两阶段目标检测算法,例如 R-CNN(regions with CNN features)^[1]、Fast R-CNN^[2]、Faster R-CNN^[3]; 2)单阶段目标检测算法,例如 SSD(single shot MultiBox detector)^[4]、YOLO(you only look once)^[5]、RetinaNet^[6]。

传统卷积神经网络通过堆叠核大小 1×1 和 3×3 的卷

积进行特征提取的方式,只能获得单一尺度的感受野特征。当目标尺度多样时,单一尺度感受野的特征提取往往无法达到较好的效果,因此,融合多尺度感受野特征信息能加强模型检测多尺度目标的能力。目前常见的多尺度感受野特征融合模块有以下 4 种:1)空间金字塔池化^[7](spatial pyramid pooling, SPP)以及快速空间金字塔池化(SPP fast, SPPF)。SPP 解决了不同尺寸的图片输入 CNN,输出特征图的尺寸不一致问题。YOLOv3^[8]引入 SPP 模块,主要通过 3 个不同大小核($k=13, 9, 5$)的最大池化(通过调整步距和填充使得输出特征图大小与输入相同)以及通道维度的拼接来实现 3 个不同尺度感受野特征的融合。

YOLOv5s^[9]的 SPPF 是对 SPP 的升级,通过串联 3 个较小核($k=5$)的最大池化以及通道维度的拼接来实现与 SPP 相同的功能,但 SPPF 具备更快的运算速度与更低的运算量。2)空洞金字塔池化(atrous spatial pyramid pooling, ASPP)。该模块来自图像分割网络 DeepLab^[10],通过多个并联的空洞卷积来提取多尺度感受野特征,并使分割结果更加准确。3)感受野模块^[11](receptive field block, RFB)。该模块采用多个并行的分支,每个分支使用包含一般卷积和空洞卷积的多重卷积操作获得丰富的特征,并通过拼接的方式融合起来。4)一次性聚合模块(one-shot aggregation block, OSAB)。该模块由 VoVNet^[12]提出,与 SPPF 相似,其通过 4 个串联的卷积并拼接每个卷积的输出来实现多尺度感受野特征的融合。

何自芬等^[13]利用 SPP 模块增强网络的表征能力,解决了数据集中目标尺寸大小不一致的问题。黄裕凯等^[14]在目标检测模型中引入 RFB 模块,增强了网络对目标多尺度特性的适应性。李成豪等^[15]利用 OSAB 模块,显著提升了模型对小目标检测的能力。郑伟等^[16]在图像分割模型中引入 RFB 模块,在增大感受野的同时避免池化操作,有效减少细节信息的丢失。张福东^[17]将 YOLOv4 目标检测算法的 SPP 模块替换为多尺度感受野特征融合能力更强的 ASPP 模块,从而提高了模型对交通标志检测的精度。

多尺度感受野特征融合虽然有效提升了模型检测精度,但多分支结构不但会增加模型运算量,还会降低模型检测速率。结构重参数化^[18](structural re-parameterization)技术的引入可以有效解决该问题,在推理阶段,结构重参数化通过合并多个并行的卷积运算,起到降低模型运算量,提高模型检测速率的作用。

本文发现 SPP 与 SPPF 使用最大池化操作虽然降低了运算量,但存在大量信息丢失的问题:最大池化每次只选择感受野内像素值最大的特征,因此在最大池化提取特征的过程中,大量特征信息遭到丢失。同时,YOLOv5s 的特征提取模块 BottleNeck 只通过单一尺度的感受野提取图像特征,缺乏多尺度感受野的特征提取能力。

为解决 YOLOv5s 中存在的上述两个问题,本文主要贡献如下:1)给出最大池化信息丢失问题的解决方案,首先利用深度卷积代替最大池化,有效地解决最大池化存在的信息丢失问题,并通过结构重参数化与跨阶段部分(cross stage partial, CSP)结构进一步降低模块在推理过程的计算量,同时提升检测速率,基于此提出了一个改进的多尺度感受野特征融合模块 CspRepSPP。相比 SPPF, CspRepSPP 在推理阶段拥有相当的计算量以及略快的运算速度,并能获得更好的检测精度。2)给出多尺度感受野特征提取的方案,不同于传统卷积的单一尺度感受野特征提取方式,基于两个不同尺度的卷积核从两个不同尺度的感受野提取图像特征并融合,设计了一个新的特征提取模块 RepBottleNeck,进一步提升了模型的检测精度,同时结

合结构重参数化,使最终模型在推理阶段的运算速度要快于原模型 YOLOv5s。目前已有大量基于 YOLOv5s 改进的模型^[19-22]被提出,本文模型相较于这些改进模型的创新点在于引入了结构重参数化技术,因此本文模型被称为基于结构重参数化的目标检测模型。

1 模型总览

本文提出的基于结构重参数化的目标检测模型以 YOLOv5s 为基准,将其中存在最大池化信息丢失问题的 SPPF 改进为特征提取更加充分、推理速度更加快速的 CspRepSPP,同时将所有 C3 模块中采用单一尺度感受野提取特征的 BottleNeck 改进为使用多尺度感受野提取特征的 RepBottleNeck。基于结构重参数化的目标检测模型的网络结构如图 1 所示,改进部分由箭头标出,图中矩形表示特征图,特征图旁的数值为对应的尺寸,c 为输出通道数,DownSample 和 UpSample 分别表示下采样和上采样。整个目标检测模型由 3 部分构成:

1) Backbone,进行下采样以及特征提取。下采样通过步距为 2 的卷积实现,特征提取通过 C3 模块实现。C3 是一个 CSP 结构模块,首先通过两个核大小为 1×1 的卷积使输入特征图的通道数减半,并在其中一个分支上使用 N 个 BottleNeck 模块(本文改进为 RepBottleNeck),将其结果与另一个分支的输出在通道维度上拼接,最终通过核大小 1×1 的卷积输出。Backbone 的末端是一个多尺度感受野特征融合模块 SPPF(本文改进为 CspRepSPP)。

2) Neck,对 backbone 中提取的信息进行更深层次的融合。该部分的结构为路径聚合网络(path aggregation network, PANet),即自上而下逐步拼接尺寸较大的特征图,再自下而上逐步拼接尺寸较小的特征图,不同尺寸特征图的融合可以提高模型的检测精度。Neck 部分同样使用 C3 模块进行特征提取,本文同样将 Neck C3 中的 BottleNeck 改进为 RepBottleNeck。

3) Head,预测结果输出。该部分通过使用 3 个核大小 1×1 的卷积,分别对 3 个不同尺度的特征图进行目标预测。

2 改进的多尺度感受野特征融合模块

基于 SPP 与 SPPF 多尺度感受野特征融合的思想,使用深度卷积代替最大池化解决信息丢失问题,并提出新的特征融合模块 RepSPP。由于深度卷积的运算量大于最大池化,为减少运算量,利用结构重参数化以及 CSP 结构,提出改进的多尺度感受野特征融合模块 CspRepSPP。

2.1 最大池化与深度卷积

本文发现,最大池化与深度卷积在计算方式上存在相似性。深度卷积每个通道的输出特征图只由对应通道的输入特征图经过图像卷积运算求得,最大池化与深度卷积类似,但不进行图像卷积运算,而是筛选感受野内的最大像素

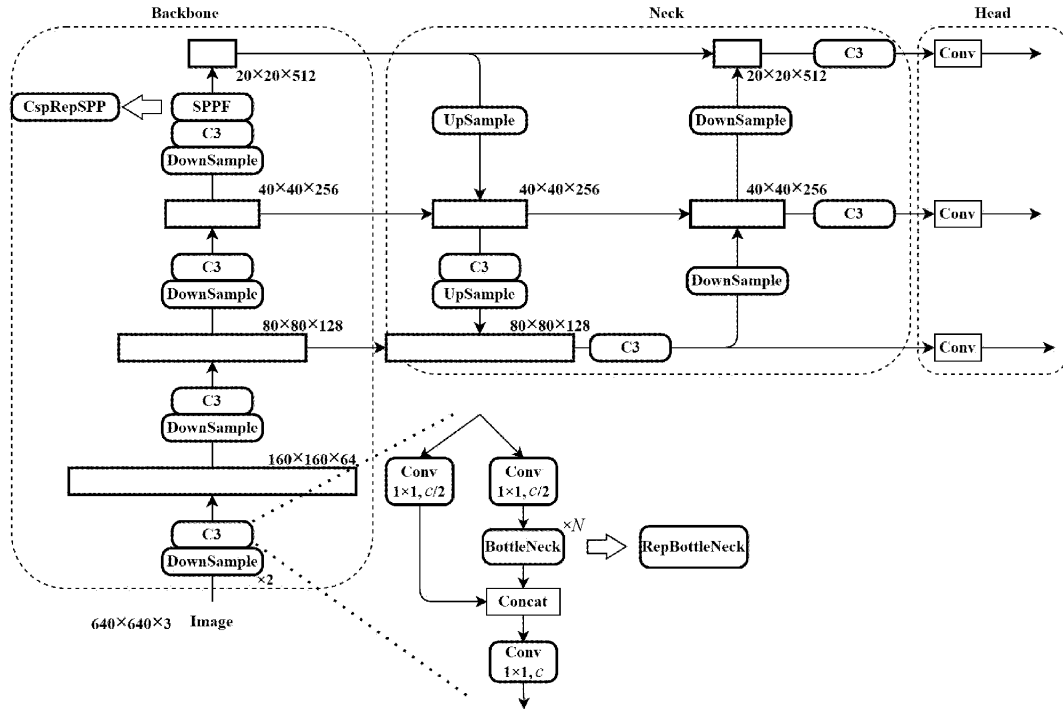


图 1 基于结构重参数化的目标检测模型的网络结构

值来提取特征。

最大池化与深度卷积类似，而深度卷积运算不存在信息丢失问题，因此，本文将 SPP 模块的最大池化替换为深度卷积(depthwise convolution, Dw-Conv)，并为节省运算量将通道维度上特征图的拼接改为按元素相加，在训练阶段，新模块 RepSPP 如图 2 所示。

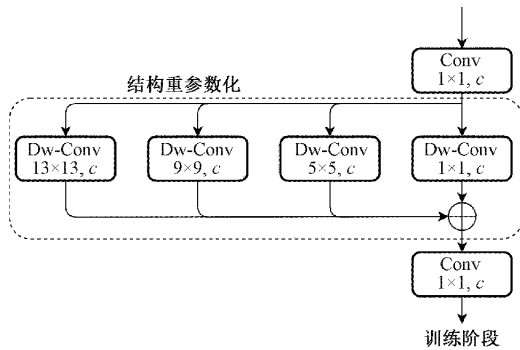


图 2 RepSPP 训练阶段示意图

2.2 结构重参数化

虽然深度卷积解决了最大池化的信息丢失问题，但深度卷积的运算量大于最大池化，表 1 提供了 YOLOv5s 加入多种 SPP 模块后模型的单张图片处理速度，由该表可知，RepSPP 训练阶段的运算速度慢于 SPP 与 SPPF。

推理阶段，结构重参数化技术可以优化多个并联结构的卷积或深度卷积，优化前后的模型完全等价，但优化后的模型推理速度更快。下面介绍结构重参数化原理。

表 1 具备不同 SPP 模块的 YOLOv5s 模型的运算速度

模块	SPP	SPPF	RepSPPTrain	RepSPP Infer	CspRep-SPP
速度/ms	36.0	35.5	37.0	35.7	35.3

RepSPP 训练阶段虚线框内的运算可用下列公式表述：

$$M^{(2)} = \sum_{i=1}^N bn(M^{(1)} * W^{(i)}, \mu^{(i)}, \sigma^{(i)}, \gamma^{(i)}, \beta^{(i)}) \quad (1)$$

其中， $M^{(1)}$ 、 $M^{(2)}$ 分别为输出、输入特征图； bn 为批量归一化； $W^{(i)}$ 为第 i 个卷积运算的卷积核，共进行 N 次卷积运算，由于 RepSPP 共进行 4 次深度卷积运算，因此此处 $N = 4$ ； $*$ 表示深度卷积运算； $\mu^{(i)}$ 、 $\sigma^{(i)}$ 为第 i 个批量归一化计算而得的均值与方差； $\gamma^{(i)}$ 、 $\beta^{(i)}$ 为批量归一化中两个可学习的变量。批量归一化的计算公式如下：

$$bn(M, \mu, \sigma, \gamma, \beta) = (M - \mu) \times \frac{\gamma}{\sigma} + \beta \quad (2)$$

$$\mu = \frac{1}{m} \times \sum_{i=1}^m x^{(i)} \quad (3)$$

$$\sigma = \sqrt{\frac{1}{m} \times \sum_{i=1}^m (x^{(i)} - \mu)^2} \quad (4)$$

其中， $x^{(i)}$ 为特征图 M 的第 i 个特征， m 为特征总数。批量归一化通过式(3)、(4)计算特征图的均值与标准差，并对特征图 M 减去均值 μ ，除以标准差 σ ，并乘以缩放因子 γ ，加上偏差 β ，其中 γ 与 β 为可学习变量。

在推理阶段，根据式(2)，结构重参数化可以对

RespSPP 在训练阶段的运算进一步化简:

$$\mathbf{M}^{(2)} = \sum_{i=1}^N ((\mathbf{M}^{(1)} * \mathbf{W}^{(i)} - \mu^{(i)}) \times \frac{\gamma^{(i)}}{\sigma^{(i)}} + \beta^{(i)}) =$$

$$\sum_{i=1}^N (\mathbf{M}^{(1)} * \mathbf{W}^{(i)} \times \frac{\gamma^{(i)}}{\sigma^{(i)}} - \mu^{(i)} \times \frac{\gamma^{(i)}}{\sigma^{(i)}} + \beta^{(i)}) =$$

$$\mathbf{M}^{(1)} * \mathbf{W}' + \beta' \quad (5)$$

$$\mathbf{W}' = \sum_{i=1}^N \mathbf{W}^{(i)} \times \frac{\gamma^{(i)}}{\sigma^{(i)}} \quad (6)$$

$$\beta' = \sum_{i=1}^N (-\mu^{(i)} \times \frac{\gamma^{(i)}}{\sigma^{(i)}} + \beta^{(i)}) \quad (7)$$

式(5)首先根据式(2)将卷积运算与批量归一化融合,接着将多个卷积融合为一个卷积。经过结构重参数化,模块虚线框内的结构可以优化为卷积核权重为 \mathbf{W}' , 偏置为 β' 的深度卷积。

RepSPP 模块在训练阶段采用图 2 结构训练模型,在推理阶段,通过上述公式,使用结构重参数化将模块优化为图 3 结构(虚线框内为结构重参数化的对象),二者的计算结果完全相同,只是训练与推理阶段的模型结构不同,并且优化后的模型推理速度更快。由表 1 知,RepSPP 的推理速度已经快于 SPP。

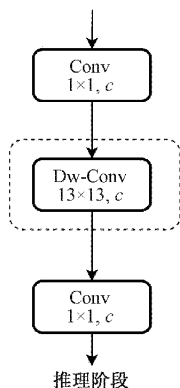


图 3 RepSPP 推理阶段示意图

2.3 CSP 结构

为进一步降低运算开销,本文对 RepSPP 进行 CSP 化,得到图 4 模块 CspRepSPP。CSP 化对特征图进行 2 个分支的处理,2 个分支均使用核大小为 1×1 的卷积将特征图通道数减半,并在一个分支上进行更深层的操作,最终将 2 个分支的结果在通道维度拼接,由于核大小 1×1 的卷积降低了特征图的通道数,因此后续的操作具备更低的运算量。由表 1 知,最终 CspRepSPP 的计算速度要快于 SPP 与 SPPF。

2.4 RepBottleNeck

由于传统卷积特征提取的感受野大小是固定的,在遇到多尺度目标时无法充分提取目标特征,这会导致检测模型出现漏检、误检等问题,为提高目标检测模型融合多尺度感受野特征的能力,本文提出了一个结合结构重参数化与深度卷积的特征提取模块 RepBottleNeck,将该模块替

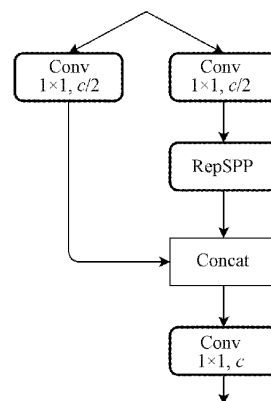


图 4 CspRepSPP 示意图

换 YOLOv5s 的传统特征提取模块 BottleNeck,能在不增加模型运算量的基础上提高模型的检测精度。BottleNeck 结构如图 5(a)所示,由 1 个核大小 1×1 与 3×3 的卷积以及残差连接(skip)组成。如图 5(b)所示,RepBottleNeck 在训练阶段使用 1 个核大小 7×7 与 3×3 的深度卷积提取图像特征,并将结果相加以实现 2 种不同尺度感受野特征的融合。由于深度卷积只在每个通道上进行,不同通道之间信息没有交互,因此在深度卷积前后使用核大小 1×1 的卷积加强特征图通道间的信息交流。最后,为防止模型在训练过程中出现梯度消失现象,RepBottleNeck 同样设置残差连接,为防止模型在推理阶段出现过拟合现象,在训练过程中加入正则手段 DropBlock。使用结构重参数化可将 RepBottleNeck 的结构进行优化,进一步降低模型部署后的运算量,最终在推理阶段,RepBottleNeck 结构示意图如图 5(c)所示。

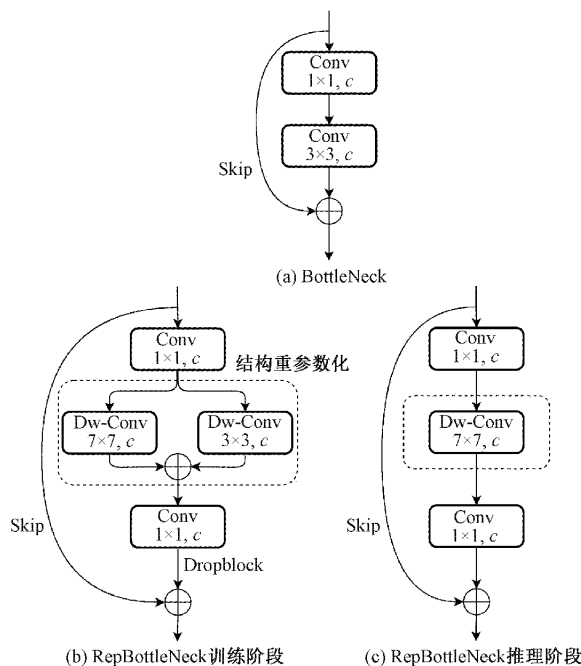


图 5 BottleNeck 和 RepBottleNeck 示意图

3 实 验

3.1 数据集与实验设置

本文使用公开数据集 PASCAL VOC 2012 进行实验, PASCAL VOC 挑战赛是一个由欧盟资助的世界级计算机视觉挑战赛。该数据集包含飞机、公交车、自行车、人、狗等 20 个类别,拥有 5 717 张训练集图片,13 609 个训练目标样本,5 283 张验证集图片,13 841 个验证目标样本。

实验在操作系统为 Windows11,处理器为 Intel (R) Core (TM) i5-11400 (F),显卡为 NVIDIA GeForce RTX 3060 的计算机上进行。采用 Python 编程语言,Pytorch 深度学习框架,CUDA 版本为 11.2,Pytorch 版本为 1.11.0。训练设置如下:训练轮次为 150,批量大小为 16,图片尺寸为 640×640,使用 Adam 优化器,初始学习率为 0.003,采用余弦退火学习率衰减策略。

3.2 评价指标

为有效评估目标检测算法的性能,选取均值平均精度 (mean average precision, mAP)、推理阶段单张图片的处理速度 (Speed) 及模型每秒 10 亿次的浮点运算数 (GFLOPs) 来衡量目标检测算法的检测精度及效率。mAP 的计算公式如下:

$$mAP = \int_0^1 P(R) dR \tag{8}$$

$$P = \frac{TP}{TP + FP} \tag{9}$$

$$R = \frac{TP}{TP + FN} \tag{10}$$

其中, P 为精确度 (precision), R 为召回率 (recall), TP 表示被预测为正例的正样本个数, FP 表示被预测为正例的负样本个数, FN 表示被预测为负例的正样本个数。选定交并比 (intersection over union, IoU) 的阈值, 根据该阈值下的精确度和召回率绘制 P-R 曲线, 并求该曲线下的面积即可得到该阈值下的 mAP。mAP_{0.5} 表示 IOU 阈值为 0.5 时的 mAP, mAP_{0.5:0.95} 表示 IOU 阈值为 0.5, 0.55, 0.6, ..., 0.95 下 mAP 的均值。

3.3 消融实验

为验证本文改进的有效性, 在 VOC 2012 数据集上采用相同的实验配置进行消融实验, 各模块对模型的精度影响如表 2 所示, 各模块对模型的推理速度和运算量的影响如表 3 所示。

表 2 本文改进对模型精度的影响

模型(YOLOv5s)	mAP _{0.5} /%	mAP _{0.5:0.95} /%
+ SPP	61.52	38.31
+ SPPF	61.52	38.31
+ CspRepSPP(ours)	62.16	40.22
+ CspRepSPP	63.66	41.53
+ RepBottleNeck(ours)		

表 3 本文改进对模型推理速度和运算量的影响

模型(YOLOv5s)	速度/ms	GFLOPs
+ SPP	36.0	16.5
+ SPPF	35.5	16.5
+ CspRepSPP(ours)	35.3	16.5
+ CspRepSPP	35.0	15.5
+ RepBottleNeck(ours)		

YOLOv5s+SPP 与 YOLOv5s+SPPF 的精度相同, 但 SPPF 的运算速度略快。YOLOv5s+CspRepSPP 相比 YOLOv5s+SPP 和 YOLOv5s+SPPF, 在 2 个精度指标上分别提升了 0.64% 和 1.91%, 在推理阶段, 模型处理单张图片的运算速度比 SPPF 快 0.2 ms, 比 SPP 快 0.7 ms, 并且 GFLOPs 与 SPP、SPPF 相当。该实验结果证明了改进的多尺度感受野特征融合模块 CspRepSPP 的有效性, 该模块使用深度卷积解决最大池化信息丢失的问题, 并利用结构重参数化以及 CSP 结构减少模型的推理速度, 因此该模块在未引入额外运算量的情况下, 提升了模型的检测精度。YOLOv5s + CspRepSPP + RepBottleNeck 相比 YOLOv5s+SPP 和 YOLOv5s+SPPF, 在 2 个精度指标上分别提升了 2.14% 和 3.22%, 在推理阶段, 模型处理单张图片的运算速度要比 YOLOv5s + SPPF 快 0.5 ms, 比 YOLOv5s + SPP 快 1 ms, 并且 GFLOPs 更低。RepBottleNeck 通过 2 个核大小不同的深度卷积加强了模型融合多尺度感受野特征的能力, 因此进一步提高了模型的检测精度, 深度卷积的应用以及结构重参数化对模块结构的优化降低了模型在推理阶段的运算量, 从而加快了推理速度。

为探究本文改进对模型训练过程的影响, 图 6、7 分别绘制了模型随迭代轮次变化的平均精度曲线, 边框定位损失曲线。由图 6 可知, CspRepSPP 与 RepBottleNeck 均在一定程度上提高了模型的平均精度, 并且改进后模型的收敛速度明显快于原模型。由图 7 可知, CspRepSPP 替换 SPPF 后的模型的边框定位损失明显更低, 因此 CspRepSPP 模块提升了目标检测模型定位目标的能力, 并且 RepBottleNeck 的引入进一步降低了模型的边框定位损失, 从而再次提升模型的目标定位能力。

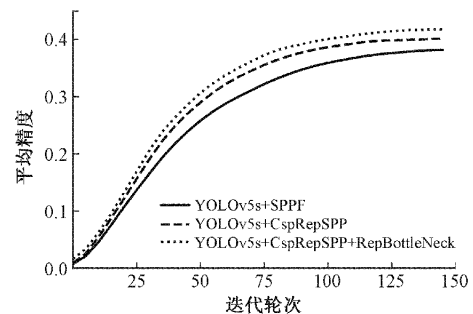


图 6 平均精度曲线图

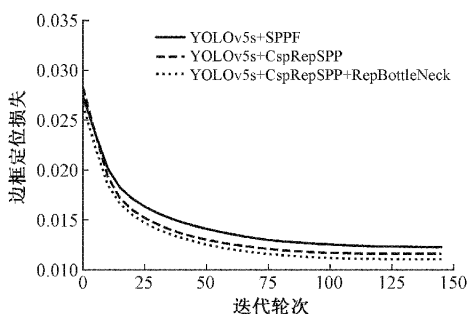


图 7 边框定位损失曲线图

3.4 CspRepSPP 的可移植性

由于本文提出的 CspRepSPP 是针对 SPP 与 SPPF 的改进,对于所有具备 SPP 与 SPPF 的目标检测模型,均可用 CspRepSPP 代替 SPP 与 SPPF,能在不引入额外运算量的情况下提升模型的精度。本文将 YOLOv3tiny、YOLOv4tiny 的 SPP 以及 YOLOXs 的 SPPF 替换成本文提出的 CspRepSPP,实验结果如表 4 所示。在 mAP0.5:0.95 指标下,CspRepSPP 几乎提升了所有模型 2% 左右的精度,同时所有模型的运算量和改进前一致。该实验证明了 CspRepSPP 具备一定的可移植性。

表 4 CspRepSPP 的可移植性

模型	mAP/%	GFLOPs
YOLOv3tiny-SPP	15.79	13.3
YOLOv3tiny-CspRepSPP	18.02	13.3
YOLOv4tiny-SPP	29.00	6.9
YOLOv4tiny-CspRepSPP	31.14	6.9
YOLOXs-SPPF	39.01	26.8
YOLOXs-CspRepSPP	40.82	26.8

3.5 和其他模型的对比

为充分证明本文改进的有效性,本文将基于结构重参数化的目标检测模型与多个基于 YOLOv5s 改进的目标检测模型对比,数据集及实验设置均与 3.1 节相同,所有模

型的主干网络相同,实验结果如表 5 所示。对比实验分别在 YOLOv5s-SE^[19]、YOLOv5s-CBAM^[20]、YOLOv5s-ASPP^[21]以及 YOLOv5s-ASFF^[22]上进行,其中 YOLOv5s-SE 和 YOLOv5s-CBAM 分别在网络中引入 SE 注意力机制和 CBAM 注意力机制,YOLOv5s-ASPP 将改进的空间金字塔池化模块 ASPP 替换原来的 SPPF,YOLOv5s-ASFF 引入了自适应空间特征融合算法 ASFF。实验结果表明,本文提出的模型分别比上述模型的 mAP0.5:0.95 高了 3.03%、2.52%、1.31%、1.52%,同时具备最快的推理速度,以及最低的参数量(parameters)。虽然 SE 注意力机制与 CBAM 注意力机制的参数量与原模型相当,但对模型精度的提升十分有限,同时,CBAM 大幅减缓了模型的推理速度。ASPP 与 ASFF 虽然显著提升了模型精度,但在推理速度和参数量上都明显高于原模型。基于结构重参数化的目标检测模型优于上述改进的 YOLOv5s 模型的原因在于,本文算法通过融合目标的多尺度感受野特征显著提升模型的精度,通过引入结构重参数化有效降低模型的推理速度与参数量,同时兼顾了模型的精度、推理速度与参数量,因此实现了在不引入额外运算量和参数量的情况下提升模型的检测精度。

表 5 和其他模型的对比

模型	mAP/%	Speed/ms	Parameters/M
YOLOv5s	38.31	35.5	7.2
YOLOv5s-SE	38.70	36.3	7.2
YOLOv5s-CBAM	39.01	50.1	7.2
YOLOv5s-ASPP	40.22	41.2	15.3
YOLOv5s-ASFF	40.01	47.7	12.6
本文	41.53	35.0	6.0

本文对验证集中的图像进行目标检测结果的可视化分析,部分图片的可视化结果如图 8 所示。由该图知,YOLOv5s 的 4 个改进模型均在一定程度上缓解了原模型的漏检现象,但图片中任然存在大量的漏检目标。本文

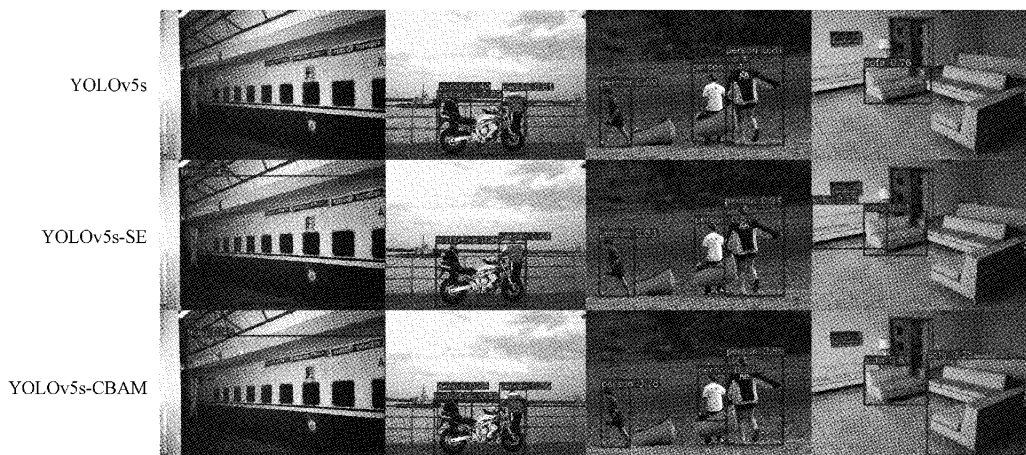




图 8 目标检测结果的可视化分析

算法通过融合多尺度感受野特征加强了模型检测不同大小尺度目标的能力,因此在第一幅图中,本文模型准确检测到了大目标(火车)以及小目标(人),而原模型未检测到任何目标,其余改进模型虽然检测到了火车但均未检测到。该实验证明了多尺度特征融合能有效提高模型检测多尺度目标的能力,从而提高模型的检测精度。

4 结 论

将深度卷积与结构重参数化结合,深度卷积解决了最大池化信息丢失的问题,结构重参数化解决了多尺度感受野特征融合运算量较大的问题,在二者有效结合的基础上,引入 CSP 结构,提出了一个特征提取充分、运算量低、具备一定可移植性的多尺度感受野特征融合模块 CspRepSPP。同时,本文将多尺度感受野特征融合的思想融入目标检测模型的特征提取模块,结合结构重参数化,提出了相比 YOLOv5s BottleNeck 更为高效的多尺度感受野特征提取模块 RepBottleNeck。实验结果表明,本文提出的改进不仅降低了模型在推理阶段的运算量与参数量,同时显著提升模型检测不同尺度目标的能力,从而有效提高 YOLOv5s 目标检测模型的检测精度。与其他基于 YOLOv5s 改进的算法相比,本文提出的基于结构重参数化的目标检测算法具备较高的检测精度,较快的推理速度,以及较小的模型参数量,因此具备良好的移动端设备部署能力。

参考文献

- [1] GIRSHICK R, JEFF D, TREVOR D, et al. Rich feature hierarchies for accurate object detection and semantic segmentation[C]. 2014 IEEE Conference on Computer Vision and Pattern Recognition, IEEE, 2014: 580-587.
- [2] GIRSHICK R. Fast r-cnn[C]. 2015 IEEE Conference on Computer Vision and Pattern Recognition, IEEE, 2015: 1440-1448.
- [3] 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, 39(6): 1137-1149.
- [4] LIU W, ANGUELOV D, ERHAN D, et al. SSD: single shot multibox detector[C]. Proceedings of the 2016 European Conference on Computer Vision, 2016: 21-37.
- [5] 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 Recognition, IEEE, 2016: 779-788.
- [6] LIN T, PRIYA G, GIRSHICK R, et al. Focal loss for dense object detection[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2020, 42(2): 318-327.
- [7] 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.
- [8] REDMON J, FARHADI A. Yolov3: An incremental improvement [J]. ArXiv Preprint, 2018, ArXiv: 1804.02767.
- [9] 李衍照, 于镭, 田金文. 基于改进 YOLOv5 的金属焊缝缺陷检测[J]. 电子测量技术, 2022, 45(19): 70-75.
- [10] CHEN L, PAPANDREOU G, KOKKINOS I, et al. Deeplab: semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2018, 40(4): 834-848.

- [11] LIU S T, HUANG D, WANG Y H. Receptive field block net for accurate and fast object detection[C]. Proceedings of the European Conference on Computer Vision, 2018: 404-419.
- [12] LEE Y, HWANG J, LEE S, et al. An energy and gpu-computation efficient backbone network for real-time object detection[C]. 2019 IEEE Conference on Computer Vision and Pattern Recognition, IEEE, 2019: 752-760.
- [13] 何自芬, 陈光晨, 陈俊松, 等. 多尺度特征融合轻量化夜间红外行人实时检测[J]. 中国激光, 2022, 49(17): 130-139.
- [14] 黄裕凯, 王青旺, 沈韬, 等. 基于 MobileNet 的多尺度感受野特征融合算法[J]. 激光与光电子学进展, 2023, 60(4): 280-288.
- [15] 李成豪, 张静, 胡莉, 等. 基于多尺度感受野融合的小目标检测算法[J]. 计算机工程与应用, 2022, 58(12): 177-182.
- [16] 郑伟, 赵金芳, 张奕婧, 等. 基于感受野扩增和注意力机制的 U-Net 脑肿瘤 MR 图像分割[J]. 河北大学学报(自然科学版), 2022, 42(5): 542-551.
- [17] 张福东. 基于深度学习的交通标志小目标检测与识别研究[D]. 天津: 天津理工大学, 2022.
- [18] DING X H, ZHANG X Y, MA N N, et al. Repvgg: making vgg-style convnets great again[C]. 2021 IEEE Conference on Computer Vision and Recognition, IEEE, 2022: 13733-13742.
- [19] 王素珍, 赵霖, 邵明伟, 等. 基于改进 YOLOv5 的输电线路绝缘子识别方法[J]. 电子测量技术, 2022, 45(21): 181-188.
- [20] 郝帅, 张旭, 马旭, 等. 基于 CBAM-YOLOv5 的煤矿输送带异物检测[J]. 煤炭学报, 2022, 47(11): 4147-4156.
- [21] 张释如, 黄综浏, 张袁浩, 等. 基于改进 YOLOv5 的煤矸识别研究[J]. 工矿自动化, 2022, 48(11): 39-44.
- [22] 陈小毛, 王立成, 张健, 等. 融合 YOLOv5 与 ASFF 算法的海产品目标检测算法研究[J]. 无线电工程, 2023, 53(4): 824-830.

作者简介

吕昌, 硕士, 主要研究方向为深度学习, 目标检测。

E-mail: 1365067150@qq.com

尹和, 硕士, 主要研究方向为深度学习, 3D 人体姿态检测。

E-mail: 1205038719@qq.com

邵叶秦(通信作者), 博士, 副教授, 研究生导师, 主要研究方向为智能图像处理。

E-mail: hnsyk@ntu.edu.cn