

基于改进 YOLOv3 的机坪工作人员反光背心检测研究

徐涛, 陈奕仁, 吕宗磊

引用本文

徐涛, 陈奕仁, 吕宗磊. 基于改进 YOLOv3 的机坪工作人员反光背心检测研究[J]. 计算机科学, 2022, 49(4): 239-246.

XU Tao, CHEN Yi-ren, LYU Zong-lei. [Study on Reflective Vest Detection for Apron Workers Based on Improved YOLOv3 Algorithm](#)[J]. Computer Science, 2022, 49(4): 239-246.

相似文章推荐 (请使用火狐或 IE 浏览器查看文章)

Similar articles recommended (Please use Firefox or IE to view the article)

[基于多级特征融合与注意力模块的场景识别方法](#)

Scene Recognition Method Based on Multi-level Feature Fusion and Attention Module

计算机科学, 2022, 49(4): 209-214. <https://doi.org/10.11896/jsjcx.210100135>

[基于改进 U-Net 网络的液滴分割方法](#)

Droplet Segmentation Method Based on Improved U-Net Network

计算机科学, 2022, 49(4): 227-232. <https://doi.org/10.11896/jsjcx.210300193>

[基于特征定位与融合的行人重识别算法](#)

Person Re-identification Based on Feature Location and Fusion

计算机科学, 2022, 49(3): 170-178. <https://doi.org/10.11896/jsjcx.210100132>

[基于空洞卷积和多特征融合的混凝土路面裂缝检测](#)

Concrete Pavement Crack Detection Based on Dilated Convolution and Multi-features Fusion

计算机科学, 2022, 49(3): 192-196. <https://doi.org/10.11896/jsjcx.210100164>

[基于图像分块与特征融合的户外图像天气识别](#)

Outdoor Image Weather Recognition Based on Image Blocks and Feature Fusion

计算机科学, 2022, 49(3): 197-203. <https://doi.org/10.11896/jsjcx.201200263>

基于改进 YOLOv3 的机坪工作人员反光背心检测研究

徐 涛 陈奕仁 吕宗磊

中国民航大学计算机科学与技术学院 天津 300000

(txu@cauc.edu.cn)

摘 要 文中提出了一种基于先验知识和改进 YOLOv3 算法的机坪工作人员反光背心检测算法。该方法针对现有目标检测方法速度偏低的问题,基于先验知识生成反光背心检测候选区域来替换初始候选区域,以减少检测区域面积,使用 Darknet-37 替代 Darknet-53 作为骨干网络进行特征提取,提高了算法的检测速度。针对反光背心在画面中所占面积偏小,且辨识难度较大的问题,在检测模型中加入空间金字塔池化结构(SPP),从而实现特征增强,并将检测尺度提升至 4 个,以进行多尺度特征融合。使用 K-means++ 算法对标注边界框尺寸重新进行聚类分析,并用聚类结果替换 YOLOv3 初始 Anchor 值。选取 GIoU 作为损失函数以提高定位精准度。实验结果表明,所提出的新型目标检测算法在自建的反光背心数据集上取得了优于 YOLOv3 的检测结果,精准率和召回率分别达到了 97.6% 和 96.1%,检测速度达到了 28.4 帧/s,有效解决了原模型中存在的定位不准、漏检、检测速度偏低等问题,在保证检测精度较高的情况下满足了机坪目标检测在实际运用中的实时性要求。

关键词: 目标检测;空间金字塔池化;特征融合;实时监测;反光背心检测

中图法分类号 TP391.4

Study on Reflective Vest Detection for Apron Workers Based on Improved YOLOv3 Algorithm

XU Tao, CHEN Yi-ren and LYU Zong-lei

School of Computer Science and Technology, Civil Aviation University of China, Tianjin 300000, China

Abstract This paper proposes a reflective vest detection algorithm for apron staff based on prior knowledge and improved YOLOv3 algorithm. Aiming at the problem of the existing target detection method with low speed, the reflective vest detection candidate region is generated based on prior knowledge to replace the initial candidate region, so as to reduce the detection area. Darknet-37 is used to replace Darknet-53 as the backbone network for feature extraction, which improves the detection speed of the algorithm. Aiming at the problem that the reflective vest occupies a small area in the picture and is difficult to identify, a spatial pyramid pooling structure (SPP) is added into the detection model to realize feature enhancement, and the detection scale is increased to four for multi-scale feature fusion. The K-means++ algorithm is used to perform cluster analysis again on the size of labeled bounding box, and the clustering result is used to replace the initial Anchor value of Yolov3. GIoU is selected as the loss function to improve the positioning accuracy. Experimental results show that the proposed new target detection algorithm in the self-built reflective vest data set is better than YOLOv3 test results, the precision rate and recall rate reach 97.6% and 96.1%, detection rate reach 28.4 frames/s, which effectively solves the problems such as inaccurate positioning, missed detection and low detection speed existing in the original model, and meets the real-time requirements in the practical application of apron target detection while ensuring a high detection accuracy.

Keywords Object detection, Spatial pyramid pooling, Feature fusion, Real-time detection, Reflective vest detection

1 引言

根据民航总局 191 号令规定,机坪工作人员工作时必须穿戴反光背心。为机坪工作人员配置穿戴反光背心,对保障机坪工作人员的安全具有重要作用,特别是在飞机出现故障或机坪上发生交通事故时,穿戴反光背心可以明显提升对工作人员的视认效果,从而减少机坪事故发生率。机坪工作人员未穿戴反光背心或者穿戴形式不正确可能会对机坪上工作

人员的安全造成巨大的威胁,许多在机坪上发生的事故都是由于工作人员未穿戴反光背心而引起的,因此,对于反光背心检测算法的研究具有重大意义。

近年来,随着计算机性能的提升与计算机存储空间的增加,以卷积神经网络为基础的目标检测算法已成为目前主流的目标检测算法之一。卷积神经网络泛化能力较好,具有强大的特征表达能力,能适应多样化背景与目标类型^[1]。Fast-RCNN^[2], Faster-RCNN^[3], RCNN (Region Convolutional

收稿日期:2021-02-19 返修日期:2021-07-02

基金项目:中央高校基本科研业务费项目中国民航大学专项(3122021088)

This work was supported by the Fundamental Research Funds for Central Universities of the Civil Aviation University of China(3122021088).

通信作者:吕宗磊(zllv@cauc.edu.cn)

Neural Network)^[4],Mask RCNN^{[5][6]}等目标检测算法的提出,在很大程度上提升了目标检测算法的准确率,然而这些方法依旧存在着检测速度较慢的问题。而YOLO(You Only Look Once)^[7],SSD(Single Shot Multi-box Detector)^[8],YOLO9000^[9],YOLOv3^[10-14],YOLOv4^[15-16],MobileNetV3^[17-18]等目标检测算法虽实时性较好,却存在检测精度相对较低的问题。针对本文中反光背心这种机坪小目标的检测与识别问题,使用现有目标检测算法无法在保证检测精度较高的同时满足机坪目标检测在实际运用中的实时性要求(25 帧/s)。因此本文针对反光背心的检测和识别问题,提出了一种基于先验知识和改进YOLOv3的反光背心目标检测算法。本文主要工作如下:

- (1)基于机坪工作人员与飞机位置有相关性的先验知识,使用预先训练好的模型对飞机进行检测,并计算出飞机停靠位置。根据飞机停靠位置生成候选区域,对反光背心数据集进行检测时只选定候选区域进行检测,以此减少了候选区域面积,从而提升了检测速度和精度。
- (2)优化了YOLOv3网络结构,使用自建的Darknet-37网络替代Darknet-53网络作为骨干网络以提取特征,引入特征增强网络空间金字塔池化(Spatial Pyramid Pooling, SPP)^[19]以获得更加丰富的局部特征信息,并将特征融合尺度提升为4尺度,以获得更多小目标特征信息。
- (3)使用K-means++算法^[20-21]对反光背心数据集重新进行维度聚类,得到了更合理的先验框尺寸,从而对反光背心数据集进行更好的预测。
- (4)使用GIoU回归损失函数^[22]替代IoU回归损失函数,以增强对反光背心定位的精度,提升算法的准确率和定位精度。

2 基于先验知识生成候选区域

通常在使用YOLOv3模型对数据集进行训练时,会使用YOLOv3模型对所输入图片的所有区域进行检测。而如图1所示,机坪工作人员的大部分工作都与飞机相关,因此所在区域相对固定,且与飞机停靠位置具有相关性。本文基于机坪工作人员所在区域与飞机停靠位置相对固定的先验知识,在使用本文改进的YOLOv3算法对反光背心数据集进行训练之前,先使用预先训练好的模型对飞机进行检测,若检测到飞机,则会对飞机位置进行追踪,经过计算得到该飞机最终停靠位置的信息,利用机坪工作人员位置与飞机停靠位置有相关性的性质,生成机坪工作人员可能会在的候选区域,之后使用针对反光背心数据集所改进的YOLOv3算法对生成的候选区域进行检测,从而减少了检测区域的面积,提升了检测速度和精度。



图1 机坪工作人员与飞机
Fig. 1 Apron crew and aircraft

如图1所示,由于飞机具有大目标的特性,更需要充分利用深层特征信息,因此需要对检测飞机的YOLOv3模型中的Darknet-53骨干网络进行修改。本文借鉴Darknet-53网络提出了专门针对大目标的Darknet-43网络,作为对飞机数据集进行特征提取的网络。Darknet-43的网络结构如图2所示,本文将主干网络中的2倍、4倍、8倍、16倍、32倍采样之后的残差网络结构数分别调整为4个、4个、4个、4个、2个,同时增加了深层残差层中残差模块数量,并减少了浅层残差层中残差模块数量与卷积核数量,从而更充分地利用深层特征信息提高对飞机的检测精度与速率。

Darknet-43				
	Type	Filters	Size	Output
	Convolutional	32	3×3	416×416
	Convolutional	64	3×3/2	208×208
4×	Convolutional	32	1×1	208×208
	Convolutional	64	3×3	
	Residual			
	Convolutional	64	3×3/2	
4×	Convolutional	32	1×1	104×104
	Convolutional	64	3×3	
	Residual			
	Convolutional	128	3×3/2	
4×	Convolutional	64	1×1	52×52
	Convolutional	128	3×3	
	Residual			
	Convolutional	256	3×3/2	
4×	Convolutional	128	1×1	26×26
	Convolutional	256	3×3	
	Residual			
	Convolutional	512	3×3/2	
2×	Convolutional	256	1×1	13×13
	Convolutional	512	3×3	
	Residual			
	AVGpool		Global	
	Connected	1000		
	Softmax			

图2 Darknet-43网络结构
Fig. 2 Darknet-43 network architecture

使用网络结构经过改进后的YOLOv3模型对飞机数据集进行训练。候选区域生成策略如下。

- (1)遍历所输入机位图片数据集,当检测到飞机时,对该飞机的位置进行追踪,计算每两帧中该飞机检测框的IoU值并保存。
- (2)根据步骤(1)中得出的IoU集计算AVG_IoU。
$$AVG_IoU = \frac{\sum_{i=1}^{sum} IoU_i}{sum}$$
- (3)计算IoU集中最接近AVG_IoU的IoU值,该IoU值所对应的图片中飞机坐标位置为该飞机最终停靠位置。
$$\min(|IoU_i - AVG_IoU|), i \in (1, sum)$$
- (4)根据飞机最终停靠区域的检测框坐标值,计算出中心点Center,以中心点向四周进行扩张,生成该机位反光背心检测的候选区域。生成的机位候选区域图与原图的对比如图3中红色框和原图所示。

如图3所示,该候选区域生成算法可有效定位机坪工作人员所在候选区域。在对反光背心数据集进行检测前,先使用该策略对数据集进行预处理操作得到候选区域范围,之后再使用本文改进的目标检测算法对候选区域进行检测,通过候选区域生成策略减少了检测区域的面积,从而提升了检测速度。同时,使用K-means++算法对经过候选区域生成后的图片进行聚类得到的AVG_IoU值比使用K-means++

算法对原图片进行聚类得到的AVG_IoU值更高,从而得到了更好的聚类效果,实现了检测精度的提升。



图3 原图和候选区域图(电子版为彩色)

Fig. 3 Original image and candidate area map

由于机坪中存在多种业务规则,且每种业务规则中机坪工作人员的位置与飞机停靠位置的相对位置存在差异,因此不同业务规则中生成的候选区域面积会有所不同。但是,根据相应业务规则可以判断哪些区域中机坪工作人员不应当存在,从而使生成的候选区域面积比原图片中候选区域面积更小,进而提升检测速度与精度。考虑到存在特殊情况(人没有在指定的位置上),可对候选区域进行适度扩张,但无论对候选区域如何扩张,都会存在机坪工作人员不可能存在的区域,所以根据该策略生成的候选区域一定会比原图片中的候选区域面积更小,因此本文算法在出现特殊情况时同样有效。

3 YOLOv3 算法改进

3.1 网络结构改进

3.1.1 改进骨干网络

传统的YOLOv3使用Darknet-53网络作为骨干网络来提取图片的特征。但是对于本文中反光背心数据集的检测来说,Darknet-53网络比较复杂、冗余,本文中的反光背心具有小目标的特性,所以需要充分地利用浅层特征信息,且本文中反光背心数据集只为一类,无需在残差层中使用大量卷积核。为了在检测反光背心时充分利用浅层特征信息并提升检测速率,本文借鉴Darknet-53网络,提出了一种运算复杂度相对较低、使用参数较少的Darknet-37作为特征提取网络。Darknet-37的网络结构如图4所示。

Darknet-37				
Type	Filters	Size	Output	
1×	Convolutional	32 3×3	416×416	
	Convolutional	64 3×3/2	208×208	
	Convolutional	32 1×1		
	Convolutional	64 3×3		
	Residual		208×208	
2×	Convolutional	64 3×3/2	104×104	
	Convolutional	32 1×1		
	Convolutional	64 3×3		
	Residual		104×104	
4×	Convolutional	128 3×3/2	52×52	
	Convolutional	64 1×1		
	Convolutional	128 3×3		
	Residual		52×52	
4×	Convolutional	256 3×3/2	26×26	
	Convolutional	128 1×1		
	Convolutional	256 3×3		
	Residual		26×26	
4×	Convolutional	512 3×3/2	13×13	
	Convolutional	256 1×1		
	Convolutional	512 3×3		
	Residual		13×13	
AVGpool			Global	
Connected			1000	
Softmax				

图4 Darknet-37 网络结构

Fig. 4 Darknet-37 network architecture

本文首先调整了网络结构中各残差层的残差模块数量,分别把Darknet-37网络中的2倍、4倍、8倍、16倍、32倍采样之后的残差模块数调整为1个、2个、4个、4个、4个,降低了深层残差层中残差模块数量,并将残差模块中的卷积层滤波器数量降低,相比传统YOLOv3中的Darknet-53网络,Darknet-37降低了参数数量,运算复杂度也随之下降,提升了对反光背心数据集的检测速度。

3.1.2 增加特征增强网络

空间金字塔池化(Spatial Pyramid Pooling,SPP)是由He等^[19]提出的一种用来解决输入神经网络中由于图像尺寸不同,导致不能满足输入要求的问题的方法。通常解决这种问题的手法是裁剪(Crop)和拉伸(Warp),如图5所示,但是此方法会改变图像的横纵比以及输入图像的尺寸,因此会扭曲原始的图像。而He等提出的SPP层可以很好地解决这些问题。如图6所示,其主要思想是将任意尺寸的特征图通过多尺度池化操作拼接成固定长度的特征向量,从而产生固定大小的特征表示。这种方法不需要限定输入图像的尺寸和比例,对于图像形变有较好的鲁棒性,且SPP可以处理多种不同横纵比和尺寸的输入图像,不仅提高了图像的尺度不变性,同时也防止了过拟合。



(a) Crop



(b) Warp

图5 裁剪与拉伸

Fig. 5 Crop and Warp

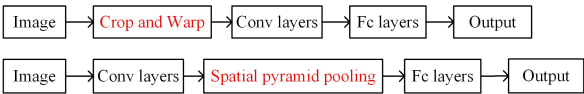


图6 CNN 一般结构和 SPP 结构

Fig. 6 CNN general structure and SPP structure

本文引入SPP结构以获得多尺度局部特征信息,通过与全局特征信息相互融合,从而得到更丰富的特征表示,以此提升预测精度。本文在Darknet-37特征提取网络最后一个特征层的卷积中加入SPP结构,SPP网络结构如图7所示。

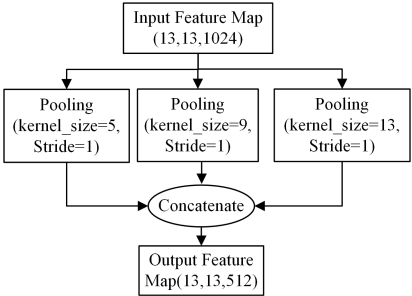


图7 空间金字塔池化网络结构

Fig. 7 Space pyramid pooling network structure

从图7可以看出,SPP网络具体步骤为:分别利用3个

不同尺度的池化层进行最大池化操作,池化层卷积核大小分别为 $5\times5,9\times9,13\times13$,步长均为 1;再堆叠经过池化处理后得到局部特征图。SPP 结构能够大幅度增强最后一个特征层的感受野,分离出其中最显著的上下文特征,因此能获得更加丰富的局部特征信息,从而提升预测精度。

3.1.3 改进多尺度检测

虽然 YOLOv3 中加入了 FPN,并且使用 3 种不同尺度的 Anchor 来增强目标检测的效果,但对于本文中反光背心这种小目标而言效果并不理想,依然会存在对反光背心的错检和漏检的情况。

在 YOLOv3 中,输入的图片经过 5 次下采样后将会依次变成大小为 208,104,52,26,13 的特征图。而 YOLOv3 的 FPN 只能利用 8 倍、16 倍和 32 倍下采样的特征图来检测目标,因此当原输入图像尺寸小于 8×8 时是无法检测到的。本文引入 4 倍下采样特征图作为特征层,如图 8 所示,本文将 YOLOv3 网络的 3 尺度检测提升成 4 尺度检测,将新增的 104×104 的特征词与其他 3 个融合层相互融合,以此获得更多浅层图像的特征信息,提高了 YOLOv3 算法对小目标的检测精度,同时也减少了对反光背心定位的错检和漏检问题。

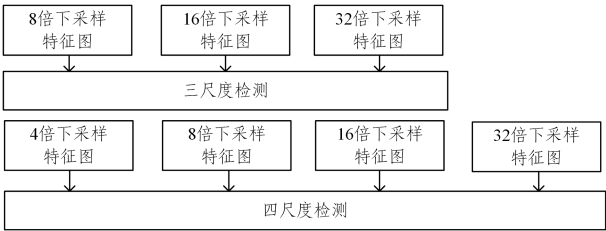


图 8 FPN 网络结构和改进的 FPN 网络结构
Fig. 8 FPN network structure and improved FPN network structure

将原始 YOLOv3 网络按照上述方法进行改进,使用 Darknet-37 替代 Darknet-53 作为特征提取网络,在 Darknet-37 特征提取网络最后一个特征层的卷积中加入 SPP 结构,并使用四尺度特征进行融合,经过改进后的总体网络结构如图 9 所示。

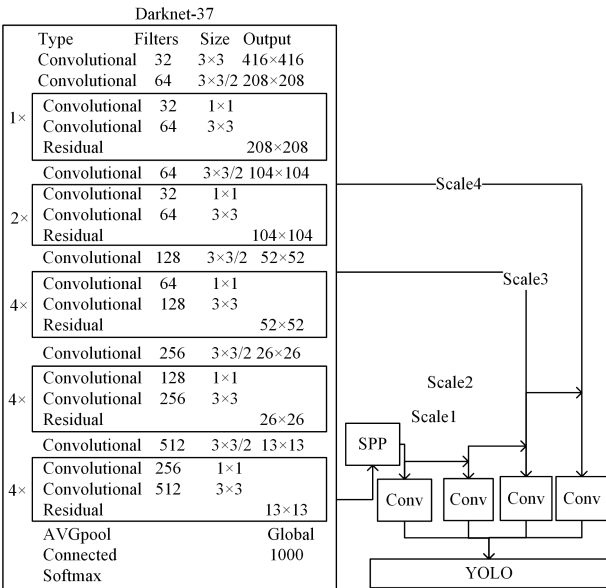


图 9 改进后网络结构示意图
Fig. 9 Improved network structure diagram

将经过候选区域生成后的图片输入改进后的网络结构后,将会在 Darknet-37 网络中进行 5 次下采样,得到最底层的 13×13 的特征图,对该特征图经过空间金字塔池化后,再进行上采样操作,然后拼接尺度为 26×26 的特征图,再对融合后的特征图进行采样,与上一尺度的特征图拼接,直到 4 个尺度特征图完成融合为止。最终,提取融合后的四尺度特征图对反光背心的位置进行预测。经过改进后的网络结构可以有效利用浅层信息,提高对小目标反光背心的检测精度和速度。

3.2 修改初始锚点参数

传统的 YOLOv3 采用 K-means 算法通过对公共数据集聚类得到原始的锚框参数。如图 10 所示,对于本文中的反光背心数据来说,YOLOv3 初始聚类获得的锚框尺寸较大,对应的参数值也较大且具有普遍性,不适用于本文中的反光背心数据集。并且相对于原始的三尺度检测,本文采用四尺度检测,因此需要重新对反光背心标签进行聚类得到 Anchor,方便对反光背心数据集进行更好的预测。

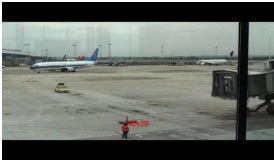


图 10 目标检测框尺寸
Fig. 10 Size of target detection frame

在原始的 YOLOv3 中,使用 K-means 算法作为聚类算法来进行聚类,从而生成锚框参数。K-means 算法使用随机选取的方法来确定初始聚类中心,通过初始聚类中心来确定一个初始划分,然后对初始划分进行优化。因此,对于初始聚类中心的选择会对聚类结果产生较大的影响,一旦初始值选择不佳,可能无法获得一个有效的聚类结果,并且 K-means 算法需要不断对样本进行分类调整,因此时间开销较大。

因此,本文使用 K-means++ 算法作为聚类算法,K-means++ 算法选择初始化 seeds 的基本思想为:使得每个起始聚类中心之间的相互距离尽可能远。K-means++ 算法的步骤如下:

- (1)随机选取输入数据点集合中的一个点作为第一个聚类中心;
- (2)对于数据集中的每个点,计算其与最近聚类中心(已选择的聚类中心)的距离 X ;
- (3)选取一个新的数据点作为新的聚类中心,其中, X 值较大的点被选取为新聚类中心的概率较大;
- (4)重复上述步骤(2)与步骤(3),直至选出 K 个聚类中心为止;
- (5)使用选出的 K 个初始的聚类中心来进行标准 K-means 聚类算法。

可以看出,在 K-means++ 聚类算法中,只有第一个聚类中心是经过随机选取产生的,相比 K-means 算法,K-means++ 聚类算法降低了随机性对聚类结果的影响。

在原始的 YOLOv3 中,K-means 聚类算法采用欧氏距离公式来初始化锚框,以预测边界框的坐标位置,因此锚框的尺寸大小实际上会对检测准确率的高低产生影响。而 IoU 表示预测框与实际框的交并比,由于其不受尺寸大小的影响,

本文采取 IoU 距离替代欧氏公式,以此对反光背心数据集中的样本进行聚类分析。使用 K-means++ 算法得到的聚类中心数量 NUM 与 AVG_IoU(平均交并比)的关系如图 11 所示。其中横坐标表示聚类中心数量 NUM,纵坐标表示 NUM 值对应的 AVG_IoU 值。

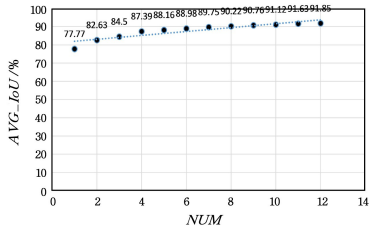


图 11 K-means++ 算法聚类分析结果

Fig. 11 K-means++ algorithm clustering analysis results

通过图 11 可以看出,AVG_IoU 随着聚类数量 NUM 的增加而逐渐上升,当 NUM 值大于 6 时,AVG_IoU 逐渐平缓。由于反光背心为一类,因此聚类结果相对较为稀疏,针对 4 种尺度的特征图,总共获得了 12 种尺寸不同的先验框,将聚类得到的 Anchor 由小到大依次排序,分别分配给从大到小排序的 4 种尺度的特征图,如表 1 所列。

表 1 反光背心数据集先验框尺寸

Table 1 Reflective vest data set prior box size

Size of feature map	Prior box size
13×13	(12×20), (13×20), (15×19)
26×26	(12×17), (13×16), (12×18)
52×52	(12×15), (14×13), (15×13)
104×104	(9×12), (13×9), (11×16)

3.3 改进损失函数

原始的 YOLOv3 使用均方误差 (Mean Square Error, MSE) 来计算目标回归损失,但是均方误差存在对目标尺度敏感的问题,回归的准确性会受到目标尺寸变化的影响。而 IoU 具有目标尺度上的鲁棒性,不仅能够用来确定正负样本,也能用来衡量边界框之间的距离,但使用 IoU 作为损失函数会存在以下几个问题:

(1) 如果标注边界框和预测边界框没有相交,根据定义, IoU 的值为 0,不能反映两者的距离大小(融合度),同时因为损失函数梯度为 0,没有梯度回传,无法进行学习训练。

(2) IoU 无法精确反映标注边界框和预测边界框的重合度大小。如图 12 所示,虽然 3 种情况的 IoU 值都相同,但能明显看出它们的重合度是不同的。其中,图 12(a)回归效果最佳,而图 12(c)回归效果最差。可见即使 IoU 相同,检测效果也可能有较大差异。因此本文提出使用 GIoU 替代 IoU 作为损失函数。

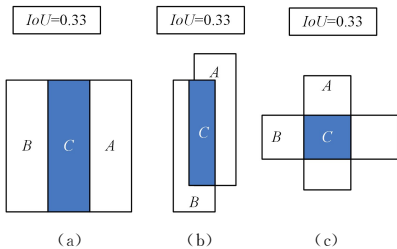


图 12 $IoU=0.33$

Fig. 12 $IoU=0.33$

与 IoU 相似,GIoU 也是一种距离度量。但与 IoU 只关注于重叠区域不同,GIoU 同时还关注其他非重合区域,能够更好地反映标注边界框和预测边界框的重合度。另外,GIoU 还有以下几个特点:

(1) GIoU 对目标尺度不敏感,回归的准确性不会受到目标尺寸变化。

(2) 对于 IoU 来说,GIoU 总是一个下限,在标注边界框和预测边界框无限重合的情况下, $IoU=GIoU=1$ 。

(3) IoU 取值区间为 $[0,1]$,而 GIoU 存在对称区间,其取值范围是 $[-1,1]$ 。GIoU 在标注边界框和预测边界框完全重合时取最大值 1,在两者无交集且无限远时取最小值 -1。GIoU 考虑到了 IoU 中未考虑到的非重叠区域,在标注边界框与预测边界框无交集时仍可以进行梯度回传,因此 GIoU 是一个比 IoU 更好的距离度量指标。

$$GIoU = IoU - \frac{|C-A \cup B|}{|C|} \quad (3)$$

GIoU 的计算式如式(3)所示,其中 A 表示预测边界框, B 表示标注边界框, C 是能涵盖它们的最小方框。

$$C = A \cup B, GIoU = IoU - 0 = 1 \quad (4)$$

$$GIoU = IoU - \frac{|C-0|}{|C|} = 0 - 1 = -1 \quad (5)$$

若 A 与 B 完全重合,则如式(4)所示, $GIoU = IoU = 1$;若 A 与 B 无交集,则如式(5)所示, $GIoU = -1$ 。使用 GIoU 算法对图 12 中 3 种 IoU 相等的情况进行计算,结果如图 13 所示。

如图 13 所示,GIoU 会关注 IOU 关注不到的标注边界框和预测边界框的非重合区域,因此学习训练效果更好,使用 GIoU 作为本文算法中的损失函数不仅可以增强对反光背心定位的精度,同时也能提升算法的准确率和定位精度。

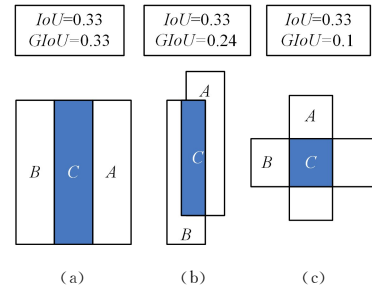


图 13 IoU 与 GIoU 的区别

Fig. 13 Differences between IoU and GIoU

4 实验结果与分析

4.1 实验环境

本文实验环境为: Intel Core i7-4790K CPU,内存 12 GB, NVIDIA GeForce GTX TITAN Black,操作系统为 Windows 10,PyTorch 版本为 1.2.0, torchvision 版本为 0.4.0, CUDA 版本为 11.1。

4.2 数据集

在反光背心检测领域,目前尚未有公开的图像数据集,本文基于所选机坪与机位所拍摄的图片,自行构建了反光背心数据集,其中共有 10 974 张图片,图片像素为 1920×1080 。

按照实验要求,仿照 COCO 公开数据集的标注格式对所采集的图片进行标注。其中,使用的图形标注工具为开源工具 LabelImg,生成对应的 XML 文件格式如图 14 所示,其中 $\langle width \rangle$ 和 $\langle height \rangle$ 分别表示输入图片的宽和高, $\langle name \rangle$ 代表标签的信息, $\langle bndbox \rangle$ 代表 Bounding Box 的具体位置信息。



图 14 数据集标记示例以及 XML 文件

Fig. 14 Example of dataset annotation and XML file

将实验数据集按照 7:2:1 的比例随机划分为训练集、测试集与验证集。在训练过程中,设置 epoch 为 60,批尺寸设置为 16,动量参数设置为 0.9,权重衰减正则项设置为 0.0005,以防止过拟合。初始学习率设置为 0.001,当网络迭代 400000 次后学习率将会衰减为原来的 1/10,经过 450000 次迭代后,学习率将会在上一个学习率的基础上再衰减 1/10。

4.3 评价标准

本文通过以下几个指标来评价反光背心检测算法的性能。

(1)查准率(Precision)与查全率(Recall)

$$Precision = \frac{TP}{TP + FP} \times 100\%$$
 (6)

$$Recall = \frac{TP}{TP + FN} \times 100\%$$
 (7)

其中, TP (True Positive), FP (False Positive), FN (False Negative) 分别代表预测框分类正确且 IoU 值大于设定阈值的预测框数量、分类错误或者分类预测正确但 IoU 值小于设定阈值的预测框数量和未被预测出来的真实框的数量。

(2)平均准确率均值(mean Average Precision, mAP)

$$mAP = \frac{\sum_{i=1}^N AP_i}{N} = \frac{\sum_{i=1}^N \int_0^1 P_i(R) dR}{N}$$
 (8)

其中, N 代表所有类别目标类型的数量和, AP 为平均准确率 (Average Precision, AP)。根据式 (6) — 式 (8) 可知, P , R ,

mAP 值越大表示该反光背心模型的定位效果和预测的准确率越高。

(3)检测速度

检测速度表示该目标检测网络每秒能检测多少数量(帧数)的图片,用 FPS(Frames Per Second)来表示。

4.4 实验结果对比及分析

4.4.1 YOLOv3 与本文算法的实验结果比较

如表 2 所列,本文算法在反光背心目标检测中取得了较好的性能。在反光背心目标检测中, YOLOv3 只获得了 85.7% 的 mAP 值,而本文算法的 mAP 值高达 96.1%,且本文算法的 FPS 也高于 YOLOv3,达到了机坪目标检测实际运用的实时性要求。YOLOv3 与本文算法的部分检测结果如图 15 所示。

表 2 YOLOv3 与本文算法比较

Table 2 YOLOv3 compared with our algorithm

Method	P/%	R/%	mAP/%	FPS
YOLOv3	85.7	85.7	85.7	21.3
Ours	97.6	96.1	96.1	28.4



(a)



(b)

图 15 YOLOv3 与本文算法比较

Fig. 15 YOLOv3 compared with our algorithm

从图 15(a)可以看出,在目标所占区域较大的情况下,对于反光背心的检测识别,本文算法与 YOLOv3 均取得了较好的检测效果,虽然置信度有所不同,但两者都可以正确地识别出机坪工作人员反光背心的穿戴情况。而在图 15(b)中,反光背心目标所占区域较小时,本文算法明显比原始的 YOLOv3 算法取得了更好的检测效果。因此无论是从识别准确率还是置信度上,本文算法相比 YOLOv3 都有显著的性能提升。

4.4.2 本文与其他算法的比较

为进一步对本文算法的有效性进行验证,在 NVIDIA GeForce GTX TITAN Black 的实验环境中,将该算法与其他算法在反光背心数据集下进行比较,并使用评价指标进行定量评估,以此来验证本文算法的有效性,对比实验结果如表 3 所列。

表 3 本文算法与其他算法的性能对比

Table 3 Performance of our algorithm compared with other algorithms

Method	P/%	R/%	mAP/%	FPS
Faster-RCNN	94.8	94.8	94.8	6.1
YOLOv3	85.7	85.7	85.7	21.3
YOLOv3-tiny	77	77.6	77	37.0
YOLOv3-spp	86.9	87.7	87.2	21.1
YOLOv3-spp3	87.2	88.1	87.8	20.9
YOLOv4-tiny	83	83.6	83.3	40.9
Ours	97.6	96.1	96.1	28.4

如表 3 所列,使用相同的反光背心测试集在以上网络中进行测试,结果表明,本文算法不仅在精确率、召回率以及平均精度上均高于 YOLOv3-spp,yolov3-spp3 等方法,而且在检测速度上也具有优势。而相比 YOLOv3-tiny,YOLOv4-tiny 等轻量级方法,本文算法虽在速度上略慢,但在精确率、召回率以及平均精度上有较大提升,并且本文算法的检测效果比 Faster-RCNN 方法的检测效果更好。Faster-RCNN 方法的平均精度为 94.8%,远高于其他 YOLO 算法,但由于 Faster-RCNN 的机制是能对同一目标进行反复检测,因此准确性本身就很高^[23],且本文算法在检测速度上有十分明显的优势,能够满足机坪目标检测在实际运用中的实时性要求。

4.4.3 本文算法在不同实验环境下的检测速度对比

为进一步探究本文算法在满足机坪目标检测实际运用中的实时性要求时(检测速度达到 25 帧/s)所需机器性能,将不同实验环境下使用本文算法最终生成的模型,在相同的反光背心数据集上进行了检测,对比结果如表 4 所列。由表 4 可知,由于实验环境中 Cuda 核心与显存等存在差异,导致不同环境下对反光背心数据集的检测速度存在明显差异。在实验环境所使用的 GPU 为 GeForce GTX TITAN Black 时,Cuda 核心数为 2880 个,显存为 6188 MiB,对反光背心数据集的检测速度能达到 28.4 帧/s,满足了机坪目标检测在实际运用中的实时性(25 帧/s)要求。根据表 4,Cuda 核心与 FPS 的对应关系如图 16 所示。根据图 17 的拟合估计,在 GPU 性能为 GeForce GTX 980 Ti(2816 个 Cuda 核心,显存 6188 MiB)及以下的实验环境中,该算法能达到 25 帧/s 以上的检测速度。

表 5 消融实验的结果对比

Table 5 Comparison of ablation test results

Group	Darknet-37	SPP	Fourscale fusion	Optimize sighting framesize	GIoU	Candidate region generation	P/%	R/%	mAP/%	FPS
Experiment 1	×	×	×	×	×	×	85.7	85.7	85.7	21.3
Experiment 2	✓	×	×	×	×	×	85.5	85.4	85.5	28
Experiment 3	✓	✓	×	×	×	×	87.1	88.8	87.3	27.8
Experiment 4	✓	✓	✓	×	×	×	90.2	92.9	90.9	27.5
Experiment 5	✓	✓	✓	✓	×	×	93.6	93.3	92.9	27.6
Experiment 6	✓	✓	✓	✓	✓	×	97.6	95.2	95.0	27.8
Experiment 7	✓	✓	✓	✓	✓	✓	97.6	96.1	96.1	28.4

从表 5 可以看出:对于第 1 组实验,原始的 YOLOv3 模型在反光背心检测上的 P 值和 R 值分别为 85.7%和85.7%,其 mAP 值为 85.7%,FPS 为 21.3;对于第 2 组实验,由于将 Darknet-37 替换了 Darknet-53 作为特征提取网络,且将卷积层的滤波器数量降低,因此检测速度增加到 28,同时也损失了一些精度;对于第 3 组实验,由于在第 2 组实验的基础上

表 4 本文算法在不同实验环境下的检测效果对比

Table 4 Detection results of our algorithm compared in different experimental environments

Experimental environment	mAP/%	Cuda core	Memory size/ MiB	FPS
GeForce 930M	96.1	384	4 096	5.3
Quadro K620	96.1	384	2 048	9.2
GeForce GTX 960M	96.1	640	4 096	17.1
GeForce GTX 950M	96.1	640	4 096	17.9
Quadro M2000	96.1	768	4 096	23.3
GeForce GTX TITAN Black	96.1	2 880	6 188	28.4
Tesla V100	96.1	5 120	16 160	131

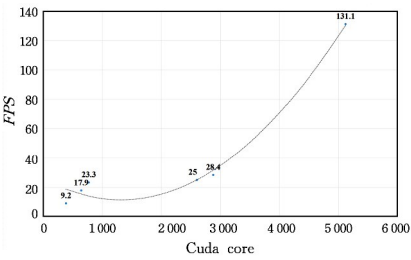


图 16 Cuda 核心与 FPS 的关系图

Fig. 16 Diagram of Cuda core and FPS

4.4.4 消融实验结果及分析

在深度学习领域,消融实验是一种常用的实验方法,主要用来分析不同的网络分支对整个实验产生的影响^[23]。为进一步分析本文改进算法对原始 YOLOv3 模型的影响,将本文算法分别裁剪成 7 组进行训练。其中,第 1 组为原始 YOLOv3 模型,第 2 组将 Darknet-37 替换 Darknet-53 作为特征提取网络,第 3 组在第 2 组的基础上加入了金字塔池化结构,第 4 组在第 3 组的基础上将特征融合尺度提升为 4 尺度,第 5 组在第 4 组的基础上使用 K-mean++ 算法对锚框值进行重新聚类,第 6 组在第 5 组的基础上使用 GIoU 损失函数替代 IoU 损失函数,第 7 组在第 6 组的基础上基于机坪工作人员与飞机停靠位置具有相关性的先验知识,提前生成候选区域替换初始候选区域,本文算法即第 7 组。7 组消融实验结果如表 5 所列,其中,“✓”表示实验中包括该结构,“×”表示实验中未包括该结构。

使用GIoU损失函数替换IoU损失函数,再次提高了FPS值与mAP值等;第7组实验为本文算法,在第6组实验的基础上基于先验知识先进行候选区域生成,再次提高了mAP值与FPS值。综上,本文针对YOLOv3的改进策略能够大幅度提升机坪工作人员反光背心的检测效果,将mAP值提升到了96.1%,且将FPS提升到了28.4帧/s。

结束语 本文基于YOLOv3算法与机坪中的先验知识,将针对反光背心数据集改进的YOLOv3目标检测算法与机坪工作人员与飞机停靠位置具有相关性的先验知识结合起来,提出了一种针对反光背心检测的新型目标检测算法。实验结果表明,本文所提算法在反光背心数据集上的检测效果比原始YOLOv3算法更精确,同时检测速度达到了28.4f/s,满足了机坪目标检测实际运用中的实时性要求。但是受限于机坪视频的保密及使用要求,本文仅针对少量的机位进行了实验,一些具体的参数还需要在实际应用中根据具体机位的情况进行现场调整;其次,本文加入的先验知识存在多样性不足的问题,对YOLOv3算法的改进还有上升的空间,今后会对本文算法进行不断改进,加入更多机坪中的先验知识,并对YOLOv3算法进行不断改进,以达到更高的检测速度与精度。

参 考 文 献

[1] ZHAO Y Q,RAO Y,DONG S P,et al. Survey on deep learning object detection [J]. Journal of Image and Graphics, 2020, 25(4):629-654.

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

[3] REN S,HE K,GRISHICK R,et al. Faster R-CNN:Towards real-time object detection with region proposal networks[C]//Advances in Neural Information Processing Systems. 2015:91-99.

[4] GRICHICK 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.

[5] HE K M,GKIOXARI G,DOLLAR P,et al. Mask R-CNN[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence,2020,42(2):386-397.

[6] CHANU M M. A deep learning approach for object detection and instance segmentation using mask RCNN[J]. Journal of Advanced Research in Dynamical and Control Systems, 2020, 12(SP3):95-104.

[7] REDMON J,DIVVALA S,GIRSHICK R,et al. You only lookonce:Unified,real-timeobject detection[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2016:779-788.

[8] LIU W,ANGUELOV D,ERHAN D,et al. SSD:single shot multibox detector[C]//Proceedings of the 2016 European Conference on Computer Vision. Cham:Springer,2016:21-37.

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

[10] REDMON J,FARHADI A. YOLOv3:an incremental improvement[J]. arXiv:1804. 02767,2018.

[11] XU Z F,JIA R S,SUN H M,et al. Light-YOLOv3:fast method for detecting green mangoes in complex scenes using picking ro-

bots[J]. Applied Intelligence,2020(6):1-18.

[12] PANG L,LIU H,CHEN Y,et al. Real-time Concealed Object Detection from Passive Millimeter Wave Images Based on the YOLOv3 Algorithm [J]. Sensors (Basel, Switzerland), 2020, 20(6):44-50.

[13] ZHAO L,LI S. Object Detection Algorithm Based on Improved YOLOv3[J]. Electronics,2020,9(3):537.

[14] DENG Z,YANG R,LAN R,et al. SE-IYOLOv3: An Accurate Small Scale Face Detector for Outdoor Security[J]. Mathematics,2020,8(1):93.

[15] DENG H,CHENG J,LIU T,et al. Research on iron surface crack detection algorithm based on improve YOLOv4 network [C]// Journal of Physics: Conference Series, 2nd International Conference on Artificial Intelligence and Computer Science, Hangzhou, Zhejiang, China, 2020:25-26.

[16] GU Y H,CAO M T,XIU J Y,et al. Violation Detection algorithm based on YOLOv4 Network [J]. Journal of Chongqing University of Technology (Natural Science),2021,35(8):114-121.

[17] HOWARD A,SANDLER M,CHU G,et al. Searching for mobilenetv3[C]// Proceedings of the IEEE International Conference on Computer Vision. 2019:1314-1324.

[18] PAN M Y,SONG H H,ZHANG K H,et al. Learning Global Guided Progressive Feature Aggregation Lightweight Network for Salient Object Detection[J]. Computer Science,2021,48(6): 103-109.

[19] 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.

[20] BACHEM O,LUCIC M,HASSANI S H,et al. Approximate k-means++ in sublinear time[C]// Proceedings of the Thirtieth AAAI Conference on Artificial Intelligence. 2016:1459-1467.

[21] BACHEM O,LUCIC M,HASSANI H,et al. Fast and provably good seedings for k-means[C]// 30th Conference on Neural Information Processing Systems,2016:55-63.

[22] 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.

[23] 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.



XU Tao, born in 1962, professor, is a member of China Computer Federation. His main research interests include intelligent information processing and so on.



LYU Zong-lei, born in 1981, Ph.D, associate professor, is a member of China Computer Federation. His main research interests include machine learning and so on.