doi:10.14132/j.cnki.1673-5439.2020.06.015

基于空洞卷积和 Focal Loss 的改进 YOLOv3 算法

许 腾,唐贵进,刘清萍,鲍秉坤

(南京邮电大学 江苏省图像处理与图像通信重点实验室,江苏 南京 210003)

摘要:为了进一步提升 YOLOv3 的小目标检测能力,文中提出将 Darknet-53 中的第 2 个残差块输出的特征图用混合空洞卷积处理后,与 YOLOv3 的 8 倍下采样特征图相融合,建立新的检测特征;同时,使用 Focal Loss 改进损失函数中的负样本置信度公式,缓解 YOLOv3 的正负样本比例失衡问题。实验结果表明,在小目标数量占比为 47.7% 的特定测试集上,改进 YOLOv3 的平均准确率和召回率分别比原 YOLOv3 提高了 8.8% 和 16%;在 VOC 测试集上,改进 YOLOv3 的平均精度均值比原 YOLOv3 提升了 3.4%。

关键词:小目标检测;样本不平衡;混合空洞卷积;Focal Loss

中图分类号:TP391 **文献标志码**:A **文章编号**:1673-5439(2020)06-0100-09

Improved YOLOv3 based on dilated convolution and Focal Loss

XU Teng, TANG Guijin, LIU Qingping, BAO Bingkun

(Jiangsu Key Lab of Image Processing & Image Communication, Nanjing University of Posts and Telecommunications, Nanjing 210003, China)

Abstract: To improve the detection performance of small targets of the YOLOv3 algorithm, a technique for producing new prediction features is proposed. It fuses the output of the hybrid dilated convolution on the second residual block in Darknet-53 and the output of the 8x down-sampling of the original network. In addition, the Focal Loss is used to compute the confidence of negative samples of the loss function, alleviating the imbalance problems of positive and negative samples of YOLOv3. Experimental results show that compared with the YOLOv3 algorithm, the improved YOLOv3 algorithm can achieve an improvement of 8.8% and 16% in terms of the average precision and the recall on a specific data set with a proportion of the number of small targets of 47.7%, and achieve an improvement of 3.4% in terms of the mean average precision on the VOC dataset.

Keywords: small targets detection; sample imbalance; hybrid dilated convolution; Focal Loss

目标检测是计算机视觉领域的重要研究方向,主要用于对特定目标进行识别和定位,近年来广泛应用于智能安防、自动驾驶等领域,所以目标检测算法引起了越来越多研究者的关注。但是在背景复杂或者光照剧烈变化的环境下,目标检测算法检测密集目标或小尺寸目标时还存在很大的问题。因此目标检测依然是当前最具挑战性的热点研究之一。

传统的目标检测方法通常是将区域选择、特征提取和机器学习中的分类器联合起来组成目标检测器。特征提取一般使用 HOG[1],SIFT[2] 等手动设计方法;分类器主要使用支持向量机(Support Vector Machine, SVM)[3],Adaboost[4] 等算法。但是传统的

收稿日期:2020-03-04;**修回日期**:2020-05-20 **本刊网址**:http://nyzr.njupt.edu.cn

基金项目:国家自然科学基金(61872424)、南京邮电大学校级科研基金(NY218001, NY219076)和模式识别国家重点实验室开放课题(201900015)资助项目

作者简介:许腾,男,硕士研究生;唐贵进(通信作者),男,博士,副教授,tanggj@njupt.edu.cn

引用本文:许腾,唐贵进,刘清萍,等.基于空洞卷积和 Focal Loss 的改进 YOLOv3 算法[J].南京邮电大学学报(自然科学版),2020,40(6):100-108.

目标检测算法存在前期特征处理复杂,鲁棒性差等缺点。随着计算机硬件和深度学习的不断发展,基于卷积神经网络(Convolutional Neural Networks,CNN)的目标检测算法正在逐渐取代传统目标检测算法。

目前,基于 CNN 的通用目标检测算法分为两种:(1) 二阶(Two-Stage)检测方法,主要代表算法有 R-CNN[5]、SPP-Net[6]、Fast R-CNN[7]、Faster R-CNN[8]、R-FCN[9]等;(2) 一阶(One-Stage)检测方法,主要代表算法有 SSD[10]、YOLOv1[11]、YOLOv2[12]、YOLOv3[13]等。2014 年,R-CNN 算法发布,该算法使用了候选区域方法(Region Proposal Method,RPM)和 CNN 算法,代替了传统的使用滑动窗口和手工设计特征的目标检测方法,使得目标检测算法取得巨大突破。接下来 SPP-Net 算法发布,该算法中提出了空间金字塔池化层(Spatial Pyramid Pooling,SPP),此方法有效地解决了传统 CNN 对输入图像的尺寸的限制。2015 年出现 Fast R-CNN 算法,在该算法中引入了多任务学习,将多个检测步骤整合到一个模型中,同时引入多任务损失函数,使得算法中的网络训练和测试流程得到了简化,因此计算速度得到了极大的提升。同年,Faster R-CNN 发布,其利用候选区域网络(Region Proposal Network,RPN)代替了 RPM,提升了特征提取效率,并降低了训练的时间和空间消耗。由于 Two-Stage 方法将检测任务分成两个阶段:候选框生成和对候选框执行检测,所以运用该方法进行目标检测时精度较高但是速度较慢。2016 年,YOLOv1 算法发布,该算法将生成候选区域和分类回归步骤合二为一,极大地提升了算法的检测速度。接下来又相继发布了 YOLOv2、YOLOv3 等改进版本,在保证检测速度不降低的同时,大大提升了检测准确率。YOLOv2 中提出了名为 Darknet-19 的新网络结构,提升了检测模型的性能。2018 年 YOLOv3 发布,其中主干网络结构从 Darknet-19 改进成为 Darknet-53。同时参考特征金字塔网络(Feature Pyramid Network,FPN)[14],将 YOLOv3 的浅层和深层特征相融合得到了不同尺寸的特征图,并将这些特征图分别进行独立检测。同年出现了 CornerNet[15],提出无锚(Anchor Free)思想,该思想抛弃了原 One-Stage 方法中的锚(Anchor),提升了检测效率的同时降低了检测时间。2019 年提出了 ExtremeNet[16] 和 CenterNet[17],这两种算法在 CornerNet 基础上进行了改进,形成了 One-Stage 方法中新的研究方向,使得目标检测的方法更加多样化,以便适应更多不同的应用场景。

本文首先在 YOLOv3 网络结构上做出改进,设计出了一种新的检测网络结构,提升了模型的小目标检测能力。(1) 将 Darknet-53 的第 2 个残差块输出的特征图用混合空洞卷积(Hybrid Dilated Convolution,HDC)[18] 和 2 倍下采样进行处理;(2) 将 YOLOv3 网络的 8 倍下采样特征图与 16 倍下采样后再进行 2 倍上采样的特征图进行融合;(3) 将(1)和(2)中的特征图再次进行融合,得到新的检测特征。新的检测特征比原特征包含了更多小目标位置信息和语义信息,使得模型对于小目标的检测精度得到了提升。同时将损失函数中负样本置信度部分的交叉熵公式用 Focal Loss[19] 代替,可以降低模型训练时负样本损失在损失函数中的权重,防止模型训练方向被负样本主导,可有效缓解 YOLOv3 的样本不平衡问题,提升 YOLOv3 的通用目标检测能力。实验结果表明,在小目标数量占比为 47.7% 的特定测试集上,改进 YOLOv3 的平均准确率和召回率分别比原 YOLOv3 提高了 8.8% 和 16%;在 VOC 测试集上,改进 YOLOv3 的平均精度均值比原 YOLOv3 提升了 3.4%。

1　YOLOv3 算法简介

YOLO 系列算法是一种端对端(End-to-End)的高效检测算法,把目标检测任务转换成回归问题来实现单幅图像直接检测。YOLOv3 将输入图片划分为 $S\times S$ 的网格,每个网格负责检测 9 个检测框的位置信息、置信度和类别概率。置信度决定了检测框中是否存在物体以及物体的位置是否准确,定义为

$$\Pr(\text{object})\times \text{IOU}_{\text{pred}}^{\text{truth}} \tag{1}$$

其中,Pr(object)为检测框中含有目标的概率,$\text{IOU}_{\text{pred}}^{\text{truth}}$为检测框与实际框的交并比。类别概率为检测框判断出的物体类别的准确率。所有被检测出有物体存在的网格,其包含的 9 个检测框的信息将分配给 3 个不同尺寸的特征图独立地进行检测。再通过非极大值抑制(Non-Maximum Suppression,NMS)[20] 算法处理后得出检测结果。

1.1　网络结构

YOLOv3 的网络结构如图 1 所示,由 YOLOv2 的 19 层增加到 53 层。图中 type 代表卷积层,filters 代表卷积核数量,size 代表卷积核尺寸,output 代表输出特征图尺寸。

	type	filters	size	output
	Convolutional	32	3×3	416×416
	Convolutional	64	3×3/2	208×208
1× Res_unit	Convolutional	32	1×1	
	Convolutional	64	3×3	
	Residual			208×208
	Convolutional	128	3×3/2	104×104
2× Res_unit	Convolutional	64	1×1	
	Convolutional	128	3×3	
	Residual			104×104
	Convolutional	256	3×3/2	52×52
8× Res_unit	Convolutional	128	1×1	
	Convolutional	256	3×3	
	Residual			52×52
	Convolutional	512	3×3/2	26×26
8× Res_unit	Convolutional	256	1×1	
	Convolutional	512	3×3	
	Residual			26×26
	Convolutional	1024	3×3/2	13×13
4× Res_unit	Convolutional	512	1×1	
	Convolutional	1024	3×3	
	Residual			13×13

图 1 YOLOv3 网络结构图

YOLOv3 还借鉴了残差网络(ResNets)[21]的思想,在网络中加入了残差块结构,解决了深层次网络的梯度问题,使得可使用的网络深度大大提升。每个残差块由多个残差单元组成,每个残差单元由两个卷积批归一化激活层(Convolution BN Leaky ReLU Layer, CBLR Layer)和输入相加组成。每个 CBLR Layer 由卷积层、批归一化(Batch Normalization, BN)[22]层和 Leaky ReLU 层组成,如图 2 所示。

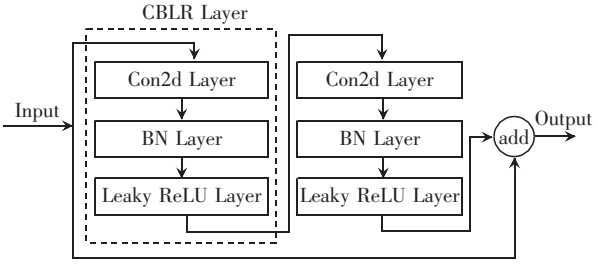


图 2 残差单元结构图

1.2 损失函数

YOLOv3 模型损失函数是由多部分损失加权求和组成,包括检测框损失、置信度损失以及类别损失。检测框损失由中心点坐标损失、宽高尺寸损失两部分组成;置信度损失为无目标检测框置信度损失和有目标检测框置信度损失共同计算求和得出;类别损失为 sigmoid 激活函数计算得到。损失函数完整公式如式(2)所示。

$$\begin{aligned}\text{loss(object)} &= \sum_{i=0}^{K\times K}\sum_{j=0}^{M} I_{ij}^{\text{obj}}(2-w_i\times h_i)\{[\hat{x}_i\log(x_i)+(1-\hat{x}_i)\log(1-x_i)]+[\hat{y}_i\log(y_i)+(1-\hat{y}_i)\log(1-y_i)]\}+\\ &\lambda_{\text{coord}}\sum_{i=0}^{K\times K}\sum_{j=0}^{M} I_{ij}^{\text{obj}}(2-w_i\times h_i)[(w_i-\hat{w}_i)^2+(h_i-\hat{h}_i)^2]-\sum_{i=0}^{K\times K}\sum_{j=0}^{M} I_{ij}^{\text{obj}}[\hat{C}_i\log(C_i)+(1-\hat{C}_i)\log(1-C_i)]-\sum_{i=0}^{K\times K}\sum_{j=0}^{M} I_{ij}^{\text{noobj}}[\hat{C}_i\log(C_i)+(1-\hat{C}_i)\log(1-C_i)]-\sum_{i=0}^{K\times K} I_{i}^{\text{obj}}\sum_{c\in\text{classes}}[\hat{p}_i(c)\log(p_i(c))+(1-\hat{p}_i(c))\log(1-p_i(c))] \end{aligned} \tag{2}$$

其中,$K\times K$ 为输入图片被模型划分后形成的所有网格单元;M 为每个网格包含的检测框数;I_{ij}^{obj} 是判断第 i 个网格单元中第 j 个检测框是否有需要检测的物体;I_i^{obj} 是模型判断物体中心是否位于第 i 个网格中;x_i, y_i, w_i, h_i, c_i 为真实框的位置,尺寸和置信度;$\hat{x}_i, \hat{y}_i, \hat{w}_i, \hat{h}_i$ 为检测框的位置和尺寸,这里注意的是 $\hat{x}_i, \hat{y}_i, \hat{w}_i, \hat{h}_i$ 并非为检测框的绝对位置和尺寸,只是相对于网格中心位置和先验框尺寸的相对值;$\hat{c}_i$ 为检测框的置信度;$p_i(c)$ 为真实框的类别概率;$\hat{p}_i(c)$ 为检测框的类别概率。

在式(2)中,第一项表示检测框的坐标损失,第二项是检测框尺寸损失,第三项是正样本检测框置信度损失,第四项是负样本检测框置信度损失,最后一项是计算类别概率损失。为了防止检测框的损失在损失函数值中占比太大,使用边界框的宽度和长度的平方根来削弱检测框损失的影响。同时在真实图片中,正样本的检测框占所有检测框总数的比例较小,因此正样本的检测框的置信度损失要比负样本的检测框的置信度损失小得多。所以,YOLOv3 添加了参数 λ_{coord} 和权重调整方程式 $(2-w_i\times h_i)$,以调整边界框损失的权重。在式(2)中,将 λ_{coord} 设置为 0.5 以直接减少边界框尺寸损失权重,用 $(2-w_i\times h_i)$ 来限制检测框损失权重。

2 YOLOv3 检测模型的改进方法

文中针对 YOLOv3 改进方法是引入 HDC 增大浅层特征图的感受野,并与深层网络特征图融合以改进检测网络结构,提升模型对小目标的检测能力。同时使用 Focal Loss 替代负样本置信度损失的交叉熵公式,调整模型训练中样本数量失衡问题,提升模型对通用目标的检测性能。本文使用的模型在检测前,会先对输入图片进行预处理使其和网络输入大小一致,检测完成后将图片恢复原状。文中使用的所有模型输入尺寸为 416×416,并且将小于模型输入尺寸 1% 的目标定义为小目标。

2.1 改进的检测网络结构

YOLOv3 在检测时进行上下尺度特征融合,深层网络特征提供目标的语义信息,浅层网络特征则

提供位置信息。然后输出 3 个不同尺寸的检测特征图,使模型在检测大目标和小目标时都有更好的表现。因为所有模型输入尺寸为 416 ×416,所以在文中 YOLOv3 模型的 3 个输出特征图分别为网络 32 倍下采样输出的 13 ×13 特征图,16 倍下采样输出特征图与 13 ×13 特征图上采样后融合的 26 ×26 特征图,8 倍下采样输出特征图与 26 ×26 特征图上采样后融合的 52 ×52 特征图。YOLOv3 检测网络结构如图 3 所示。

虽然 YOLOv3 模型通过特征融合提升了对小目标的检测能力,但是依然存在一些问题没有解决。YOLOv3 模型中用于检测小目标的 8 倍下采样输出的特征图在没有融合浅层网络特征图的情况下,对小目标位置信息的获取能力有限。所以检测的目标像素小于 8 ×8 时,准确率会有明显下降。为了让网络获取到更多小目标的特征信息,提高对小目标的检测准确率,文中将 Darknet-53 中的第 2 个残差块输出的 104 × 104 特征图与 52 × 52 检测特征图融合。

虽然 104 ×104 特征图含有更多的目标位置信息,但是感受野比 52 ×52 的检测特征图小,这两个特征图直接融合不利于检测同一尺寸范围内的目标。所以为了增大感受野,同时保持特征图不丢失小目标的定位信息,引入一种新的卷积操作:空洞卷积(Dilated Convolution)[23],即将 CBLR Layer 改为空洞卷积批归一化激活层(Dilated Convolution BN Leaky ReLU Layer,DCBLR Layer)。空洞卷积是在标准卷积的基础上通过填零操作增大感受野,同时只增加了一个超参数:空洞率(Dilation Rate,DR),不会大幅增加计算量。

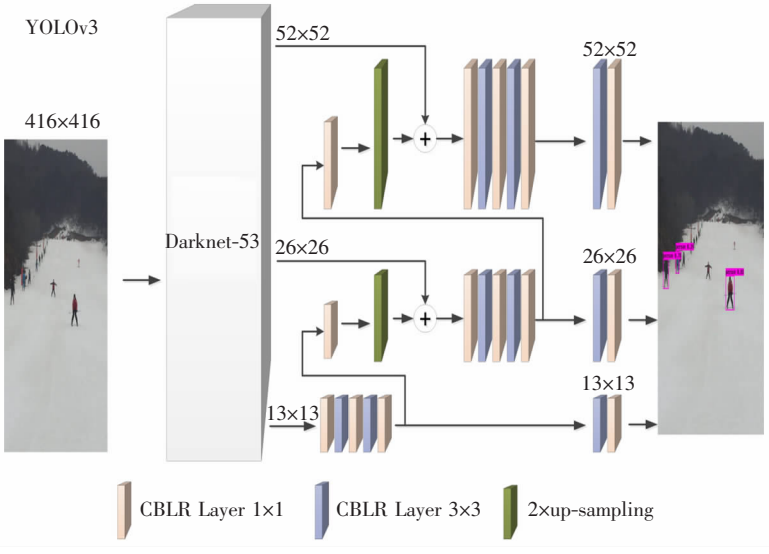
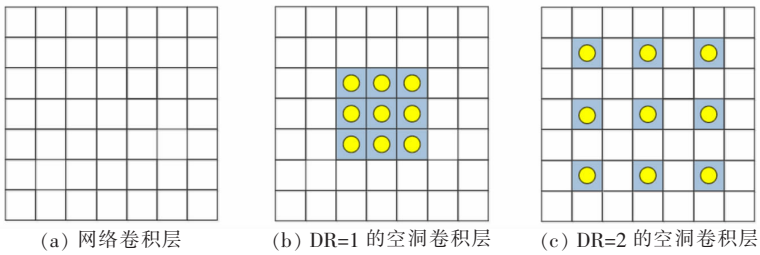


图 3　YOLOv3 检测网络结构

使用单一空洞卷积存在网格效应(Gridding Effect)[18],即空洞卷积在卷积核两个采样像素之间插入 0 值,如果 DR 过大,卷积会过于稀疏,捕获信息能力差。本文采用 HDC:使用 DR 为 1 和 2 的 3 × 3 空洞卷积(如图 4 所示),串联构成块(Block),可以在扩大感受野的同时减轻网格效应的弊端。改进的 YOLOv3 检测网络结构如图 5 所示,将 104 × 104 特征图,通过 HDC block 处理后,经过 2 倍下采样与 52 ×52 检测特征图融合,组成新的52 ×52检测特征图,提升检测小目标的准确率。



(a) 网络卷积层　(b) DR=1 的空洞卷积层　(c) DR=2 的空洞卷积层

图 4　混合空洞卷积组成

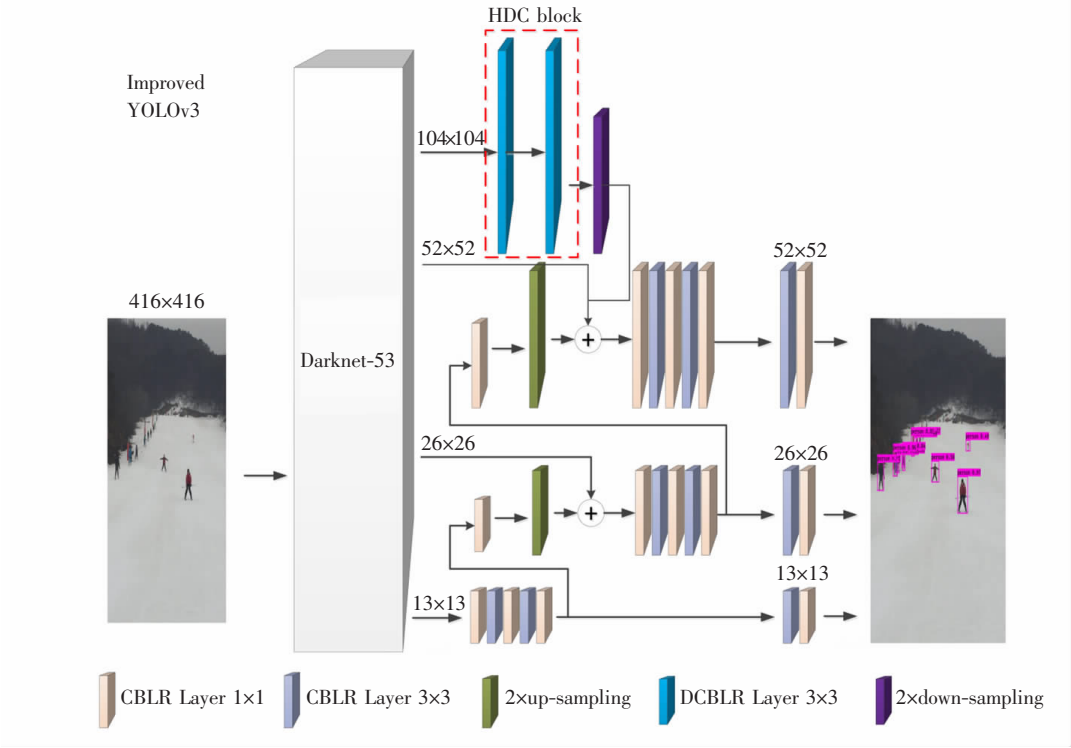


图 5 改进的 YOLOv3 检测网络结构

2.2 改进的损失函数

在 YOLOv3 模型中,作者采用多尺度特征输出来对不同尺寸的目标进行检测。根据像素越高特征图感受野越小的特点,让大尺寸特征图检测较小目标,中等尺寸特征图检测中等尺寸目标,小尺寸特征图检测较大目标,用以改善模型对小目标检测不佳的问题。但是,多尺度检测加剧了正负样本数量的不平衡问题。模型将同时输出 13×13,26×26,52×52三个检测特征图,特征图中的每个像素为一个单元格,每个单元格有三个检测框,即同时输出 10 647 个检测框。但是只有极少部分的检测框包含目标,尤其是在用于检测小目标的高像素特征图中,含有目标的检测框的比重更小。所以在模型训练过程中,大部分的候选框会被标记为负样本,导致训练过程中正负样本的数量差距巨大,使得负样本主导梯度更新的方向,网络无法学习到有用的信息,降低了模型检测准确率。本文引入 Focal Loss 替换损失函数中负样本置信度部分的交叉熵公式,降低了模型训练时负样本损失在损失函数中的权重,以此纠正模型梯度的更新方向,使网络学到更有用的信息,对目标进行更准确的检测。

在模型中原本使用交叉熵公式来计算负样本置信度损失,二元分类交叉熵如式(3)所示。

$$L_{ce} = -y\log y' - (1-y)\log(1-y') = \begin{cases} -\log y', & y=1 \\ -\log(1-y'), & y=0 \end{cases} \tag{3}$$

其中,y 表示真值,1 为正样本,0 为负样本。y'表示检测值,范围从 0 到 1。对于正样本来说,检测概率越高,交叉熵损失越小。而对于负样本,检测概率越低,交叉熵损失越小。所以用大量负分类样本进行迭代训练,模型无法优化到理想状态。

Focal Loss 是基于交叉熵公式改进而来,主要是为了解决 YOLOv3 中正负样本比例严重失衡的问题。Focal Loss 公式如下:

$$L_{fl} = -\alpha y(1-y')^{\gamma}\log y' - (1-\alpha)(1-y)y'^{\gamma}\log(1-y') = \begin{cases} -\alpha(1-y')^{\gamma}\log y', & y=1 \\ -(1-\alpha)y'^{\gamma}\log(1-y'), & y=0 \end{cases} \tag{4}$$

如式(4)所示,首先在原有的交叉熵公式基础上加了一个因子 γ。当 $\gamma=0$ 时,Focal Loss 就为交叉熵公式。当 $\gamma>0$ 时,会让损失函数减少对易分类样本的计算,增加对难分类样本的计算。此外,加入了平衡因子 α,用来调整正负样本数量失衡的问题。实验发现当 $\gamma=2$ 和 $\alpha=0.9$ 时,实验效果较好。

3 实验结果与分析

3.1 实验准备

本文将 VOC2007 和 VOC2012 的训练验证集合作为本文实验的训练集,同时随机将其中 10% 的训练集图片作为验证集。为了验证模型在小目标检测和通用目标检测两方面都有明显提升,通过两个实

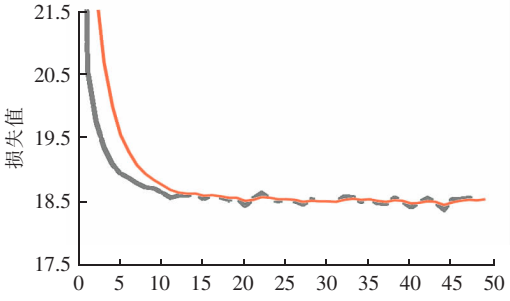
验来验证改进后的模型性能:(1) 使用网络图库中选取的 295 张图片作为测试集 1,测试集 1 中标注了 1 171 个单类目标,其中包含 559 个小目标,占比 47.7%,然后将原模型和改进后的模型进行实验对比;(2) 选取 VOC2007 测试集作为测试集 2,对所有 20 类目标进行检测,最后将改进后的算法与原算法以及其他典型算法在相同数据集上进行实验对比。

本文实验所使用配置环境:深度学习框架为 Keras + Tensorflow,操作系统为 Windows 10,CPU 为 i7-8700k,运行内存为 16 GB,显卡内存为 11 GB,GPU 为 NVIDIA(R) GTX(R) 1080Ti。

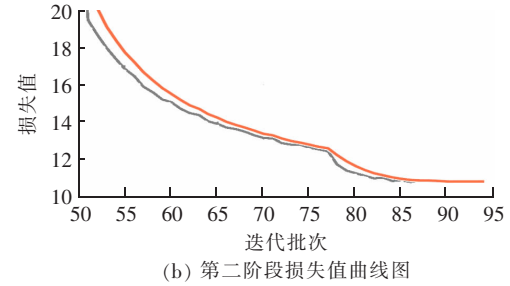
3.2　模型训练

实验中采用迁移学习的方式,分两个阶段训练。第一阶段冻结整个网络中的前 260 层,使用原始的 YOLOv3 权重进行迁移学习,只训练最后 3 层网络。第二阶段解冻前 260 层,对全网络进行训练。第一阶段学习率设为 0.001,使损失值快速达到一个较低的水平;第二阶段学习率设为 0.000 1,减少曲线振荡,使损失值平稳减小到最低,然后得到改进后的 YOLOv3 网络权重。

训练过程中损失函数值的曲线图如图 6 所示,深色为训练损失曲线,浅色为验证损失曲线,横坐标代表迭代批次(Epoch),纵坐标代表损失值。



(a) 第一阶段损失值曲线图



(b) 第二阶段损失值曲线图

图 6　模型训练两阶段损失函数曲线图

从图 6 可以看出第一阶段一开始损失值下降较快,迭代到 20 个 epoch 之后下降变缓,最后收敛至 18.5 左右,如图6(a)所示。然后开启第二阶段训练,迭代至 90 个 epoch 时比较稳定,最后收敛至 11 左右得出整个网络的权重,如图 6(b)所示。

3.3　实验结果及对比

3.3.1　小目标实验

为了验证改进后的 YOLOv3 的小目标检测能力,将原 YOLOv3 和改进后的 YOLOv3 在小目标占比为 47.7% 的测试集 1 中进行对比实验。首先计算模型对目标检测的准确率和召回率。目标检测的准确率(Precision,P)和召回率(Recall,R)定义可分别表示为:

$$P = \frac{D_{TP}}{D_{TP} + D_{FP}} \tag{5}$$

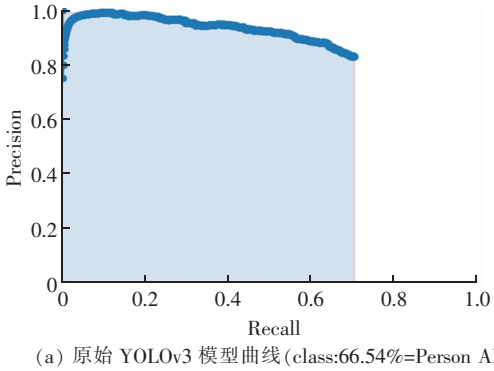
$$R = \frac{D_{TP}}{D_{TP} + D_{FN}} \tag{6}$$

式(5)和式(6)中,D_{TP}表示检测正确的目标数;D_{FN}表示未检测出的目标数;D_{FP}表示检测错误的目标数。

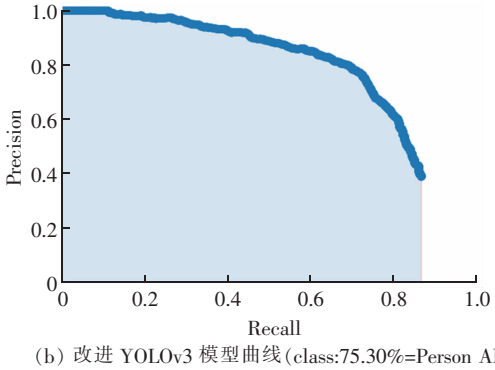
对于同类小目标来说,使用 Precision-Recall 曲线图,可以表示模型对 Precision 与 Recall 两个指标的权衡优化。平均准确率(Average Precision,AP)是评价模型检测性能的常用指标,表示为 Precision-Recall 曲线的积分计算结果,如式(7)所示

$$AP = \int_0^1 P(r)\,dr, r \in (0,1) \tag{7}$$

在测试集 1 中分别对两个模型进行测试,平均准确率 AP 和召回率 R 如图 7 所示。



(a) 原始 YOLOv3 模型曲线(class:66.54%=Person AP)



(b) 改进 YOLOv3 模型曲线(class:75.30%=Person AP)

图 7　两个模型的 Precision-Recall 曲线

两个模型的平均准确率 AP、召回率 R 和每秒帧数(Frames Per Second,FPS)的数值如表 1 所示。

表 1 两个模型 AP、R 和 FPS 实验结果对比

Detection model	AP/%	R/%	FPS
YOLOv3	66.5	70.5	23
Improved YOLOv3	75.3	86.5	21

与原 YOLOv3 相比,改进的 YOLOv3 对测试集 1 的检测平均准确率由 66.5% 提高到了 75.3%,召回率由 70.5% 提高到了 86.5%。实验结果表明,在测试集 1 中,改进的 YOLOv3 检测性能与原 YOLOv3 相比,有明显的提升,FPS 略有下降。

原 YOLOv3 和改进的 YOLOv3 在测试集 1 下的检测结果如图 8 所示。每一对图片的左边为 YOLOv3 检测结果,右边为改进的 YOLOv3 的检测结果。可以看出,改进后的 YOLOv3 减少了 YOLOv3 中存在的错检和漏检问题。

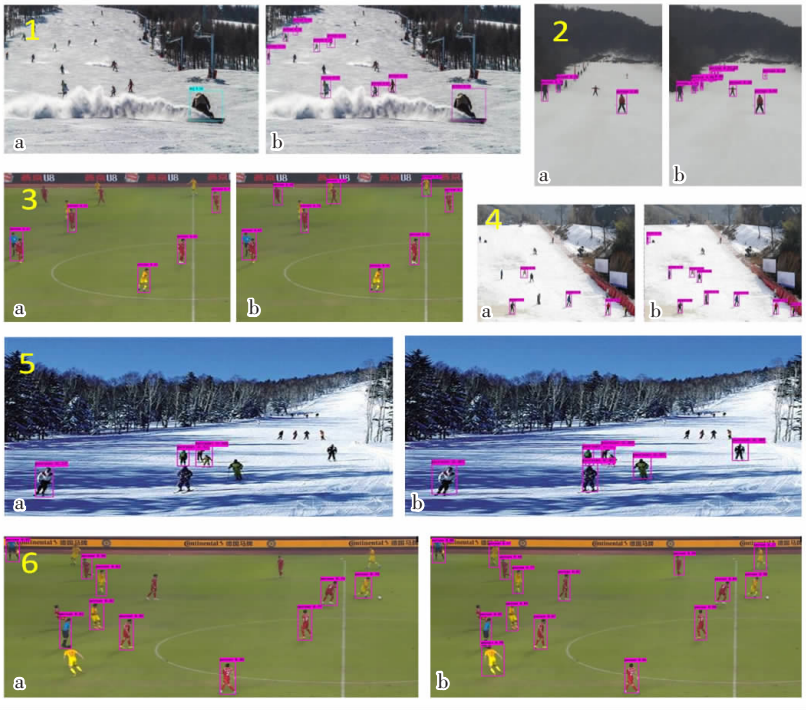


图 8 YOLOv3 模型和改进的 YOLOv3 模型在测试集 1 中的检测结果(a 为 YOLOv3 结果,b 为改进的 YOLOv3 结果)

3.3.2 通用目标实验

在测试集 2 中分别对改进的 YOLOv3 和 YOLOv3 进行测试,计算各类目标的平均准确率,同时计算所有目标的 mAP。mAP 是目标检测算法中最常用且最重要的评价指标,是用来衡量模型综合性能的最佳评判标准之一,其计算公式如式(8)所示:

$$\mathrm{mAP} = \frac{\sum \mathrm{AP}_C}{N(\mathrm{Classes})} \tag{8}$$

其中,AP_C 为每个类别的平均准确率,$N(\mathrm{Classes})$ 为数据集中的类别数量。

(1) 本文改进方法增益对比

本文为提升 YOLOv3 算法检测性能,采用空洞卷积和 Focal Loss 对 YOLOv3 进行改进。两种改进方法对模型的检测性能提升对比如表 2 所示。NM 表示原 YOLOv3 算法,HDC 表示只使用空洞卷积改进的 YOLOv3 算法,FL 表示只使用 Focal Loss 改进的 YOLOv3 算法,HDC + FL 表示使用了两个方法改进的 YOLOv3 算法。

由表 2 可知,两种方法对 YOLOv3 改进都是有效的,但是 Focal Loss 对于模型检测性能的增益比空洞卷积要高。在两种方法同时使用的情况下,检测结果比单独使用之和更好。在加入 Focal Loss 后对模型检测速度基本没有影响,但是由于空洞卷积的加入,改进的 YOLOv3 的网络结构层数有所增加,所以检测速度会略有下降。

表 2 改进的两种方法对性能提升的比较

Method	mAP/%	Improvement Ratio/%	FPS
NM	78.1	0	23
HDC	79.3	1.2	21
FL	80.1	2.0	23
HDC + FL	81.5	3.4	21

(2) 不同改进方式性能对比

将本文使用的改进方法与文献[24 - 25]中的

方法进行对比,展示其在同一测试集上对 YOLOv3 检测性能的提升能力,具体结果如表 3 所示。其中文献[24]包含两种方法 M1、M2,分别使用 IOU、GIOU 进行改进。而 M3 方法来自文献[25],对网络和损失函数进行了改进。由表 3 可知,在相同条件下,文中采用的方法对模型性能提升效果明显。

表 3　不同算法相比 YOLOv3 的性能提升对比

Method	M1[24]	M2[24]	M3[25]	ours
Improvement Ratio/%	2.02	4.08	0.80	3.40

(3) 模型性能对比

改进后 YOLOv3 与原 YOLOv3 及其他典型的算法的实验结果对比如表 4 所示。Faster R-CNN、SSD_300、YOLOv2_416、MLKP、DetNAS、Auto-FPN 的实验结果分别来自文献[8]、[10]、[12]、[26]、[27]、[28],YOLOv3_416 与改进的 YOLOv3 实验结果由本文提供。由表 4 可知,改进后的 YOLOv3 比原 YOLOv3 的 mAP 值高 3.4%,说明本文采用的改进措施可以有效提升模型的通用目标检测性能,而且改进后的 YOLOv3 检测性能高于表中大部分的算法。

表 4　不同算法实验结果对比

Method	Network	Dataset	mAP/%
Faster R-CNN[8]	VGG	VOC 07 + 12	73.2
SSD_300[10]	VGG	VOC 07 + 12	72.4
YOLOv2_416[12]	Darknet-19	VOC 07 + 12	76.8
MLKP[26]	ResNet-101	VOC 07 + 12	80.6
DetNAS[27]	ResNet-101	VOC 07 + 12	80.1
Auto-FPN[28]	ResNet-50	VOC 07 + 12	81.8
YOLOv3_416	Darknet-53	VOC 07 + 12	78.1
Ours	Darknet-53	VOC 07 + 12	81.5

原 YOLOv3 和改进的 YOLOv3 在测试集 2 下的检测结果如图 9 所示。每一对图片的左边为原 YOLOv3 检测结果,右边为改进的 YOLOv3 检测结果。由图可知,改进的 YOLOv3 减少了 YOLOv3 中存在的错检和漏检问题。

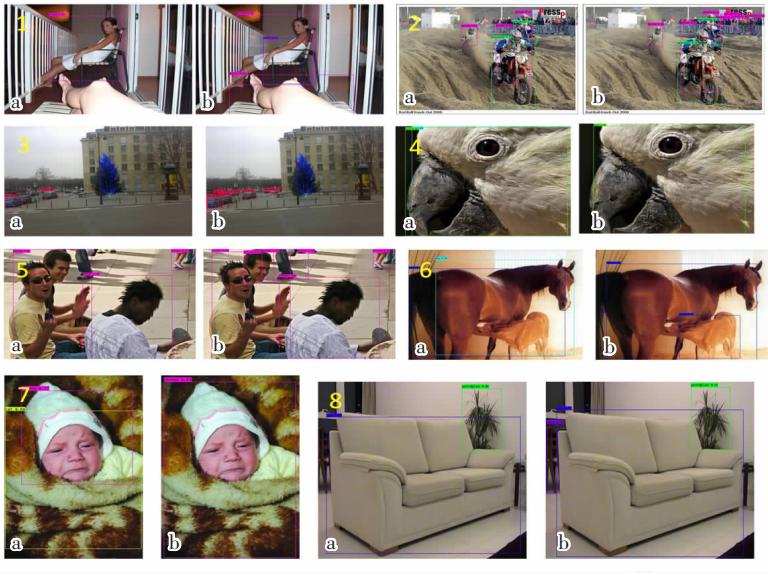


图 9　YOLOv3 模型和改进的 YOLOv3 模型在测试集 2 下的检测结果(a 为 YOLOv3 结果,b 为改进后的 YOLOv3 结果)

4　结束语

本文提出了一种 YOLOv3 改进方法,提高了 YOLOv3 模型的检测性能。文中首先改进了检测网络结构,使用混合空洞卷积模块和 2 倍下采样处理 Darknet-53 中的第 2 个残差块输出的 104 × 104 特征图,然后与 52 × 52 检测特征图进行融合,建立新的预测特征图。同时引进 Focal Loss 来改进损失函数,替换负样本置信度中的交叉熵公式。实验结果表明,改进后的 YOLOv3 模型与原 YOLOv3 模型相比,对通用目标和小目标的检测能力都有明显的提高,但是检测速度有略微下降。所以如何在不降低现有检测性能的情况下,简化网络结构,降低计算量,是未来还需要继续探索的方向。

参考文献:

[1] VIOLA P, JONES M. Rapid object detection using a boosted cascade of simple features[C]// Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition. 2001:511 - 518.

[2] LOWE D G. Distinctive image features from scale-invariant keypoints[J]. International Journal of Computer Vision, 2004,60(2):91 - 110.

[3] CORTES C, VAPNIK V. Support-vector networks[J]. Machine Learning, 1995, 20(3): 273 – 297.

[4] FERREIRA A J, FIGUEIREDO M A T. Boosting algorithms: a review of methods, theory, and applications[M] // Ensemble Machine Learning. Boston: Springer, 2012: 35 – 85.

[5] GIRSHICK R, DONAHUE J, DARRELL T, et al. Rich feature hierarchies for accurate object detection and semantic segmentation[C] // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2014: 580 – 587.

[6] 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.

[7] GIRSHICK R. Fast R-CNN[C] // Proceedings of the IEEE International Conference on Computer Vision. 2015: 1440 – 1448.

[8] 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.

[9] DAI J, LI Y, HE K, et al. R-FCN: object detection via region-based fully convolutional networks[C] // Advances in Neural Information Processing Systems. 2016: 379 – 387.

[10] LIU W, ANGUELOV D, ERHAN D, et al. SSD: single shot MultiBox detector[C] // Proceedings of the European Conference on Computer Vision. 2016: 21 – 37.

[11] REDMON J, DIVVALA S, GIRSHICK R, et al. You only look once: unified, real-time object detection[C] // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2016: 779 – 788.

[12] REDMON J, FARHADI A. YOLO9000: better, faster, stronger[C] // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2017: 7263 – 7271.

[13] REDMON J, FARHADI A. Yolov3: an incre-mental improvement[EB/OL]. [2020-02-18]. https: // arxiv. org/abs/1804. 02767.

[14] LIN T Y, DOLLAR P, GIRSHICK R, et al. Feature pyramid networks for object detection[C] // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2017: 2117 – 2125.

[15] LAW H, DENG J. Cornernet: detecting objects as paired keypoints[C] // Proceedings of the European Conference on Computer Vision. 2018: 734 – 750.

[16] ZHOU X, ZHOU J, KRAHENBUHL P. Bottom-up object detection by grouping extreme and center points[C] // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2019: 850 – 859.

[17] DUAN K, BAI S, XIE L, et al. CenterNet: keypoint triplets for object detection[C] // International Conference on Computer Vision. 2019: 6569 – 6578.

[18] WANG P, CHEN P, YUAN Y, et al. Understanding convolution for semantic segmentation[C] // IEEE Winter Conference on Applications of Computer Vision. 2018: 1451 – 1460.

[19] LIN T Y, GOYAL P, GIRSHICK R, et al. Focal loss for dense object detection[C] // Proceedings of the IEEE International Conference on Computer Vision. 2017: 2980 – 2988.

[20] NEUBECK A, VAN G L. Efficient non-maximum suppression[C] // 18th International Conference on Pattern Recognition. 2006: 850 – 855.

[21] HE K, ZHANG X, REN S, et al. Deep residual learning for image recognition[C] // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2016: 770 – 778.

[22] IOFFE S, SZEGEDY C. Batch normalization: accelerating deep network training by reducing internal covariate shift[C] // International Conference on Machine Learning. 2015: 448 – 456.

[23] YU F, KOLTUN V, FUNKHOUSER T. Dilated residual networks[C] // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2017: 472 – 480.

[24] REZATOFIGHI H, TSOI N, GWAK J Y, et al. Generalized intersection over union: a metric and a loss for bounding box regression[C] // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2019: 658 – 666.

[25] 赵琼, 李宝清, 李唐薇. 基于改进 YOLO v3 的目标检测算法[J]. 激光与光电子学进展, 2020, 57(12): 313 – 321.
ZHAO Qiong, LI Baoqing, LI Tangwei. Target detection algorithm based on improved YOLO v3[J]. Laser & Optoelectronics Progress, 2020, 57(12): 313 – 321. (in Chinese)

[26] WANG H, WANG Q, GAO M, et al. Multi-scale location-aware kernel representation for object detection[C] // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2018: 1248 – 1257.

[27] CHEN Y, YANG T, ZHANG X, et al. DetNAS: backbone search for object detection[C] // Advances in Neural Information Processing Systems. 2019: 6638 – 6648.

[28] XU H, YAO L, ZHANG W, et al. Auto-FPN: automatic network architecture adaptation for object detection beyond classification[C] // Proceedings of the IEEE International Conference on Computer Vision. 2019: 6649 – 6658.

卷 终