基于改进 YOLOv5 的护目镜佩戴检测算法

聂壮壮,汪　军,黄翔翔
安徽工程大学 计算机与信息学院,安徽 芜湖 241000

摘　要:目的 解决目前危化实验室、工厂等危险环境下护目镜佩戴情况检测存在的人工检查效率低下、无法有效保障人员眼部安全等问题。**方法** 首先构建出护目镜佩戴检测数据集,其中包含 4 个真实场景图片与部分网络爬取数据集,并通过数据增强等手段将原始的 3 383 张扩充至 5 462 张图片,构成最终数据集,使各个样本数量达到均衡,有效预防了因样本不均衡导致的模型精度低的问题;接着提出改进型 YOLOv5 目标检测算法来实现对护目镜佩戴情况的自动检测,算法在 YOLOv5 中添加 SPD 小目标检测模块,该模块完全消除了传统卷积模块中导致信息丢失的步长卷积和池化操作,使网络保留更多信息,引入坐标注意力机制解决了因添加 SPD 带来的相邻位置关系无法有效提取的问题;同时,将原本的损失函数替换为 SIoU 损失函数,有效解决了真实框与目标框相互包含情况下的 IoU 计算问题,减少了计算自由度,降低了模型计算量,提升了模型准确率。**结果** 在护目镜配戴检测数据集上的实验结果表明:改进型的 YOLOv5 模型在护目镜佩戴检测数据集上的平均精度为 72.7%,相较于原始 YOLOv5 模型平均精度提高了 5.6%。**结论** 该模型实现了对复杂环境下护目镜佩戴情况的基本检测。

关键词:护目镜佩戴检测;数据增强;目标检测;SPD;坐标注意力机制
中图分类号:TP391.41　　**文献标识码:**A　　doi:10.16055/j.issn.1672-058X.2025.0001.007

Goggle Wearing Detection Algorithm Based on Improved YOLOv5

NIE Zhuangzhuang, WANG Jun, HUANG Xiangxiang
School of Computer and Information, Anhui Polytechnic University, Anhui Wuhu 241000, China

Abstract: Objective In response to the problems of low efficiency in manual inspection and the inability to effectively ensure the eye safety of personnel in hazardous environments such as chemical laboratories and factories, this study aims to address these issues in the detection of goggles wearing. **Methods** Firstly, a dataset for goggles wearing detection was constructed, including four real-scene images and a portion of data obtained by web crawling. By means of data augmentation, the original dataset of 3383 images was expanded to 5462 images to form the final dataset, ensuring the balance of sample quantities and effectively preventing the problem of low model accuracy caused by sample imbalance. Then, an improved YOLOv5 object detection algorithm was proposed to automatically detect the wearing status of goggles. In the YOLOv5 algorithm, an SPD small target detection module was added to completely eliminate the stride convolution and pooling operations that lead to information loss in traditional convolution modules, allowing the network to retain more information. Subsequently, a coordinate attention mechanism was introduced to address the problem of ineffective extraction of neighboring position relationships caused by the addition of SPD. Moreover, the original loss function was replaced with the SIoU loss function to effectively solve the IoU calculation problem when the real box and the target box contain each other, reducing the degrees of freedom in calculations, decreasing the model' s computational complexity,

收稿日期:2023-08-31　**修回日期:**2023-10-15　**文章编号:**1672-058X(2025)01-0048-09
基金项目:安徽省自然科学基金(2208085QD106).
作者简介:聂壮壮(1998—),男,安徽阜阳人,硕士研究生,从事计算机视觉与智能感知研究.
通讯作者:汪军(1975—),男,安徽芜湖人,教授,硕士研究生导师,从事智能感知与控制研究. Email:wangjun@ahpu.edu.cn.
引用格式:聂壮壮,汪军,黄翔翔. 基于改进 YOLOv5 的护目镜佩戴检测算法[J]. 重庆工商大学学报(自然科学版),2025,42(1):48-56.
NIE Zhuangzhuang, WANG Jun, HUANG Xiangxiang. Goggle wearing detection algorithm based on improved YOLOv5[J]. Journal of Chongqing Technology and Business University (Natural Science Edition), 2025, 42(1): 48-56.

and improving model accuracy. **Results** Experimental results on the goggles wearing detection dataset show that the improved YOLOv5 model has an average precision of 72.7% on the goggle wearing detection dataset, which is 5.6% higher than the average precision of the original YOLOv5 model on the same dataset. **Conclusion** This model realizes the basic detection of goggles wearing status in complex environments.

Keywords: goggles wearing detection; data augmentation; object detection; SPD; coordinate attention mechanism

1 引　言

南方医科大学统计[1]显示,在 1 851 例眼外伤病例中,一线生产人员 856 例,占比高达 46.25%;锐器及金属碎片 569 例,占比 30.74%;金属熔液 481 例,占比 25.99%;碱化学伤 328 例,占比 17.22%。正确佩戴护目镜可有效预防眼部意外伤害。现阶段实验室等危险场景的护目镜佩戴情况检测以人工检查为主,相关科研院校安全监管人员无法做到对护目镜佩戴情况的有效检查,同时人工监管无法保证实时性,实际佩戴往往依靠实验人员自身自觉,无法切实有效保证人员眼部安全。因此,利用计算机视觉和深度学习技术代替传统人工护目镜佩戴情况检查,可以实现对实验室等场景下护目镜佩戴的高效检测,从而更有效地保证相关人员眼部安全。

目标检测技术在计算机视觉领域中具有重要地位,自深度学习技术发展以来,该领域取得了巨大的进步。目标检测技术分为基于区域的方法(Region-based methods)和基于回归的方法(Regression-based methods)。基于区域的方法首先生成候选区域,然后将这些区域输入到深度神经网络中进行分类,最典型的代表是 R-CNN 系列(R-CNN[2]、Fast R-CNN[3]、Faster R-CNN[4]。基于回归的方法直接预测目标的边界框和类别,无需生成候选区域,SSD(Single Shot MultiBox Detector)[5]和 YOLO(You Only Look Once)[6-7]是这类方法的代表。

在护目镜佩戴检测方面,张[8]首先提出一个检测方法:该算法利用图像合成技术,将护目镜通过特征匹配等手段合成到人脸上,接着再将合成后的图片作为最终的数据集进行模型训练。模型训练采用的 ResNet-50 作为主干网络,通过在不同阶段的 CONV_X 层之间增加池化 MAXPOOL 层和非线性激励 REUL 层,来避免过拟合现象和减少计算量;接着针对护目镜佩戴检测调整特征向量维度;最后添加 DROPOUT 机制,即在每一轮训练过程中,通过随机选择网络内一半的特征检测器使其停止工作,可以提高模型的泛化能力。作者指出,通过使每轮训练中参与的子网络结构略有不同,但层数相同,即可阻止特征之间的协同作用,从而减缓过拟合现象。

本文针对护目镜配戴检测出现的数据集不充分不合理做出相应调整,采用真实场景收集等手段,有效防止因数据集不丰富所造成的检测模型过拟合现象。接着通过优化损失函数,在主干网络中引入 Space-to-depth(SPD)模块以提升小目标检测精度,同时为了解决因引入 SDP-Conv 所带来的位置信息丢失问题,在主干网络中增加坐标注意力机制,使相邻位置的信息有效提取,能够实现对护目镜佩戴情况的有效检测。

2 YOLOv5 网络结构

YOLOv5 算法是一款基于 Pytorch 框架的目标检测算法,根据模型参数量大小分为 s、m、l、x 等 4 个版本,4 个版本的网络宽度和深度以及网络参数量依次增加。本文将采用 YOLOv5l 检测模型。

YOLOv5 目前已迭代至 6.0 版本,较之前版本结构更加便捷,并采用全新的空间金字塔池化(Spatial Pyramid Pooling-Fast,SPPF)替代传统的空间金字塔池化(Spatial Pyramid Pooling,SPP)层,以减少模型计算量并提升速度。YOLOv5 是一个包含主干网络、颈部、头部、输入和输出端的网络模型。它的架构如图 1 所示。

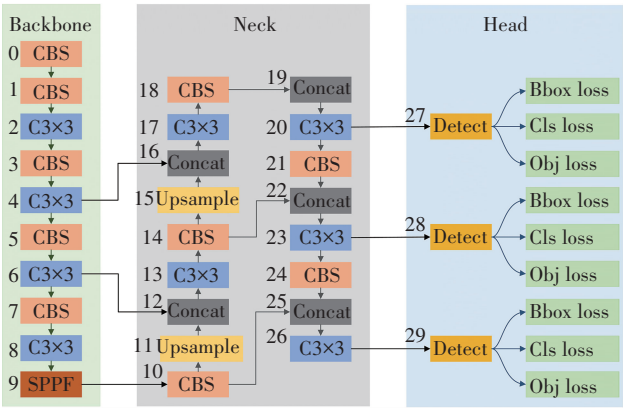


图 1　YOLOv5 网络结构图

Fig. 1　Structure diagram of YOLOv5 network

2.1 输入端

YOLOv5 的输入端包括马赛克数据增强、自适应计算锚框和自适应缩放图像。马赛克数据增强通过随机缩放、随机裁剪和随机排布的方式将 4 张图片拼接在一起,从而增强了模型的检测能力[19]。

在模型训练过程中,YOLOv5 可以启动自适应锚框计算功能,自动计算每个类别训练集中的最佳锚框大小,然后将最佳锚框作为输出的预测框,并与真实框进行对比,计算两者的匹配程度,从而为更新网络参数提

供依据。

实验使用 K 均值算法(K-Means)对护目镜佩戴检测数据集中的目标边界框进行聚类,以产生适应实际目标框的不同数量和大小的先验框。通过多次的聚类实验,得到了 9 个具体的先验锚框尺寸。实验结果显示,在选取聚类数目为 9 时,先验框个数与平均交并比(MIoU)的增长率逐渐下降,并最终达到最优,具体情况如图 2 所示。因此,本研究建议使用这 9 个先验锚框尺寸来提升模型的收敛效率。具体先验框尺寸见表 1。

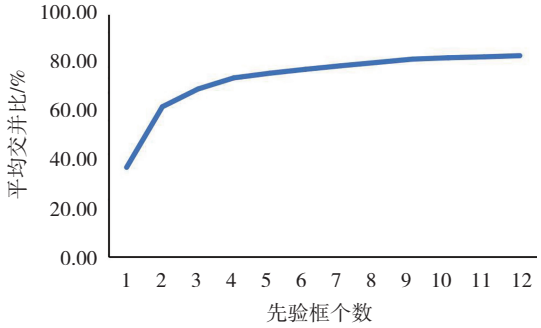


图 2 先验框聚类平均交并比图

Fig. 2 Average intersection and union ratio of prior frame clustering

表 1 先验框尺寸

Table 1 Dimensions of the prior frame

特征图尺度	锚框大小		
小	(19,34)	(26,35)	(31,56)
中	(40,82)	(45,56)	(45,56)
大	(55,114)	(80,162)	(156,198)

2.2 主干网络

YOLOv5 网络模型的 Backbone 主要由 Focus、C3(改进后的 BottleneckCSP)和 SPPF(Spatial Pyramid Pooling-Fast)模块组成。其中,对图片切片操作通过 Focus 模块完成,将图片的宽度和高度信息整合到通道维度,同时缩小高度和宽度为原来的 1/2,通道数增加 4 倍。这个过程可以通过将距离为 2 的 4 个位置堆叠到一个通道上来实现。Focus 模块的设计旨在降低 FLOPS 并提高检测速度。Focus 结构示意图如图 3 所示。

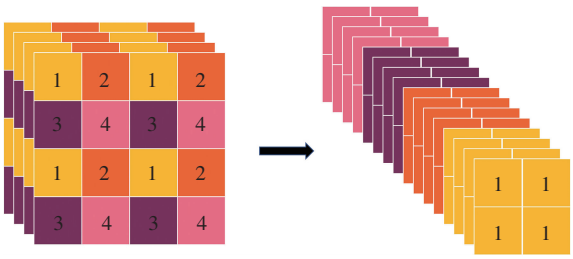


图 3 Focus 结构示意图

Fig. 3 Schematic diagram of Focus structure

2.3 Neck

YOLOv5 的网络架构巧妙地融合了特征金字塔网络(Feature Pyramid Networks,FPN)和路径聚合网络(Path Aggregation Network,PAN),这两种网络结构的融合优化了算法对各种尺度目标的检测能力[10],从而提高了整体的检测性能。

FPN 是一种多尺度特征融合方法,它通过融合网络各层特征来实现对小目标物体的有效检测[11]。这是因为在传统卷积网络中,底层特征图包含丰富的细节信息,而高层特征图包含更多的语义信息。FPN 通过自上而下,自下而上的多尺度融合,使得每一层特征图都包含了丰富的细节和语义信息,从而提高了小目标的检测效果。

PAN 则在 FPN 的基础上,增加了自下而上的路径,使得底层的定位信息能更有效地传递到高层,从而进一步提高了对密集目标的检测性能。这是因为在目标检测中,底层特征包含的定位信息对于高精度的目标检测是非常重要的,而传统的 FPN 结构在这一点上存在一定的不足。

通过结合 FPN 和 PAN,YOLOv5 实现了不同主干层对不同检测层进行特征融合。这样的设计使得 YOLOv5 在处理多尺度、多密度的目标检测任务时,具有更高的效率和准确性。

总的来说,YOLOv5 通过 FPN 和 PAN 的融合,实现了对于多尺度、多密度目标的高效准确检测。这种设计使得 YOLOv5 在目标检测任务中,尤其是对于小目标和密集目标的检测,具有显著的优势。

2.4 输出端

YOLOv5 的输出端采用的是 GIoU,在改进的算法中使用 SIoU 损失函数(SIoU loss)替换原始算法的 GIoU 损失函数(GIoU loss),在目标检测后处理过程中,采用非极大抑制(Non-Maximum Suppression,NMS)来对生成的所有目标框进行筛选得到最终的检测结果,以获得更高的目标检测准确率。

3 数据集

在深度学习研究领域,科学合理的数据集是课题研究的基础。针对护目镜佩戴的场所特点,本文构建一个护目镜佩戴检测数据集,包含金属件加工的车工车间、铣工车间,以及化学实验室等一些实际场景。构建数据集主要分为以下 3 个过程:数据收集、数据筛选、数据增强。

3.1 数据收集

本文数据集所需图片来自于实际场景的现场监控视频抽帧,以及少量网络爬取图片。构建的数据集包含 5 个场景,分别为实验室 1、实验室 2、化学实验室、工厂、网络爬虫。图 4 为场景展示。

(a) 实验室 1　　(b) 实验室 2

(c) 化学实验室　　(d) 工厂

图 4　数据集场景图

Fig. 4　Dataset scenario

3.2　数据清洗

从监控抽帧获得的图片和网络爬取的图片,有相当数量是重复图片和无关键人员目标物的图片,这些图片对本文研究无任何意义,因此需要删除。

3.3　数据增强

本文得到的原始数据集共包含 3 383 张图片,根据护目镜的佩戴规定和场景特点,将护目镜佩戴检测目标分为 5 个类别:正确佩戴(goggles)、侧脸(side_head)、未佩戴(no_goggles)、错误佩戴(error_goggles)、背脸(back_head),具体分类如图 5 所示。

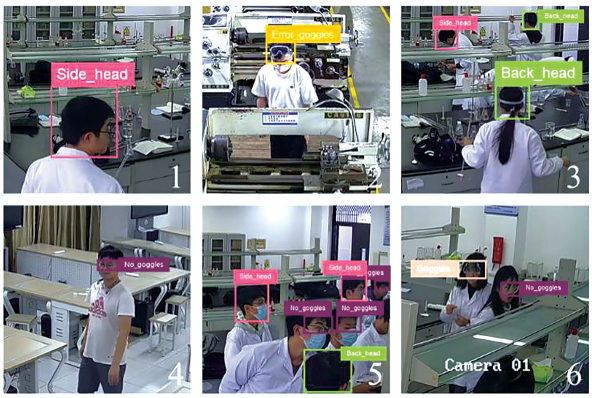


图 5　目标分类图

Fig. 5　Target classification diagram

通过分析原始数据集可以发现错误佩戴这一分类的目标过少,样本不均衡情况严重,可能导致该类别在检测中的准确率偏低。为了避免这种情况的发生,本文采取包含 AutoContrast(调整图像对比度)、Posterize_Color(减少每个颜色通道的 bits)、Brightness(调整图像亮度)、Sharpness(调整图像清晰度)、Solarize_add(调整图像色彩)在内的多种图像增强手段,最终使各个类别的目标数目均满足训练的需要。其中每个类别的目标数据增强前后的数目如图 6 所示。

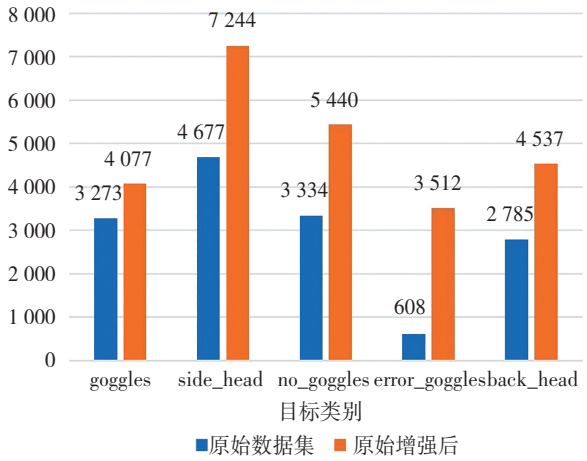


图 6　数据增强前后对比图

Fig. 6　Comparison before and after data enhancement

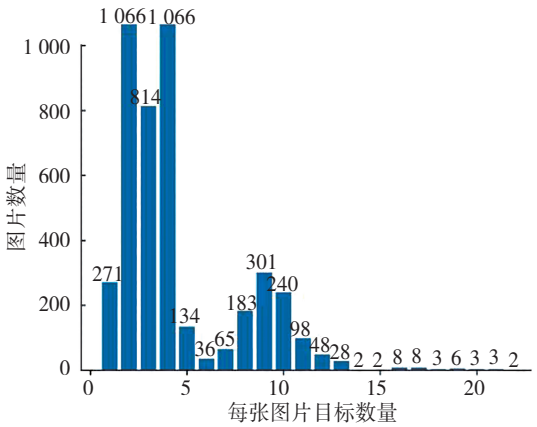
表 2 显示了数据集的划分情况。其中,实验室 1、实验室 2、工厂和网络爬虫这 4 个场景按照 9:1 的比例分别划分为训练集和验证集。剩下的工厂场景的所有图片都被用作测试集。训练集、验证集和测试集使用不同的场景,这样可以通过测试集的结果来评估模型的泛化能力。

表 2　数据集划分情况表

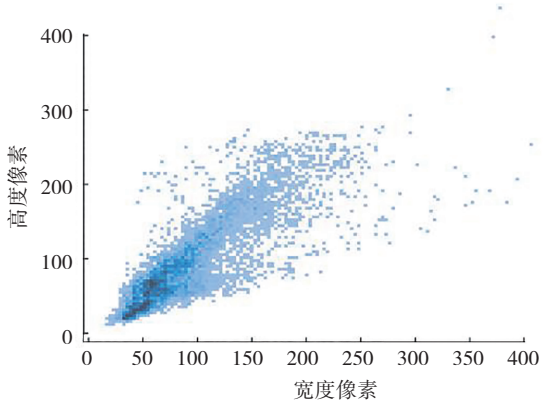
Table 2　Dataset breakdown table

场　景	各场景图片总数	训练集	验证集	测试集
实验室 1	760	684	76	0
实验室 2	680	612	68	0
化学实验室	3 200	2 880	320	0
工厂	588	0	0	588
网络爬虫	234	211	23	0
合计	5 462	4 387	487	588

图 7 为最终数据集的分析可视化结果。图 7(a)为每个目标数量图片的数量统计,图 7(b)为锚框大小分布,可直接反映出目标大小分布情况。



(a) 目标数量与图片数量的统计



(b) 锚框大小分布

图 7 数据集分析可视化结果图

Fig. 7 Visualization results of dataset analysis

4 模型的改进与优化

4.1 SPD 小目标检测模块

传统的卷积神经网络在目标检测等任务中已经被大规模采用,检测效果也逐步提升,足以应对大对数场景。然而,在面临本文所研究的护目镜佩戴检测之类的较低分辨率的复杂任务时,传统卷积的效果迅速降低。

传统卷积的步长卷积和池化操作直接导致了细粒度信息的丢失和较低效率的特征表示学习。为此,Raja Sunkara 等[12]提出了新的卷积模块 SPD-Conv,与传统的卷积模块相比,SPD-Conv 用 SPD 层和非跨步卷积层组成,从而完全消除了传统卷积模块中导致信息丢失的步长卷积和池化操作(图 8)。

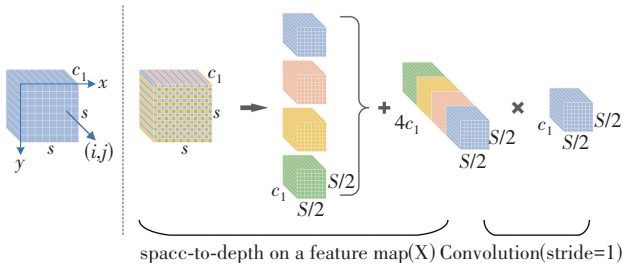


图 8 SPD 结构示意图

Fig. 8 SPD structure diagram

SPD 模块应用图像转化技术,对神经网络中的特征图进行下采样。对于任意大小为 $S\times S\times C_1$ 的中间特征图 X,可以切出一系列子特征图,如式(1)所示:

$$
\begin{aligned}
&f_{0,0}=X[0:S:scale,0:S:scale],\\
&f_{1,0}=X[1:S:scale,0:S:scale],\cdots,\\
&f_{scale-1,0}=X[scale-1:S:scale,0:S:scale]\\
&f_{0,1}=X[0:S:scale,1:S:sacle],\\
&f_{1,1}=X[1:S:scale,1:S:scale],\cdots,\\
&f_{scale-1,1}=X[scale-1:S:scale,1:S:scale]\\
&\vdots\\
&f_{0,scale-1}=X[0:S:scale,scale-1:S:scale],\\
&f_{1,scale-1}=X[1:S:scale,scale-1:S:scale],\cdots,\\
&f_{scale-1,scale-1}=X[scale-1:S:scale-1,scale-1:S:scale-1]
\end{aligned}
\tag{1}
$$

通常情况下,对于任意给定的原始特征图 X,可以用 $i+x$ 和 $i+y$ 的比例因子切出子图 $f_{x,y}$,因此每个子图都按比例因子进行下采样。例如,当 $scale=2$,可以得到 4 个形状为 $\left(\frac{S}{2},\frac{S}{2},C_1\right)$ 的子图 $f_{0,0}$、$f_{1,0}$、$f_{0,1}$、$f_{1,1}$,并将 X 下采样 2 倍。

随后,将 $f_{0,0}$、$f_{1,0}$、$f_{0,1}$、$f_{1,1}$ 沿着通道维度连接起来得到征图 X',其空间维度减少了一个比例因子,而通道维度增加了比例因子 2。简单来说,SPD 模块将特征图 $X(S,S,C_1)$ 转换为中间特征图 $X'\left(\frac{S}{scale},\frac{S}{scale},scale^2C_1\right)$。在进行 SPD 特征变换后,添加一层步长为 1,且具有 C_2 个过滤器的卷积层,其中 $C_2<cale^2C_1$,并使 $X'\left(\frac{S}{scale},\frac{S}{scale},scale^2C_1\right)$ 转化为 $X''\left(\frac{S}{scale},\frac{S}{scale},C_2\right)$。

4.2 Coordatt 坐标注意力机制

SPD-Conv 卷积虽可消除传统卷积模块中步长卷积、池化操作引起的信息丢失,但可能无法有效提取相邻位置关系。为克服此缺点,网络中引入了坐标注意力机制(Coordinate Attention, CA)[13],以捕获远程依赖关系并保留位置信息。

坐标注意力机制将通道注意力拆分为两个正交空间方向上的一维特征编码,然后沿两个空间方向进行特征融合。这样可以获得包含两个方向和位置信息的注意力图,从而增强对感兴趣目标的信息提取。

相比仅考虑注意力信息的机制,坐标注意力机制将位置信息嵌入其中,因此可以更好地捕捉局部区域之间的关系。具体而言,它分别在水平和垂直方向上聚合特征,以有效保留空间坐标信息在注意力图中。坐标信息嵌入时,每个通道对水平和垂直坐标进行编码,使用尺寸分别为 $(H,1)$ 和 $(1,W)$ 的池化核。h 和 w 分别表示当前注意力模块输入特征图的高度和宽度[14],它们的值会随着向下采样而变化。坐标注意力机制通过融合坐标信息,可以更好地建模位置关系,增强网络的特征表达能力。以第 c 个通道的输出为例,其高度为 h 的结果如式(2)所示;同样地,以第 c 个通道的输出为例,其宽度为 ω 的值如式(3)所示:

$$z_c^h(h)=\frac{1}{W}\sum_{0\leqslant i\leqslant W}x_c(h,i) \tag{2}$$

$$z_c^\omega(\omega)=\frac{1}{H}\sum_{0\leqslant j\leqslant H}x_c(j,\omega) \tag{3}$$

式(2)、式(3)中,h 和 ω 代表输入特征图的高度和宽度,H 和 W 为卷积核的大小,$z_c^h(h)$ 表示高度为 h 的第 c 个通道的输出,$z_c^\omega(\omega)$ 表示宽度为 ω 的第 c 个通道的输出。

为了聚合水平和垂直输入特征,采用两个一维全局池操作,生成两个独立方向的感知特征图和注意力图。每个注意力图包含远程依赖关系和位置信息。

在注意力生成阶段,将两幅特征图像拼接,并采用卷积操作和激活函数处理,然后分解成水平和垂直注意张量。再次使用卷积和激活函数处理后,将两个注意力图像与输入特征图像相乘得到最终结果。坐标注意力机制结构图如图 9 所示。

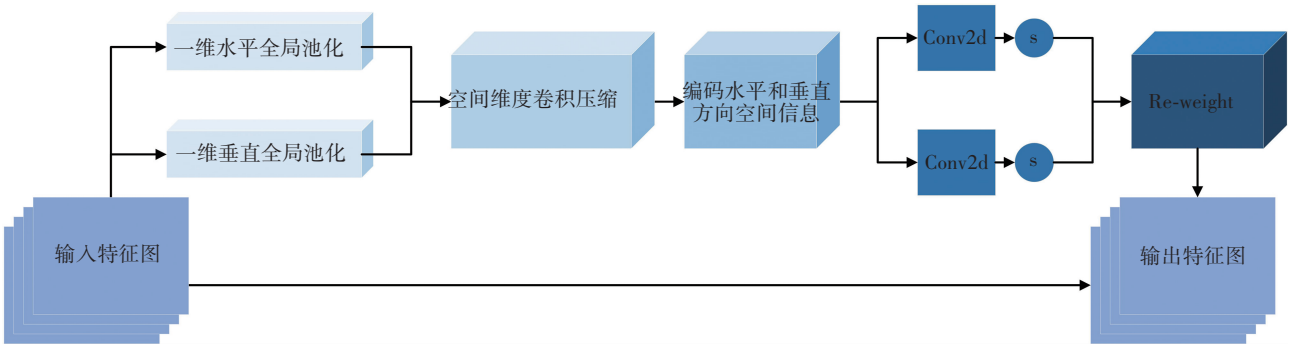


图 9　坐标注意力机制结构图

Fig. 9　Coordinate attention structure diagram

4.3　优化损失函数

YOLOv5 模型的损失函数由 3 个部分组成,包括分类损失函数、定位损失函数和目标置信度损失函数。其中,Bounding box 的损失采用 GIoU Loss[15]。GIoU Loss 计算如式(4)所示:

$$L_{\mathrm{GIoU}}=1-L_{\mathrm{IoU}}+\frac{|C-B\cup B^{gt}|}{|C|} \tag{4}$$

式(4)中,定义了预测框 $B=\{x,y,w,h\}$ 和真实框 $B^{gt}=\{x^{gt},y^{gt},w^{gt},h^{gt}\}$,同时引入变量 C 来表示能够囊括 B 和 B^{gt} 的最小区域面积。需要注意的是,当预测框与真实框之间的偏离变大时,C 的值也会相应增大。此外也可以通过将 C 减去预测框与真实框的面积来得到一个具体的数值,而这个数值将会趋近于 1。虽然 GIoU 有效解决了真实框与预测框不相交时 IoU 无法衡量的问题,但无法有效解决真实框与目标框相互包含的 IoU 计算问题。

因此,为了改进边界框回归的定位损失函数,本文引入了 SIoU Loss。SIoU 损失函数由 4 个部分构成,分别是 Angle cost、Distance cost、Shape cost 和 IoU cos。,SIoU Loss 回归损失函数如式(5)所示:

$$\begin{aligned}
&L_{\mathrm{SIoU}}=1-L_{\mathrm{IoU}}+\frac{\Delta+\Omega}{2}\\
&\wedge=1-2\sin^{2}\left(\arcsin(x)-\frac{\pi}{4}\right)\\
&\Delta=\sum_{t=x,y}(1-\mathrm{e}^{-(2-\wedge)\rho_t})\\
&\Omega=\sum_{t=w,h}(1-\mathrm{e}^{-\omega_t})^{\theta}
\end{aligned} \tag{5}$$

式(5)中,$\wedge$ 为 Angle cost、Δ 为考虑 Angle cost 条件下的 Distance cost,Ω 为 Shape cost。SIoU Loss 考虑了回归之间的角度惩罚,重新定义了角度惩罚度量,使预测框能够快速漂移到最近的轴,从而减少了自由度数目,提高了推理的速度和准确性,相比于 GIoU Loss 更加有效。

4.4　优化后的网络结构

YOLOv5 目标检测算法是目前较为稳定且被大量工程实际应用的优秀算法,但针对护目镜佩戴检测这一具体问题,还是需要进一步优化改进。根据前几章理论分析研究,结合护目镜佩戴检测的特点,可以得到引入 SPD-Conv 模块与优化损失函数的 YOLOv5 算法 YOLOv5l-SC(YOLOv5l-SPD-CA),其模型结构如表 3 所示。

表 3　YOLOv5l-SC 模型结构

Table 3　Structure of YOLOv5l-SC model

序　号	from	module	arguments
0	-1	Focus	[64,3]
1	-1	Conv	[128,3,1]
2	-1	space_to_depth	[1]
3	-1	C3	[128]
4	-1	Conv	[256,3,1]
5	-1	space_to_depth	[1]
6	-1	C3	[256]
7	-1	CoordAtt	[256]
8	-1	Conv	[512,3,1]
9	-1	space_to_depth	[1]
10	-1	C3	[512]
11	-1	CoordAtt	[512]
12	-1	Conv	[1 024,3,1]
13	-1	space_to_depth	[1]
14	-1	C3	[1 024]
15	-1	CoordAtt	[1 024]
16	-1	SPPF	[1 024,5]
17	-1	Conv	[512,1,1]
18	-1	nn. Upsample	[None,2,nearest]
19	[-1,10]	Concat	[1]
20	-1	C3	[512,False]
21	-1	Conv	[256,1,1]
22	-1	nn. Upsample	[None,2,nearest]
23	[-1,6]	Concat	[1]
24	-1	C3	[256,Fales]

续表(表3)

序 号	from	module	arguments
25	-1	Conv	[256,3,1]
26	-1	nn. Upsample	[1]
27	[-1,21]	Concat	[1]
28	-1	C3	[512,False]
29	-1	Conv	[512,3,1]
30	-1	nn. Upsample	[1]
31	[-1,17]	Concat	[1]
32	-1	C3	[1 024,False]
33	-1	Detect	[nc,anchors]

表 3 中,from 代表了当前层的输入层,而-1 则表示上一层,module 指代了当前层所使用的模块,arguments 包含了模块参数的信息,其中包括输入通道数、输出通道数、卷积核大小以及步长等。

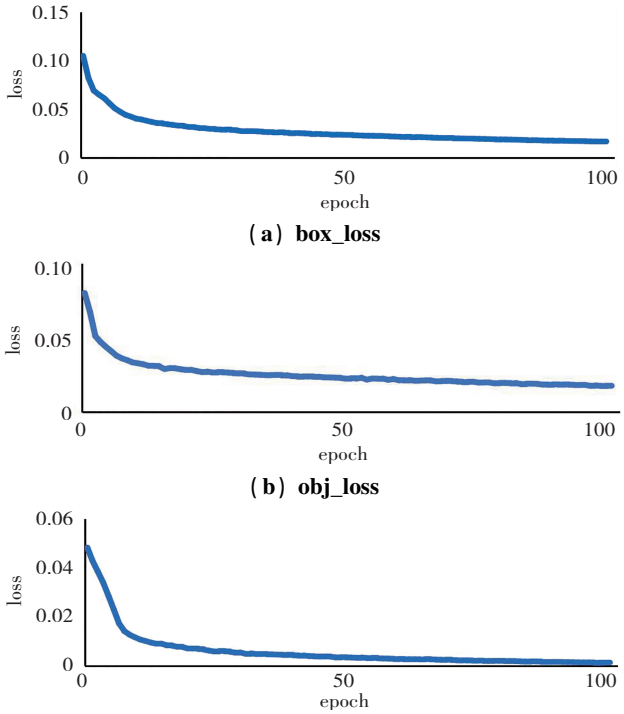
5 仿真实验与结果分析

5.1 实验环境

本文实验使用的操作系统为 Ubuntu20.4,CPU 型号为 Intel © Core™ i9-10980XE CPU @ 3.00GHz,运行内存为 64 GB。GPU 型号为 NVIDIA GeForce RTX3090,显存大小为 24 GB。采用的是基于 Pytorch 的深度学习框架,编程语言为 Python3.8,CUDA 版本为 11.4。

5.2 模型训练算法

在 YOLOv5 模型训练中,模型损失函数 L 值越低,模型越优秀,期望值是 0。为了达到模型的最佳性能,学习率设置为 0.01,学习率动量设置为 0.937,设置 epoch 为 100。训练 Loss 变化如图 10 所示。



(a) box_loss

(b) obj_loss

(c) cls_loss

图 10 训练损失函数变化图

Fig. 10 Plot of training loss function variation

5.3 评价指标

本文的实验使用了精确度、召回率、F_1-Score 和平均精度均值这几个评价指标来评估模型的准确性,并且使用检测帧率来评价模型的实际应用情况。在介绍评价指标之前,需要先了解一些概念。首先是正样本被正确分配的数量,即 N_{TP};其次是负样本被正确分配的数量,即 True Negatives(N_{TN});接着是被错误分配为正样本的数量,即 N_{FP};最后是被错误分配为负样本的数量,即 False Negatives (N_{FN})。

根据式(6),精确度 P 是指被正确分配的正样本数占总分配的正样本数的比例。

$$P=\frac{N_{TP}}{(N_{TP}+N_{FP})} \tag{6}$$

召回率 R 是指被正确分类的正样本数与总正样本数的比例。具体如式(7)所示。

$$R=\frac{N_{TP}}{(N_{TP}+N_{FN})} \tag{7}$$

F_1:精度和召回率的调和平均数。对于单个类别的 F_1 分数,可使用如下公式计算,如式(8)所示:

$$F_1^k=2\frac{R_k\times P_k}{R_k+P_k} \tag{8}$$

接着计算全部类别的平均值 F_1,如式(9)所示:

$$F_1=\left(\frac{1}{n}\sum_{i=1}^{n}F_1^k\right)^2 \tag{9}$$

R_{mAP} 是一种用于评估准确性的指标,它计算了所有类别 R_{AP} 的平均值。R_{AP} 的计算是根据差值平均准确率的方法进行的,即通过计算 Precision-Recall 曲线下的面积来得出。具体的计算方法可以参考公式(10)。

$$R_{AP}=\left(\frac{1}{n}\sum_{\left(r\in\frac{1}{n},\frac{2}{n},\cdots,\frac{n-1}{n},n\right)}P_i(r)\right) \tag{10}$$

式(10)中,n 代表示检测点的个数,$P_i(r)$ 表示召回率为 r 时准确率的数值。通过计算 R_{AP},可最终获得 R_{mAP} 的值,如公式(11)所示:

$$R_{mAP}=\frac{1}{n}\sum_{i=1}^{n}R_{AP} \tag{11}$$

5.4 实验结果与分析

为了验证本文所提出的算法(YOLOv5l-SC)具有更好的效果,使用控制变量法,在保证实验训练集、验证集、测试集完全相同的情况下,使用包含 Faster-RCNN、Cascade-RCNN、YOLOv3 目前较为流行的多种目标检测网络以及 YOLOv5 低参数版本 YOLOv5s 4 种模型进行对比实验。经过实验评价,本研究以 R_{mAP} 和 R_{Fps} 作为主要评价指标,并结合 F_1 值进行评估。实验结果如表 4 所示。

表 4　对比实验表

Table 4　Table of comparative experiments

模型	N_{GFLOPs}	参数量大小	准确率	召回率	F_1	mAP. 5	R_{fps}/fps
Faster-RCNN	322. 33	44. 14M	—	—	—	0. 622	22. 3
Cascade-RCNN	350. 12	68. 94M	—	—	—	0. 714	19. 4
YOLOv3	310. 29	61. 55M	—	—	—	0. 506	54. 0
YOLOv5s	15. 8	7M	0. 614	0. 588	0. 600	0. 583	77. 4
YOLOv5_SPD_CA	107. 7	53. 65M	0. 728	0. 711	0. 719	0. 727	35. 44

根据表 4 可得出结论,本文所提出的改进算法能够有效地检测和识别护目镜佩戴情况。对各类目标的检测 mAP 达到 72. 7%(当 IoU 阈值值设为 0. 5 时,对所有类别的平均精度),远高于 YOLOv3 和 YOLOv5s,高于 Faster-RCNN,略高于 Cascade-RCNN;正向推理计算量(N_{GFLOPs})和推理帧率(R_{fps})相较于 Faster-RCNN 和 Cascade-RCNN 两阶段检测算法明显更有优势,更小的参数量可以保证模型在低算力边缘设备部署时,也能有很好的检测效果。综上所述,本文提出的护目镜佩戴检测算法在综合测试中性能最优,能够满足实际复杂场景下的护目镜佩戴检测。

5. 5　消融实验

该消融对比实验是为了验证各个改进模块的优化作用,实验结果如表 5 所示。以 YOLOv5l 为基础模型,分别改进损失函数和添加 SPD 模块,在仅优化损失函数或仅添加 SPD 小目标检测模块时,各项指标均略有提升,对比表明同时优化损失函数和添加 SPD 小目标检测模块提升更为显著。在此基础上添加坐标注意力机制有效解决了因引入 SPD 卷积所造成的位置信息丢失问题,使得模型的准确率进一步提升。消融实验结果表明本文所改进的 YOLOv5l-SC 检测模型比原始模型显著提升。

表 5　消融实验对照表

Table 5　Comparison table of ablation experiments

模型	SIoU	SPD	CA	准确率	召回率	F_1	mAP. 5	R_{fps}/fps
YOLOv5l	×	×	×	0. 709	0. 631	0. 668	0. 671	44. 66
YOLOv5l-SIoU	√	×	×	0. 712	0. 68	0. 696	0. 681	44. 02
YOLOv5-SPD	×	√	×	0. 743	0. 668	0. 703	0. 682	35. 92
YOLOv5-SPD_SIoU	√	√	×	0. 74	0. 693	0. 716	0. 708	34. 64
YOLOv5l-SC	√	√	√	0. 728	0. 711	0. 719	0. 727	35. 44

5. 6　检测结果分析

为了测试本文网络模型的实际场景的可行性,对数据集图片进行测试,图 11 为本文模型在不同场景的检测效果图。由图可以看出:本文训练出的检测模型,在应对人员密集场景时也拥有较高的检测准确率。子图 3 是复杂背景的检测情况,可以看出在一些背景任务干扰情况下,可以准确识别真实的人,而不会错误识别显示屏中的人,子图 4 为远处人员目标,模型可做到准确识别较远处目标。综上所述,YOLOv5l-SC 模型可以应对大多数场景下的护目镜佩戴检测任务。

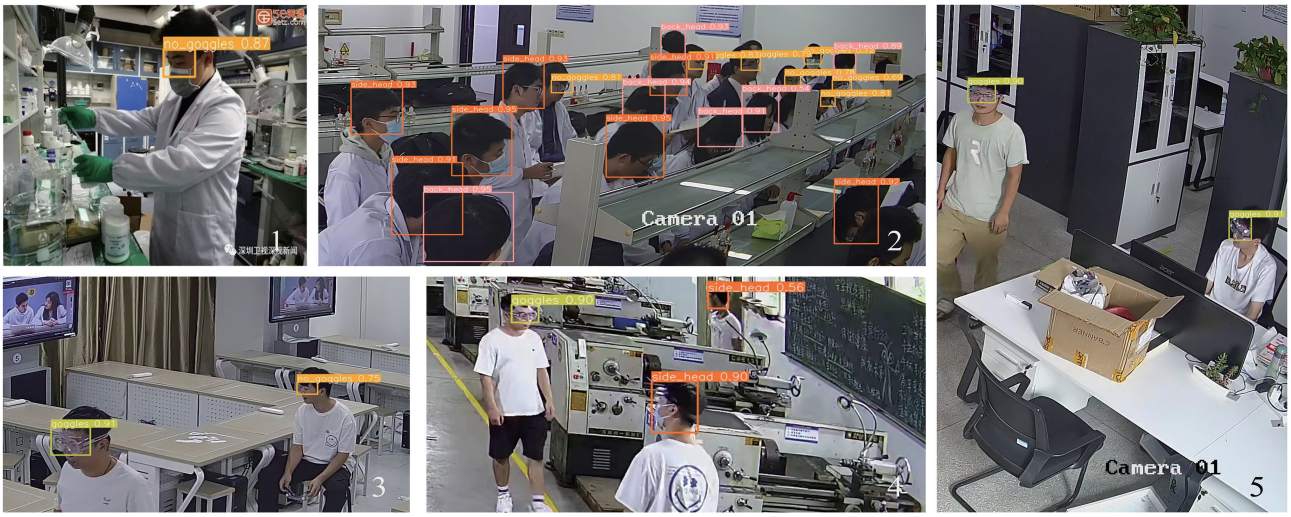


图 11　实验场景测试结果图

Fig. 11　Test results of experimental scenarios

6 结 论

基于实验室工业场景的人员眼部安全防护需求,首先构建了一个包含多个场景的护目镜佩戴检测数据集,通过数据集分析和数据增强技术使得数据集各个检测类别达到均衡;接着提出了一种基于 YOLOv5 模型的改进算法 YOLOv5l-SC 模型。该模型通过添加 SPD 小目标检测模块和改进的 SIoU 损失函数,随后引入坐标注意力机制解决了因添加 SPD-Conv 所造成的位置信息丢失问题,最终,提高了模型检测护目镜佩戴情况的准确性。实验结果表明:YOLOv5l-SC 模型在复杂背景、远景、密集人群等场景均可实现护目镜佩戴情况检测。得益于构建的数据集场景丰富,训练出的 YOLOv5l-SC 模型泛化能力较优,可有效应对其他场景的护目镜佩戴检测任务。

针对护目镜佩戴检测所实际应用的场景进行分析,可以看出下一步的研究方向是模型轻量化,在保证检测精度的前提下,尽可能优化提升模型的检测速度,以实现护目镜佩戴检测设备的分布式部署,最终低成本、大规模实现护目镜佩戴检测项目的落地。

参考文献(References):

[1] 黄勤,王敬伟,黄旺兴,等.眼外伤住院患者 1 851 例临床分析[J].广东医学,2019,40(S1):251-253.
HUANG Qing, WANG Jin-wei, HUANG Wang-xing, et al. Clinical analysis of 1 851 hospitalized patients with ocular trauma [J]. Guangdong Medical Journal, 2019, 40(S1): 251-253.

[2] 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. Piscataway: IEEE Press, 580-587.

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

[4] SHAOQING R, KAIMING H, ROSS G, 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.

[5] BERG A C, FU C Y, SZEGEDY C, et al. SSD: Single shot multibox detector: 10.1007/978-3-319-46448-0 _ 2 [P]. 2015. 1. 1

[6] 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 Pecognition, Riscataway: IEEE Press, 2016: 779-788.

[7] REDMON J, FARHADI A. YOLO9000: better, faster, stronger [C]//Proceedings of the IEEE Conference on Computer Vision & Pattern Recognition. Piscataway: IEEE Press, 2017: 6517-6525.

[8] 张登奎. 基于深度学习的护目镜佩戴检测系统的设计与实现[D].北京,北京邮电大学,2021.
ZHANG Deng-kui. Design and implementation of goggles wear detection system based on deep learning[D]. Beijing: Beijing University of Posts and Telecommunications, 2021.

[9] 杨国亮,杨浩,余帅英,等.改进 YOLOv5 的交通标志检测算法[J].计算机工程与应用,2023,59(10):262-269.
YANG Guo-liang, YANG Hao, YU Shuai-ying, et al. Improved YOLOv5 traffic sign detection algorithm [J]. Computer Engineering and Applications, 2023, 59(10): 262-269.

[10] 黎国溥,陈升东,王亮,等.基于改进 YOLOv5 的车辆端目标检测[J].计算机系统应用,2022,31(12):127-134.
LI Guo-po, CHENG Sheng-dong, WANG Liang, et al. Vehicle-end target detection based on improved YOLOv5[J]. Computer System Application, 2022, 31(12): 127-134.

[11] 邵海琳,季怡,刘纯平,等.基于增强特征金字塔网络的场景文本检测算法[J].计算机科学, 2022, 49(2): 248-255.
SHAO Hai-ling, JI Yi, LIU Chun-ping, et al. Scene text detection algorithm based on enhanced feature pyramid network[J]. Computer Science, 2022, 49(2): 248-255.

[12] 赵应武.图像超分辨率重建[J].中国新通信,2020, 22(2): 99-100.
ZHAO Ying-wu. Image super-resolution reconstruction [J]. China New Telecommunications, 2020, 22(2): 99-100.

[13] HOU Q, ZHOU D, FENG J. Coordinate attention for efficient mobile network design [C]//Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. Riscataway: IEEE Press, 2021: 13713-13722.

[14] 彭成,张乔虹,唐朝晖,等.基于 YOLOv5 增强模型的口罩佩戴检测方法研究[J].计算机工程,2022,48(4):39-49.
PENG Cheng, ZHANG Qiao-hong, TANG Zhao-hui, et al. Research on mask wearing detection method based on YOLOv5 enhancement model[J]. Computer Engineering, 2022, 48(4): 39-49.

[15] 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/CVF conference on computer vision and pattern recognition. Piscataway: IEEE Press, 2019: 658-666.

责任编辑:李翠薇