文章编号:1671-6833(2022)04-0016-07

基于改进 YOLOv4 的自然人群口罩佩戴检测方法

薛均晓, 武雪程, 王世豪, 田萌萌, 石 磊

(郑州大学 网络空间安全学院,河南 郑州 450002)

摘 要: 针对自然场景下的人群口罩佩戴检测常常会受到口罩样式、颜色,佩戴者肤色以及天气等多种因素的影响,提出在原 YOLOv4 的基础上引入协调注意力机制,进而提升主干特征提取网络对于浅层次特征图像位置信息的利用进而更好地捕获小物体——口罩,同时能够丰富浅层次特征图像的语义信息和加强远距离依赖关系,更精准地定位和识别目标区域;对 YOLOv4 的网络结构进行改进以提升整体网络的容量以及深度,进而扩大感受野并提升算法的鲁棒性;引入 DIoU-NMS 在于缓解目标存在遮拦而被错误抑制的现象,DIoU-NMS 从 IoU 指标及检测框的中心点距离两个方面进行非极大值抑制,使得对于 IoU 阈值的选取没有那么苛刻。实验结果表明,改进 YOLOv4 的平均精度均值达到 95.81%,相较于原 YOLOv4 平均精度均值提升了 4.62%。改进后的 YOLOv4 具有良好的性能,能够在自然场景下准确地完成口罩佩戴检测任务。

关键词: 深度学习; 口罩佩戴检测; YOLOv4; 协调注意力机制; 神经网络

中图分类号: TP39 **文献标志码**: A **doi**:10.13705/j.issn.1671-6833.2022.04.020

0 引言

近期,国内新冠疫情防控态势依然很严峻,在人流量密集的场所极易出现病毒传播及感染现象。当前首要的防控措施仍为佩戴口罩,目前公共场所的防疫工作主要是人为地检查行人的口罩佩戴情况,这样做极易造成漏检及误检现象的发生。随着深度学习技术的进步,计算机视觉领域也取得了非常大的进展,利用飞速发展的计算机视觉技术来进行自然场景下的人群口罩佩戴检测任务能在很大程度上取得更好的疫情防控结果。

自然场景下的人群口罩佩戴检测任务实则为目标检测任务。近年来,应用神经网络来完成目标检测任务的方法层出不穷,诞生了 Faster R-CNN[1] 等诸多基于候选区域的目标检测算法。之后相关研究者提出了改进的单阶目标检测算法,与此同时 YOLO[2] 系列的目标检测算法也取得了迅速发展。

YOLO 系列算法最早由 Joseph Redmon 等于 2015 年提出,并在 YOLO 算法的基础上开发了 YOLOv2[3],YOLOv3[4]。2020 年,YOLO 系列算法的研究者 Alexey Bochkovskiy 提出 YOLOv4[5],该算法基于 YOLOv3,应用了大量的新兴手段,实现了速度与精度的完美均衡。

现有的主流检测算法应用到自然人群口罩佩戴检测任务中,会受到口罩样式颜色各异、佩戴者肤色以及天气等多种因素的影响,从而导致检测算法的准确率降低、鲁棒性较差以及不能够满足算法实时性要求等。其中,Faster R-CNN 虽具有较高的准确率,但由于其网络结构的限制使得检测速率达不到基本的实时性要求;YOLOv3 对于小目标物体的检测效果不够理想,并且整体的检测效果也相对较差。基于此,本文基于 YOLOv4 提出了一种优化的口罩佩戴检测算法。本文在原 YOLOv4 主干特征提取网络中引入协调注意力机制,并对网络结构进行改进,在后处理阶段使用 DIoU-NMS 来进行非极大值抑制。实验结果表明,本文提出的算法可以更好地满足疫情防控的实际需求。

1 YOLOv4 算法

YOLOv4 算法可以分成输入端、主干特征提

收稿日期:2021-11-12;**修订日期**:2022-03-09

基金项目:国家自然科学基金资助项目(62006210);河南省高等学校青年骨干教师培养计划(22020GGJS014)

通信作者:石磊(1967—),男,河南郑州人,郑州大学教授,博士,博士生导师,主要从事人工智能、网络空间安全方面的研究,E-mail: shilei@zzu.edu.cn。

取网络、加强特征提取网络以及输出端 4 部分。输入端包含图像预处理、将输入图像的大小变换为规定的输入大小以及对图像信息进行归一化等多种处理;其次是主干特征提取网络,YOLOv4 加入融合 CSPNet[6] 后组成的 CSPDarknet53 作为主干特征提取网络;YOLOv4 的加强特征提取网络由 PANet[7] 结构组成,用于提升特征表达的多样性以及加强特征信息的融合;最后为输出端,用于进行分类与回归任务以及最终预测结果的输出。

YOLOv4 在经过主干特征提取网络之后共有 3 个输出,分别为 L3、L4 以及 L5。其中 L3 与 L4 在经过一次 1×1 的卷积之后会输出至加强特征提取网络进行相应的特征融合;L5 则会经过 3 次卷积后输入至空间金字塔池化层 SPP,之后会进入加强特征提取网络进行特征融合。最后送入输出端便可得到最终的检测结果。

2　改进的 YOLOv4 算法

2.1　协调注意力机制

协调注意力机制是由 Hou 等[8] 提出的一种新颖的注意力机制。协调注意力机制创新性地将空间位置信息嵌入到通道注意力中,进而解决 SE(squeeze-and-excitation)中存在的只考虑内部通道信息而忽略位置信息的问题,同时也解决了 CBAM(convolutional block attention module)无法获取远距离依赖关系的问题,并且避免引入大的计算开销。由于复杂自然场景下的人群口罩佩戴检测任务极易受物体遮拦或自然环境的影响,并且特征图像的分辨率越高,其感受野相对越小,对于特征的敏感程度也越高并且拥有更为丰富的空间位置信息,因此本文将协调注意力模块放在主干特征提取网络的前端进行指导。协调注意力模块的引入能够进一步提升主干特征提取网络对于小目标物体的捕捉能力,丰富浅层次特征图像的语义信息以及获取到更大区域的位置信息[9]。结合协调注意力机制后的主干特征提取网络能够捕获跨通道信息以及对方向、位置敏感的信息从而更精准地定位和识别目标区域。

协调注意力模块如图 1 所示,该模块分为两个步骤:坐标信息嵌入以及协调注意力生成。

在坐标信息嵌入这一部分,协调注意力模块将二维的全局池化转换为分别只对单一维度的特征编码,促使协调注意力模块能够捕捉到具有精确位置信息的远程空间交互。给定输入 $\boldsymbol{X}$,使用

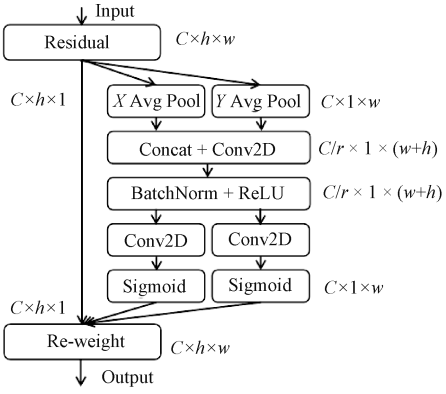

图 1　协调注意力模块

Figure 1　Coordinate attention module

尺寸为 $(h,1)$ 和 $(1,w)$ 的池化核,分别沿水平坐标方向以及垂直坐标方向对每个通道进行编码。如式(1)及式(2)所示,这两种变换沿着水平方向和垂直方向进行特征聚合并返回一组方向感知的特征图像。

$$\boldsymbol{Z}_c^h(h)=\frac{1}{w}\sum_{0\leqslant i<w}\boldsymbol{X}_c(h,i);\tag{1}$$

$$\boldsymbol{Z}_c^w(w)=\frac{1}{h}\sum_{0\leqslant j<h}\boldsymbol{X}_c(j,w)。\tag{2}$$

式中:$\boldsymbol{Z}_c^h(h)$、$\boldsymbol{Z}_c^w(w)$ 分别为沿着垂直方向和水平方向进行平均池化操作后得到的输出结果;h 和 w 分别为特征图像的高度及宽度。

在协调注意力生成部分,如式(3)所示,首先对该组方向感知的特征图像进行堆叠并通过卷积压缩其通道数,此时特征通道数为 C/r,之后通过 BatchNorm 以及 ReLU 对垂直方向和水平方向的位置信息进行编码。接着沿空间维度将 $\boldsymbol{f}$ 切分为两个单独的张量 $\boldsymbol{f}^h\in\mathbf{R}^{C/r\times h}$ 和 $\boldsymbol{f}^w\in\mathbf{R}^{C/r\times w}$,再分别利用两次卷积将 $\boldsymbol{f}^h$ 和 $\boldsymbol{f}^w$ 变换为和输入 $\boldsymbol{X}$ 相等的特征通道数并进行归一化加权处理,得到 $\boldsymbol{g}^h$ 和 $\boldsymbol{g}^w$,如式(4)、(5)所示。

$$\boldsymbol{f}=\delta(F_1([\boldsymbol{Z}_c^h,\boldsymbol{Z}_c^w]));\tag{3}$$

$$\boldsymbol{g}^h=\sigma(F_h(\boldsymbol{f}^h));\tag{4}$$

$$\boldsymbol{g}^w=\sigma(F_w(\boldsymbol{f}^w))。\tag{5}$$

式中:$F_1(\cdot)$ 表示首先对输入的两个张量进行堆叠,接着进行 1 次 1×1 的卷积,再进行 BatchNorm 数据归一化处理;$\delta(\cdot)$ 为 ReLU 激活函数;$\boldsymbol{g}^h$ 为生成的垂直方向权重;$\boldsymbol{g}^w$ 为生成的水平方向权重;$F_h(\cdot)$ 以及 $F_w(\cdot)$ 分别代表一次 1×1 的卷积,用于调整通道数;$\sigma(\cdot)$ 代表 Sigmoid 归一化加权处理。

最后对 $\boldsymbol{g}^h$ 和 $\boldsymbol{g}^w$ 进行扩展,作为协调注意力权重,协调注意力模块的最终输出如式(6)所示:

$$\boldsymbol{y}_c(i,j)=\boldsymbol{x}_c(i,j)\times \boldsymbol{g}_c^h(i)\times \boldsymbol{g}_c^w(j)。\quad (6)$$

式中：$\boldsymbol{x}_c(i,j)$ 为原始输入；$\boldsymbol{g}_c^h(i)$ 为垂直方向的权重；$\boldsymbol{g}_c^w(j)$ 为水平方向的权重；$\boldsymbol{y}_c(i,j)$ 为经过加权后得到的特征图像。

2.2 网络结构改进

原 YOLOv4 中的加强特征提取网络虽然在一定程度上可以提升深层次特征图像的语义表征及特征融合能力[10]，但在自然场景下的人群口罩佩戴检测任务中，因目标较小及天气等多种因素的影响而导致实际检测结果并不理想。因此本文将空间金字塔池化层 SPP 前后的卷积层数均提升为 5 层。受到原始 YOLOv4 网络结构的启发，本文对 L3 及 L4 输出至加强特征提取网络之前的卷积层数进行提升，由原先的 1 层卷积提升为 3 层卷积。改进后的 YOLOv4 网络结构图如图 2 所示。这样做可进一步提升整体网络的容量及深度，提取到更深层次的特征，进一步扩大感受野，同时提升语义表征能力以及算法的鲁棒性[11]。

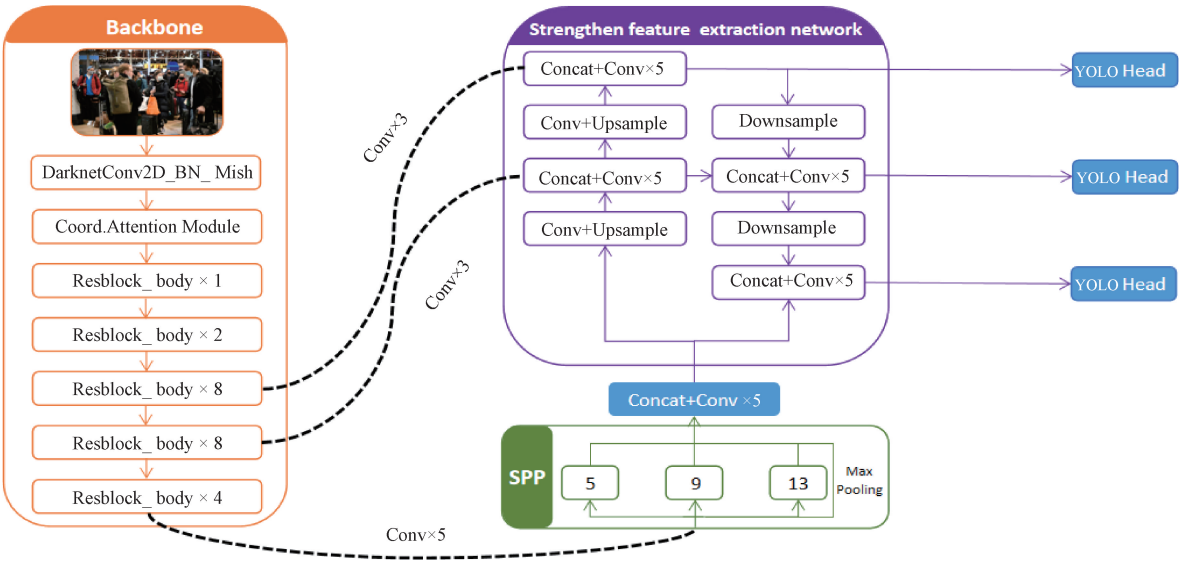

图 2 改进 YOLOv4 网络结构图

Figure 2 Improved YOLOv4 network structure diagram

2.3 DIoU-NMS

NMS(non maximum suppression)是目标检测任务中后处理阶段必备的算法，其目的在于去除冗余检测框，保留最为准确的检测框。

传统 NMS(traditional-NMS)只使用 IoU 指标来抑制冗余框，但是这种方法对于目标重叠以及遮拦的情况经常出现错误的抑制；soft-NMS 与传统 NMS 的区别仅在于对于目标所得预测分数的处理，并不能避免由于目标重叠以及遮拦导致的错误抑制的发生。基于上述问题，本文选取 DIoU-NMS(Distance-IoU-NMS)来替代传统 NMS。DIoU-NMS 将 DIoU 作为 NMS 的准则，其在进行抑制的过程中不仅只考虑 IoU 指标，而且还考虑两个检测框之间的中心点距离。DIoU-NMS 的优势在于如果两个检测框的 IoU 值相同则会对两个检测框的中心点距离进行比较：当两个检测框中心点距离较远时，DIoU-NMS 则判定它们可能位于不同的目标上，并且 DIoU-NMS 对于 IoU 阈值的选取也并不苛刻。

DIoU-NMS 的评价公式如式(7)及式(8)所示：

$$S_i=0,\ f_{\mathrm{DIoU}}(M,B_i)\geqslant thresh;\quad (7)$$

$$S_i=S_i,\ f_{\mathrm{DIoU}}(M,B_i)<thresh。\quad (8)$$

式中：$thresh$ 为设定的 IoU 阈值；$f_{\mathrm{DIoU}}(M,B_i)$ 为 DIoU-NMS 的判定结果；$S_i=0$ 表示该检测框为冗余框，$S_i=S_i$ 表示该框为不同的目标框。

由于 DIoU 的特性使得本文对于 IoU 阈值的选取不再那么苛刻，并且将冗余框与得分值最高的检测框的中心点距离考虑进来，更有助于缓解目标存在遮拦而被错误抑制的现象。

3 实验及结果分析

本文实验基于 PyTorch 1.8.1 框架，编程语言为 python3.7，操作系统为 Ubuntu16.04，GPU 为 NVIDIA RTX2080Ti，集成开发环境为 PyCharm。网络的输入大小为 416×416，优化器使用 Adam 优化算法。权重衰减值 decay 设置为 0.000 5，共训练 120 个 epoch，其中在冰冻训练阶段共训练 30 个 epoch，初始学习率设置为 0.001；在解冻训练阶段共训练 90 个 epoch，初始学习率设置为

0.000 1,采用余弦退火衰减学习率调整策略[12]。

3.1　数据集介绍

目前有关自然场景下人群口罩佩戴检测任务的数据集很少,且普遍缺少口罩佩戴错误这一类别的数据。因此本文的数据集主要为利用网络爬虫技术爬取到的符合本文要求的图片,从 RMFD 等开放数据集中选取符合本文要求的一部分图片,以及通过网络视频抽帧获得的部分图片,除此之外又自制了部分图片进而构成了本文的原始数据集。原始数据集中共计 7 854 张图片,包含佩戴口罩、未佩戴口罩、口罩佩戴错误 3 个类别,涉及公共街道、地铁站等多个公共自然场景。口罩佩戴错误的情况主要包含未遮住鼻子、未遮住口鼻以及口罩在下巴处这 3 种最常见的口罩佩戴错误示例。数据集依照 Pascal VOC 格式进行标注,使用的标注工具为 LabelImg。由于算力的限制以及 Mosaic 数据增强方法的不稳定性,针对原始数据集中数据量相对较小及所含口罩佩戴错误类别数据数量较少的情况,本文采用包括仿射旋转变换、高斯模糊、高斯滤波等多种方法在内的随机几何数据增强方法。图 3(b)~3(d)均为本文使用上述随机几何数据增强方法的示例。数据增强之后数据集中图片的总数量为 11 447,佩戴口罩、未佩戴口罩以及口罩佩戴错误这 3 个类别的数目分别为 13 458、8 123、7 164。

图 3　原图与数据增强后图片对比

Figure 3　Comparison between the original image and the images after random data augmentation

3.2　评价指标及实验结果分析

在本文的实验中所选取的算法性能评价指标为平均精度(*AP*)和平均精度均值(*mAP*)。*AP* 值从准确率 *P* 和召回率 *R* 两个方面来衡量模型的性能。准确率 *P* 表示实际是正类并且被预测为正类的样本占所有预测为正类的样本的比例;召回率 *R* 表示实际是正类并且被预测为正类的样本占所有实际为正类样本的比例,公式如式(9)、(10)所示:

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

$$R = \frac{TP}{TP + FN}。 \tag{10}$$

式中:*TP* 为被检测为正样本的正样本;*FN* 为被误检为负样本的正样本;*FP* 为被误检为正样本的负样本。

AP 值由准确率-召回率曲线积分计算。*AP* 的值越高,模型表现越好。*mAP* 是各个类别 *AP* 值的平均值,用于衡量多个目标类别的平均检测精度。*mAP* 值的大小可体现出模型对所有类别综合检测性能的高低。

$$AP = \int PR\mathrm{d}R; \tag{11}$$

$$mAP = \frac{\sum_i^c C_i}{c}。 \tag{12}$$

式中:C_i 代表各个类别的 *AP* 值;*c* 为任务中所需要检测的类别数目。

改进 YOLOv4 的平均精度如图 4 所示。最终算法的 *mAP* 值为 95.81%。与原 YOLOv4 相比,*mAP* 值提升了 4.62%。

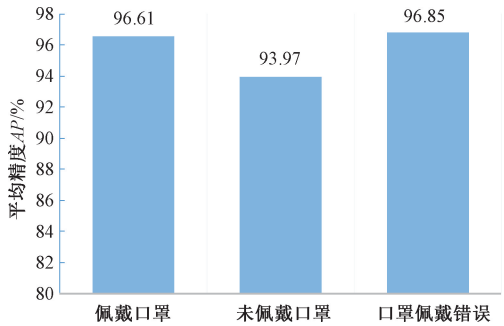

图 4　平均精度图

Figure 4　Diagram of average precision

为综合评估本算法的性能,本文将改进 YOLOv4 的实验结果与其他主流目标检测算法的实验结果进行对比,具体实验结果如表 1 所示。由表 1 可以看出,改进 YOLOv4 算法相较于 Faster R-CNN、SSD、YOLOv3 以及原 YOLOv4 的 *AP* 值均有较大提升;同时,改进 YOLOv4 算法的 *mAP* 值也优于以上 4 种基线方法。本文提出的 YOLOv4 算法改进方案加强了对图像特征的提取能力,进一步提升了对小目标物体的检测能力,以及特征图像的语义表征能力,扩大了深层次特征图像的感受野,同时对细节特征的处理也有所提升。通过对比目前主流的检测算法,进一步证实了本文

表1 不同主流检测算法的性能对比

Table 1 Performance comparison of different mainstream detection algorithms

算法	AP/%			mAP/%
	佩戴口罩	未佩戴口罩	口罩佩戴错误	
Faster R-CNN[1]	88.32	89.16	91.84	89.77
SSD[13]	84.75	81.68	93.48	86.63
YOLOv3[4]	87.12	89.28	93.56	89.99
YOLOv4[5]	90.28	89.75	93.54	91.19
改进 YOLOv4	96.61	93.97	96.85	95.81

针对自然人群口罩佩戴检测任务提出的算法设计方案的有效性。

对于自然场景下的人群口罩佩戴检测任务而言,要求算法具有较好的实时性。因此本文对算法的实时性也进行了测试,设置样本数量为5,在同一台设备上对改进YOLOv4算法以及主流的目标检测算法的实时性进行了测试,测试结果如表2所示。由表2可以看出,本文提出的算法在较大幅度提升检测精度的前提下,其检测速率为24.40帧/s,比原YOLOv4算法仅下降1.29帧/s,仍能够满足基本的算法实时性要求。

表2 检测速率对比

Table 2 Comparison of detection rates of different mainstream detection algorithms

算法	检测一张图片所需的时间/s	帧率/(帧·s^{-1})
Faster R-CNN[1]	0.054	18.44
SSD[13]	0.016	62.50
YOLOv3[4]	0.024	41.06
YOLOv4[5]	0.039	25.69
改进 YOLOv4	0.041	24.40

3.3 消融实验

本文提出了一系列的消融实验方案来验证算法改进方案的有效性。消融实验中共设置有7组实验,具体方案如下所示:第1组,原YOLOv4算法,设置该组实验的目的在于作为对照以确定改进是否有效;第2组,RFB(receptive fields block)模块+网络结构改进;第3组,SE模块+网络结构改进;第4组,CBAM模块+网络结构改进;第5组,CA(coordinate attention)模块+网络结构改进;第6组,CA模块+网络结构改进+soft-NMS;第7组,CA模块+网络结构改进+DIoU-NMS,即改进YOLOv4。

此处设置第3组及第4组实验的目的在于确定本文所使用的CA模块的有效性。设置第6组实验的目的在于与第7组实验进行对比,以确定后处理阶段DIoU-NMS的有效性。

第2组实验只关注对于深层次特征图像的语义表征能力的提升。文中使用RFB模块来替代原先网络结构中的空间金字塔池化层,得以进一步扩大感受野,目的在于进一步提升深层次特征图像的语义表征能力,进而与第5组实验形成鲜明对照。消融实验结果如表3所示。

本文通过消融实验,证实了CA模块相较于SE模块以及CBAM模块在该任务中的有效性及网络结构改进的有效性,同时也证实了DIoU-NMS相较于传统NMS以及soft-NMS在处理目标重叠以及遮拦而发生错误抑制问题时的有效性。除此之外,本文还证实了同时加强深层次特征图像以及浅层次特征图像的语义表征能力的效果,要优于仅加强深层次特征图像的语义表征能力的效果。最终在没有过多增加参数量的情况下,本文提出的算法改进方案的*mAP*值比原YOLOv4算法提高了4.62百分点。

分析可知,本文提出的改进YOLOv4算法性能提升的原因在于将CA模块应用于主干特征提取网络的前端进行指导,致使浅层次特征图像的语义表征能力更强;同时,网络结构的改进进一步提升深层次特征图像的语义表征能力。因此本文提出的算法改进方案使得3个中间特征图像L3、

表3 消融实验结果

Table 3 Results of ablation studies

算法	AP/%			mAP/%	模型参数量/10^6
	佩戴口罩	未佩戴口罩	佩戴口罩错误		
YOLOv4	90.28	89.75	93.54	91.19	64.36
RFB+网络结构改进	96.22	93.51	96.23	95.31	69.87
SE+网络结构改进	91.48	92.72	92.11	92.10	64.85
CBAM+网络结构改进	92.22	91.87	93.44	92.51	65.33
CA+网络结构改进	96.24	93.97	96.69	95.63	64.82
CA+网络结构改进+soft-NMS	96.28	93.88	96.64	95.60	64.88
改进 YOLOv4	96.61	93.97	96.85	95.81	64.84

L4、L5 在输入至加强特征提取网络之前含有比原 YOLOv4 更加丰富的语义信息,并且使网络可以更加精准地定位和识别目标区域。

3.4　检测效果对比

实际检测效果对比如图 5 所示,其中图 5(b)为改进 YOLOv4 检测效果图,图 5(a)为原 YOLOv4 检测效果图。在复杂的自然人群场景下原 YOLOv4 存在着较为严重的漏检以及错检现象,并且对于人物侧脸以及受物体遮挡的情况处理效果不佳,而改进 YOLOv4 的检测结果中目标漏检以及错检的现象均有所缓解,同时算法提升了对于人物侧脸以及受物体遮挡情况的处理效果。

(a) 原YOLOv4检测效果图

(b) 改进YOLOv4检测效果图

图 5　改进前后 YOLOv4 检测效果图

Figure 5　YOLOv4 test results before and after improvement

4　结论

本文提出一种基于改进 YOLOv4 的口罩佩戴检测算法。该算法针对主干特征提取网络提取的浅层次特征拥有更丰富的空间位置信息但匮乏语义信息这一特性,通过引入协调注意力机制来加强主干特征提取网络在复杂自然场景下对于小目标物体的检测能力;增加空间金字塔池化层 SPP 前后的卷积层数,以及 L3 和 L4 在输出至 PANet 之前的卷积层数进而提升整体网络的容量和深度;在后处理阶段应用 DIoU-NMS 来缓解错误抑制的现象。实验结果表明,本文提出的改进 YOLOv4 具有良好的性能,能够满足疫情防控场景下的实际需求。下一步将会在保证高准确率的基础上进一步对模型进行剪枝以及优化,以期获得更高的检测速率,进而更好地提升算法的实时性。

参考文献:

[1] 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.

[2] 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. Piscataway: IEEE, 2016: 779-788.

[3] REDMON J, FARHADI A. YOLO9000: better, faster, stronger[C]//2017 IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2017: 7263-7271.

[4] REDMON J, FARHADI A. YOLOv3: an incremental improvement[EB/OL]. (2018-04-08)[2021-11-01]. https://arxiv.org/abs/1804.02767.

[5] BOCHKOVSKIY A, WANG C Y, LIAO H Y M. YOLOv4: optimal speed and accuracy of object detection[EB/OL]. (2020-04-23)[2021-11-01]. https://arxiv.org/abs/2004.10934.

[6] WANG C Y, MARK LIAO H Y, WU Y H, et al. CSPNet: a new backbone that can enhance learning capability of CNN[C]//2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops. Piscataway: IEEE, 2020: 390-391.

[7] 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. Piscataway: IEEE, 2018: 8759-8768.

[8] HOU Q B, ZHOU D Q, FENG J S. Coordinate attention for efficient mobile network design[C]//2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2021: 13713-13722.

[9] 张坚鑫, 郭四稳, 张国兰, 等. 基于多尺度特征融合的火灾检测模型[J]. 郑州大学学报(工学版),

2021, 42(5): 13-18.
ZHANG J X, GUO S W, ZHANG G L, et al. Fire detection model based on multi-scale feature fusion[J]. Journal of Zhengzhou university (engineering science), 2021, 42(5): 13-18.

[10] 薛均晓, 程君进, 张其斌, 等. 改进轻量级卷积神经网络的复杂场景口罩佩戴检测方法[J]. 计算机辅助设计与图形学学报, 2021, 33(7): 1045-1054.
XUE J X, CHENG J J, ZHANG Q B, et al. Improved efficient convolutional neural networks for complex scene mask-wearing detection[J]. Journal of computer-aided design & computer graphics, 2021, 33(7): 1045-1054.

[11] 李润川, 张行进, 陈刚, 等. 基于多特征融合的心搏类型识别研究[J]. 郑州大学学报(工学版), 2021, 42(4): 7-12.
LI R C, ZHANG X J, CHEN G, et al. Research on heartbeat type recognition based on multi-feature fusion [J]. Journal of Zhengzhou university (engineering science), 2021, 42(4): 7-12.

[12] LOSHCHILOV I, HUTTER F. SGDR: stochastic gradient descent with warm restarts[EB/OL]. (2017-05-03) [2021-11-01]. https://arxiv.org/abs/1608.03983v5.

[13] LIU W, ANGUELOV D, ERHAN D, et al. SSD: single shot MultiBox detector[C]//European conference on computer vision. Cham: Springer, 2016: 21-37.

A Method on Mask Wearing Detection of Natural Population Based on Improved YOLOv4

XUE Junxiao, WU Xuecheng, WANG Shihao, TIAN Mengmeng, SHI Lei

(School of Cyber Science and Engineering, Zhengzhou University, Zhengzhou 450002, China)

Abstract: The mask wearing detection in natural scenes is often affected by various factors such as the style and color of the mask, the skin color of the wearer, and the weather. In this study, based on the original YOLOv4, the coordinate attention mechanism was introduced to improve the utilization of the backbone network for spatial information of shallow feature maps and better capture small objects-masks. At the same time, it could enrich the semantic information of shallow feature maps and strengthen the long-distance dependencies to more accurately locate and identify object regions. This paper improved the network structure of YOLOv4 to enhance the capacity and depth of the overall network, so as to expand the receptive fields and improved the robustness of the algorithm. The introduction of DIoU-NMS could alleviate the phenomenon that the object was blocked and incorrectly suppressed. DIoU-NMS could perform NMS from the two aspects of IoU and center point distance of bounding boxes, so that the selection of the IoU threshold was not so harsh. The experimental results showed that the average precision of the improved YOLOv4 was 95.81%, which was 4.62% higher than the average precision of the original YOLOv4. The improved YOLOv4 had exciting performance and could complete the task of comprehensive and accurate mask wearing detection in natural scenarios.

Keywords: deep learning; mask wearing detection; YOLOv4; coordinate attention mechanism; neural network