

文章编号: 2095-2163(2023)02-0179-08

中图分类号: TP391.4

文献标志码: A

基于通道注意力与空间金字塔的改进型 YOLOv3 及其应用

游 双¹, 张著洪^{1,2}

(1 贵州大学 大数据与信息工程学院, 贵阳 550025; 2 贵州大学 贵州省系统优化与科学计算特色重点实验室, 贵阳 550025)

摘要: 针对 YOLOv3 存在行人遮挡、漏检下检测精度低、漏检率高的问题, 本文提出一种基于通道注意力及空间金字塔的改进型 YOLOv3。将通道注意力机制嵌入残差网络中, 强化关键信息的特征提取; 利用能增大感受野的空间金字塔融合多尺度特征图, 增强相互遮挡目标的特征提取能力; 利用改进型非极大抑制模块消除冗余预测框, 避免重叠目标漏检。比较性的数值实验表明, 相较于 YOLOv3, 改进型 YOLOv3 的检测准确率及综合评估指标值分别提高了 9.33% 和 5.01%, 且行人检测的鲁棒性和泛化能力更强。

关键词: 特征提取; YOLOv3; 注意力机制; 空间金字塔

Channel attention and spatial pyramid-based improved YOLOv3 and its application

YOU Shuang¹, ZHANG Zhuhong^{1,2}

(1 College of Big Data and Information Engineering, Guizhou University, Guiyang 550025, China; 2 Guizhou Provincial Characteristic Key Laboratory of System Optimization and Scientific Computation, Guizhou University, Guiyang 550025, China)

[Abstract] Aiming at the problems of pedestrian occlusion, low detection accuracy and high misdetection rate in YOLOv3, an improved YOLOv3 is proposed based on channel attention and spatial pyramid. In the design of model, the channel attention mechanism is embedded into the residual network to strengthen the feature extraction of key information. The spatial pyramid that can enlarge the receptive field is used to fuse multi-scale feature maps in order to enhance the detection of mutually occluded targets. The non-maximum suppression model is improved and used to eliminate redundant prediction frames and avoid detection failures of overlapping targets. The comparative experiments have validated that compared with YOLOv3, the detection accuracy and comprehensive evaluation index value of the improved YOLOv3 are increased by 9.33% and 5.01% respectively, and meanwhile the robustness and generalization abilities of pedestrian detection are stronger.

[Key words] feature extraction; YOLOv3; attention mechanism; spatial pyramid

0 引言

随着中国社会经济的快速发展和科技水平的快速提升, 行人检测在智能监控、人脸识别、自动驾驶等各方面具有广泛的应用和重要的研究价值。由于行人图像受背景遮挡、姿态及拍摄角度等影响, 给行人检测增加了难度。目前, 行人检测主要包括传统检测法和深度学习法两类。传统检测法是在确定输入图像中目标所在候选区域基础上提取目标特征, 并依据目标分类器对目标进行识别和类别划分^[1]。大致可分为基于全局特征的检测法和基于运动检测的方法, 前者采用图像的静态特征, 如边缘、形状、纹理等, 通过分类器对图像的目标进行识别和类别划

分, 但在出现分辨率低、目标遮挡和拥挤情形时, 分类准确率低^[2]; 后者由背景建模法和分类器两部分组成, 其中背景建模法利用当前帧与前一帧获取运动目标, 其检测速度快, 但目标检测易于受光照、阴影等影响, 导致目标分类准确率低^[3-5]。例如, Gong 等^[6]提出一种基于高斯背景建模、方向梯度直方图和支持向量机的分类算法, 通过前景分割、特征降维和信息更新 3 个步骤进行行人检测, 具有较强的实时性和高检测准确率, 但计算效率低, 易于产生过多冗余窗口, 进而导致检测速度慢且检测精度低。

深度学习法通常利用卷积神经网络 (Convolution neural network, CNN) 提取图像特征, 利用大样本图像训练网络参数, 进而对行人进行检测^[7]。其

基金项目: 国家自然科学基金 (62063002)。

作者简介: 游 双 (1996-), 男, 硕士研究生, 主要研究方向: 智能优化算法; 张著洪 (1966-), 男, 博士, 教授, 博士生导师, 主要研究方向: 数据科学与计算智能、深度学习。

通讯作者: 张著洪 Email: zhzhong@gzu.edu.cn

收稿日期: 2022-09-13

能克服传统检测法易于出现检测速度慢、鲁棒性弱等缺陷,且能满足实时性与准确性的要求^[8]。涉及的主要方法包括候选区域选择法和逻辑回归法,候选区域选择法首先利用滑动窗口获取参考运动目标,利用 CNN 提取运动目标的特征,进而利用分类器对特征进行分类,该方法虽能高精度地检测行人,但检测速度较慢,实时性较差^[9-10]。例如,任少卿等^[11]利用区域生成网络(Region proposal network, RPN)生成感兴趣区域,进而与 CNN 结合,获得能提高检测速度的 Fast R-CNN 目标检测模型,但检测效率有待提高。逻辑回归法利用图像卷积提取特征,并通过回归获得目标对象所在边框和类别信息,其目标检测速度快、实时性强且应用范围广,但存在计算量较大,占用内存大等问题^[12-13]。例如,刘建国等^[14]将局部二值模式表示的纹理特征作为 YOLOv3 的前置模块,获得改进型 YOLOv3,有效提高行人检测的准确率,但特征提取能力较差且耗时较长;黄凤琪等^[15]利用 K-means++ 算法重构目标锚框,并将其与 YOLO 算法融合,获得的 DCD-YOLO 算法,能提高小目标行人检

测精度,但实时检测速度较慢,占用内存较大。

综上,传统检测算法存在检测精度低、鲁棒性弱等缺点,难以满足实时性要求,而深度学习与行人检测的结合,能高效率检测视频序列中的行人。为此,本文针对密集场景中行人检测问题,在传统 YOLOv3 基础上,借助梯度下降法并结合通道注意力机制和空间金字塔模型,尝试性地提出一种改进型 YOLOv3(Improved YOLOv3, IYOLOv3)行人检测模型,并应用于密集场景的重叠行人检测。

1 YOLOv3 模型

Joseph Redmon 等于 2018 年提出一种适用于行人检测的端对端 CNN,即 YOLOv3。以尺度为 $M \times M$ 的灰度图为输入,经 CNN 输出图像中行人所在矩形框的中心位置 (x, y) 、宽 ω 及高 h ,即 4 维向量 (x, y, ω, h) 。经由逻辑回归获取边界框的位置,利用特征金字塔网络(Feature Pyramid Networks, FPN)以及不同尺度图像框提取行人的特征,最终实现对行人的检测^[16],其网络结构如图 1 所示。

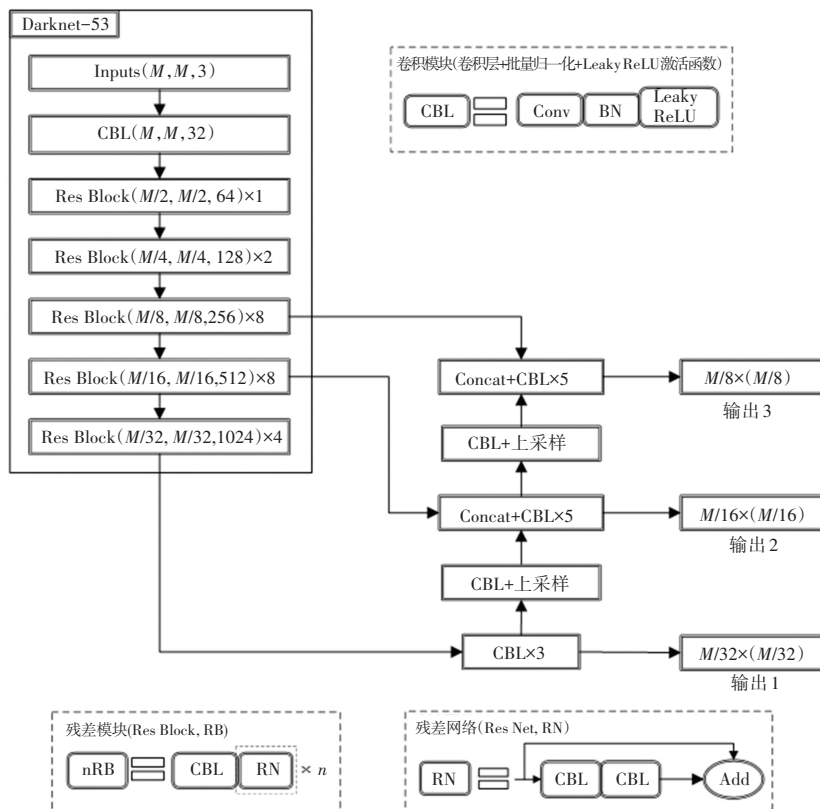


图 1 YOLOv3 网络结构

Fig. 1 Network structure of YOLOv3

YOLOv3 由两部分构成,即主干特征提取网络 Darknet-53 和多尺度检测网络。Darknet-53 由 5 个

残差结构组成,依次包含 1、2、8、8、4 个残差模块。每个残差结构经由 3×3 图像框进行下采样,即依次对输入

图片进行5次下采样,使得每一次图像框的大小压缩一半,但通道数翻倍。每个残差块(Res Block, RB)由卷积块(Convolutional Batch Normalization Leaky ReLU, CBL)和残差网络(Res Net, RN)构成;每个残差网络由2个CBL和1个快捷链路层堆叠而成。在此,每个CBL主要由特征提取的卷积层、特征归一化处理的批量归一化层(Batch Normalization)、激活函数层(Leaky ReLU)构成;另外,多尺度检测网络由3个输出层、上采样层、若干卷积层和Concat层组成。

YOLOv3以Darknet-53作为基础网络提取特征,利用FPN网络融合不同层级的特征信息,进而在3个不同的尺度上进行行人检测。以416×416尺寸的输入图像为例,图中输出1是将Darknet-53的第五个残差结构输出的特征图经过一系列卷积后,以13×13的维度进行目标检测;输出2是通过Darknet-53中第五个残差结构输出的特征图,并经由上采样后与第四个残差结构输出的特征图融合,按26×26维度进行目标检测;输出3是将第三个残差结构的特征图与后两个残差结构输出的特征图融合,按52×52维度进行目标检测。如此3种输出分别对应深层、中层和浅层特征,其中深度特征图具有尺寸小、感受野大的特点,有利于大尺度行人的检测。反之,浅层特征图则便于小尺度行人的检测。因此,该模型既可以检测大行人目标,也可以检测小行人目标。

对于给定尺寸为 $M \times M$ 的灰度图 A_i ,将其划分为 $S \times S$ 个子图;对于第 j 个子图,依据尺度13×13、26×26、52×52获得参考图像框的真实中心位置、宽度、高度分别为 $(\hat{x}_{ij}, \hat{y}_{ij})$ 、 $\hat{\omega}_{ij}$ 及 \hat{h}_{ij} 。

另一方面,YOLOv3作用于第 j 个子图,经由尺度分别为13×13、26×26、52×52的特征图、权值矩阵 \mathbf{W} 、阈值向量 \mathbf{b} ,输出包含行人的图像框的中心位置、宽度、高度,即 (x_{ij}, y_{ij}) 、宽 ω_{ij} 及高 h_{ij} ,进而经由梯度下降法求解式(1)优化问题:

$$\min(\mathbf{W}, \mathbf{b}) = \frac{1}{K} \sum_{i=1}^K \sum_{j=1}^{S \times S} ((x_{ij} - \hat{x}_{ij})^2 + (y_{ij} - \hat{y}_{ij})^2 + (\omega_{ij} - \hat{\omega}_{ij})^2 + (h_{ij} - \hat{h}_{ij})^2) \quad (1)$$

2 改进型YOLOv3的设计与算法

尽管YOLOv3用于行人目标检测具有检测速度快、通用性强等优势,但其自身存在的固有缺陷使得目标检测效果仍需作进一步改善。主要表现如下:

(1) 主干特征网络(Darknet-53)的多尺度检测能力不足,导致网络难以提取远距离小目标及重叠

行人的特征;

(2) 金字塔结构的多尺度特征层所包含的语义信息匮乏,感受野较窄,直接影响后续行人的检测效果;

(3) 在相互遮挡场景下,行人的先验预测框与真实框互相重叠,将对网络损失函数值计算造成误差;

(4) 在检测阶段,非极大值抑制(Non-maximum suppression, NMS)模型容易将被遮挡的行人先验预测框判定为虚假预测框,使得网络无法识别被遮挡的行人,从而出现漏检现象,导致检测精度降低。

基于以上因素,在此提出IYOLOv3网络结构,并用于解决YOLOv3存在行人检测精度低、定位目标的误差大和漏检严重等问题,如图2所示。

IYOLOv3由3部分构成,即基于注意力机制中压缩与激励(Squeeze and Excitation, SE)与Darknet-53中残差网络ResNet的SE-ResNet、空间金字塔池化(Spatial pyramid pooling, SPP)及改进型非极大值抑制模型(Improved nonmaximum suppression, INMS)。对比图1和图2获知,IYOLOv3是在YOLOv3基础上增加SPP、INMS及Concat+CBM×3。在图2中,SE-ResNet被用于增强特征提取和表示能力。与YOLOv3的主干特征提取网络的区别原ReLU激活函数被Mish函数取代,同时原残差模块(Res Block)被扩展为含通道注意力机制的SE-Res Block,进而获得改进型主干网络SE-Darknet-53;另外,SPP由尺度不同的3个最大池化层构成,是一个经由3个卷积块(CBM)与主干网络(SE-Darknet-53)连接的空间金字塔池化模块,将池化后获得的特征图传递给通道拼接层(Concat+CBM×3),扩大主干特征接收范围,提升上下文的特征分离能力,提取的特征图经由通道拼接层处理后传送给INMS检测层。

2.1 模型结构设计

2.1.1 SE-ResNet

鉴于Darknet-53较难提取视觉图像中模糊或者较小行人目标,在此将SE与Darknet-53中的残差网络ResNet结合,获得的基于SE的残差网络(SE-ResNet)更加关注行人特征,使得YOLOv3的行人特征判别能力得到增强,SE模块示意图如图3所示。SE-ResNet是一种注意力机制下的深度学习方法,通过压缩与激励操作得到一组权重,然后对各通道产生的特征图加权。即通过压缩与激励操作获得各特征图所述特征通道的重要程度,增强有用信息所述特征通道的输出能力,以便提升特征提取能力。

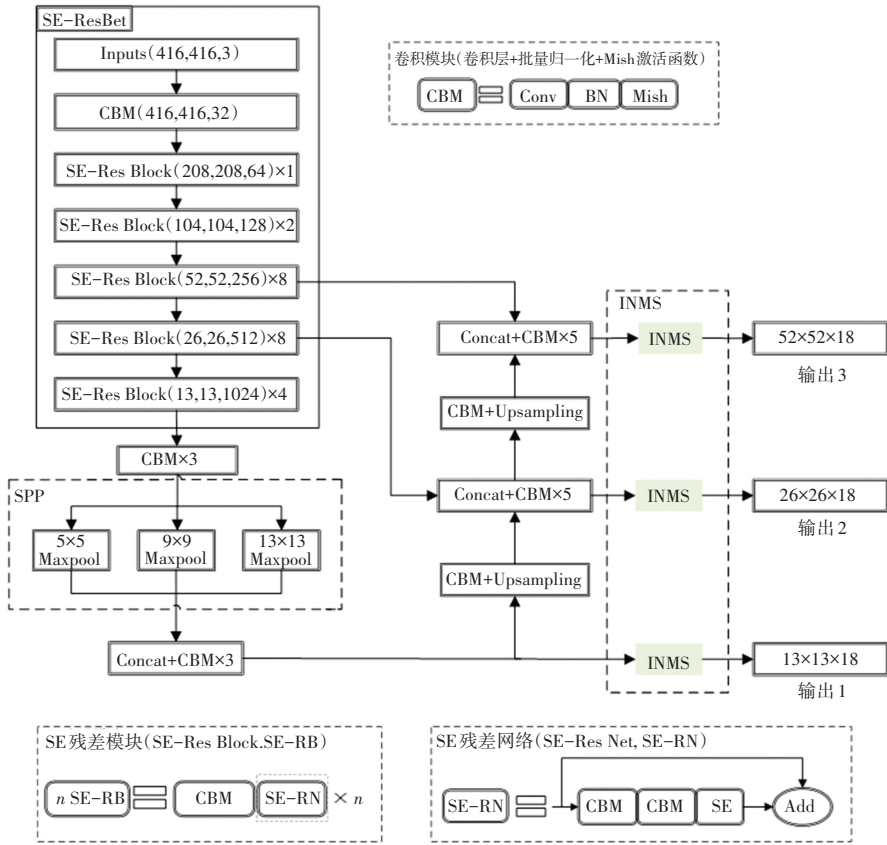


图2 IYOLOv3 网络结构

Fig. 2 Network structure of IYOLOv3

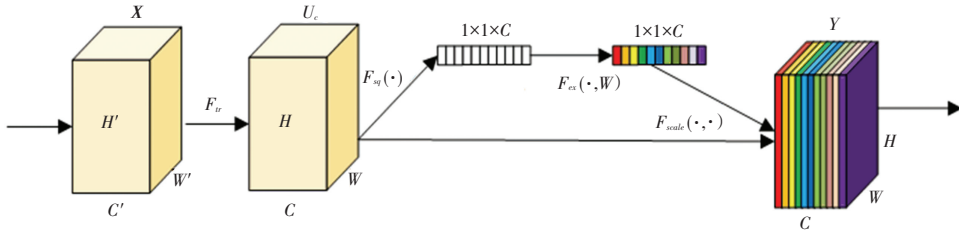


图3 SE 模块示意图

Fig. 3 Schematic diagram of SE module

图3中,对于第 m 通道的特征图 X ,经由 n 个卷积 F_r 变换获得尺度为 $H \times W$ 的 n 个特征图 $U_{ci}, 1 \leq i \leq n$;随后,依次经由压缩、激励、加权 3 步确定特征图的权重。

首先,借助全局平均池化及特征压缩操作 $F_{sq}(\cdot)$,依据式(2)将第 i 个 $H \times W$ 特征图退化为一个实数,用于表征第 i 特征通道上响应的全局分布。

$$Z_{ci} = F_{sq}(U_{ci}) = \frac{1}{H \times W} \sum_{k=1}^H \sum_{l=1}^W U_{ci}(k, l) \quad (2)$$

其次,经由式(2)将 i 个大小为 $H \times W$ 的特征图变成 C 个实数构成的向量表示为 Z_c 。激励操作包含两个全连接层及 $ReLU$ 激活函数和 $Sigmoid$ 函数。利用两个全连接层对特征图降维与升维;第一个全连接

层将特征维度降低到原来的 C/r 倍,再利用 $Relu$ 激活函数实现对输入特征的非线性表达,其中 r 为缩放系数,一般设置为 16;第二个全连接层用来恢复特征图的维度,进而经过 $Sigmoid$ 激活函数获得归一化权重向量, A_c 为特征激励后得到的权重向量,即每个特征图 U_{ci} 经激励后得到的重要程度构成的向量,其计算过程如式(3):

$$A_c = F_{ex}(Z_c, W) = \rho(W_2 \delta(W_1 Z_c)) \quad (3)$$

其中, $F_{ex}(\cdot)$ 表示压缩操作; W_1 表示第一个全连接层的权重矩阵; $\delta(\cdot)$ 表示 $ReLU$ 激活函数; W_2 表示第二个全连接层的权重; $\rho(X)$ 表示向量 X 中每个分量经由 $Sigmoid$ 函数作用后获得的向量。

最后,对 A_c 中的分量作归一化处理,得到的第 i

个分量作为第 i 个特征图 U_i 所在通道的权重。

2.1.2 空间金字塔池化(SPP)

复杂场景中行人目标尺度变化大,YOLOv3 的多尺度预测侧重于多尺度卷积层的全局特征串联,忽略了同一卷积层上多尺度局部特征的融合。在 Darknet-53 的最后一个特征层增加空间金字塔池化(Spatial Pyramid Pooling, SPP)模块,SPP 模块结构如图4所示。SPP 通过不同尺度的最大池化操作融合全局、局部特征通道,获取特征图中不同尺度的感受野,增强特征图的表达能力。

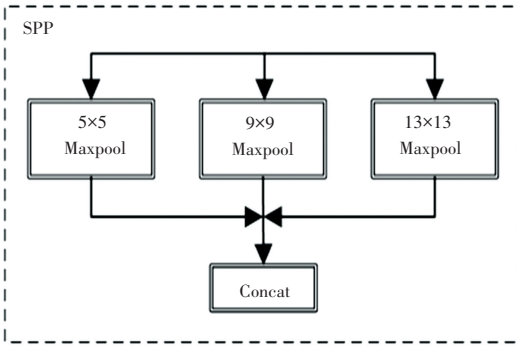


图4 SPP 模块结构

Fig. 4 SPP module structure

为了增强多尺度特征的适应性,SPP 模块由3个尺度分别为 5×5 、 9×9 、 13×13 的最大池化层和 Concat 通道组成;Concat 接收此3个池化层输出的特征图,并经处理后得到3个具有相同尺度的特征图。SPP 使具有不同尺度的特征图融合,可增强网络对图像几何空间信息的表达能力,提高小目标行人的检测精度。

2.1.3 改进型非极大抑制(INMS)

非极大值抑制模型应用于冗余行人检测框的剔除时,经由式(4)删除行人检测框:

$$s_i = \begin{cases} s_i, IoU(M, b_i) < \sigma \\ 0, else \end{cases} \quad (4)$$

其中, b_i 及 s_i 分别表示第 i 个检测框及其得分; M 表示当前得分最高的框; σ 为设定的阈值。

交并比(Intersection over Union, IoU) 表示真实框和网络生成的预测框的重叠度,式(5):

$$IoU(A, B) = \frac{A \cap B}{A \cup B} \quad (5)$$

其中, A 为预测框, B 为真实框。

两个不同目标所在检测框相互重叠时,因抑制因素,得分较低的框直接被删除,从而导致漏检,造成被遮挡的行人难以被检测。为能检测到被遮挡的行人,在式(4)中引入惩罚和渐进抑制思想,由式(6)确定

检测框的得分:

$$S_i = \begin{cases} s_i, IoU(M, b_i) < \sigma \\ s_i e^{-\frac{IoU(M, b_i)^2}{r}}, else \end{cases} \quad (6)$$

其中, r 为高斯系数。

2.2 改进型损失函数

给定第 k 幅尺度为 $M \times M$ 的输入图像, IYOLOv3 的损失函数值为检测框的位置均方差(Mean Square Error, MSE)及置信度损失值之和。

2.2.1 改进型 MSE 函数

虽然 MSE 的求导较易,但却将检测框的中心位置坐标和宽与高视为独立变量,忽略了两者之间的关系,从而易导致检测框与真实框的位置差距较大。因此,由式(7)度量检测框与真实框在重叠面积、中心点距离、长、宽方面的差异。

$$L_{CloU}(k, \mathbf{W}, \mathbf{b}) = 1 - IoU(A, B) + \frac{dist^2(A_{ctr}, B_{ctr})}{c^2} + \lambda \left(\frac{\omega^{gt}}{h^{gt}} - \frac{\omega}{h} \right) \quad (7)$$

其中, \mathbf{W} 和 \mathbf{b} 分别是 IYOLOv3 的权值矩阵和阈值向量; A 和 B 分别表示预测框与真实框; (A_{ctr}, ω, h) 及 $(B_{ctr}, \omega^{gt}, h^{gt})$ 分别为预测框、真实框的中心点位置、高及宽; λ 为常数; $dist(\cdot)$ 为欧氏距离函数; c 为 A_{ctr} 及 B_{ctr} 的对角线长度中的最小值。

2.2.2 改进型置信度损失函数

检测框中是否存在行人由置信度判断,若存在行人,则置信度值为1,否则为0。为此, IYOLOv3 产生的误检损失由式(8)计算:

$$L_{conf}(k, \mathbf{W}, \mathbf{b}) = \sum_{i=0}^3 \sum_{j=0}^D I(i, j) (c_i - \hat{c}_i)^2 + \sum_{i=0}^3 \sum_{j=0}^D (1 - I(i, j)) (c_i - \hat{c}_i)^2 \quad (8)$$

其中, D 表示锚框数,一般设置为3; $I(i, j)$ 表示第 i 个栅格的第 j 个预测框是否含有目标,含有目标时,取值为1,否则取值为0;在此,第1~3个栅格的尺寸依次为 13×13 、 26×26 、 52×52 ; c_i 为预测框的置信度值;第 i 个矩形框若含有目标,则 \hat{c}_i 的取值为1,否则为0。

在给定的 K 幅输入图像下, IYOLOv3 的批量损失误差函数由式(9)确定:

$$L(\mathbf{W}, \mathbf{b}) = \frac{1}{K} \sum_{k=1}^K (L_{CloU}(k, \mathbf{W}, \mathbf{b}) + L_{conf}(k, \mathbf{W}, \mathbf{b})) \quad (9)$$

2.3 算法描述

结合 YOLOv3 的结构与模块设计,其算法描述如下:

- (1) 参数设置:图像尺寸 $M \times M$, 学习率 η , 最大迭代次数 T , 卷积核大小 Q 及数量 N ;
- (2) 输入尺度为 $M \times M$ 、规模为 K 的灰度图像集 P ;
- (3) 置 $n \leftarrow 1$, 随机初始化 W 和 b ;
- (4) 置 $k \leftarrow 1$;
- (5) 计算损失函数值:由 SE-Darknet-53 提取图像集 P 中第 k 幅图的特征;通过 SPP、FPN 层对获得的特征信息进行融合;由 INMS 层计算第 k 幅图产生的损失值;若 $k \leq K$, 则 $k \leftarrow k + 1$; 否则, 计算 P 产生的损失值;
- (6) 在给定的学习率 η 下, 由式(10)和式(11)更新网络的权值和阈值:

$$W(n+1) = W(n) - \eta \frac{\partial L}{\partial W} \quad (10)$$

$$b(n+1) = b(n) - \eta \frac{\partial L}{\partial b} \quad (11)$$

- (7) $n \leftarrow n + 1$; 若 $n \leq T$, 则返回(3); 否则, 则结束。

YOLOv3 的计算复杂度主要由(4)~(6)步确定。(4)~(5)步的计算量由 YOLOv3 的输入图像大小 $M \times M$ 确定;(6)步的计算复杂度由灰度图像训练集规模 K 以及卷积核大小 Q 和数量 N 确定。因此, YOLOv3 的计算效率主要依赖于 M, K, Q 及 N 。

3 数值实验

在 Windows11 (CPU/i5-12500H 3.10 GHz, RAM/16 GB)/Python 3.7 环境下展开数值实验。为测试 YOLOv3 检测行人的有效性, 选取两阶段目标检测模型(Faster R-CNN)、单阶段目标检测模型(SDD、YOLOv2 和 YOLOv3)参与比较, 各自的最大迭代次数均为 93 750。YOLOv3 的参数设置: 训练

轮数 $epoch$ 取 100, 学习率 η 为 0.001, 批处理数据 batch size 是 16, 动量参数 $momentum$ 取 0.9。选取 CrowdHuman 数据集中人群密度大、人群遮挡较为突出的 15 000 幅图像按 9:1 的比例随机选取图像, 构成训练集和测试集, 图像种类只有 1 个行人类别, 图像大小裁剪为 416×416。

3.1 评价指标

选取准确率 ($Precision, P$)、召回率 ($Recall, R$)、准确率与召回率综合评估指标 F_1 、平均精度 ($Average Precision, AP$) 作为模型的评价指标。 P 为评估模型的全局准确程度; R 表示被正确预测为正样本的概率; F_1 越大, 说明试验方法越有效。 P, R 及 F_1 经由下式确定, 式(12)~式(14):

$$P = \frac{TP}{TP + FP} \quad (12)$$

$$R = \frac{TP}{TP + FN} \quad (13)$$

$$F_1 = 2 \frac{PR}{P + R} \quad (14)$$

其中, TP 为将正样例预测为正样例的个数, 即真正例; FP 表示假正例的个数; FN 为将正样本预测为负样例的个数, 即假反例。

AP 是以召回率为横坐标、准确率 R 为纵坐标形成的 $P-R$ 曲线与坐标轴围成的面积, 其值越大, 则模型的检测效果越好。

3.2 实验结果与分析

情形 1 YOLOv3 的性能测试

为了获知 YOLOv3 中各模块的性能优劣, 将 YOLOv3 中的 Darknet53、MP、损失函数 L_{IoU} 逐次用 SE-Darknet53、SPP 及改进型损失函数 L_{CIoU} 取代, 由此获得的模型依次记作 YOLOv3-I、YOLOv3-II、YOLOv3-III; 随后, 将 YOLOv3、YOLOv3-I 及 YOLOv3-II 3 种模型均作用于 CrowdHuman 行人测试图像集, 功能模块的性能测试结果见表 1。

表 1 功能模块的性能测试结果

Tab. 1 Performance test results of functional modules

实验	主干网络	池化层	损失函数	非极大值抑制	AP/%
YOLOv3	Darknet53	MP	IoU	NMS	56.95
YOLOv3-I	SE-Darknet53	MP	IoU	NMS	63.26
YOLOv3-II	SE-Darknet53	SPP	IoU	NMS	66.77
YOLOv3-III	SE-Darknet53	SPP	CIoU	NMS	68.48
IYOLOv3	SE-Darknet53	SPP	CIoU	INMS	70.12

由表 1 可知, 随着 YOLOv3 的 Darknet53、MP、IoU、NMS 逐次被 SE-Darknet53、SPP、CIoU、INMS

取代, 获得的 YOLOv3-I、YOLOv3-II、YOLOv3-III 及 IYOLOv3 相比于 YOLOv3 在平均精度 AP 方面,

依次提升 6.31%、9.82%、11.53% 及 13.17%;说明随着 YOLOv3 中的模块逐渐被本文设计的模块取代,模型得到的平均精度逐渐提升,因此 SE-Darknet53、SPP、CIoU 及 INMS 均有助于提升 YOLOv3 的性能,尤其是 IYOLOv3 的行人检测平均精度达到 70.12%。

情形2 模型的性能比较

基于 Crowdhuman 行人数据测试集,依据以上评价指标,各模型获得的实验结果见表 2。

表2 模型的评价指标值比较

模型	$P / \%$	$R / \%$	$F_1 / \%$	$AP / \%$
Faster R-CNN	43.44	49.86	46.38	44.63
SSD	84.63	36.04	51.36	51.55
YOLOv2	79.14	44.29	57.19	52.83
YOLOv3	83.84	43.94	58.37	56.95
IYOLOv3	85.03	52.05	65.42	70.12

由表 2 可知,IYOLOv3 相较于参与比较的模型,准确率 P 、召回率 R 、综合评估准确率与召回率指标 F_1 及平均精度均值 AP 均有明显提升。相较于 YOLOv2 及 IYOLOv3,IYOLOv3 的 (AP , F_1) 值依次提升 (17.29%, 8.23%) 及 (9.33%, 5.01%);相较于 SSD, AP 、 F_1 值依次提升 16.79% 及 12.05%。特别地,相比于 Faster R-CNN,其 AP 、 F_1 值依次提升 25.49% 及 19.04%,表明注意力机制、空间金字塔池化以及改进型非极大值抑制模型能强有力地提升 YOLOv3 的行人检测性能,使得 IYOLOv3 的 AP 达到 70.12% 且 F_1 值为 65.42%。另外,YOLOv3 相比于其它参与比较的 3 种模型,其行人检测性能有明显提升。

情形3 实际场景下检测结果的可视化

为测试模型的泛化能力,将 YOLOv3 与 IYOLOv3 作用于测试集中 3 种实际场景的事例,各自的检测效果如图 5 所示。

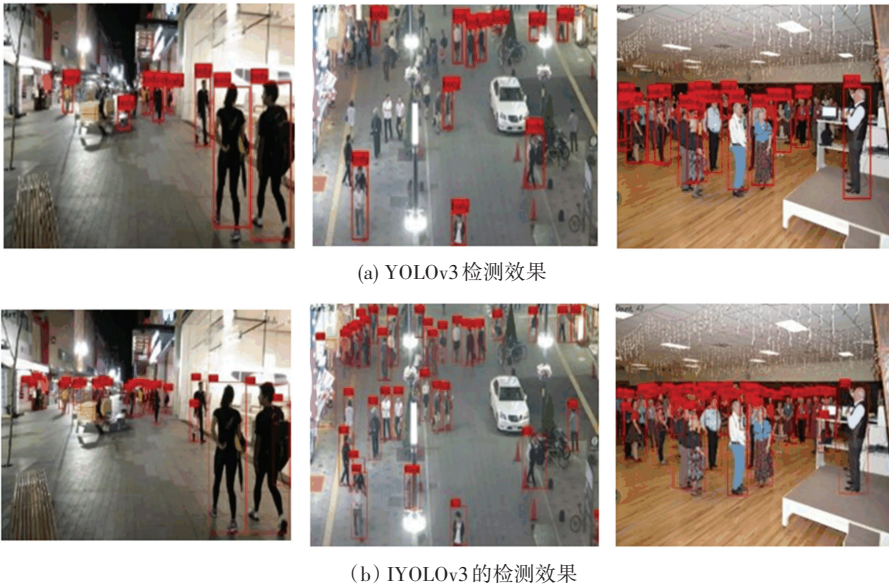


图5 行人检测效果比较

Fig. 5 Comparison of pedestrian detection results

图 5(a) 表明,Darknet-53 难以提取远距离的小目标行人的特征,进而导致 YOLOv3 对于较远处以及互相遮挡的行人出现漏检现象,检测效果欠佳。另外,图 5(b) 表明, IYOLOv3 能克服 YOLOv3 存在漏检小尺度或遮挡行人目标的问题,能显著地降低漏检率,且对于背景遮挡、光照变化等干扰, IYOLOv3 也具有较好的稳定性。因此, IYOLOv3 极大地改善了行人检测的性能。

4 结束语

针对 YOLOv3 存在行人检测漏检及检测准确率低的问题,本文在其模型结构基础上,利用注意力机制改进其主干网络,并引入空间金字塔池化增强模型的特征提取能力,同时引入惩罚和渐进抑制思想改进非极大值抑制模型,获得能有效提升 YOLOv3 模型的行人检测能力的改进型 IYOLOv3 模

型。模型比较证实, YOLOv3 应用于密集场景下行人的检测问题, 能有效回避漏检现象且检测精度高。进一步的工作中, 将展开 YOLOv3 的灵敏度及检测效率研究, 以便其适应于各种复杂场景的行人检测。

参考文献

- [1] 姬东飞, 丁学明. 基于特定区域去相关的行人检测算法[J]. 信息与控制, 2020, 49(4): 10-11.
- [2] 潘明远, 宋慧慧, 张开华, 等. 学习全局引导渐进特征聚合轻量级网络的显著性目标检测[J]. 计算机科学, 2021, 48(6): 7-8.
- [3] 张杰. 基于深度学习的行人检测与重识别方法及系统实现[D]. 山东: 山东大学, 2021.
- [4] KAEWTRAKULPONG P. An improved adaptive background mixture model for real-time tracking with shadow detection[J]. Video Based Surveillance Systems Computer Vision & Distributed Processing, 2002, 1(1): 135-144.
- [5] 邓炜, 刘秉瀚. 联合语义的深度学习行人检测[J]. 计算机系统应用, 2018, 27(6): 6-7.
- [6] 龚露鸣, 徐美华, 刘冬军, 等. 基于混合高斯和 HOG+SVM 的行人检测模型[J]. 上海大学学报(自然科学版), 2018, 24(3): 341-351.
- [7] 陈奇华. 针对小目标的深度学习行人检测算法研究[D]. 广州: 华南理工大学, 2019.
- [8] 张顺, 龚怡宏, 王进军. 深度卷积神经网络的发展及其在计算机视觉领域的应用[J]. 计算机学报, 2019, 42(3): 453-482.
- [9] GIRSHICK R, DONAHUE J, DARRELL T, et al. Rich Feature Hierarchies for Accurate Object Detection and Semantic Segmentation[J]. IEEE Computer Society, 2014, 2(2): 580-587.
- [10] REN S, HE K, GIRSHICK R, et al. Faster R-CNN: Towards real-time object detection with region proposal networks[J]. IEEE Transactions on Pattern Analysis & Machine Intelligence, 2017, 39(6): 1137-1149.
- [11] REN S, HE K, 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.
- [12] REDMON J, DIVVALA S, GIRSHICK R, et al. You only look once: unified, real-time object detection[J]. IEEE, 2016, 1(1): 779-788.
- [13] REDMON J, FARHADI A. YOLO9000: better, faster, stronger[C]// IEEE Conference on Computer Vision & Pattern Recognition. IEEE, 2017: 6517-6525.
- [14] 刘建国, 罗杰, 王帅帅, 等. 基于 YOLOv2 的行人检测方法研究[J]. 数字制造科学, 2018, 16(1): 50-54.
- [15] HUANG F Q, CHEN M, FENG G F. Improved YOLO object detect algorithm based on deformable convolution[J]. Computer Engineering, 2021, 47(10): 269-275, 282.
- [16] SHI T C, LIU M Y, YUN NIU, et al. Underwater targets detection and classification in complex scenes based on an improved YOLOv3 algorithm[J]. Journal of Electronic Imaging, 2020, 29(4): 043013.
- [4] WOODRUFF R B. Customer value: The next source for competitive advantage[J]. Journal of the Academy of Marketing Science, 1997, 25(2): 139-153.
- [5] 傅辉子. 网络经济中信息产品定价影响因素研究[J]. 图书情报导刊, 2020, 5(1): 45-49.
- [6] 赖茂生, 张梦雅. 网络情报产品的定价策略研究[J]. 情报科学, 2011, 29(10): 1457-1461.
- [7] 张亚明, 苏妍嫒. 双寡头企业信息消费产品定价策略[J]. 控制工程, 2019, 26(3): 542-548.
- [8] 关涛, 高晶, 张雪桐. 顾客感知价值对网络信息产品定价的影响研究[J]. 财经理论与实践, 2017, 38(4): 97-102.
- [9] 钟凯. 网络消费者感知价值对购买意愿影响的研究[D]. 辽宁大学, 2013.
- [10] 李舒娴. UGC 移动短视频平台用户感知价值与用户忠诚的关系研究[D]. 暨南大学, 2018.
- [11] 王珊珊. 消费者感知价值对智能新产品购买意愿的影响研究[D]. 天津商业大学, 2022.
- [12] 屈慧杰. 基于 Netlogo 仿真的电子商务平台定价模式研究[D]. 郑州大学, 2018.
- [13] 王欣悦, 曾晖, 刘通飞. 基于 Netlogo 的工程造价纠纷解决因素仿真研究[J]. 建筑经济, 2021, 42(5): 113-116.
- [14] 刘小虎, 张恒巍, 张玉臣, 等. 基于博弈模型与 NetLogo 仿真的网络攻防态势研究[J]. 系统仿真学报, 2020, 32(10): 1918-1926.
- [15] 田水承, 董威松, 沈小清, 等. 基于 Netlogo 的矿工不安全行为传播仿真研究[J]. 安全与环境学报, 2019, 19(6): 2016-2022.
- [16] 平健. 基于 Netlogo 的政府应急组织合作关系演化趋势仿真研究[J]. 软科学, 2018, 32(2): 124-129.

(上接第 178 页)