

改进 YOLO v3 的安全帽佩戴检测方法

施 辉, 陈先桥, 杨 英

武汉理工大学 计算机科学与技术学院, 武汉 430063

摘 要:在生产和作业场地中,工人由于不佩戴安全帽而引发的安全事故时有发生。为了降低由于未佩戴安全帽而引发的安全事故发生率,提出了一种基于改进 YOLO v3 算法的安全帽佩戴检测方法。通过采用图像金字塔结构获取不同尺度的特征图,用于位置和类别预测;使用施工现场出入口监控视频作为数据集进行目标框维度聚类,确定目标框参数;在训练迭代过程中改变输入图像的尺寸,增加模型对尺度的适应性。理论分析和实验结果表明,在安全帽佩戴检测任务中,mAP(Mean Average Precision)达到了 92.13%,检测速率提高到 62 f/s,其检测准确率与检测速率相较于 YOLO v3 均略有提高,所提算法不仅满足安全帽佩戴检测中检测任务的实时性,同时具有较高的检测准确率。

关键词:图像处理;深度学习;YOLO v3;安全帽佩戴检测

文献标志码:A **中图分类号:**TP391.41 **doi:**10.3778/j.issn.1002-8331.1811-0389

施辉,陈先桥,杨英.改进 YOLO v3 的安全帽佩戴检测方法.计算机工程与应用,2019,55(11):213-220.

SHI Hui, CHEN Xianqiao, YANG Ying. Safety helmet wearing detection method of improved YOLO v3. Computer Engineering and Applications, 2019, 55(11):213-220.

Safety Helmet Wearing Detection Method of Improved YOLO v3

SHI Hui, CHEN Xianqiao, YANG Ying

College of Computer Science and Technology, Wuhan University of Technology, Wuhan 430063, China

Abstract: In production and operation sites, safety accidents caused by workers not wearing safety helmets occur from time to time. In order to reduce such incidence, a safety helmet wearing detection method based on improved YOLO v3 algorithm is proposed. Firstly, feature maps of different scales are acquired by using an image pyramid structure for position and category prediction. Secondly, the target frame parameters are clustered by using the construction site entrances and exits monitoring video as the data set. Finally, during the training process, the size of the input image is changed during the iterative process to increase the adaptability of the model to the scale. Theoretical analysis and experimental results show that in the safety helmet wearing detection task, the mAP(Mean Average Precision) reaches 92.13%, and the detection rate increases to 62 frame/s. The detection accuracy and detection rate are slightly improved compared with YOLO v3. It not only meets the real-time performance of the detection task in the safety helmet wearing test, but also has a high detection accuracy.

Key words: image processing; deep learning; YOLO v3; safety helmet wearing detection

1 引言

近年来,随着国家大力发展基础建设,全国建筑业安全生产理念也逐步进行推广。2018年上半年,国务院安委办通报了全国建筑业安全生产形势,通报显示,建筑业事故总量依旧不容乐观。上半年建筑行业生产安

全事故共 1 732 起,死亡 1 752 人,其中,由于施工现场违章违规引起事故频繁发生。在施工过程中,佩戴安全帽是一项基本的规章制度,但由于监管不力,工人不佩戴安全帽而引发的安全事故时有发生。为了改善这种情况,有必要对施工人员的安全帽佩戴进行检测,降低

基金项目:国家重点研发计划项目(No.2018YFC0810400)。

作者简介:施辉(1994—),男,硕士研究生,研究领域为深度学习,数字图像处理,E-mail:630235654@qq.com;陈先桥(1961—),男,博士,教授,研究领域为图形图像处理,计算机仿真;杨英(1993—),女,硕士研究生,研究领域为深度学习,数字图像处理。

收稿日期:2018-11-29 **修回日期:**2019-01-14 **文章编号:**1002-8331(2019)11-0213-08

CNKI 网络出版:2019-03-13, <http://kns.cnki.net/kcms/detail/11.2127.TP.20190311.1538.012.html>

由于作业人员没有佩戴安全帽而造成的事故发生率。

传统的目标检测需要通过手工设计特征来实现,该方法检测准确率低,不具备鲁棒性。近年来,深度学习凭借卷积神经网络(CNN)在提取图像特征时无需手工设计特征的优势^[1],渐渐获得学者青睐,相应地,很多研究人员提出了一系列基于深度学习的目标检测算法。Girshick等^[2]在2014年使用候选区域(region proposal)+CNN代替传统目标检测使用的滑动窗口+手工设计特征,设计了区域卷积神经网络(R-CNN),在VOC 2012数据集上,将目标检测的平均准确度(mAP)提升了30%,达到53.3%。Girshick^[3]和Ren等^[4]分别提出了快速区域卷积神经网络(Fast R-CNN)和超快区域卷积神经网络(Faster R-CNN),不仅提高了准确率,还增加了检测速度,帧速率可以达到5 f/s。2015年,Redmon J等^[5]提出了YOLO检测算法,该算法达到了可以检测视频的速度(45 f/s)。2016年,W Liu等人^[6]提出了SSD(Single Shot MultiBox Detector)检测算法,该算法在检测精度和检测时间上均取得了良好的效果,与此同时,在YOLO基础上,Redmon又相继提出了YOLO v2^[7]和YOLO v3^[8]检测算法,其中YOLO v3的检测效果更好,在COCO数据集上实现了在51 ms时间内mAP达到57.9%的效果,与RetinaNet在198 ms内mAP达到57.5%效果相当,性能相似但速度快3.8倍,由此可见,YOLO v3在目标检测领域,能够同时保证准确率和检测速率,取得较好的检测效果。

由于深度学习在目标检测领域取得了快速发展,许多学者开始将深度学习技术用于实际应用场景中,并取得了很好的效果。姚群力等^[9]采用深度卷积神经网络对目标检测进行了相关研究。谢林江等^[10]将卷积神经网络用于行人检测领域,提高了行人检测准确率。彭清等^[11]采用卷积神经网络提取特征,将其用于车辆识别领域,实现了对车辆的快速识别。由此可见,深度学习方法目前受到了越来越多的关注,将其与实际应用场景相结合,是目前的一大热门研究方向。

安全帽佩戴检测也是一类目标检测问题。目前已经有一些学者对安全帽的自动识别技术进行了相关研究,冯国臣等^[12]采用机器视觉的相关方法对安全帽自动识别进行深入研究。胡恬^[13]在着重分析小波变换和BP神经网络在安全帽识别的基础上,提出了安全帽识别神经网络模型。刘晓慧等^[14]采用肤色检测的方法定位人脸,再利用支持向量机(SVM)实现安全帽的识别。现有方法主要采用传统的目标检测方法,大多存在准确率低,对环境要求较高等问题,同时检测速度较慢,不能满足生产环境中的实时检测要求。

本文使用施工现场监控视频数据,对其进行相应的数据处理,制作了安全帽佩戴检测数据集。基于YOLO v3模型为主体,首先采用图像金字塔结构,将不

同层次特征进行融合,获取不同尺度的特征图,用于位置和类别预测;然后对目标框维度进行聚类,增加先验框(anchor box)个数,使得模型能够获取更多的物体边缘信息;最后,在训练过程中,采用多尺寸图片进行训练,使得模型能够适应不同分辨率的图片。实验表明,改进的YOLO v3算法能够在保证检测精度的同时,提高检测速率。

2 YOLO v3原理

2.1 特征提取网络Darknet-53

YOLO v3使用新的网络来实现特征提取。从网络结构上看,相比YOLO v2中的Darknet-19网络,添加了残差单元,使用连续的 3×3 和 1×1 卷积层,YOLO v3将其扩充为53层并称该网络为Darknet-53^[8]。该网络结构包含53个卷积层以及5个最大池化层,同时,在每一个卷积层后增加了批量归一化操作(batch normalization)和去除dropout操作,防止出现过拟合现象,该网络比Darknet-19功能强大,同时比ResNet-101或ResNet-152^[15]更有效,在ImageNet^[16]数据集下的性能测试结果如表1所示。

表1 特征网络性能对比表

Backbone	Top-1/%	Top-5/%	Bn	Ops/ 10^9	BFLOP/s	识别帧率/ (f·s ⁻¹)
Darknet-19	74.1	91.8		7.29	1 246	171
ResNet-101	77.1	93.7		19.70	1 039	53
ResNet-152	77.6	93.8		29.40	1 090	37
Darknet-53	77.2	93.8		18.70	1 457	78

由表1可以看出,Darknet-53比ResNet-101检测效果更好,且速度提高了1.5倍,Darknet-53比ResNet-152具有相似的性能,速度提高了2倍,与Darknet-19相比,检测准确率有较大提升,虽然速度不及Darknet-19,但仍然满足实时性要求。

2.2 YOLO v3设计思想

YOLO v3采用多个尺度融合的方式来进行预测。使用类似FPN^[17](Feature Pyramid Networks)的融合做法,在多个尺度的特征图(feature map)上进行位置和类别预测,提高了目标检测准确率。

在YOLO v3中,同样使用了维度聚类(dimension clusters)作为先验框来预测边界框。通过 k -means^[18]方法对数据集中的目标框进行维度聚类,得到9个不同大小的先验框,并将其均分到多个尺度的特征图上,尺度更大的特征图使用更小的先验框,该方案使得最终获得的先验框数量比YOLO v2多,特征提取效果更好。

在类别预测中,YOLO v3不使用softmax^[19]对每个框分类,而是采用多个独立的逻辑(logistic)分类器,在训练过程中,使用二元交叉熵损失(binary cross-entropy loss)来进行类别预测。

基于以上设计思想,YOLO v3与其他目标检测算法相比,在检测准确率和速度上,均取得了较好的效果,YOLO v3网络与其他目标检测框架性能对比如表2所示。

表2 YOLO v3与其他网络性能对比表

Method	mAP-50/%	time/ms
SSD321	45.4	61
DSSD321	46.1	85
R-FCN	51.9	85
SSD513	50.4	125
DSSD513	53.3	156
FPN FRCN	59.1	172
RetinaNet-50-500	50.9	73
RetinaNet-101-500	53.1	90
RetinaNet-50-800	57.5	198
YOLO v3-320	51.5	22
YOLO v3-416	55.3	29
YOLO v3-608	57.9	51

3 改进方法

YOLO v3在目标检测领域已经取得了很好的检测效果,但针对YOLO v3所做的物体检测均在常见数据集上实现,而对于特定的安全帽佩戴数据集,需要对YOLO v3进行一些改进,以适应于特定的安全帽佩戴检测任务,其改进方法如图1所示。

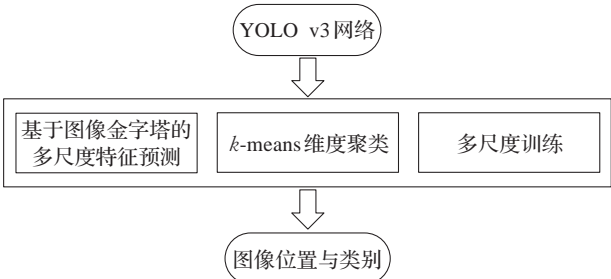


图1 改进方法示意图

基于图像金字塔结构,改进YOLO v3算法模型,将不同层次的特征图进行融合连接,得到3组预测特征图,并在这3组预测特征图上进行位置和类别的预测。

在自制的安全帽佩戴数据集上使用k-means算法,确定先验框维度。YOLO v3算法中得到的先验框维度是在COCO数据集上训练得到,且将其参数均分到三种不同尺度上,而对于安全帽佩戴检测这一具体任务,需要针对其特定数据集进行聚类操作,得到相应的聚类中心。

由于YOLO v3中去除了全连接层,因此在训练过程中,可以随时改变输入尺寸,使得训练后的模型能适应不同尺度的输入图像,其中,输入图像为自制的安全帽佩戴数据集。

3.1 基于图像金字塔的多尺度特征检测

YOLO v3中使用Darknet-53网络来提取特征,利

用特征可视化技术,能够清楚地看到各个层级的特征效果。

图2选取了Darknet-53中的conv1_1层以及conv53_1层的特征图形进行分析,可以看出,低层级的特征具有丰富的细节和定位信息,较高层级的特征具有丰富的语义特征,从低层级到高层级,细节信息不断减少,而语义信息不断增加,对于位置预测而言,需要更多的低层特征信息,对于类别预测而言,需要更多高层的局部信息,因此,可以考虑基于图像金字塔的模式,使用上采样将高层特征与低层特征进行融合,获取不同尺度的特征图,用于位置和类别预测。

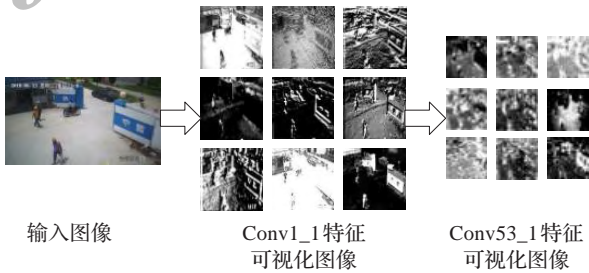


图2 Darknet-53特征可视化

图3中右边的特征金字塔是由左边的特征金字塔生成的,整个过程为:首先对输入图像进行深度卷积,然后对layer2上面的特征进行卷积操作,对layer4中的特征进行上采样操作,使其具有相同的尺寸,然后对处理后的layer2和layer4执行卷积和操作,将获得的结果输入到layer5层中。同理,相应地对多个层之间进行特征融合,得到用于预测的多组特征图。基于此方案,将处理过的低层特征和高层特征进行累加,这样做的目的是,由于低层特征可以提供更加准确的位置信息,而多次下采样和上采样操作使得深层网络的定位信息存在误差,因此将其结合起来使用,构建了一个更深的特征金字塔,融合了多层特征信息,并在不同的特征图上进行预测。

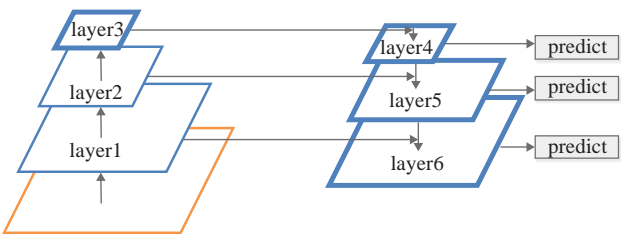


图3 特征融合金字塔

借鉴上述特征融合思路,对YOLO v3算法进行改进,使用上采样将高层特征与低层特征进行融合,最终得到3组特征图,并利用这3组不同尺度的特征图进行预测。改进的网络结构如图4所示。

具体网络结构改进细节:首先通过Darknet-53得到特征金字塔,将conv53层进行连续的1×1和3×3卷积操作,得到一组待处理yolo层,然后将该层进行一组1×1

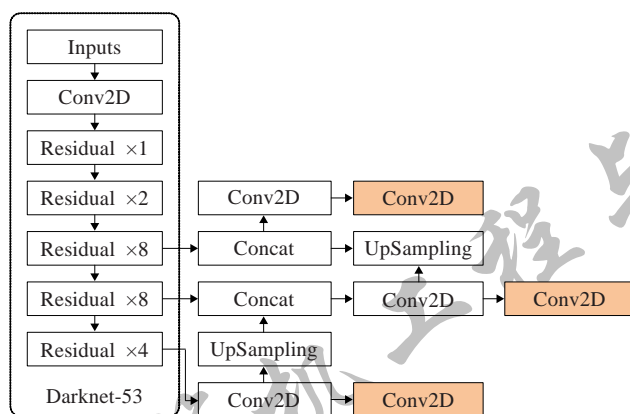


图4 改进的网络结构图

和 3×3 卷积操作,得到小尺度yolo层;同时,将该层进行上采样,与Darknet-53中的conv45层进行卷积和操作,同样使用连续 1×1 和 3×3 卷积操作,得到第二组待处理yolo层,将该层进行一组 1×1 和 3×3 卷积操作,得到中尺度yolo层;同时,将该层进行上采样,与Darknet-53中的conv29层进行卷积和操作,同样使用连续 1×1 和 3×3 卷积操作,得到第三组待处理yolo层,将该层进行一组 1×1 和 3×3 卷积操作,得到大尺度yolo层。经过上述操作,得到3组不同尺度的yolo特征层,并使用这3组特征层进行位置和类别预测。

3.2 k -means 维度聚类算法

YOLO v3借鉴了YOLO v2中使用先验框来预测边界框坐标的方法,不同的是,YOLO v3使用 k -means算法预测得到9个先验框,并将其均分到3个尺度特征图,其中,尺度更大的特征图使用更小的先验框,以此来获得更多的目标边缘信息。

YOLO v3中计算得到的9组先验框维度分别为 $(10, 13), (16, 30), (33, 23), (30, 61), (62, 45), (59, 119), (116, 90), (156, 198), (373, 326)$,但是在实际安全帽佩戴检测任务中,YOLO v3算法中计算得到的先验框维度不适合安全帽佩戴检测场景。待检测目标框尺寸主要分布如图5所示,因此使用YOLO v3算法中原有的先验框维度很难得到精确的目标框信息。



图5 待检测目标框的尺寸

因此,在安全帽佩戴检测场景下,使用 k -means算法对自制的安全帽佩戴数据集进行聚类分析,得到9组

先验框维度中心,分别为: $(59, 22), (68, 30), (75, 35), (88, 37), (90, 40), (99, 45), (108, 54), (119, 57), (128, 63)$,将其面积从小到大排列,均分到3个不同尺度的特征图上,其中,尺度更大的特征图使用更小的先验框。最后,将用此聚类中心进行安全帽佩戴检测实验。

3.3 多尺度训练

在YOLO检测算法中采用卷积网络来提取特征,然后使用全连接层来得到预测值。但由于全连接层的存在,网络在训练过程中输入图片的尺寸必须固定,因此最终训练得到的网络对不同尺寸的测试图片不具备鲁棒性。

针对自制的安全帽数据集,其输入图像具有不同大小的尺寸,因此,为了增强模型对图像尺寸的鲁棒性,可以采取多尺度训练的策略。具体来说,在YOLO v3网络中移除全连接层,改为全卷积操作。图6显示了将全连接层转化为卷积层的过程。

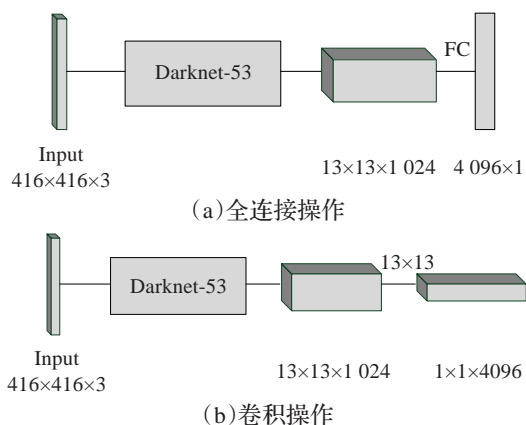


图6 全连接层转化为卷积层

图6(a)使用全连接层后进行预测,图6(b)将全连接层转化为卷积层进行预测。当输入图片大小为 416×416 ,经过Darknet-53网络后,输出 $13\times 13\times 1\ 024$ 的特征图,图6(a)通过一个包含4 096个神经元的全连接层后,得到了一组 $4\ 096\times 1$ 的特征图;图6(b)则使用4 096个 13×13 的卷积核,最终得到 $1\times 1\times 4\ 096$ 的特征图,其实质等价于全连接得到的4 096个神经元。对于上述两种网络结构,图片输入尺寸为 416×416 时,网络均能正常运行,但安全帽自制数据集中还存在其他尺寸的图片,例如当输入 608×608 大小的图片时,经过Darknet-53网络后,输出 $19\times 19\times 1\ 024$ 的特征图,对于图6(a)结构,接下来需要将该特征图与4 096个神经元进行全连接,由于原架构尺寸大小是 13×13 ,现在是 19×19 ,因此网络在传播过程中,无法使用之前的参数矩阵,网络不能正常运行。而对于图6(b)结构,全连接改为卷积操作后,网络能够继续运行,最终得到 $7\times 7\times 4\ 096$ 的正确输出。因此,将全连接层改为全卷积操作后,用不同尺寸的输入图片进行训练,改进算法能够适应不同尺寸的测试图片。

除此之外,将全连接层改为全卷积操作后,对于网络性能也有较好的提升。当使用图6(a)中的全连接层,网络进行前向传播运算时,经过计算,需要708 837 377个参数,而使用图6(b)中的全卷积操作,只需要696 320个参数,由此可见,将全连接层转化为卷积操作后,能够减少参数个数,降低网络运算量,提升网络性能。

基于上述分析,本文将采用多尺度训练策略对自制的安全帽数据集进行训练。由于整个网络具有5个最大池化层,因此网络下采样倍率为32,在训练过程中,将安全帽佩戴数据集训练图片输入大小分为一系列32倍数的值,其尺寸计算公式为:

$$S_{n+1}=32+S_n, n\leq 9 \tag{1}$$

其中 S_n 为第 n 组输入图像的尺寸。在网络初始化过程中, S_1 为320×320。

通过式(1)可得出输入图片尺寸大小为:{320, 352, ..., 608},在训练过程中,每隔10轮随机选择一种输入图片大小,以达到模型可以适应不同尺寸图片的效果。多尺度训练检测过程如图7所示。

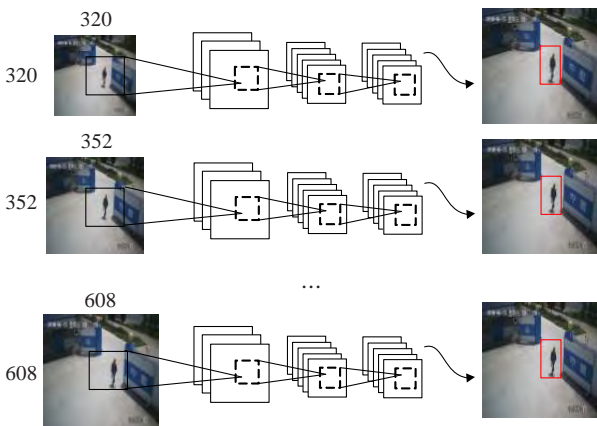


图7 多尺度训练图

4 实验结果及对比分析

4.1 实验数据集制作

对深度学习检测任务而言,实验数据集是基础条件。而针对安全帽佩戴检测一直没有公开数据集可供使用,因此,本文自制了一套安全帽佩戴大规模数据集,在建设该数据集的过程中,主要包含四个方面:数据收集、数据预处理、数据筛选以及数据标记。

4.1.1 数据收集

安全帽佩戴数据集的数据来源主要分为两部分:施工现场视频监控数据和互联网爬虫数据。其中施工现场视频监控数据为主要数据来源,互联网爬虫数据为辅助数据来源。

本文主要使用中建三局二公司麻城医院项目部施工现场视频监控数据,通过调取工地大门摄像头视频数据,获取不同时间段内的工地视频监控数据。

由于视频监控数据主要来源于特有的工地监控数据,数据具有一定的规律性。为避免这种问题,同时保证数据集的丰富性,本文利用互联网搜索引擎对安全帽佩戴数据进行爬取,作为辅助数据。

4.1.2 数据预处理

数据预处理过程主要包含两部分:数据格式转换以及归一化操作。

由于工地视频监控数据为视频格式文件,首先需要使用OpenCV开发库将视频格式文件转化为图像格式文件,作为初筛数据。

从工地视频中获取到的图片数据和从网络中爬取到的数据具有不同的分辨率,因此需要将图片随机格式化为式(1)中所计算得到的图片尺寸。

4.1.3 数据筛选

从视频监控录像分帧得到的图片数据,很多图片中没有行人以及安全帽等主体,全部为背景图片,该部分图片对于本文研究没有实际意义,故需要将该部分图片数据剔除,本文使用Faster R-CNN算法进行数据筛选,挑选出有行人的图片作为标注数据集。

4.1.4 数据标记

本文在安全帽佩戴检测数据库建设过程中,采用labelImg工具对图像进行标记,利用该标记工具,在人工操作的时候,只需要在图片中标记出自定义的各种目标,该工具可自动生成相应的配置文件。其中labelImg操作界面如图8所示。

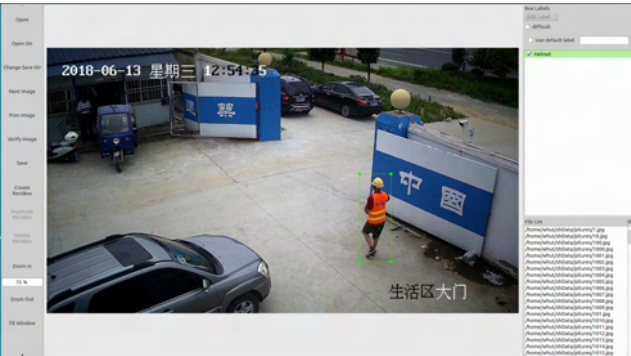


图8 labelImg 操作界面

4.2 实验方案

在实验过程中,将数据集分为两类:一类是网络训练时需要的检测数据,由于需要使用多尺度训练方法,将该部分数据集分为10组,即10种不同分辨率,每组图片分辨率由公式(1)计算所得。通过旋转、加入噪声等方法来扩大样本数,最终选取26 000张图片作为检测训练数据集样本,不同分辨率的样本数量均匀分配,以保证在训练过程中,适应不同尺度的图像数据;另一类是测试数据集,使用1 000张图片作为测试集,测试集样本数据随机归一化为任意分辨率。在训练集中,正样本分为A、B、C三类。其中,A类是佩戴安全帽的作业人员刚

进入摄像头区域;B类是佩戴安全帽的作业人员全部进入摄像头区域;C类是佩戴安全帽的作业人员即将离开摄像头区域。负样本分为D、E、F、G四类。D为将安全帽拿在手上未佩戴的作业人员;E为没有携带安全帽的作业人员;F为夏天时佩戴草帽的作业人员;G为冬天时佩戴棉帽的作业人员。即数据集总共分为7类,如图9所示。



图9 数据集样本类别

4.3 实验平台及网络训练

4.3.1 实验平台

本文在实验过程中对硬件环境要求较高,需要使用GPU进行运算,表3为本实验硬件环境配置说明。同时,本文在服务器上搭建了用于进行实验的环境,包括Ubuntu16.04、CUDA、python、OpenCV、gcc等常用环境,关于YOLO v3和CNN,分别使用了Darknet-53框架和Caffe框架。

表3 实验硬件环境配置说明

品名	型号	数量
CPU	INTEL I7 8700K	1
主板	技嘉Z370 AORUS GAMING5	1
内存	海盜船 16 GB/3 200	2
显卡	技嘉GTX1080Ti gaming OC 11G	4
固态硬盘	三星 256 GB	1
硬盘	西数 1 TB	2

4.3.2 网络训练

使用YOLO v3官网上提供的权重参数作为网络训练的初始化参数,随机使用自制安全帽佩戴检测训练数据集中的图片进行网络参数微调(fineturne),使得整个网络检测效果达到最优,部分实验参数设置如表4所示。

表4 网络参数说明表

参数名	参数值
学习率(learning rate)	0.002
迭代(epoch)	80
批量大小(batch size)	32
动量(momentum)	0.9
权重衰减(weight_decay)	0.000 5
学习率衰减步长(lr_step)	40
学习率衰减因子(lr_factor)	0.1
非极大值抑制(nms)	0.5

4.4 安全帽佩戴检测模型性能分析

4.4.1 小目标场景下的性能测试

3.1节提出的基于图像金字塔的多尺度检测方法,对于小目标的检测效果有较大提升。为了验证其对小目标检测性能的检测效果,本文将测试数据集按照目标尺寸大小进行排序,分别将其目标尺寸中0~10%,10%~30%,30%~70%,70%~90%,90~100%的目标划分为XS、S、M、L、XL 5个子类,分别代表了不同目标的尺寸。图10中列出了原始YOLO v3算法和改进算法对不同尺寸目标的检测性能效果。

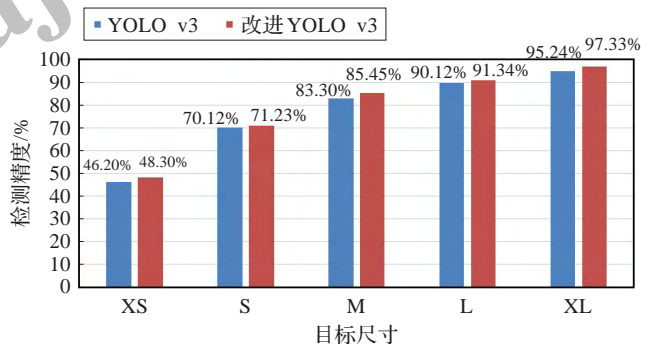


图10 两种算法的目标敏感性分析

由图10可以看出,改进YOLO v3算法对不同目标尺寸的检测准确率均比原始YOLO v3算法更高,因此,使用多特征融合的策略,在对安全帽佩戴进行检测时,结合高层特征与低层特征,使得整个算法在对小目标的检测上取得了较好的效果。

4.4.2 不同分辨率图像的性能测试

3.3节提出的多尺度训练方法,可以增强模型对于不同分辨率图像检测的鲁棒性。为了验证模型对不同分辨率输入图像的检测效果,本文在测试过程中,将测试集图片均分为3种不同分辨率大小,即{320, 608, 1024},分别代表了低分辨率、中分辨率、高分辨率三类图像,图11代表了两种算法针对不同分辨率图片的检测准确率性能效果。

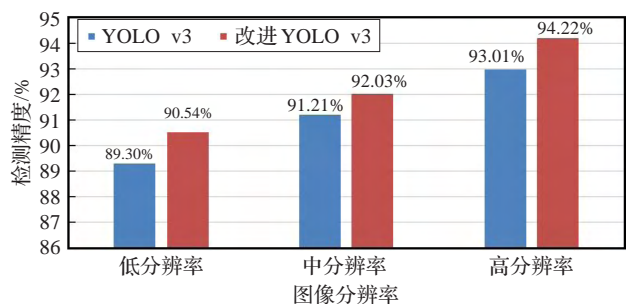


图11 两种算法的目标图像分辨率敏感性分析

由图11可以看出,改进YOLO v3算法对不同分辨率图像的检测准确率均比原始YOLO v3算法高,由此可见,本文采用多尺度训练检测策略,能够增强模型对不同分辨率图像的适应性。

4.4.3 安全帽佩戴检测模型对比实验

本文主要采用改进型的YOLO v3算法进行实验,改进点有:基于图像金字塔的多尺度特征检测、目标框维度聚类以及多尺度训练。同时,使用文献[6]中的SSD、文献[4]中的Faster R-CNN以及文献[8]中的YOLO v3等算法进行对比,以mAP以及每秒识别帧数作为检测效果评价指标,结果如表5所示。

表5 实验结果对比表		
算法	mAP/%	识别帧率/(f·s ⁻¹)
SSD	84.87	41.0
Faster R-CNN	94.72	0.4
YOLO v3	90.84	56.0
改进型YOLO v3	92.13	62.0

4.5 结果分析

从实验结果可以看出,Faster R-CNN的mAP值最高,达到了94.72%,改进的YOLO v3算法mAP值为92.13%,虽然检测准确率不及Faster R-CNN,但检测速率比其快155倍,因此,对于实时检测任务而言,改进的YOLO v3表现更好。而SSD算法以及原始的YOLO v3算法在mAP以及识别帧率两项指标上都不及改进型YOLO v3算法,由此可见,改进型YOLO v3算法同时兼顾了检测准确率以及检测速率,能够较好地完成安全帽佩戴检测任务。

另外,为了更加直观地感受不同算法之间的检测区别,本文选取了一些检测图像进行对比分析,其中图12为SSD算法检测效果图,图13为Faster R-CNN检测效果图,图14为YOLO v3检测效果图,图15为改进YOLO v3检测效果图。



图12 SSD检测效果图

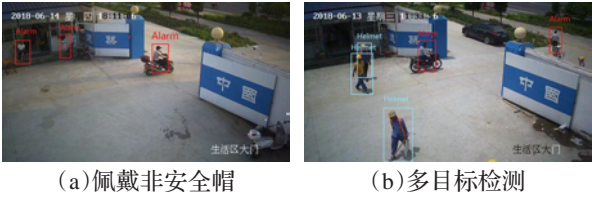


图13 Faster R-CNN检测效果图



图14 YOLO v3检测效果图

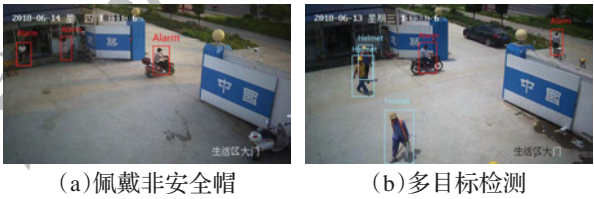


图15 改进YOLO v3检测效果图

在图12~图15中,图(a)为目标未佩戴帽子或者佩戴其他帽子的情况,可以看出,Faster R-CNN、YOLO v3以及改进YOLO v3算法均能够正确检测出目标所属类别,但SSD将目标佩戴草帽的情况误检为佩戴安全帽。图(b)为图像中出现多目标情况,且存在小目标的场景,可以看出,SSD算法以及YOLO v3算法对远处的小目标出现了漏检情况,而Faster R-CNN和改进YOLO v3算法对多目标检测表现较好。综合可见,在目标准确率上,Faster R-CNN以及改进YOLO v3均优于其他两种算法,但由于Faster R-CNN在目标检测过程中,需要构建RPN网络,涉及大量计算,因此在检测速度上不及改进YOLO v3,因此,对于安全帽佩戴检测任务,本文提出的改进YOLO v3算法能够在保持较高检测速率的同时,满足实时检测要求。

5 结束语

本文提出了一种基于改进YOLO v3算法的安全帽佩戴检测方法。使用施工现场出入口监控视频作为数据集进行安全帽佩戴检测实验,通过采用基于图像金字塔的多尺度特征检测、目标框维度聚类以及多尺度训练等方法改进YOLO v3网络,在保证较高检测准确率的同时仍然能够有较快的检测速度,基本满足作业环境监控视频中安全帽佩戴检测的准确性和实时性需求。

参考文献:

[1] Le Cun Y,Bengio Y,Hinton G.Deep learning[J].Nature, 2015,521(7553):436-444.

[2] Girshick R,Donahue J,Darrell T,et al.Rich feature hierarchies for accurate object detection and semantic segmentation[C]//IEEE Conference on Computer Vision and Pattern Recognition,2014:580-587.

[3] Girshick R.Fast R-CNN[C]//IEEE International Conference on Computer Vision,2015:1440-1448.

[4] Ren S,He K,Girshick R,et al.Faster R-CNN:towards real-time object detection with region proposal networks[C]// International Conference on Neural Information Processing Systems,2015:91-99.

[5] 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 Recognition,2015:779-788.

- [6] Liu W, Anguelov D, Erhan D, et al. SSD: single shot multibox detector[C]//European Conference on Computer Vision, 2016: 21-37.
- [7] Redmon J, Farhadi A. YOLO9000: better, faster, stronger[C]//IEEE Conference on Computer Vision and Pattern Recognition, 2017: 6517-6525.
- [8] Redmon J, Farhadi A. YOLOv3: an incremental improvement[J]. IEEE Conference on Computer Vision and Pattern Recognition, 2018: 89-95.
- [9] 姚群力, 胡显, 雷宏. 深度卷积神经网络在目标检测中的研究进展[J]. 计算机工程与应用, 2018, 54(17): 1-9.
- [10] 谢林江, 季桂树, 彭清, 罗恩韬. 改进的卷积神经网络在行人检测中的应用[J]. 计算机科学与探索, 2018, 12(5): 708-718.
- [11] 彭清, 季桂树, 谢林江, 等. 卷积神经网络在车辆识别中的应用[J]. 计算机科学与探索, 2018, 12(2): 282-291.
- [12] 冯国臣, 陈艳艳, 陈宁. 基于机器视觉的安全帽自动识别技术研究[J]. 机械设计与制造工程, 2015, 44(10): 39-42.
- [13] 胡恬. 利用几何分析法和BP神经网络进行人脸识别的研究[J]. 计算机工程与设计, 2002, 23(9): 18-21.
- [14] 刘晓慧, 叶西宁. 肤色检测和Hu矩在安全帽识别中的应用[J]. 华东理工大学学报, 2014, 40(3).
- [15] He K, Zhang X, Ren S, et al. Deep residual learning for image recognition[C]//IEEE Conference on Computer Vision and Pattern Recognition, 2015: 770-778.
- [16] Krizhevsky A, Sutskever I, Hinton G E. ImageNet classification with deep convolutional neural networks[C]//International Conference on Neural Information Processing Systems, 2012: 1097-1105.
- [17] Lin T Y, Dollár P, Girshick R, et al. Feature pyramid networks for object detection[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017: 2117-2125.
- [18] 张素洁, 赵怀慈. 最优聚类个数和初始聚类中心点选取算法研究[J]. 计算机应用研究, 2017, 34(6): 1617-1620.
- [19] Liu W, Wen Y, Yu Z, et al. Large-margin softmax loss for convolutional neural networks[C]//International Conference on International Conference on Machine Learning, 2016: 507-516.

(上接第198页)

- [20] Malpica N, de Solórzano C O, Vaquero J J, et al. Applying watershed algorithms to the segmentation of clustered nuclei[J]. Cytometry Part A, 1997, 28(4): 289-297.
- [21] Kimmel R, Kiryati N, Bruckstein A M. Sub-pixel distance maps and weighted distance transforms[J]. Journal of Mathematical Imaging & Vision, 1996, 6(2/3): 223-233.
- [22] Zhang D, Lu G. A comparative study on shape retrieval using Fourier descriptors with different shape signatures[C]//International Conference on Intelligent Multimedia and Distance Education, 2003: 1-9.
- [23] Goshtasby A, Mokhtarian F, Mackworth A. Scale-based description and recognition of planar curves and two-dimensional shapes[J]. IEEE Trans on Pattern Anal Mach Intell, 2009, PAMI-8(1): 34-43.
- [24] Mokhtarian F, Mackworth A K. A theory of multiscale, curvature-based shape representation for planar curves[J]. IEEE Transactions on Pattern Analysis & Machine Intelligence, 1992, 14(8): 789-805.
- [25] 刘寅, 滕晓龙, 刘重庆. 复杂背景下基于傅里叶描述子的手势识别[J]. 计算机仿真, 2005, 22(12): 158-161.
- [26] 陈书贞, 张俊军. 基于曲率尺度空间的指纹特征提取算法[J]. 电子技术, 2008, 45(1): 101-105.
- [27] Han Y, Hara A, Kuzuya A, et al. Automatic recognition of DNA pliers in atomic force microscopy images[J]. New Generation Computing, 2015, 33(3): 253-270.
- [28] Hu M K. Visual pattern recognition by moment invariants[J]. IRE Transactions on Information Theory, 1962, 8(2): 179-187.

(上接第208页)

- [15] 陈琳. 基于机器学习的软件缺陷预测研究[D]. 重庆: 重庆大学, 2016.
- [16] Jindal R, Malhotra R, Jain A. Software defect prediction using neural networks[C]//International Conference on Reliability, 2015.
- [17] Ryu D, Choi O, Baik J. Value-cognitive boosting with a support vector machine for cross-project defect prediction[J]. Empirical Software Engineering, 2016, 21(1): 43-71.
- [18] Polson N G, Scott J G, Windle J. Bayesian inference for logistic models using polya-gamma latent variables[J]. Journal of the American Statistical Association, 2012, 108(504).
- [19] Heinze G, Ploner M. logistf: Firth's bias-reduced logistic regression[EB/OL]. [2019]. <https://rdrr.io/cran/logistf/man/logistf.html#heading-3>.
- [20] Meyer D, Dimitriadou E, Hornik K, et al. Misc functions of the department of statistics[J]. Probability Theory Group, 2015, 23(4): 189-205.
- [21] Friedman J, Hastie T, Tibshirani R, et al. Regularization paths for generalized linear models via coordinate descent[J]. Journal of Statistical Software, 2010, 33(1): 1-22.
- [22] Dejaeger K, Verbraken T, Baesens B, et al. Toward comprehensible software fault prediction models using Bayesian network classifiers[J]. IEEE Transactions on Software Engineering, 2013, 39(2): 237-257.