计算机系统应用 ISSN 1003-3254, CODEN CSAOBN
Computer Systems & Applications,2023,32(8):221−229 [doi: 10.15888/j.cnki.csa.009197]

E-mail: csa@iscas.ac.cn
http://www.c-s-a.org.cn
Tel: +86-10-62661041

面向小目标的 YOLOv5s 安全帽佩戴检测①

李冰涛, 李大海

(江西理工大学 信息工程学院, 赣州 341001)
通信作者: 李冰涛, E-mail: 920887943@qq.com

摘　要: 本文介绍了一种新的基于 YOLOv5s 的目标检测方法, 旨在弥补当前主流检测方法在小目标安全帽佩戴检测方面的不足, 提高检测精度和避免漏检. 首先增加了一个小目标检测层, 增加对小目标安全帽的检测精度; 其次引入 ShuffleAttention 注意力机制, 本文将 ShuffleAttention 的分组数由原来的 64 组减少为 16 组, 更加有利于模型对深浅、大小特征的全局提取; 最后增加 SA-BiFPN 网络结构, 进行双向的多尺度特征融合, 提取更加有效的特征信息. 实验表明, 和原 YOLOv5s 算法相比, 改善后的算法平均精确率提升了 1.7%, 达到了 92.5%, 其中佩戴安全帽和未佩戴安全帽的平均精度分别提升了 1.9% 和 1.4%. 本文与其他目标检测算法进行对比测试, 实验结果表明 SAB-YOLOv5s 算法模型仅比原始 YOLOv5s 算法模型增大了 1.5M, 小于其他算法模型, 提高了目标检测的平均精度, 减少了小目标检测中漏检、误检的情况, 实现了准确且轻量级的安全帽佩戴检测.

关键词: YOLOv5s; 安全帽佩戴检测; ShuffleAttention 注意力机制; SA-BiFPN

引用格式: 李冰涛,李大海.面向小目标的 YOLOv5s 安全帽佩戴检测.计算机系统应用,2023,32(8):221–229. http://www.c-s-a.org.cn/1003-3254/9197.html

YOLOv5s-based Helmet Wearing Detection for Small Targets

LI Bing-Tao, LI Da-Hai

(School of Information Engineering, Jiangxi University of Science and Technology, Ganzhou 341001, China)

Abstract: In this study, a new target detection method based on YOLOv5s is introduced to make up for the deficiencies of the current mainstream detection methods in terms of detection precision and missed detection of small target helmet wearing. Firstly, a small target detection layer is added to increase the detection precision of the small target helmet. Secondly, the ShuffleAttention mechanism is introduced. The number of ShuffleAttention groups is reduced from 64 to 16 in this study, which is more conducive to the global extraction of the depth and size of the model. Finally, the SA-BiFPN network structure is added to carry out the bidirectional multi-scale feature fusion to extract more effective feature information. Experiments show that compared with the original YOLOv5s algorithm, the average precision of the improved algorithm is increased by 1.7%, reaching 92.5%. The average precision of the algorithms with and without helmets is increased by 1.9% and 1.4% respectively. The proposed detection algorithm is compared with other target detection algorithms. The experimental results show that the SAB-YOLOv5s algorithm model is only 1.5M larger than the original YOLOv5s algorithm model, which is smaller than other algorithm models. It improves the average precision of target detection, reduces the probability of missing and false detection in small target detection, and achieves accurate and lightweight helmet wearing detection.

Key words: YOLOv5s; helmet wearing detect; ShuffleAttention mechanism; SA-BiFPN

① 收稿时间: 2023-02-08; 修改时间: 2023-03-08; 采用时间: 2023-03-14; csa 在线出版时间: 2023-05-22
CNKI 网络首发时间: 2023-05-24

始终把安全生产放在第 1 位, 尤其是在建筑施工领域这样长期有较高风险的环境中. 目前, 建筑行业仍属于安全事故高频发的行业, 参加施工人数的逐年递增, 是一个不可忽视的庞大群体[1]. 而安全帽作为生产工作者的必备安全工具, 它能够起到缓冲和减震的作用, 还能分散一定的压力, 对于保护人的头部来说作用很大. 如果没有佩戴安全帽, 就会失去对头部的保护, 佩戴人员非常容易受到伤害. 因此, 在施工现场, 要求现场人员必须要佩戴安全帽, 并且要正确佩戴, 最大程上减少安全风险[2]. 所以, 对安全头盔的配戴状况进行监测, 对于保障工人的生命健康有着十分重大的现实意义.

现阶段, 目标检测算法可以分为两个主流方向: 一种是基于回归策略的单阶段检测算法, 如 SSD[3]、YOLO 系列[4–7]、RetinaNet[8] 等. 这类算法的核心理论是将图像输入模型, 直接返回目标的边界锚框、位置和类别信息. 另一种是基于优化候选区域的两阶段检测算法, 主要有 R-CNN[9]、Fast R-CNN[10]、Faster R-CNN[11] 和 R-FCN[12] 等. 这类算法主要是在第 1 阶段从图像中生成一个候选区域, 然后在第 2 阶段将候选区域输入到卷积神经网络中, 利用分类器对结果进行类别的判定. 这两种方法都有各自的优点, 其中单阶段的检测方法具有更高的时效性, 在实时性上有很大的优越性, 而两阶段的检测方法平均精度更高, 识别的结果更准确.

上述两类算法在安全帽佩戴检测中均有大量运用, 但存在对小目标安全帽检测效果不佳的问题. 2021 年李鹏[13] 在 Faster R-CNN 的基础上增加可变形卷积和可转换空洞卷积, 获得了一个能够自适应目标尺度转换的多尺度安全帽佩戴检测网络结构, 其平均准确度较当时最优的模型有所改善, 但对于小目标安全帽以及遮挡目标的检测效果并不理想. Song[14] 在 YOLOv3 的基础上提出压缩激励的 RSSE 模块用于加强特征提取, 采用四尺度特征预测代替三尺度特征预测, 并且改进了 CIOU 损失函数, 该方法在提高了探测准确率和速度的同时, 也提高了探测的速度, 但是在小物体上, 其探测效率不高, 而且容易受到光线的干扰. 杨贞等[15] 基于 YOLOv4 并且采用深层次的网络模型替换传统的深度特征网络模型, 具有良好的健壮性, 但对于远距离小目标安全帽佩戴检测的误检率较高, 存在人员密集场景下检测不出的情况.

此外, 由于图像中的小目标具有较低的特征表示和较低的分辨率, 以及在人群中很容易发生遮挡等特点, 针对小目标安全帽检测也成为了研究的重难点. 吕宗喆等[16] 基于 YOLOv5 算法对损失函数的计算方法进行了优化, 并且加入了切片辅助微调和推理, 让小目标能够产生更大的像素面积, 从而提高了网络的微调与推断能力. 李嘉信等[17] 提出了一种多维空间注意力模型, 对各维度的空间特性进行融合, 同时结合特征提取过程中的多种特征, 实现了在特征抽取中对多维信息的有效获取. 朱玉华等[18] 在 Faster R-CNN 的基础上将 ResNet101 和 FPN 进行了融合, 形成了一个多尺度的特征提取网络. 在安全帽识别过程中, 对安全帽识别器的大小进行了适当的调节, 使得安全帽识别器能够覆盖全部的目标区. 这些方法都旨在保留小目标的特征, 加强特征提取过程. 受此启发, 本文也将其作为切入点进行研究.

本文基于安全帽佩戴检测任务, 并且针对小目标检测, 提出了一种改进的 YOLOv5s 算法. 在确保检测准确率的同时, 尽量加快检测的效率, 本文使用了基于 YOLOv5s 的轻型算法. 首先, 在小物体探测中加入一个小型物体探测层次, 以提高其探测准确率. 其次, 针对特征提取不充分问题, 在 YOLOv5s 中加入了改进的 ShuffleAttention 注意力机制, 其对特征通道进行分组, 并对每个子特征同时使用空间和通道注意力机制, 最终让不同组的特征进行融合. 最后, 为了提取更加有效的特征信息, 提高目标的检测精度, 提出了 SA-BiFPN 网络结构, 将改进的 ShuffleAttention 注意力机制和 BiFPN 结构融合, 采用双向多尺度特征融合技术, 可以增强模型的学习性能, 减少漏检率.

1 YOLOv5s 的网络结构

YOLOv5 是目前 YOLO 系列算法中最优秀的一代, 其精度高, 运行速率快, 而且体积较少. YOLOv5 共有 5 种模型: YOLOv5n、YOLOv5s、YOLOv5m、YOLOv5l 和 YOLOv5x. 它们之间最大的差异在于, 在特定的网络中, 特征提取模块的数量是和卷积核有差别的. 模型大小和参数数目在 5 种不同的版本中依次递增. 鉴于安全帽佩戴检测对实时性和轻量化方面的需求, 本文从模型的大小、效率和准确度等角度出发, 选择 YOLOv5s 作为改进的算法框架.

YOLOv5s 网络主要由 4 部分组成: input、backbone、neck 和 head. Input 中使用了 Mosaic 的数据强

化技术, 将 4 幅照片采用随机放缩、随机剪切以及随机排列的方式进行拼接, 大大丰富了检测数据集, 特别是随机缩放增加了很多小目标, 让网络的鲁棒性更好[19]. Backbone 主干网络主要的作用是用来提取图片特征, 其主要由 CBS、C3、SPPF 等模块组成. CBS 从图像中提取特征, 然后使用 C3 来减少计算量和内存, 有利于加快推理速度, SPPF 为空间金字塔池化层, 通过不同池化核大小的最大池化进行特征提取, 提高网络的感受野. Input 还引入了基于图像的自适应性调整和基于图像扩展的算法. Backbone 部分是 YOLOv5s 的主干网络, 是用来提取特征的网络. 主要由 CBS、C3、SPPF 等模块组成, CBS 对图像提取特征, 之后用 C3 降低计算量和内存, 有利于加快推理速度, SPPF 为空间金字塔池化层, 通过不同池化核大小的最大池化进行特征提取, 提高网络的感受野. Neck 中使用了 PANet 的构造方法, 它可以提高模型在不同比例下的检测器, 从而可以区分出相同尺寸和尺寸的相同目标. Head 部分是 3 个 Detect 检测器, 利用基于网格的 anchor 在不同尺度的特征图上进行目标检测, 输入大小为 640×640 的图像, 那么输出的检测层尺度为 80×80、40×40 和 20×20.

2 改进的 YOLOv5s 算法

2.1 添加改进的 ShuffleAttention

针对图像中较小的物体, 由于其所占的像素较少, 且容易受到周围环境等因素的影响, 使得原有 YOLOv5 模型在进行卷积取样时, 小目标的特征信息很容易丢失, 因此本文引入 ShuffleAttention (SA)[20] 注意力机制, 告知模型应该更多地注意哪些地方. 对于小型和集中的对象, 该方法能有效地进行特征信息的提取, 从而进一步提高了检测的精度.

注意力机制主要分为两类, 一类是空间注意力机制, 另一类是通道注意力机制. 通过不同的聚合策略、变换和增强函数, 将不同位置的相同特征聚合起来, 分别去获取像素对之间的关系和通道依赖关系. 虽然将二者结合起来可能会获得更优的表现, 但还是会不可避免地增加算力消耗, 比如 CBAM[21], 而 SA 可以解决此类问题. SA 可以将两种注意力机制有效地结合起来, 并且模型的复杂度较低, 计算量小.

首先对于给定的特征图 $X \in R^{C\times H\times W}$, 其中 C, H, W 分别是通道数、空间高度和宽度. SA 首先沿着通道维度将 X 划分为 G 个组, 即 $X=[X_1,\cdots,X_G]$, $X_k \in R^{C/G\times H\times W}$, 在训练过程中每个子特征 X_k 逐渐地获取一个语义响应. 然后利用注意力模块, 让每个子特征都生成一个对应的重要度系数. 在任意一个注意力单元的开始时刻, X_k 的输入会沿着通道维度, 使其分为两个分支, $X_{k1}, X_{k2} \in R^{C/2G\times H\times W}$.

如图 1, 一个分支利用通道间的相互关系, 输出通道注意力图, 另一个分支则会利用特征的空间关系, 输出空间注意力图. 这样, 模型就可以关注在那些有意义的信息上. 然后两个分支的结果会被拼接起来, 使得通道个数与输入的个数相同. 最后, 所有的特征会被聚合起来, 沿着通道维度实现跨组信息交流. 通道注意力和空间注意力的输出表示分别如式 (2)、式 (3) 所示:

$$s=\frac{1}{H\times W}\sum_{i=1}^{H}\sum_{j=1}^{W}X_{k1}(i,j) \tag{1}$$

$$X'_{k1}=\sigma(W_{1s}+b_1)\cdot X_{k1} \tag{2}$$

$$X'_{k2}=\sigma(W_2\cdot GN(X_{k2})+b_2)\cdot X_{k2} \tag{3}$$

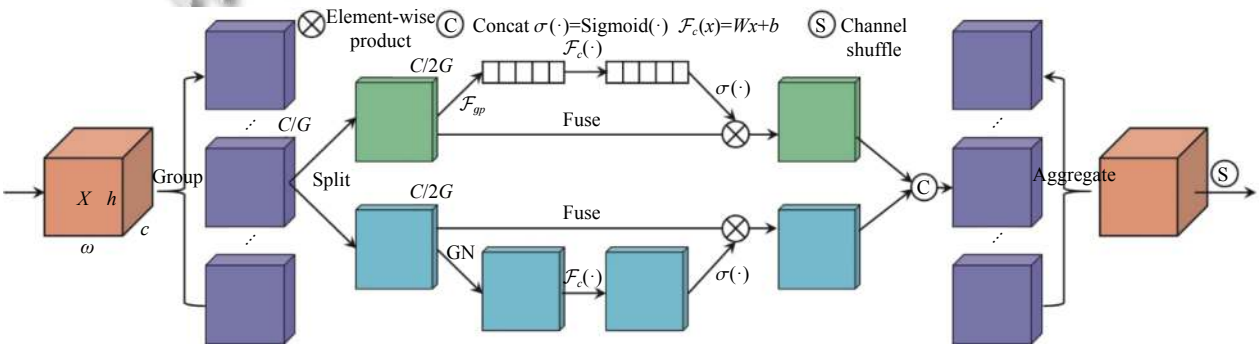


图 1 SA 模块结构图

其中, $W_1 \in R^{C/2G\times1\times1}$和$b_1 \in R^{C/2G\times1\times1}$用于缩放和平移 s, GN 为 Group Norm, 用于获取空间统计数据, 而 W_2, b_2是形状为$R^{C/2G\times1\times1}$的参数.

值得注意的是, W_1, b_1, W_2, b_2和 GN 超参数为 SA 中引入的参数. 在单个 SA 模块中, 每个分支中的通道数为$C/2G$, 因此单个 SA 模块中的总参数为$3C/G$. 但其实 SA 模块的总参数相对于 YOLOv5s 网络百万级别的参数量来说微不足道, 而且参数G的不同取值对通道注意力有较大的影响, 因此本文考虑从分组数作为切入点进行改进.

SA 的作者设置默认的G为 64, 也就是将特征图分为 64 组, 这样虽然让参数量有所降低, 但过多的分组会降低通道注意力的效果, 使得组与组之间的通信变弱, 影响小目标的检测精度. 经过实验探究 (如图 2 所示), 在可能不添加更多参数的情况下, 同时尽可能地提高检测的平均精确度, 本文选择将特征图分成 16 组, 即设置G为 16, 这样在增加少量参数的情况下, 让 SA 注意力机制更适合于本文的小目标检测, 检测精度更高. 算法流程如算法 1 所示.

算法 1. 改进 G 为 16 的 SA 注意力机制

1) 输入特征映射X, 沿着通道尺寸将其分为 16 组, 即$X=[X_1,\cdots,X_{16}]$, 并且每组X逐渐捕获训练过程中的特定语义响应;
2) 将每组X分为两部分, 一部分使用通道注意力, 另一部分使用空间注意力;
3) 对两个部分按通道数进行叠加, 实现组内的信息融合;
4) 对所有的X进行随机混合操作, 使不同组之间进行信息流通;
5) 输出特征图.

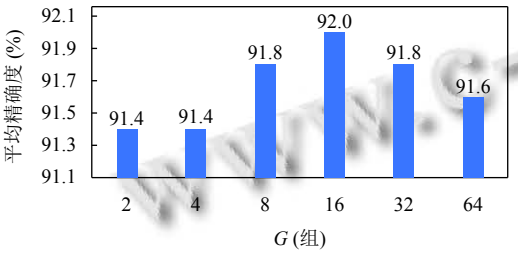


图 2 G 的不同取值对平均精确度的影响

SA-16 (将 SA 中对特征图的分组设置为 16) 模块不但可以提取图像更多的基本特征, 还可以加强特征间的相关信息交流, 有利于模型对深浅、大小特征的全局采集, 从而提高模型的准确率. 并且 SA-16 引用的参数量并不多, 保证了模型的轻量性.

2.2 改进的多尺度特征融合网络 SA-BiFPN

低层神经网络具有较低的感知范围和较好的特征表达, 具有很好的解析度, 但对语义的描述能力不够全面. 与之相反, 随着网络层次的增加, 感知范围会越来越大, 语义表达也会越来越好, 但同时也会影响到图像的解析度. 在进行多层次的卷积运算之后, 许多细节特性会逐渐模糊, 对图像的描述也会减弱, 这也是小目标检测困难的原因之一. 因此, 为了获得强语义性的特征, 更好地检测出小目标对象, 多尺度特征融合应运而生.

目前常见的特征融合网络有 FPN[22]、PANet[23] 和 BiFPN[24] 等. FPN 特征金字塔结构首先对特征点进行不断的下采样后, 拥有了一堆具有高语义内容的特征层, 然后重新进行上采样, 使得特征层的长宽重新变大, 用大尺寸的特征图去检测小目标. PANet 相较于 FPN 增加了一条自底向上的通道, 用于缩短层之间的路径, 还提出了自适应特征池和全连接融合两个模块, 保证特征的完整性和多样性. BiFPN 给各个特征层赋予了不同权重去进行融合, 让网络更加关注重要的层次.

为了提取更加有效的特征信息, 本文提出了 SA-BiFPN 网络结构, 将 SA-16 和 BiFPN 结构相结合. 首先从骨干网络 SA-16 层中提取不同大小的特征图, 然后通过水平连接与下采样层进行第一次特征融合, 再通过跳跃连接与相同尺度下的上采样层和下采样层进行第 2 次特征融合, 最后得到多尺度融合特征图[25]. SA-BiFPN 网络结构图如图 3 所示, 算法流程如算法 2 所示.

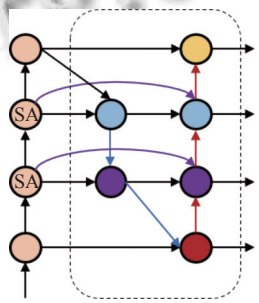


图 3 SA-BiFPN 网络结构图

本文将 YOLOv5 中的特征金字塔网络 PANet 改进为 SA-BiFPN 结构, 增强了特征金字塔的表达能力, 使网络的特征融合能力达到更优. 将 backbone 中经过 SA 模块处理的特征图被反复地输入到 BiFPN 结构中, 进行双向多尺度的特征融合, 从而增强了模型对特征的提取能力, 并且提取的特征更加有效, 提升了模型的准确率.

算法 2. 结合 SA 注意力机制的 BiFPN

1) 输入带有 SA 注意力的特征结点;
2) 增加一个跳跃连接, 由于它们在相同层, 在不增加过多计算量的同时, 可以融合更多的特征;
3) 对同一层进行多次重复, 以达到更高层次的特征融合. 将每一条双向 (自上而下和自下而上) 路径视为一个特征的网络层;
4) 增加一个额外的权重, 让网络可以学习每个输入特征的重要性;
5) 输出特征结点.

2.3 增加小目标检测层

原始的 YOLOv5s 网络的 head 部分共有 3 个尺度的检测层, 如果输入图像的尺寸为 640×640, 那么输出的检测层尺寸分别为 80×80, 40×40, 20×20, 分别用于小、中、大目标的检测. 然而, 由于实际场景中情况复杂, 施工现场的工人较多, 且大多在远处或高空作业, 在视频或图像中的显示较小, 为了提高佩戴安全帽检测的准确率, 本文在原 YOLOv5s 网络的 head 部分再增加一个尺度为 160×160 的检测层, 满足识别小目标安全帽的需求. 这样在 head 部分共输出 4 个尺度的检测层, 而在模型训练时输出层每层都有 3 个先验框, 那么所有先验框的数量就增加到了 12 个. 增加的检测层更有利于检测小目标安全帽, 减少漏检、误检的情况.

最终改进的 SAB-YOLOv5s 模型如图 4 所示.

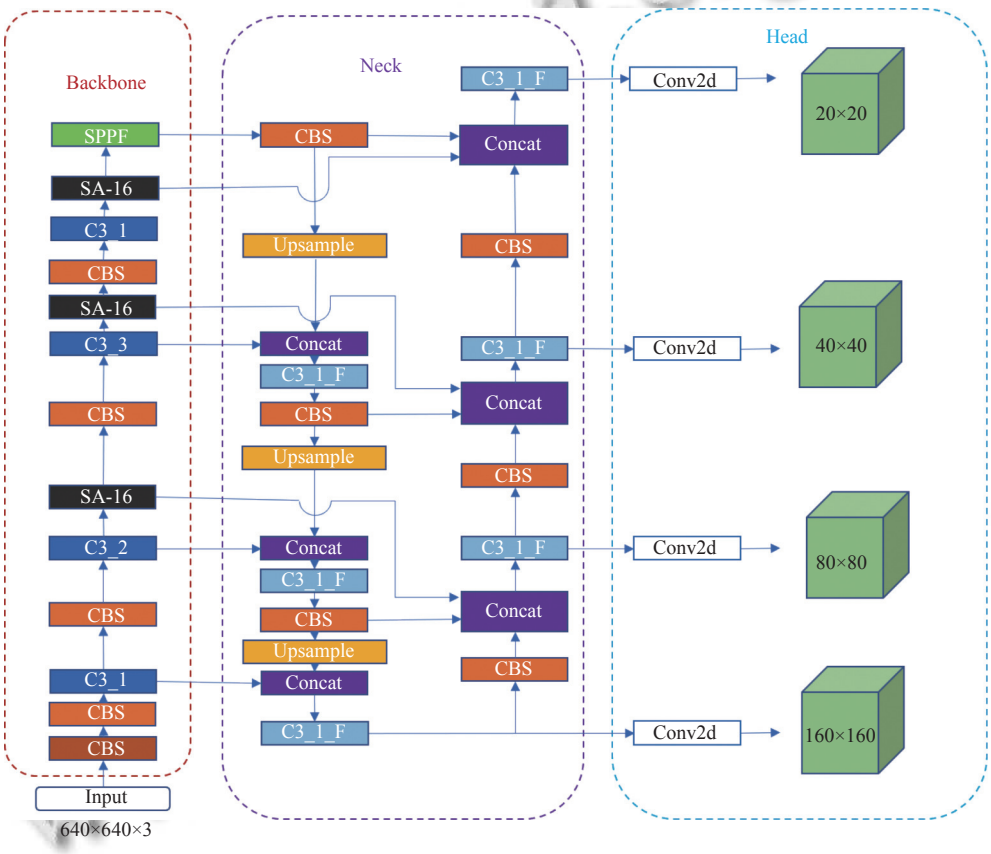


图 4 SAB-YOLOv5s 模型结构

3 实验结果及分析

3.1 实验环境及数据集

本文的实验环境为单张 Tesla V100GPU, 内存 32 GB, 基于 Linux 操作系统, 使用 PyTorch 1.10.2 作为深度学习框架, Python 版本为 3.7.6, CUDA 版本为 10.2. 消融实验硬件配置环境相同.

本文所使用的安全帽佩戴检测数据集是 SHWD (safety helmet wearing-dataset). 该数据集一共包括 7 581 张图像, 其中包含了 9 044 个戴着安全帽的对象, 11 514 个没有戴安全帽的对象, 并且本文还把数据集中的 VOC 格式用 Python 脚本批量地转化成 YOLO 格式, 方便利用 YOLOv5s 模型进行训练. 标注的内容包括两类, 分别是 hat 和 person. 其中 hat 代表佩戴安全帽, person 代表未佩戴安全帽. 该数据集编写脚本随机

分成训练集、验证集和测试集, 其中训练集有 5 957 张图片, 验证集有 812 张图片, 测试集有 812 张图片, 训练集、验证集和测试集都是独立的.

3.2 训练策略与评估指标

进入网络训练的图片大小都被设置为 640×640, 初始学习率 lr0 设置为 0.01, 并且使用余弦退火动态调整学习率, 学习率动量为 0.937, 权重衰减系数为 0.000 5, batch size 为 24, 为增加训练样本的多样性, 开启 Mosaic 数据增强, 采用 SGD 函数优化函数, 训练 100 个 epoch.

对实验中的真实情况和预测情况的分类如表 1 所示. 其中 *TP* 表示预测情况与真实情况都为正例, *FP* 表示预测为正例而真实为反例, *FN* 表示预测为反例而真实情况为真例, *TN* 表示预测情况的真实情况均为反例. 本文采用的衡量指标包括平均精度 *AP* (average precision) 和均值平均精度 *mAP* (mean average precision), 平均精度综合考虑了精确率 *P* (precision) 和召回率 *R* (recall). 同时为了实际部署的需要, 也应考虑模型的体积, 也就是训练完成后生成的模型权重大小. 以 *P* 为纵轴, *R* 为横轴构成 *P*-*R* 曲线, 那么 *AP* 就是 *P*-*R* 曲线围成的面积, 而 *mAP* 就是所有类别 *AP* 的平均值, 一般取 *IoU*=0.5 来计算 *mAP* 也就是 *mAP*@0.5. 上述所提指标计算如下:

$$P = \frac{TP}{TP+FP} \tag{4}$$

$$R = \frac{TP}{TP+FN} \tag{5}$$

$$AP = \int_0^1 P(r)dr \tag{6}$$

$$mAP = \frac{\sum_{i=1}^{n} AP_i}{n} \tag{7}$$

其中, $n=2$, 因为本次实验的检测类别数为 2, 而 AP_i 则表示第 i 个类别的准确率.

表 1 真实情况和预测情况的分类

真实情况	预测情况	
	True	False
Positive	*TP* (true positive)	*FN* (false positive)
Negative	*FP* (false positive)	*TN* (true negative)

3.3 损失函数对比

YOLOv5 网络的损失函数共包含 3 个部分: box_loss 表示定位损失函数, 用于衡量预测框与标定框之间的误差; obj_loss 表示置信度损失函数, 反映了网络的置信度误差; cls_loss 表示分类损失函数, 用于计算锚框与对应的标定分类是否正确.

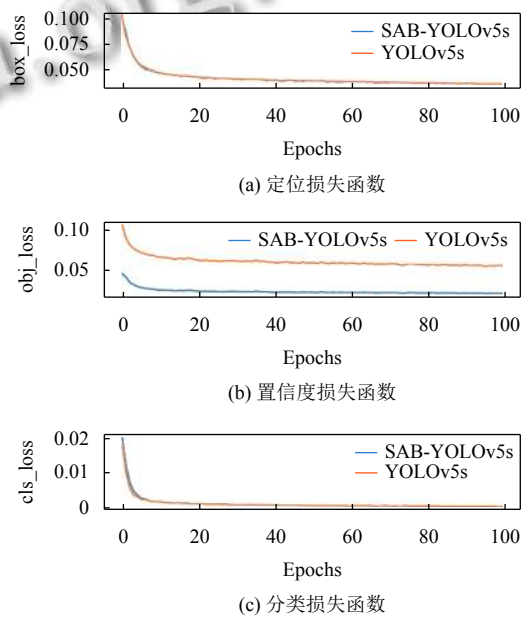
$$IoU = \frac{(A \cap B)}{(A \cup B)} \tag{8}$$

$$GIoU\_Loss = IoU - \frac{C-(A \cup B)}{C} \tag{9}$$

$$L = -\frac{1}{n}\sum_{x}[y\ln a + (1-y)\ln(1-a)] \tag{10}$$

box_loss 采用 *GIoU_Loss* 函数, *GIoU_Loss* 较于 *IoU_Loss* 具有更快的收敛速度, 并且更加稳定. 计算公式可表示为式 (9), 其中 A 是真实目标框, B 是预测框, C 是这两个区域的闭包 (包含这两个矩形区域并且平行于坐标轴的最小矩形). obj_loss 和 cls_loss 采用交叉熵损失函数, 计算公式可表示为式 (10), 其中 x 代表样本, y 代表标签, a 表示预测的输出, n 表示样本总量.

图 5 从上到下表示 box_loss、obj_loss 和 cls_loss, 它们可以度量神经网络预测信息与期望信息 (标签) 的距离, 预测信息越接近期望信息, 损失函数值越小. 由图可以明显看出 SAB-YOLOv5s 模型的 obj_loss 较于 YOLOv5s 模型来说收敛速度更快, 并且损失值也更小, 说明本文提出的模型提升了原网络的收敛能力, 并且网络的置信度更高, 识别更准确.



(a) 定位损失函数

(b) 置信度损失函数

(c) 分类损失函数

图 5 损失函数对比图

3.4 实验结果对比和分析

为了能更加为清晰地展现检测结果和评价性能, 本文进行了模块的消融实验, 其实验结果如表 2 所示.

表 2 消融实验结果

方法	第1组	第2组	第3组	第4组
YOLOv5s	√	√	√	√
检测层	—	√	√	√
SA-16	—	—	√	√
SA-BiFPN	—	—	—	√
mAP@0.5 (%)	90.8	91.4	92.0	92.5
Weight (M)	14.3	15.1	15.2	15.8

从实验结果可以看出, 相较于原始的 YOLOv5s, 增加了小目标检测层后, *mAP* 从 90.8% 提高到了 91.4%. 实验结果显示增加检测层后提高了网络对小尺寸目标的检测准确性, 减少了小目标安全帽的漏检率. 而在增加了 SA-16 模块后, *mAP* 从 91.4% 提升到了 92.0, 结果表明网络在增加 SA 注意力机制后提高了对图像浅层和深层特征信息的提取能力, 检测精确度有一定程度的提高. 在最终增加 SA-BiFPN 网络进行多尺度特征融合后, *mAP* 达到了 92.5%, 有效提高了网络模型的检测精确度. 最终模型的 *mAP* 相较于原始的 YOLOv5s, 提高了 1.7%, 而模型权重仅增加了 1.5M, 验证了本文改进算法的可行性.

同时为了探究本文算法对数据集中佩戴安全帽和未佩戴安全帽两种图像的具体提升情况, 表 3 列出了 YOLOv5s 算法和本文算法的详细参数对比. 从表中可以看出, 本文算法对于原始 YOLOv5 算法的 *AP* 值均有所提升, 其中佩戴安全帽的情况提升了 1.9%, 而未佩戴安全帽的情况提升了 1.4%.

从图 6 可以看出, 原始的 YOLOv5s 模型存在着相当一部分漏检的情况, 尤其是小目标安全帽佩戴检测, 而这恰恰是实际应用中安全帽佩戴检测的难点、痛点. 本文提出的算法能够有效解决密集小目标未识别的问题, 优化安全帽检测中漏检的问题. 但本文算法的检测结果如图 7 所示也存在一些瑕疵, 大部分目标物体检测精度提高的同时, 少部分有非常微小的降低.

表 3 SAB-YOLOv5s 与其他主流算法的性能对比

算法	*mAP* (%)	模型大小 (M)	FPS	*AP* (%)		*P* (%)		*R* (%)	
				hat	person	hat	person	hat	person
Faster-RCNN	90.9	108	74	87.8	94.0	89.6	93.7	82.1	90.2
SSD300	83.4	97.5	89	80.5	86.3	82.3	86.2	78.3	83.1
YOLOv4	86.8	203	101	84.3	89.3	85.5	88.3	79.6	86.6
YOLOv5s	90.8	14.3	119	87.5	94.1	89.8	93.6	81.9	89.9
SAB-YOLOv5s	92.5	15.8	116	89.4	95.5	90.3	94.0	82.4	90.9

图 6 YOLOv5s 算法检测结果图

图 7 SAB-YOLOv5s 算法检测结果图

3.5 对比实验

为了证明 SAB-YOLOv5s 模型在安全帽佩戴检测方面的优势, 我们将其与当前常用的一些目标探测方法作了对比, 如 YOLOv4, SSD300, Faster-RCNN 和原

始的 YOLOv5s 算法.

由表 3 可知, 与其他主流的目标检测方法相比, 本文提出的 SAB-YOLOv5s 网络模型其 *mAP* 值都有一定的提升, 达到了 92.5%. Faster-RCNN 在安全帽佩戴检测中精度比较高, 达到了 90.9%. 但是其计算量太大, 生成的模型远大于 YOLOv5s 目标检测算法模型.

YOLOv4 算法模型平均精度较低, 生成的模型也较大, 不适合实际部署. 本文提出的算法模型仅比原始 YOLOv5s 算法模型增加了 1.5M, *mAP* 却达到了 92.5%, 佩戴安全帽检测和未佩戴安全帽检测的平均精度都有所提升.

为了对模型的检测效率进行比较, 文章将不同算法的检测速度进行了单独的测试, 并且在相同的 GPU 上进行了试验. 结果表明, 经过改进后的 YOLOv5 算法的检测速度要比原 YOLOv5 算法慢一些, 但与 YOLOv4 相比提高了 15 FPS, 与 SSD300 相比提高了 27 FPS, 同时也证明了改进后的算法是有效的, 可以保证模型快速地对安全帽佩戴进行检测.

4 结论

针对高密度小像素物体在安全帽佩戴检测中出现的漏检和误检问题, 本文提出了一种改进 YOLOv5s 的算法. 在 YOLOv5s 的基础上, 首先增加了一个尺度为 160×160 的小目标检测层, 提高了小目标安全帽的检测精度, 然后引入 SA-16 注意力机制, 在此基础上加入了一种基于 SA-BiFPN 的新型拓扑结构, 实现了多维的双方向特征的融合, 从而提高了识别的性能. 改进后的算法模型较为轻量化, 能够集成到实际应用的安全检查框架中, 可广泛部署于生产环境中. 但是 SHWD 数据集对未佩戴安全帽的情形标注地还是不够全面, 有些头像识别不出来, 如何准确全面地检测出头像是下一步的研究方向.

参考文献

1 王亚光. 基于安全生产对建筑施工的管理预防. 现代商贸工业, 2022, 43(1): 186–188.

2 吕谋贵. 分析施工现场安全帽佩戴情况监控技术. 低碳世界, 2018, (7): 205–206.

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

4 Redmon J, Divvala S, Girshick R, *et al*. You only look once: Unified, real-time object detection. Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas: IEEE, 2016. 779–788.

5 Redmon J, Farhadi A. YOLO9000: Better, faster, stronger. Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition. Honolulu: IEEE, 2017. 7263–7271.

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

7 Bochkovskiy A, Wang CY, Liao HYM. YOLOv4: Optimal speed and accuracy of object detection. arXiv:2004.10934, 2020.

8 Lin TY, Goyal P, Girshick R, *et al*. Focal loss for dense object detection. Proceedings of the 2017 IEEE International Conference on Computer Vision. Venice: IEEE, 2017. 2980–2988.

9 Girshick R, Donahue J, Darrell T, *et al*. Rich feature hierarchies for accurate object detection and semantic segmentation. Proceedings of the 2014 IEEE Conference on Computer Vision and Pattern Recognition. Columbus: IEEE, 2014. 580–587.

10 Girshick R. Fast R-CNN. Proceedings of the 2015 IEEE International Conference on Computer Vision. Santiago: IEEE, 2015. 1440–1448.

11 Ren SQ, He KM, Girshick R, *et al*. Faster R-CNN: Towards real-time object detection with region proposal networks. Proceedings of the 28th International Conference on Neural Information Processing Systems. Montreal: MIT Press, 2015. 91–99.

12 Dai JF, Li Y, He KM, *et al*. R-FCN: Object detection via region-based fully convolutional networks. Proceedings of the 30th International Conference on Neural Information Processing Systems. Barcelona: Curran Associates Inc., 2016. 379–387.

13 李鹏. 基于目标检测与深度估计的施工现场安全预警关键技术研究与实现 [硕士学位论文]. 成都: 电子科技大学, 2021.

14 Song HR. Multi-scale safety helmet detection based on RSSE-YOLOv3. Sensors, 2022, 22(16): 6061. [doi: 10.3390/s22166061]

15 杨贞, 朱强强, 彭小宝, 等. 基于深度级联模型工业安全帽检测算法. 计算机与现代化, 2022, (1): 91–97, 119.

16 吕宗喆, 徐慧, 杨骁, 等. 面向小目标的 YOLOv5 安全帽检测算法. 计算机应用: 1–9. http://kns.cnki.net/kcms/detail/51.1307.TP.20220929.1425.005.html

17 李嘉信, 胡杨, 黄协舟, 等. 面向小目标的多空间层次安全帽检测. 计算机工程与应用: 1–9. http://kns.cnki.net/kcms/detail/11.2127.TP.20221129.1158.004.html

18 朱玉华, 杜金月, 刘洋, 等. 基于改进 Faster R-CNN 的小目标安全帽检测算法研究. 电子制作, 2022, 30(19): 64–66, 83.

19 杨晓玲, 江伟欣, 袁浩然. 基于 YOLOv5 的交通标志识别检测. 信息技术与信息化, 2021, (4): 28–30.

20 Zhang QL, Yang YB. SA-Net: Shuffle attention for deep convolutional neural networks. Proceedings of the 2021 ICASSP IEEE International Conference on Acoustics, Speech and Signal Processing. Toronto: IEEE, 2021. 2235–2239.

21 Woo S, Park J, Lee JY, *et al*. CBAM: Convolutional block attention module. Proceedings of the 15th European Conference on Computer Vision. Munich: Springer, 2018. 3–19.

22 Lin TY, Dollár P, Girshick R, *et al*. Feature pyramid networks for object detection. Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition. Honolulu: IEEE, 2017. 2117–2125.

23 Liu S, Qi L, Qin HF, *et al*. Path aggregation network for instance segmentation. Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City: IEEE, 2018. 8759–8768.

24 Tan MX, Pang RM, Le QV. EfficientDet: Scalable and efficient object detection. Proceedings of the 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Seattle: IEEE, 2020. 10781–10790.

25 马燕婷, 赵红东, 阎超, 等. 改进 YOLOv5 网络的带钢表面缺陷检测方法. 电子测量与仪器学报, 2022, 36(8): 150–157.

(校对责编: 牛欣悦)