计算机系统应用 ISSN 1003-3254, CODEN CSAOBN
Computer Systems & Applications,2023,32(1):288−295 [doi: 10.15888/j.cnki.csa.008899]
©中国科学院软件研究所版权所有.
E-mail: csa@iscas.ac.cn
http://www.c-s-a.org.cn
Tel: +86-10-62661041

基于改进 YOLOv3 的智慧足球场行人检测①

徐克圣, 崔效魁

(大连交通大学 软件学院, 大连 116028)
通信作者: 崔效魁, E-mail: 1162894409@qq.com

摘　要: 由于足球比赛场景中密集人群、移动小目标居多, YOLOv3 算法存在检测精确度较低且模型参数量较大等问题, 使其无法部署在资源算力有限的移动设备上, 本文提出了一种基于改进 YOLOv3 的行人检测方法, 将 Darknet-53 主干特征提取网络替换为更加高效且轻量化的 GhostNet 网络; 同时选取了 4 个尺度的检测分支层并采用 K-means++算法改善 anchor box 的聚类效果; 添加空间金字塔池化对输入图像实现相同大小的输出; 提出 CIoU 损失函数来计算目标定位损失值; 添加 heatmap 热力图可视化并在训练中使用 Mosaic 数据增强. 实验结果表明, YOLOv3-GhostNet 在 VOC 融合数据集上 *mAP* 达到 90.97% 的同时相比 YOLOv3 算法提高了 1.75%, 参数量减少了约 81.4% 且实时检测速率提高了约 1.5 倍, 在小型移动设备上表现出不错的检测效果.

关键词: 智慧足球场; 行人检测; 深度学习; YOLOv3; GhostNet; 深度可分离卷积

引用格式: 徐克圣,崔效魁.基于改进 YOLOv3 的智慧足球场行人检测.计算机系统应用,2023,32(1):288–295. http://www.c-s-a.org.cn/1003-3254/8899.html

Pedestrian Detection in Intelligent Football Field Based on Improved YOLOv3

XU Ke-Sheng, CUI Xiao-Kui

(Software Technology Institute, Dalian Jiaotong University, Dalian 116028, China)

Abstract: Football match scenes are featured with dense crowds and many mobile targets, and YOLOv3 algorithm has low detection accuracy and requires massive model parameters, which makes it unable to be deployed on mobile devices with limited computing power. In view of these problems, this study proposes a pedestrian detection method based on improved YOLOv3. Specifically, the study replaces the Darknet-53 backbone feature extraction network with a more efficient and lightweight GhostNet network, selects detection branch layers with four scales, and adopts the K-means++ algorithm to improve the clustering effect of the anchor box. Furthermore, the study adds spatial pyramid pooling to achieve an output with the same size as the input image, puts forward the CIoU loss function to calculate the loss value of target positioning, adds heatmap visualization, and uses Mosaic data enhancement in training. The experimental results show that YOLOv3-GhostNet achieves a *mAP* of 90.97% on the VOC fusion dataset, with an improvement of 1.75% compared with the YOLOv3 algorithm. In addition, it reduces the number of parameters by about 81.4% and increases the real-time detection rate by about 1.5 times, which shows a positive detection effect on small mobile devices.

Key words: intelligent football field; pedestrian detection; deep learning; YOLOv3; GhostNet; depth separable convolution

足球是一项风靡全世界且在各个国家都非常具有影响力的球类运动, 像亚洲杯和世界杯这样的大型足球赛事未来几年时间内将会在我们国家接连举办, 以往传统大型国际足球赛事会经常出现双方球迷暴乱、

① 基金项目: 辽宁省教育厅科研经费项目 (JDL2019025)
收稿时间: 2022-05-30; 修改时间: 2022-06-27; 采用时间: 2022-07-06; csa 在线出版时间: 2022-08-26
CNKI 网络首发时间: 2022-11-16

© 中国科学院软件研究所 http://www.c-s-a.org.cn

运动员打架事件的发生, 将目标检测算法部署在高清摄像头、无人机等小型移动设备上应用于智慧体育场馆, 对场内外观众、场上比赛运动员进行实时监测, 具有一定的实际应用意义.

目标检测被广泛应用于行人检测、图像识别、自动驾驶以及许多日常生活领域. 传统行人检测方法一般采用类似于 Haar[1]、HOG[2] 等的图像分割技术或类似穷举的滑动窗口方式来提取图像中行人特征, 随后将其传递给如 SVM[3]、AdaBoost[4] 等分类器, 进而判断目标的类别.

深度学习在图像分类任务中取得广泛成功之后, 将图像领域中各个问题的处理精度都提升到了一个更高的水平, 基于区域选择和逻辑回归的两大类方法被应用于目标检测任务中. 基于区域选择的两阶 (two-stage) 目标检测算法被 R-CNN 家族系列算法独占鳌头. 最初的 R-CNN[5] 算法训练与测试速度较慢. Fast R-CNN[6] 仍不能满足实时检测的需求. Faster R-CNN[7] 则大幅提升了检测精度与速度. 随后, Mask R-CNN[8]、Cascade R-CNN[9] 等模型的出现, 使得 R-CNN 家族不断壮大, R-CNN 系列算法虽然检测精度较高, 但由于其网络复杂度的问题, 使得实时检测仍然是一个问题.

为更好地平衡检测精度与速度, 基于逻辑回归的一阶 (one-stage) 目标检测方法被提出. Redmon 等人[10–12] 相继提出了 YOLO、YOLO9000 和 YOLOv3 算法. YOLO 拥有非常快的检测速度, 可以轻松地实时运行, 但其存在严重的定位错误. Liu 等人[13] 提出基于 VGG 网络的 SSD 方法, 但其检测精度并不高. YOLO9000 弥补了 YOLO 的不足, 借鉴 Faster R-CNN 的算法思想, 引入 anchor 机制和 K-means 聚类方法对模型进一步深入优化, 其检测性能与 SSD 持平, 增强了对小尺寸目标的检测. YOLOv3 则利用深度残差网络构成 Darknet-53 分类网络代替 YOLO9000 中的 Darknet-19 提取图像特征, 同时借鉴 ResNet 并融入特征金字塔网络 (feature pyramid network, FPN) 结构[14] 中, 在传统公共场所的行人检测问题中具备较强的多尺度特征提取能力.

YOLO 系列算法凭借较为优秀的算法思想在准确性上略强于上述经典目标检测算法, 但仍无法在高速变化且复杂度较高的场景下满足多数物体的类别检测. 黄同愿等人[15] 通过层次敏感度分析对 YOLOv3 主干网络进行精简, 并通过引入空间金字塔池化来加强小目标的检测; 张路达等人[16] 提出利用多尺度融合结构, 通过更好的提取特征信息进而实现特征增强; Zheng 等人[17] 提出 DIoU 和 CIoU 两种评价标准, 其中 CIoU (complete intersection over union) 则为当前目标检测领域公认最优秀的评价标准.

本次研究中, 考虑到足球场行人检测相较于传统公共场所的技术难点主要在于场上的运动员不断地在移动且速度较快, 这就要求模型需要持有较高的实时检测速率. 卷积神经网络存在庞大的参数量和计算量也是在设计计算模块时需要考虑的重点, 在有限的存储空间和计算资源的情况下采用轻量化的网络结构设计是非常有必要的. 为了降低网络复杂度, 本文对 YOLOv3 主干网络进行调整, 采用更加轻量化的 GhostNet 网络对 Darknet-53 主干特征提取网络进行替换, 使之更加适配足球场内行人检测任务, 更利于在资源算力不足的小型移动设备上完成模型部署.

1 YOLOv3 目标检测算法原理

如图 1 所示, YOLOv3 将输入图像变为 416×416 大小后输入到 Darknet-53 主干特征提取网络中, 同时池化层和最后的全连接层被取消, 采用步长为 2 的卷积下采样, 不仅减少计算量, 还保留了图像更多的相关信息. 经过一系列下采样、卷积等操作, 可以获得图片不同层次的位置和语义信息. 网络输出层采用 3 层来进行预测, 最终使用非极大值抑制确定结果.

2 改进的 YOLOv3 目标检测算法

2.1 网络结构优化

原 YOLOv3 网络采用 3 种尺度检测不同尺寸的目标, 将网络输入图像统一缩放为 416×416 尺寸, 然后划分为 $N\times N$ 的网格, 每个网格使用 3 个先验框预测目标物体, 输出为 $N\times N\times 3\times(4+1+C)$, 先验框的中心坐标、宽高的 4 个值和置信度值包含在其中, N 代表网格数量, C 代表类别数量. 如图 2 所示, 由于数据集中存在大量小目标与遮挡严重的目标, YOLOv3 虽然通过 3 个不同尺度的分支对目标进行检测, 但因为输出特征图维度较低, 对于小目标仍然存在检测困难的问题, 漏检、误检或重复检测的状况时有发生.

本文综合考虑最小目标检测的底限和网络的大小, 将原始 YOLOv3 尺度为 13×13、26×26、52×52 的分

© 中国科学院软件研究所 http://www.c-s-a.org.cn

支再增加一个 104×104 尺度, 通过 4 倍、8 倍、16 倍、32 倍下采样操作来细化特征提取网络输出的特征图, 实现 4 个尺度的相互融合, 提升对小目标的检测效果. 增加一个检测层会使模型的复杂度升高, 在训练时会影响网络的检测速度, 但对模型的检测精度会有一定的提升.

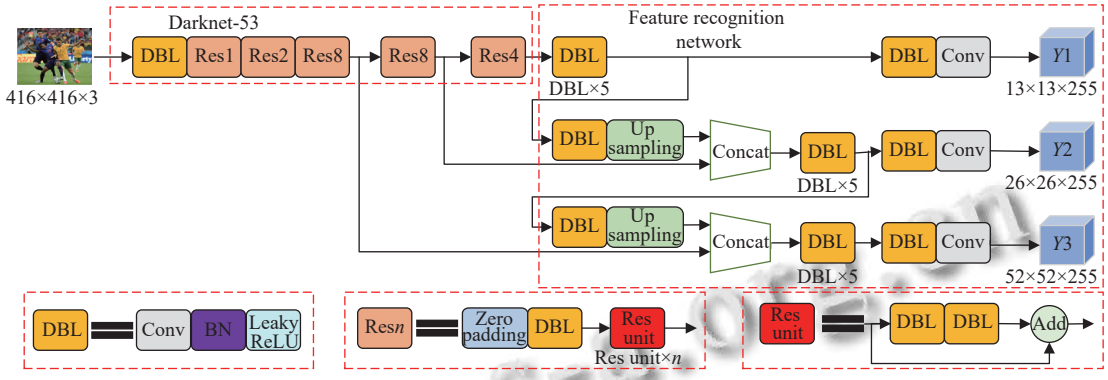

图 1 YOLOv3 网络结构图

(a) 漏检 (b) 误检 (c) 重复检测

图 2 YOLOv3 在本文数据集上的错误检测实例

2.2 K-means++先验框预测

K-means++算法相比于 K-means 算法重点优化了初始聚类中心的选取, 使得聚类时间和效果都有所改善. 原 YOLOv3 网络 3 个特征输出层上的每个网格单元只能预测 3 组先验框, 对于复杂场景下遮挡严重、密集且距离较远的小目标容易发生漏检的情况. 如图 3 所示, 改进后的 YOLOv3 网络使用 K-means++聚类算法聚类 12 组先验框, 然后均匀分布在 4 个特征输出层上, 通过增加 anchor box 的数量来提升网络检测的整体精度, 而模型的浮点计算量基本上没有太大浮动. 其中 13×13、26×26、52×52、104×104 的特征层分别对应大、中、小和较小感受野, 其对应的 12 组先验框的尺寸分配结果依次为 (188, 315)(313, 212)(364, 371)、(93, 106)(102, 238)(170, 150)、(37, 32)(53, 157)(62, 59)、(13, 21)(17, 48)(30, 89).

2.3 空间金字塔池化

YOLOv3 网络要求输入图像尺寸固定, 但在行人检测问题中, 这一要求会使得网络对遮挡目标和小目标的漏检率偏高. 因此本文将空间金字塔池化 (spatial pyramid pooling, SPP) 模块[18] 添加到 YOLOv3 网络的检测分支内, 该模块由一个跳跃连接层和多个尺寸不同的最大池化层构成, 针对输入图像尺寸不统一的问题, 使用固定分块的池化操作对不同尺寸的输入实现相同尺寸的输出. 多种尺寸的池化操作可以扩大特征图对应的感受野, 从而应对多尺度目标表示的困难. 如图 4 所示, 特征图分别经过各个分支处理后重新合并起来传到下一层网络中. 当网络输入为 416×416 时, 为了保证局部特征与全局特征在特征图上能够良好的进行融合, 进而提升网络的检测精度, 最大池化层最大设计为 13×13, 针对行人检测问题中, 大、中、小 3 个尺寸的行人目标均常见的情况, 最终添加的 SPP 模块选用尺寸为 5×5、9×9、13×13 的最大池化层和一个跳跃连接层组成.

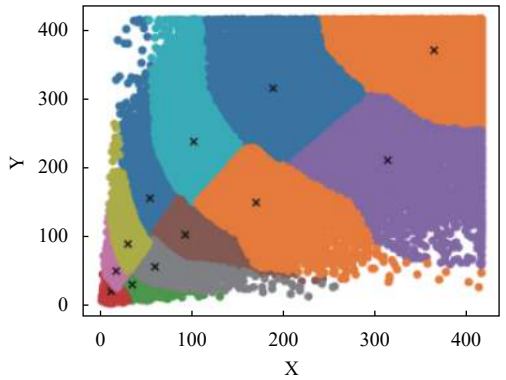

图 3 K-means++先验框聚类图

2.4 损失函数改进

YOLOv3 采用交并比损失 *IoU* loss 评价预测框与

© 中国科学院软件研究所 http://www.c-s-a.org.cn

真实框重合程度, 其在进行网络训练时, 在得到相同损失值时会出现不同的结果, 很难正确反映预测框与真实框之间的具体位置情况. 其计算式如式 (1) 所示:

$$IoU = \frac{|A \cap B|}{|A \cup B|} \tag{1}$$

为了消除这个不稳定影响并优化检测精度, 本文引入完全交并比 CIoU 代替 *IoU* 作为边界框回归损失函数, 使得边界框回归更加稳定, 收敛精度更高, 且没有 *IoU* 的缺陷, 其损失函数公式如式 (2) 所示:

$$Loss_{\mathrm{CIoU}} = 1 - IoU + \frac{\rho^2(b, b^{gt})}{c^2} + \alpha v \tag{2}$$

其中, $\rho^2(b, b^{gt})$表示预测框与真实框中心点 (b)、(b^{gt}) 的欧式距离, c 则表示二者的最小外接矩形的对角线距离, αv为惩罚因子, 其计算式分别如式 (3) 和式 (4) 所示:

$$\alpha = \frac{v}{(1 - IoU) + V} \tag{3}$$

$$v = \frac{4}{\pi^2}\left(\arctan\frac{w^{gt}}{h^{gt}} - \arctan\frac{w}{h}\right)^2 \tag{4}$$

其中, w、h 分别为预测框的宽、高, w^{gt}、h^{gt}为真实框的宽、高.

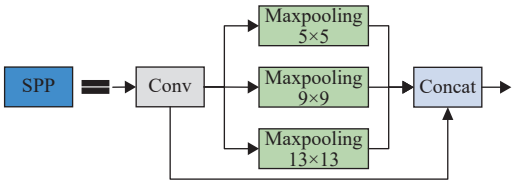

图 4 SPP 模块结构

3 YOLOv3-GhostNet 网络

3.1 特征提取网络改进

智慧足球场馆使用的大多数是成本较为低廉的嵌入式设备, 这就使得网络部署需要较少的参数量和计算量, 为了降低模型复杂度, 减少网络参数和计算量, 现对算法进行轻量化设计, 本文使用华为诺亚方舟实验室在 2020 年度 CVPR 上提出的 GhostNet 网络结构[19], 将 YOLOv3 算法中的 Darknet-53 主干特征提取网络替换为 GhostNet 网络结构, 并将其进行迁移学习.

Ghost 模块使用更少的参数生成相同的特征, 其网络特征层中的冗余部分很可能包含重要特征信息, 所以 Ghost 中保留了这些冗余信息, 来用更低的计算量成本获取更多的特征信息. 如图 5 所示, Ghost 模块将传统的卷积操作分为两个步骤进行, 首先使用计算量较少的普通卷积操作对输入的特征图生成部分真实特征图, 接着再利用 DWConv 操作对真实特征图的各个通道进行深度卷积得出 Ghost 特征图, 然后再对二者进行 Concat 拼接, 得出最后的输出特征图.

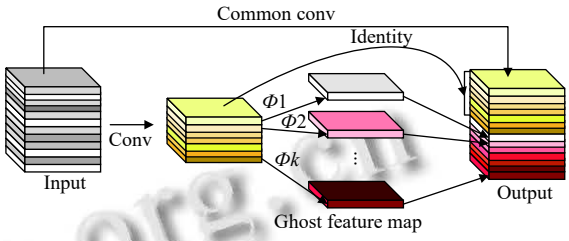

图 5 Ghost 模块结构图

若输入特征图表示为 $H \times W \times C$, 输出为 $H' \times W' \times M$, 把输入分为 n 层, 卷积核大小为 $k \times k$, 则普通卷积和 Ghost 卷积的计算量分别如式 (5) 和式 (6) 所示:

$$H' \times W' \times M \times k \times k \times C \tag{5}$$

$$H' \times W' \times \frac{M}{n} \times k \times k \times C + (n-1) \times H' \times W' \times \frac{M}{n} \times k \times k \tag{6}$$

从 Ghost 模块的计算量可以得出结论, 利用普通卷积和深度卷积两部分来计算的方式有效降低了网络计算复杂度. 与此同时, GhostNet 网络还引入了 SE 注意力机制模块于 Ghost BottleNeck 模块结构中来使提取的特征针对性更强, 特征利用更加充分.

3.2 YOLOv3-GhostNet

YOLOv3-GhostNet 网络中的 Ghost BottleNeck 瓶颈层由两个功能不同的 Ghost Module 构成, 如图 6 所示, GhostNet 网络中的第 1 个 Ghost Module 的主要功能是为了增加通道数; 第 2 个 Ghost Module 可使通道数减少至与输入相连接的通道数量相匹配. Ghost 模块分为 *Stride*=1 和 *Stride*=2 两种不同的步长, 本文在该模块中使用 *Stride*=2 的 DWConv 深度可分离卷积.

4 相关工作

4.1 数据集准备

本文使用开源的 VOC2007+2012 的 train+val 数据集 (含 16551 张图片)、VOC2007 的 test 数据集 (含 4952 张图片) 进行观众行人检测实验. 除此之外, 为了增加测试集数量, 本实验还选取了 1 000 张自制世界杯足球比赛数据集, 主要针对足球比赛时场上足球运动员的检测, 同时对图片背景观众席进行虚化, 只保留场上运

© 中国科学院软件研究所 http://www.c-s-a.org.cn

动员移动时的图片进行模型训练. 通过对 PASCAL VOC 数据集进行数据清洗只保留 20 种类别标签信息中的 person 一种类别标签信息, 然后按照 9:1 划分为训练集和验证集. 合并后的数据集中训练集中包含 person 类别标签数量为 6182, 验证集数量为 689, 测试集数量则为 3097.

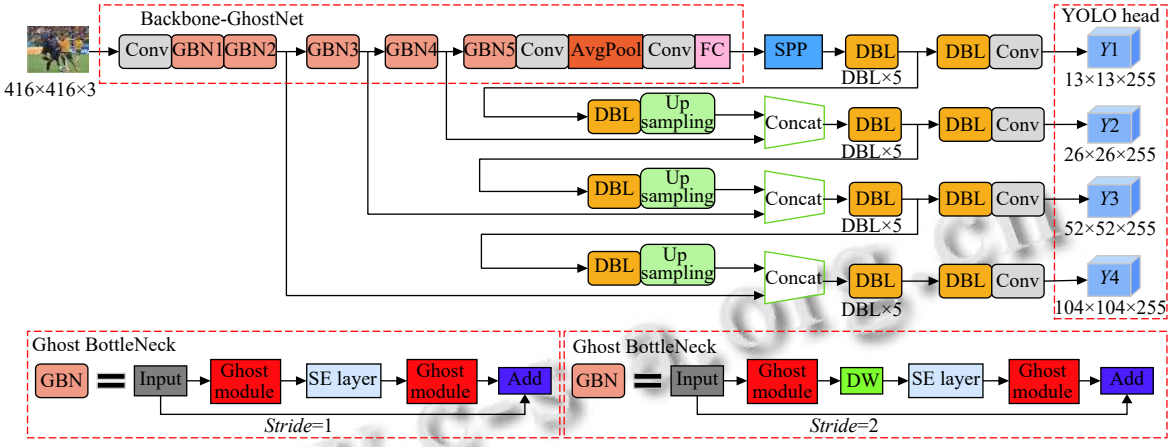

图 6 YOLOv3-GhostNet 结构图

4.2 数据预处理

在网络训练任务中对原始数据集中图像的标注非常关键. 图 7 所示为 LabelImg 图像标注软件, 通过对自制数据集进行 VOC 格式标注, 将标注框的位置信息保存到对应图像的 XML 文件中, 从而进行网络训练.

图 7 LabelImg 图像标注示意图

5 实验与结果分析

本实验平台采用 Windows 10 操作系统, CPU: Intel(R) Core(TM) i7-10875H CPU @ 2.30 GHz, 内存 16 GB, GPU: NVIDIA GeForce RTX 2060 6 GB 显存, 深度学习框架 PyTorch, cuda 10.2+OpenCV, cudnn 7.6.5.

5.1 训练过程与结果

本文对基于改进 YOLOv3 的智慧足球场行人检测算法模型进行训练. 在训练过程中对训练数据使用 Mosaic 数据增强, 使用随机缩放、翻转、平移等操作, 采用 4 张图片拼接的方式来提高训练中每个批次输入图片的数量, 较好地提高了网络的鲁棒性和泛化能力.

实验采用迁移学习的思想将训练分为冻结阶段和解冻阶段, 起初冻结模型主干进行训练, 可以防止权值被破坏并加快训练效率, 之后进行解冻训练. 训练过程中, 平滑标签 label_smoothing 设置为 0.01; 模型参数更新方式为 sgd; batch-size 为 16; gamma 为 0.92; weight_decay 为 5E–4; 最大学习率 Init_lr 设置为 1E–2, 最小学习率则为 Init_lr×0.01, 采用学习率检测机制对其进行动态调整. 训练总轮次设置为 100 个 epoch, 每训练完一个 epoch 保存一次训练好的模型, 最终针对本文特定场景选用 loss 值最低的模型进行检测.

如图 8 所示, 由训练过程中的损失函数值曲线收敛变化情况分析得出, 初始化权重为随机值导致在前期的学习训练集中 loss 迅速下降; 后期 loss 缓慢下降, 在训练到第 70 个 epoch 时还有一个小幅度下降, 直到训练结束, 模型逐渐达到收敛状态, 几乎不再发生变化, 最小 loss 值降为 0.032, YOLOv3-GhostNet 模型已经达到预期的训练效果.

5.2 评价指标

本文选用平均精确度均值 mAP 和每秒检测帧数 FPS 来综合反映模型的性能, FPS 是实时检测速率的直接体现, 引入准确率 P (precision) 和召回率 R (recall) 对平均精确度值进行计算. Precision、recall 的计算方式如式 (7) 和式 (8) 所示:

$$P = \frac{TP}{TP + FP} \tag{7}$$

© 中国科学院软件研究所 http://www.c-s-a.org.cn

$$R=\frac{TP}{TP+FN} \tag{8}$$

对于正样本, 网络判断为正样本和负样本分别记为 *TP*、*FN*; 而对于负样本, 则分别记为 *FP*、*TN*. 同时采用 *F*1 指标来综合衡量准确率 *P* 与召回率 *R*, 其值越接近 1 则效果越好. *mAP*、*F*1 的计算方式分别如式 (9) 和式 (10) 所示:

$$mAP=\int_0^1 P(R)d(R) \tag{9}$$

$$F1=\frac{2PR}{P+R}=\frac{2TP}{2TP+FP+FN} \tag{10}$$

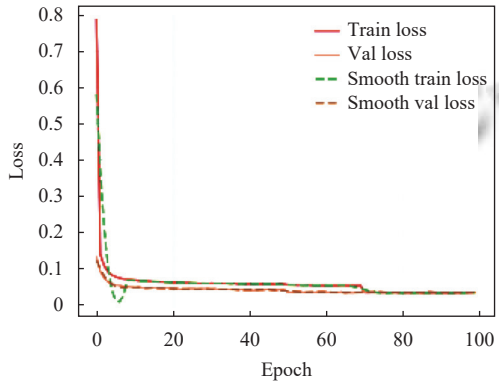

图 8 损失函数曲线图

5.3 实验结果分析

本文选用 VOC2007 测试集与自制数据集来验证改进后的 YOLOv3-GhostNet 网络的检测效果, 如图 9 所示, 本文算法能够对图像中的行人以及足球运动员进行精准地识别并且定位, 很少出现漏检、误检的情况, 具有良好的检测效果.

(a) 随机检测结果

(b) Heatmap 热力图可视化

图 9 改进算法检测结果图

本次实验对 Faster R-CNN、SSD、YOLOv3、改进的 YOLOv3 和 YOLOv3-GhostNet 网络进行训练与测试, 并得到如表 1 所示的 *P*、*R*、*mAP* 和 *F*1 等技术指标. YOLOv3-GhostNet 在 *mAP* 达到 90.97% 的同时持有 45.32 fps 的实时检测速率, 满足实时检测的场景需求.

表 1 不同方法技术指标对比

Model	Input-size	*P*	*R*	*mAP* (%)	*F*1
Faster R-CNN	600×600	0.868	0.814	87.40	0.84
SSD	300×300	0.804	0.663	76.81	0.73
YOLOv3	416×416	0.887	0.791	89.22	0.84
改进的YOLOv3	416×416	0.948	0.811	92.80	0.87
YOLOv3-GhostNet	416×416	0.929	0.789	90.97	0.85

在测试中 Faster R-CNN 算法获得了 87.40% 的 *mAP*, 比本文算法低了 3.57%. 由于网络复杂度的问题, 模型在进行计算时耗时较长, 在检测速率方面该算法仅保持着 13.46 fps 的实时速率, 难以满足实际场景中对检测速度的要求; 由于 SSD 算法其网络计算复杂度相对比较低, 在本文数据集上检测的 *mAP* 仅为 76.81%, 持有 45.21 fps 的实时检测速率, 满足实际检测任务中对速率的要求, 但其检测的准确率相对较低, 大概率出现误检、漏检的情况, 难以应对密集人群场景. 两者与本文所提的算法相比均没有太明显的优势.

YOLOv3 算法检测的 *mAP* 为 89.22%, 改进 YOLOv3 的 *mAP* 为 92.80%, YOLOv3-GhostNet 的 *mAP* 则为 90.97%, 相比 YOLOv3 网络分别提升了 3.58%、1.75%. 由于改进的 YOLOv3 算法在检测网络中提取了更加丰富的特征信息, 因而在检测精度方面的 *mAP* 比原始 YOLOv3 网络高出了 3.58%, 但由于其网络复杂度的提升, 检测速率比原始 YOLOv3 网络低了 3.27 fps, 在保持 26.98 fps 的实时检测速率的同时网络检测的平均精度要远远优于原 YOLOv3 网络. 由于主干特征提取网络的轻量化改进, 使得网络复杂度大幅降低, YOLOv3-GhostNet 网络检测的 *mAP* 相比于原 YOLOv3 网络虽然仅提升了 1.75%, 和改进的 YOLOv3 算法相比, 检测的平均精确度要略低一筹, 但其在模型参数量和实时检测速率方面均有很大的改善, 更加符合智慧体育场馆对模型的需求.

由表 2 可以看出, 改进的 YOLOv3 网络模型参数量显著提高, 由于网络复杂度变大致使 FPS 略有下降. YOLOv3-GhostNet 相比于 YOLOv3 网络模型参数量

© 中国科学院软件研究所 http://www.c-s-a.org.cn

减少了约 81.4%, 仅为 43.6 MB, 且 FPS 提高了约 49.8%, 在保证检测精度的同时持有 45.32 fps 的实时检测速率, 较好的平衡了检测精度与速度.

表 2 模型参数对比

算法	模型大小 (MB)	FPS (fps)
YOLOv3	234	30.25
改进的YOLOv3	250.8	26.98
YOLOv3-GhostNet	43.6	45.32

图 10 显示了原始 YOLOv3、改进的 YOLOv3 和 YOLOv3-GhostNet 三种网络对相同图像数据的检测结果. 若在发生遮挡严重或人群相对密集的复杂情况下, 不难看出三者都有出现漏检的情况, 改进的 YOLOv3 算法表现得更为优秀, 可以更好地将被遮挡部分的行人检测出来, YOLOv3-GhostNet 网络的检测效果和改进的 YOLOv3 网络相比, 由于采用了轻量化网络的原因致使其检测效果有所下降, 但和 YOLOv3 网络相比, 其检测效果也有了较好的改变.

(a) YOLOv3

(b) 改进的 YOLOv3

(c) YOLOv3-GhostNet

图 10 行人检测效果对比图

6 结束语

本文提出了一种基于改进 YOLOv3 的智慧足球场行人检测方法, 旨在解决大型智慧足球场馆比赛中对观众行人以及场上运动员目标检测任务中存在检测精度较低、网络实时性较差的问题. 采用 GhostNet 网络取代原 YOLOv3 的主干特征提取网络, 有效减少网络参数量的同时使得模型更加轻量化, 大大降低了网络计算复杂度. YOLOv3-GhostNet 网络检测的 *mAP* 达到了 90.97%, 相较于 YOLOv3 网络模型参数量减少了约 81.4%, 检测速率可以达到 45.32 fps, 在兼顾精度与速度的同时能够较快速的达到良好的检测效果, 满足智慧足球场馆内进行比赛时对场上快速移动的运动员进行实时检测的应用场景. 同时也适用于传统公共场所内对行人的实时检测, 便于在小型移动嵌入式设备上完成轻量化模型的部署.

参考文献

1 Viola P, Jones MJ. Robust real-time face detection. International Journal of Computer Vision, 2004, 57(2): 137–154. [doi: 10.1023/B:VISI.0000013087.49260.fb]

2 Wang XY, Han TX, Yan SC. An HOG-LBP human detector with partial occlusion handling. Proceedings of the IEEE 12th International Conference on Computer Vision. Kyoto: IEEE, 2009. 32–39. [doi: 10.1109/ICCV.2009.5459207]

3 Chen PH, Lin CJ, Schölkopf B. A tutorial on *v*-support vector machines. Applied Stochastic Models in Business and Industry, 2005, 21(2): 111–136. [doi: 10.1002/asmb.537]

4 金立生, 王岩, 刘景华, 等. 基于 Adaboost 算法的日间前方车辆检测. 吉林大学学报 (工学版), 2014, 44(6): 1604–1608. [doi: 10.13229/j.cnki.jdxbgxb201406011]

5 Girshick R, Donahue J, Darrell T, *et al*. Rich feature hierarchies for accurate object detection and semantic segmentation. Proceedings of 2014 IEEE Conference on Computer Vision and Pattern Recognition. Columbus: IEEE, 2014. 580–587. [doi: 10.1109/CVPR.2014.81]

6 Girshick R. Fast R-CNN. Proceedings of 2015 IEEE International Conference on Computer Vision (ICCV). Santiago: IEEE, 2015. 1440–1448. [doi: 10.1109/ICCV.2015.169]

7 Ren SQ, He KM, Girshick R, *et al*. Faster R-CNN: Towards real-time object detection with region proposal networks. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2017, 39(6): 1137–1149. [doi: 10.1109/TPAMI.2016.2577031]

8 He KM, Gkioxari G, Dollár P, *et al*. Mask R-CNN. Proceedings of 2017 IEEE International Conference on Computer Vision (ICCV). Venice: IEEE, 2017. 2980–2988.

© 中国科学院软件研究所 http://www.c-s-a.org.cn

[doi: 10.1109/ICCV.2017.322]

9 Cai ZW, Vasconcelos N. Cascade R-CNN: Delving into high quality object detection. Proceedings of 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City: IEEE, 2018. 6154–6162. [doi: 10.1109/CVPR.2018.00644]

10 Redmon J, Divvala S, Girshick R, *et al*. You only look once: Unified, realtime object detection. Proceedings of 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Las Vegas: IEEE, 2016. 779–788. [doi: 10.1109/CVPR.2016.91]

11 Redmon J, Farhadi A. YOLO9000: Better, faster, stronger. Proceedings of 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Honolulu: IEEE, 2017. 6517–6525. [doi: 10.1109/CVPR.2017.690]

12 Redmon J, Farhadi A. YOLOv3: An incremental improvement. arXiv:1804.02767, 2018.

13 Liu W, Anguelov D, Erhan D, *et al*. SSD: Single shot MultiBox detector. Proceedings of the 14th European Conference on Computer Vision (ECCV). Amsterdam: Springer, 2016. 21–37.

14 Lin TY, Dollár P, Girshick R, *et al*. Feature pyramid networks for object detection. Proceedings of 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Honolulu: IEEE, 2017. 936–944. [doi: 10.1109/CVPR.2017.106]

15 黄同愿, 杨雪姣, 向国徽, 等. 基于 YOLOv3 的改进模型在行人检测中的应用. 重庆理工大学学报 (自然科学), 2020, 34(8): 155–164.

16 张路达, 邓超. 多尺度融合的 YOLOv3 人群口罩佩戴检测方法. 计算机工程与应用, 2021, 57(16): 283–290. [doi: 10.3778/j.issn.1002-8331.2103-0505]

17 Zheng ZH, Wang P, Liu W, *et al*. Distance-IoU loss: Faster and better learning for bounding box regression. Proceedings of the 34th AAAI Conference on Artificial Intelligence. New York: AAAI, 2020. 12993–13000.

18 He KM, Zhang XY, Ren SQ, *et al*. Spatial pyramid pooling in deep convolutional networks for visual recognition. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2015, 37(9): 1904–1916. [doi: 10.1109/TPAMI.2015.2389824]

19 Han K, Wang YH, Tian Q, *et al*. GhostNet: More features from cheap operations. Proceedings of 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Seattle: IEEE, 2020. 1577–1586. [doi: 10.1109/CVPR42600.2020.00165]

(校对责编: 牛欣悦)

© 中国科学院软件研究所 http://www.c-s-a.org.cn