doi:10. 16055/j. issn. 1672-058X. 2022. 0004. 005

基于 YOLO-CDF 神经网络的安全帽检测

张学锋, 王子琦, 汤亚玲
(安徽工业大学 计算机科学与技术学院, 安徽 马鞍山 243000)

摘 要:针对当前安全帽检测准确性低和适应性差的问题,提出一种以 YOLOv3 网络为基础,进行相应改进的安全帽检测方法;为了保证安全帽检测的准确度和增大对图片中安全帽的关注度,采用注意力机制增强了从图片提取出的空间信息和语义信息,减少了图像细节的丢失,再使用可变卷积来适应人的姿态变化,增强了模型对目标的适应性,减少了一定量的训练样本,最后通过改变输出特征图的尺寸,融合浅层的网络特征,提升了人头等小目标的识别率;采用自制的 HELMET 数据集对方法进行训练与测试,并通过对比实验表明:方法相较于其他检测方法能够提取到更多的目标特征,达到更高的平均精度均值,同时在实际应用中适应性较好。

关键词:安全帽检测; 注意力机制; 可变卷积;特征图
中图分类号:TP183 **文献标志码**:A **文章编号**:1672-058X(2022)04-0032-10

0 引 言

在化工厂或者建筑工地中往往有着种种令人难以防范的危险,所以工厂一般会制定严格的作业操作规范来对工人进行安全培训,其中最基本的就是进入工地戴好安全帽。但工人却往往会忽视掉这些危险,操作不规范,忘戴安全帽,这些都带来了很大的安全隐患。传统的方法是通过人工监控的方式对工人发出告警,但是这种方法难以起到作用,而深度学习目标检测由于其直观,检测速度快被越来越多的应用到安全帽检测中。双步目标检测算法如 R-CNN[1],Fast R-CNN[2],Faster R-CNN[3] 等先采用 Selective Search[4],RPN(Region Proposal Network)等方法找出固定尺寸的矩形框(Anchor),再利用深度学习网络对目标位置和类别进行识别,准确率虽然理想,但是检测时间长,难以部署在工程上。而单步目标检测算法,如 YOLO[5],YOLOv2[6],YOLOv3[7],SSD[8] 等算法都是通过网络直接进行回归产生目标的类别和位置信息,单步算法相比双步算法省略了一步,检测速度快,准确度相比双步目标检测算法要差一些,但是实现了端到端的目标检测,具有很高的实用价值。

利用目标检测网络进行安全帽检测研究已经成为热点,方明等[9] 在 YOLOv2 上引入密集连接网络[10] 和轻量化网络结构,减少了参数和计算量,易于部署,但是当背景颜色与安全帽颜色相近时,会出现漏检情况,徐守坤等[11] 在 Faster R-CNN 上运用多尺度训练和增加锚点数量的方式来实现对安全帽小目标检测的优化,但该方法所需时间长,难以部署

收稿日期:2021-01-19;**修回日期**:2021-03-10.
基金项目:安徽省教育厅重大课题基金项目(KJ2017ZD05);安徽高校协同创新项目(GXXT-2019-008); 结合实物机器人的化工企业救援仿真系统(TZJQR002-2021).
作者简介:张学锋(1978—),男,河北行唐人,教授,博士,从事计算机仿真研究.
通讯作者:王子琦(1996—),女,安徽宿州人,硕士生,从事计算机视觉研究. Email:1073910633@ qq. com.

在实际. 王兵等[12]使用 GIOU[13]计算方法,与 YOLOv3 的目标函数相结合,解决了评价指标与目标函数不一致的问题,提升了安全帽佩戴检测的准确率,但当目标出现一定的形状变化时,难以识别,适应性不强。

为了实现对工人安全帽佩戴情况的准确以及快速地检测,选择了在检测速度和准确率都有不错效果的 YOLOv3 算法作为基础网络,将自制的 HELMET 数据集中的安全帽和人作为训练和检测的目标,针对 YOLOv3 网络安全帽检测中人员姿态变化难以识别,以及人头等小目标检测能力不足的问题,对其结构进行改进。改进后的方法可以在提升图像信息关注度的同时对人员和安全帽形状变化检测有更好的准确度。将融合改进后的模型(YOLO-CDF)进行训练与测试,并与 YOLOv3,SSD300,以及 Fast R-CNN 在同一环境下进行对比实验,实验结果表明 YOLO-CDF 对安全帽和人的检测精度有明显的提升。

1　基础网络 YOLOv3

1.1　网络结构

YOLOv3 的网络如图 1 所示,其基本组成单元 DBL 是由卷积层(Conv)、批标准化层(BN)和激活层(ReLU)组成(图 2),2 个 DBL 加上残差操作形成残差单元(ResNet unit)(图 3),利用了 ResNet[14]的残差思想,增大了网络的深度,有利于解决深层次的网络梯度消失问题。再由残差单元和降采样操作形成 YOLOv3 的骨干网络即 DarkNet-53 网络(图 4),不再用池化来减小图像尺寸来增大感受野,因为这样会使得图像的一部分信息损失,而是采用步长为 2 的卷积来代替池化,可以提取更高级特征。

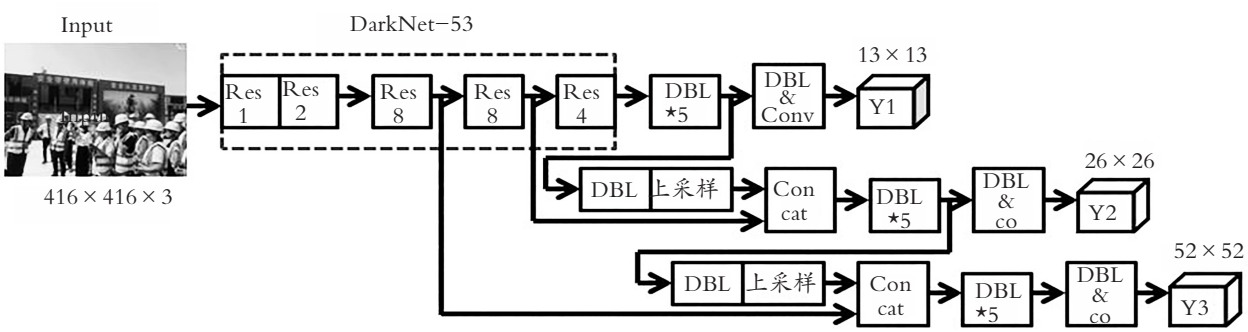


图 1　YOLOv3 网络结构

Fig. 1　Network structure of YOLOv3

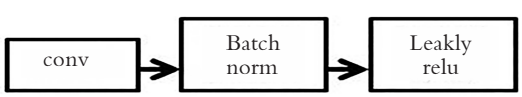


图 2　DBL 结构

Fig. 2　Structure of DBL

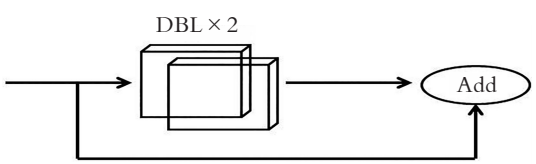


图 3　ResNet Unit 结构

Fig. 3　Structure of ResNet Unit

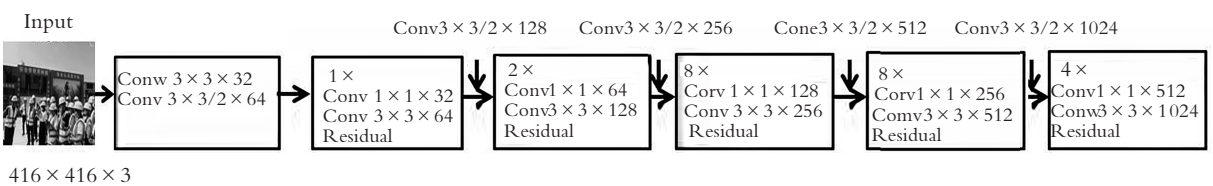


图 4　DarkNet-53 网络

Fig. 4　Network of DarkNet-53

1.2　特征金字塔多尺度融合

CNN(Convolutional Neural Networks)网络对于尺度变化的表现不敏感表现为当网络越深,特征图的每个单元格的感受野就越大,虽然语义信息更强,但是分辨率下降,位置信息模糊,对小目标的检测产

生影响。特征金字塔网络[15]里提到在网络较深处得到比较高级抽象的特征,不断进行上采样与浅层特征融合,这样融合后的特征图再用来预测不同大小规模的目标就会更准确。YOLOv3 借鉴这一思想,共进行了 5 次降采样,并在最后 3 次降采样进行目标预测,输出 3 个尺寸的特征图,分别是 13×13,26×26,52×52,尺寸越小,则对应的感受野越大,52×52 的特征图,感受野最小,可以获得更多的细节,适用于检测小目标;26×26 的特征图,则是用于检测中等大小的目标;13×13 的特征图尺寸最小,感受野最大,将图中的全局信息聚合在一起,适用于检测大目标。

1.3 边界框预测

YOLOv3 将图片划分成 $N\times N$ 个网格,在每个单元格上进行预测目标框的中心点坐标,长宽以及类别,采用了 K 均值聚类算法,3 个不同尺寸特征图的网格分别设置 3 个先验框,用 K-means 聚类出 9 种边界框大小。YOLOv3 输出预测框相对值的相关信息,需要结合偏移量计算,计算公式如式(1)—式(4):

$$b_x=\sigma(t_x)+c_x \tag{1}$$

$$b_y=\sigma(t_y)+c_y \tag{2}$$

$$b_w=p_w\times e^{t_w} \tag{3}$$

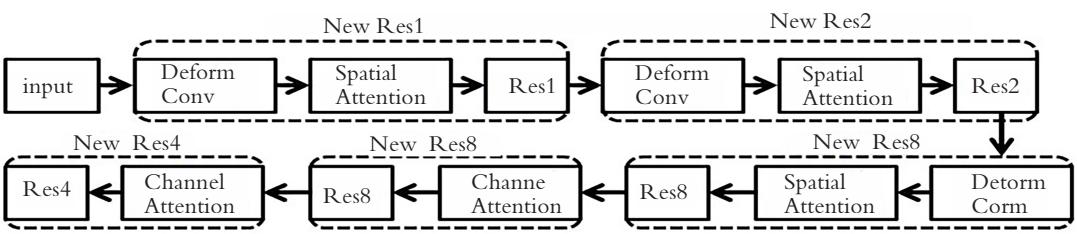
$$b_h=p_h\times e^{t_h} \tag{4}$$

其中 $\sigma(\cdot)$ 为 sigmoid 激活函数,(b_x,b_y,b_w,b_h) 为预测框(boundingbox)的中心点坐标和长宽,(t_xt_x,t_y,t_w,t_h) 为经过网络学习相对先验框(anchor)的偏移量,c_x,c_y 为单元格的左上点坐标,p_w,p_h 为先验框相对特征图的宽和高。

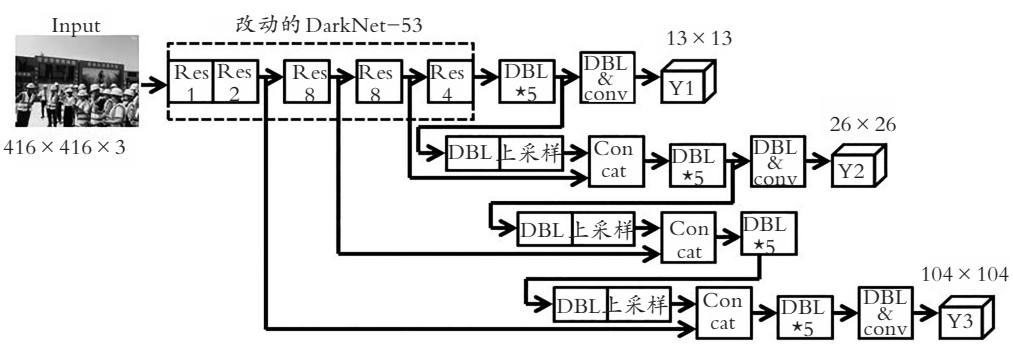
2 YOLOv3-CDF 框架

2.1 网络结构

YOLOv3-CDF 的整体框架如图 5 所示,其中 Deform conv 表示可变卷积模块,Spatial Attention 表示空间注意力模块,Channel Attention 表示通道注意力模块,Res 表示上文提到的残差单元,New Res 表示添加模块后的残差单元,如图 5(a)所示。经过本文利用可变卷积模块来减少目标因形状变化导致的误识别率,在较少的数据集上就有很好的效果,一定程度上减少工作量。而在安全帽检测之前的工作中,对图像细节部分在深层的网络中不能很好地保留,为了更好的提取特征以及保留背景的纹理细节,分别将空间注意力和通道注意力引入到网络的不同层中,以自适应细化特征。最后根据观察到所用数据集的目标里有很多的小目标从而对输出特征图在尺度上变化,来提高小目标的识别率。



(a) Darkenet-53 网络变动



(b) 多尺度特征图变动

图 5 YOLOv3-CDF 网络

Fig. 5 Network of YOLOv3-CDF

2.2　双注意力机制

注意力机制模拟人类视觉注意力,重点强调与周围变化大,令人关注的部分,在图像处理中就是将全局像素之间的相互依赖作为特征的加权,对主要特征进行重点关注,并抑制不必要的特征。SENet[16]专注于图片通道,全面获取通道间的相关性,主要是在每个通道上做了全局的池化处理,使在某种程度上具有全局的感受野,并进行了特征选择,完成了通道维度上的原始特征的重新标定。受到启发,Woo 提出 CBAM [17]模块,结合了空间注意力和通道注意力,考虑空间信息和通道信息的共同作用,相比仅使用一种注意力有更好的效果,且 CBAM 是一个轻量级的通用模块,可以无缝地集成到任何 CNN 架构中。

金字塔特征注意网络[18]里提到低级特征往往是点、线等边缘信息,而高层特征往往是语义信息。根据不同层次特征的特点,采用通道方式关注高层特征,空间注意力为低层特征需要选择有效特征。本文将 CBAM 的空间注意力模块放在网络浅层,增强特征的位置信息,将通道注意力模块放在网络深处,对抽象语义信息关注,如图 5 所示。

通道注意力和空间注意力模块的具体操作如图 6 所示。空间注意力模块需要清楚图片的哪些位置应该有更高的反馈,通过聚合平均池化和最大池化的特征图,送入到 7×7 的卷积核中卷积,按通道维度产生 2 维的空间特征图。而在通道注意力模块中,输入的特征图首先经过平均池化(AvgPool)和最大池化(MaxPool)的操作计算出特征,相比只进行单一池化,丢失的信息会减少。然后将特征送入共享的多层感知机模型(含有一个隐藏层)产生通道注意力图。通道和空间注意力都在平均池化和最大池化两个方面对特征进行提取聚合,进一步提高网络的表征能力。空间注意力计算如式(5),通道注意力计算如式(6):

$$M_s(F)=\sigma(f^{7\times7}(\text{Avgpool}(F);\text{Maxpool}(F)))=\sigma(f^{7\times7}([F^s_{\text{avg}};F^s_{\text{max}}])) \tag{5}$$

其中 $\sigma(\cdot)$ 为 sigmoid 激活函数,f 为卷积操作,[;]表示拼接操作,F^s_{avg} 为每个通道上的平均池化,F^s_{max} 为每个通道上的最大池化,7×7 是卷积核大小,F 为输入特征图。

$$M_C(F)=\sigma\begin{pmatrix}MLP(\text{Abgpool}(F))+\\MLP(\text{Maxpool}(F))\end{pmatrix}=\sigma(W_1(W_0(F^c_{\text{avg}}))+W_1(W_0(F^c_{\text{max}}))) \tag{6}$$

其中 $\sigma(\cdot)$ 为 sigmoid 激活函数,MLP 为两层的神经网络,权重为 $W_0\in R^{C/r\times C}$,$W_1\in R^{C\times C/r}$,且权重共享,r 为下降率,F^c_{avg} 为平均池化,F^c_{max} 为最大池化,F 为输入特征图。

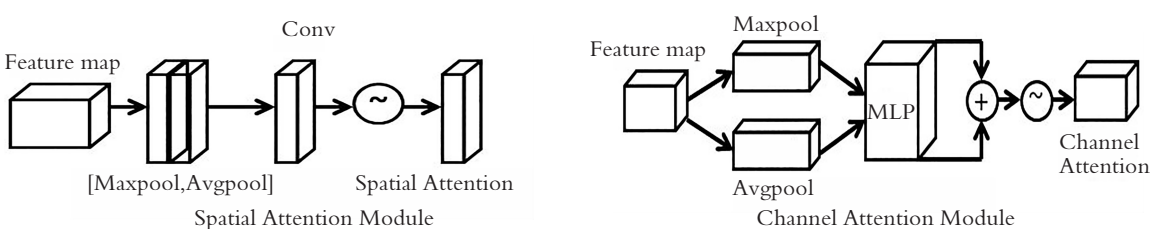


图 6　空间注意力模块和通道注意力模块

Fig. 6　Spatial attention module and channel attention module

2.3　可变卷积

对于安全帽检测来说,目标主要是人和人的头部,在大部分的图片里,人都是正常站立,比较容易检测到,但是进行作业的工人所呈现的身体状态是多种形式的,例如弯腰、蹲、坐等形态,并且由于摄像头的拍摄角度问题,人的身体在姿态、大小、和角度上变化多样,这使得网络难以辨认出目标物。

可变卷积网络[19]在神经网络中引入了学习目标空间几何形变的能力,相比于标准卷积的卷积核受限于固定的形状造成的采样能力有限,对于解决具有空间变化的目标识别任务更加有效,如图 7 所示。普通卷积是直接学习权重,而可变卷积是学习

了根据不同特征提取该特征所需要的相应偏移的能力,这样使得对物体的形状更加敏感,普通卷积计算如式(7),可变卷积计算如式(8):

$$\boldsymbol{y}(p_0)=\sum_{p_n\in\mathbb{R}}\boldsymbol{w}(p_n)\times\boldsymbol{x}(p_0+p_n) \tag{7}$$

$$\boldsymbol{y}(p_0)=\sum_{p_n\in\mathbb{R}}\boldsymbol{w}(p_n)\times\boldsymbol{x}(p_0+p_n+\Delta p_n) \tag{8}$$

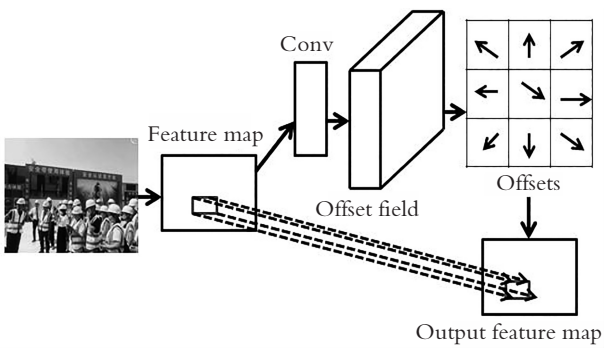


图 7 可变卷积结构图

Fig. 7 Structure of deformable convolution

其中:式(7)表达普通卷积,式(8)表达可变卷积。$\boldsymbol{y}$ 为输出的特征矩阵,$\boldsymbol{x}$ 为输入的特征矩阵,p_0 为其上的任意位置,$\boldsymbol{w}$ 为采样值的权重矩阵,$\boldsymbol{R}$ 为单元格,且 $\boldsymbol{R}$ 覆盖了整个输入特征矩阵,p_n 为 $\boldsymbol{R}$ 上的任意为位置,Δp_n 为偏移量,通常不是整数,应用双线性插值来确定偏移后采样点的值。

综上,可变卷积有更好地适应目标形变的能力,相对标准卷积,会多一部分计算开销,以便自适应地进行卷积。对于安全帽检测,需要精准的目标位置信息,所以本文考虑将可变卷积放在基础网络的浅层,如图 5(a)所示。

2.4 特征图多尺度变化

YOLOv3 使用了 3 个尺度的特征图融合,但是在安全帽检测中,人头在图像中的占比往往很小,属于小目标,而 YOLOv3 对网络浅层信息利用得不够充分,会导致人头检测效果欠佳。而添加新的尺度会增加模型的复杂度,造成训练和检测时间加长。经过考虑,决定将输出的 52×52 尺度的特征图再进行上采样与 DarkNet 里产生的 104×104 的特征图进行拼接。这样做可以找到早期特征映射中的细粒度特征,并获得更有意义的语义信息。其余操作与其他尺度的操作相同,从而形成 104×104 尺度的检测,如图 5(b)所示。

3 实验与分析

实验在自行标注的数据集 HELMET 上进行,其介绍在 3.2 节给出。为证明方法的有效性,选取双阶段检测算法 Faster R-CNN 和单阶段目标检测算法 YOLOv3 和 SSD300 进行对比实验。

3.1 实验环境与实验参数

实验环境配置如表 1 所示。对官方推荐的实验数值与自己的实验环境相结合,做出调整后,实验参数设置如表 2 所示。

表 1 实验环境配置

Table 1 Experimental environment configuration

名 称	相关配置
操作系统	Ubuntu 16.04
中央处理器/GHz	Intel(R) Xeon(R) CPU E5-2620v4 @2.10GHz
显卡 GPU	NVIDIA RTX2080Ti(12GB)
GPU 加速库	CUDA10.1
内存/GB	26
深度学习框架	Keras

表 2 实验参数设置

Table 2 Experimental parameter settings

参数名称	参数值	备 注
Learning_Rate	0.001	初始学习率
batch_size	6	批处理大小
classes	6	标签类别
epoch	600	迭代次数
score_threshold	0.6	框置信度阈值
iou_thredhold	0.5	同类别 IoU 阈值

3.2 实验数据集

选取网络中搜集到的安全帽以及工人的图片共 3 174 张作为实验数据,数据集包含各种颜色以及不同场景下的安全帽和人,且有个体差异,光照差异,视角变化以及不同程度的遮挡,数据信息丰富,其中测试集和验证集都是随机从数据集中抽取一定比例

进行划分。测试集比例为 0.3,验证集比例为训练集的 0.3 倍,包含 1 556 张训练集图片,666 张验证集图片,952 张图片用于测试。并对图片用 LabelImg 工具进行人工标注,标注示例如图 8 所示。

安全帽按颜色分为 4 个类,还有 person 类以及 none(没有戴安全帽的人头)类,将安全帽的特征细分,方便工程的部署,类别标签见表 3。

图 8　标注图片示例

Fig. 8　Example of annotation picture

表 3　标签类别

Table 3　Label categories

检测框位置	标签名	描　述
头部	blue	佩戴蓝色安全帽
	red	佩戴红色安全帽
	white	佩戴白色安全帽
	yellow	佩戴黄色安全帽
	none	未佩戴安全帽
人	person	人

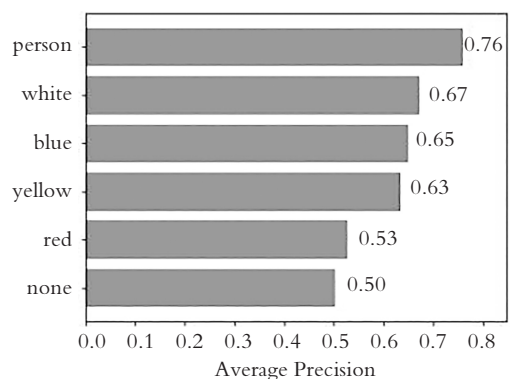
3.3　网络训练与结果分析

将单个模块加入的算法以及融合后的算法(YOLO-CDF)使用相同的训练数据文件,训练过程中每经过 100 个批次,将学习率乘以 0.2,加速模型收敛,且聚类文件相同,聚类后的先验框如表 4 所示,并将相关设置修改一致进行训练。使用相同的测试集进行测试,最后对实验结果进行对比分析,并与 YOLOv3,SSD300 以及 Fast R-CNN 等经典算法做比较,如图 9 所示。

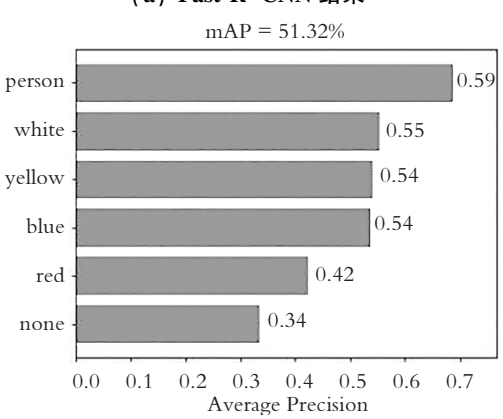
表 4　先验框聚类结果

Table 4　Priori-box clustering results

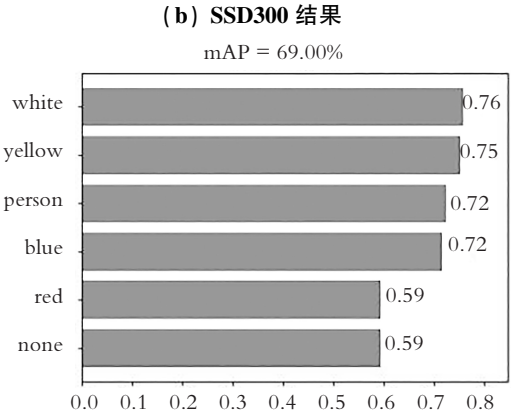
特征图尺寸/mm	尺　度	先验框
13×13	大	86×220
		162×307
		364×666
26×26	中	35×47
		49×72
		57×127
104×104	小	11×14
		18×23
		26×33



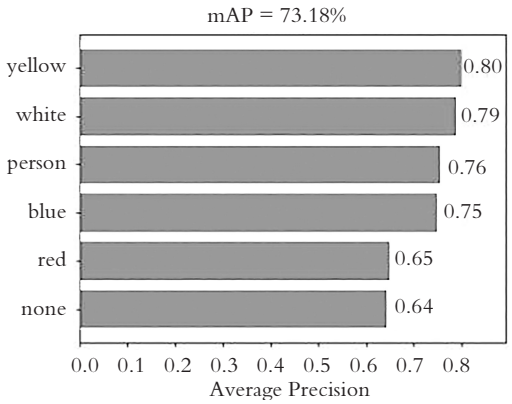
(a) Fast R-CNN 结果



(b) SSD300 结果



(c) YOLOv3 结果



(d) YOLO-CDF 结果

图 9 各模型测试每个目标类别的 AP 值以及 mAP 结果

Fig. 9 Each model testing the AP value of each target category and the mAP results

3.3.1 模型选择与评价标准

在验证集上损失最小的认为是最优模型,作为模型的代表进行测试,loss 的计算和 YOLOv3 的 loss 计算相同(式(9)),主要包括坐标误差(式(10)),置信度误差(式(11))和分类误差(式(12))。

$$\mathrm{Loss}=\mathrm{Loss}_{\mathrm{coord}}+\mathrm{Loss}_{\mathrm{IOU}}+\mathrm{Loss}_{\mathrm{class}} \tag{9}$$

$$Loss_{\mathrm{coord}}=\lambda_{\mathrm{coord}}\sum_{i=0}^{S^2}\sum_{j=0}^{K}\ell_{ij}^{\mathrm{obj}}\begin{bmatrix}(x_{ij}-\hat{x}_{ij})^2+\\(y_{ij}-\hat{y}_{ij})^2\end{bmatrix}+$$

$$\lambda_{\mathrm{coord}}\sum_{i=0}^{S^2}\sum_{j=0}^{K}\ell_{ij}^{\mathrm{obj}}[(w_{ij}-\hat{w}_{ij})^2+(h_{ij}-\hat{h}_{ij})^2] \tag{10}$$

$$Loss_{\mathrm{IOU}}=-\sum_{i=0}^{S^2}\sum_{j=0}^{K}\ell_{ij}^{\mathrm{obj}}\begin{bmatrix}\hat{c}_{ij}\log_e(c_{ij})\\+(1-\hat{c}_{ij})\\\log_e(1-c_{ij})\end{bmatrix}-$$

$$\lambda_{\mathrm{noobj}}\sum_{i=0}^{S^2}\sum_{j=0}^{K}\ell_{ij}^{\mathrm{noobj}}\begin{bmatrix}\hat{c}_{ij}\log_e(c_{ij})\\+(1-\hat{c}_{ij})\\\log_e(1-c_{ij})\end{bmatrix} \tag{11}$$

$$Loss_{classes}=-\sum_{i=0}^{S^2}\sum_{c\in classes}l_{ij}^{\mathrm{obj}}\begin{bmatrix}\hat{p}_{ij}\log_e(p_{ij})\\+(1-\hat{p}_{ij})\\\log_e(1-p_{ij})\end{bmatrix} \tag{12}$$

其中 $\mathrm{Loss}_{\mathrm{coord}}$ 为坐标误差,$\mathrm{Loss}_{\mathrm{IOU}}$ 为置信度误差,$\mathrm{Loss}_{\mathrm{classes}}$ 为分类误差,λ_{coord} 为坐标损失权重,λ_{noobj} 为置信度损失权重,S^2 为网格大小,K 为每个网格中候选框的个数,$\ell_{ij}^{\mathrm{nobj}}=1$ 表示单元格 j 中的边界框 i 里有目标对象出现,$l_{ij}^{\mathrm{noobj}}=1$ 则是单元格 j 中的边界框 i 里没有目标出现.(x,y,w,h,c,p) 为预测框横坐标、纵坐标、宽度、高度,置信度以及类别预测值,$(\hat{x},\hat{y},\hat{w},\hat{h},\hat{c},\hat{p})$ 为其真实框所对应的值。

测试使用类别平均精度(AP_{50})和各类别的平均精度均值(mAP_{50})作为评价指标。

3.3.2 测试结果

模块测试参数设置如表 5 所示。不同模块加入后进行测试的 mAP 如表 6 所示。表 6 中加粗字体为每列最优值。YOLO-C 表示仅添加 CBAM 模块的网络,YOLO-D 表示仅添加 Deform conv 模块的网络,YOLO-F 表示特征图变化后的网络。表 6 中可以看到融合所有模块方法的 YOLO-CDF 算法相比单个模块的加入的平均精度均值更高。

表 5 测试参数设置

Table 5 Test parameter settings

参数名称	参数值	备 注
score	0.2	得分阈值
iou	0.5	交叠框阈值
image_size	416×416	图片输入尺寸
gpu_num	1	gpu 数量

表 6 各模块加入后的 mAP

Table 6 mAP after each module is added

模块/模型	CBAM	Deform conv	特征图变化	mAP/%
YOLO-C	√	×	×	70.72
YOLO-D	×	√	×	71.68
YOLO-F	×	×	√	71.15
YOLO-CDF	√	√	√	**73.18**

为显示方法的有效性,经典算法 YOLOv3,SSD300,Fast R-CNN 测试参数采用官方推荐的数值,对同样的训练数据进行训练,得到训练模型后,对同一测试数据集行测试各目标的 AP 值以及 mAP,其结果如图 9,并整理成表 7。可以看到:改进的 YOLO-CDF 算法在测试集上的表现相比其他算法要更突出,mAP 相比较表现较好的 YOLOv3 算法还提高了 4.18%,这样的精度提升对于检测工地安全帽和工人的安全具有重要的意义。

表 7　AP 值以及 mAP 结果

Table 7　AP values and mAP results

模型	none	person	blue	red	white	yellow	mAP
Fast R-CNN	0.50	0.76	0.65	0.53	0.67	0.63	0.624 5
SSD300	0.34	0.69	0.54	0.42	0.55	0.54	0.513 2
YOLOv3	0.59	0.72	0.72	0.59	0.76	0.75	0.690 0
YOLO-CDF	**0.64**	**0.76**	**0.75**	**0.65**	**0.79**	**0.80**	**0.731 8**

注:加粗字体为每列最优值。

本文还对 YOLOv3 和 YOLO-CDF 的检测速度进行了测试对比,如表 8 所示。结果表明:虽然 YOLO-CDF 算法相比其基础网络 YOLOv3 的检测速度有些许降低,但是牺牲部分速度可以有更好的精准度,且这个检测速度依然可以满足实时检测的应用需要,安全帽实时检测的实用性完全可以得到保障。

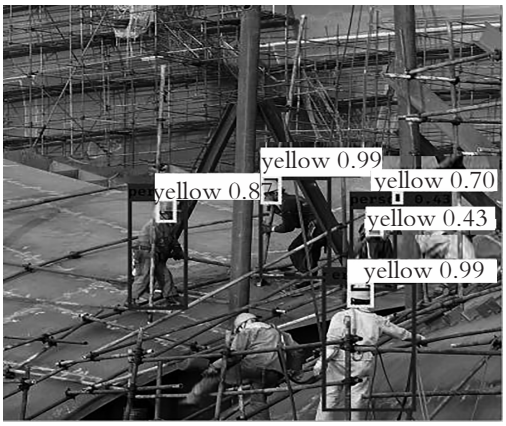
表 8　速度对比

Table 8　Speed comparison

模　型	FPS/(frame/s)
YOLOv3	55
YOLO-CDF	50

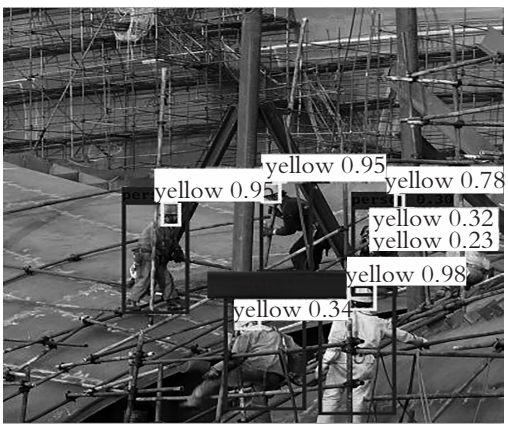
图 10 为 SSD300,Fast R-CNN 以及 YOLOv3 算法和本文提出的 YOLO-CDF 算法的检测效果对比图。展示的图像为工地某一场景,人员集中、姿态随意且有遮挡,其选自实验所用数据集的测试集。可以看出 YOLO-CDF 算法相比较于其他算法对于图中跨越栏杆的人及其头部识别更加精准,更加贴近目标轮廓,得分也较高;而其他算法有漏识别以及检测框贴近目标不到位的情况存在。证明提出的 YOLO-CDF 算法可以更加高效地检测到目标的特征信息,改善目前安全帽检测算法性能。

(a) 测试图片原图

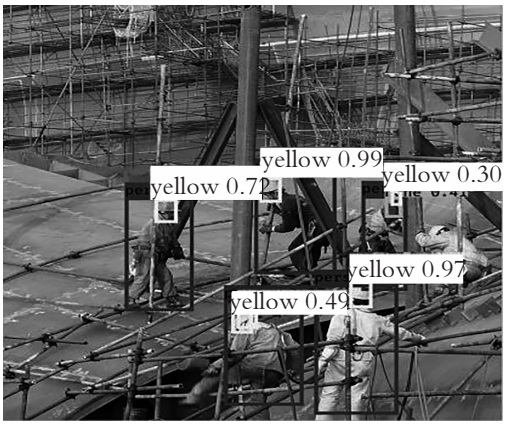


(b) SSD300 预测图

(c) Fast R-CNN 预测图



(d) YOLOv3 预测图



(e) YOLO-CDF 预测图

图 10 对比实验结果

Fig. 10 Results of comparative experiments

4 结 语

选取 YOLOv3 作为基础网络,针对安全帽检测进行了注意力机制和可变卷积的增加以及多尺度特征图的检测改进,使对安全帽的检测更加精准,并自行标注了安全帽数据集进行了对比实验。实验结果表明:进行改进后的算法 YOLO-CDF 在准确度上有更加良好的表现,在测试数据集上达到 73.18%的 mAP,高于其他目标检测网络算法,且检测速度也能够完成对安全帽进行实时检测的要求,能够在安全帽检测领域发挥出可靠性和实用价值。在之后的工作中,将进一步研究目标检测模型的精度与速度之间的关系,对 YOLOv3 网络的 loss 函数以及非极大值抑制部分进行改进,在增大准确度的基础上提升速度,同时增加样本数据多样性,并进一步优化缩小剪枝网络模型,减少训练时间。

参考文献(References):

[1] GIRSHICK R, DONAHUE J, DARRELL T, et al. Rich feature hierarchies for accurate object detection and semantic segmentation[C]//Proceedings of the 2014 IEEE Conference on Computer Vision and Pattern Recognition. Columbus: IEEE, 2014: 580—587.

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

[3] REN S, HE K, 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.

[4] UIJLINGS J, SANDE K E, GEVERS T, et al. Selective search for object recognition[J]. International Journal of Computer Vision, 2013, 104(2): 154—171.

[5] REDMON J, DIVVALA S K, GIRSHICK R, et al. You only look once: unified, real-time object detection[C]// Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas: IEEE, 2016: 779—788.

[6] REDMON J, FARHADI A. YOLO9000: better, faster, stronger[C]//Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition. Honolulu: IEEE, 2017: 6517—6525.

[7] REDMON J, FARHADI A. YOLOv3: an incremental improvement[EB/OL]. [2018-04-08]. arXiv: Computer Vision and Pattern Recognition, 2018. https://arxiv.org/abs/1804.02767.

[8] LIU W, ANGUELOV D, ERHAN D, et al. SSD: single shot multibox detector[C]//Proceedings of the 14th European Conference on Computer Vision Amsterdam. The Netherlands: Springer, 2016: 21—37.

[9] 方明,孙腾腾,邵桢. 基于改 YOLOv2 的快速安全帽佩戴情况检测[J]. 光学精密工程,2019,27(5):1196—1205.
FANG Ming, SUN Teng-teng, SHAO Zhen. Fast helmet wearing detection based on improved YOLOv2[J]. Optical Precision Engineering, 2019, 27(5): 1196—1205.

[10] HUANG G, LIU Z, LAURENS V, et al. Densely connected convolutional networks[C]//Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition. Honolulu: IEEE, 2017: 2261—2269.

[11] 徐守坤,王雅如,顾玉宛,等. 基于改进 Faster RCNN 的安全帽佩戴检测研究[J]. 计算机应用研究,2020, 37(3):901—905.
XU Shou-kun, WANG Ya-ru, GU Yu-wan, et al. Research on helmet wearing detection based on improved faster RCNN[J]. Computer Application Research,2020, 37(3):901—905.

[12] 王兵,李文璟,唐欢. 改进 YOLOv3 算法及其在安全帽检测中的应用[J]. 计算机工程与应用,2020,56(9): 33—40.
WANG Bin, LI Wen-jing, TANG Huan. Improved YOLOv3 algorithm and its application in helmet detection [J]. Computer Engineering and Application, 2020, 56(9): 33—40.

[13] REZATOFIGHI H, TSOI N, GWAK J, et al. Generalized intersection over union: a metric and a loss for bounding box regression[C]//Proceedings of the 2019 IEEE Conference on Computer Vision and Pattern Recognition Long Beach: IEEE, 2019: 658—666.

[14] HE K, ZHANG X, REN S, et al. Deep residual learning for image recognition[C]//Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas: IEEE, 2016: 770—778.

[15] LIN T, DOLLAR P, GIRSHICK R, et al. Feature pyramid networks for object detection[C]//Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition. Honolulu: IEEE, 2017: 936—944.

[16] HU J, SHEN L, ALBANIE S, et al. Squeeze-and-excitation networks[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2019(12): 1—7.

[17] WOO S, PARK J, LEE J Y, et al. CBAM: convo iutional block attention module[C]//Proceedings of the 2018 European Conference on Computer Vision (ECCV). Munich: Springer, 2018: 3—19.

[18] ZHAO T, WU X. Pyramid feature attention network for saliency detection[C]//Proceedings of the 2019 IEEE Conference on Computer Vision and Pattern Recognition. Los Angeles: IEEE, 2019: 3085—3094.

[19] DAI J, QI H, XIONG Y, et al. Deformable convolutional networks[C]//Proceedings of the 2017 IEEE Conference on International Conference on Computer Vision Venice. Italy: IEEE, 2017: 764—773.

Helmet Detection Based on YOLO-CDF Neural Network

ZHANG Xue-feng, WANG Zi-qi, TANG Ya-ling

(School of Computer Science and Technology, Anhui University of Technology, Anhui Maanshan 243000, China)

Abstract: In order to solve the problem of low accuracy and poor adaptability of helmet detection, the author proposed an improved helmet detection method based on YOLOv3 network. Aiming at ensuring the accuracy of helmet detection and increasing the attention to the helmet in the picture, this method used attention mechanism to enhance the spatial information and semantic information extracted from the picture, and reduced the loss of image details. And then the author used deformable convolution to adapt to the variety of human posture, enhance the adaptability of the model to the target, and reduce a few of training samples. In the end, the author changes the size of output feature map and the shallow network features are integrated to improve the recognition rate of small targets such as human head. The self-made HELMET dataset is used to train and test the method, and the comparative experiments show that compared with other detection methods, this method can extract more target features, achieve higher mean average accuracy, and has better adaptability in practical application.

Key words: safety helmet detection; attention mechanism; deformable convolution; feature map

责任编辑:田　静

引用本文/Cite this paper:

张学锋,王子琦,汤亚玲. 基于 YOLO-CDF 神经网络的安全帽检测[J]. 重庆工商大学学报(自然科学版),2022,39(4):32—41.

ZHANG Xue-feng, WANG Zi-qi, TANG Ya-ling. Helmet detection based on YOLO-CDF neural network[J]. Journal of Chongqing Technology and Business University (Natural Science Edition),2022,39(4):32—41.