

文章编号:1673-9469(2020)04-0078-09

DOI: 10.3969/j.issn.1673-9469.2020.04.013

YOLOv3 在安全帽佩戴检测中的应用研究

马小陆,王明明,王 兵

(安徽工业大学 电气与信息工程学院,安徽 马鞍山 243002)

摘要: 针对施工现场环境复杂,基于 YOLOv3 的安全帽佩戴检测算法存在精度低、鲁棒性差等问题,提出一种改进 YOLOv3 的安全帽佩戴检测算法。使用 K-means 算法聚类出先验框,改进了网络输出尺度;并在输出端引入了跳跃连接构成残差模块;同时改进分类损失函数以平衡正负样本、难易样本对模型的影响。为验证该方法的有效性,在 NVIDIA GTX1660Ti 平台上进行了验证,实验结果表明,改进后的 YOLOv3 安全帽佩戴检测算法平均准确率提高了 4.84%,提升了对被遮挡的目标以及小目标的检测能力,具有较强的鲁棒性。

关键词: 安全帽;YOLOv3;特征融合;损失函数

中图分类号: TP29

文献标识码: A

Application of YOLOv3 in Safety Helmet Wearing Detection

MA Xiaolu, WANG Mingming, WANG Bing

(School of Electrical and Information Engineering, Anhui University of Technology, Ma'anshan, Anhui 243002, China)

Abstract: In view of the complex construction site environment and the problems of the YOLOv3 helmet wearing detection algorithm such as low accuracy and poor robustness, an improved YOLOv3 helmet wearing detection algorithm was proposed. The K-means algorithm was used to cluster out anchor boxes, so as to improve the network output scales. The residual block was introduced at the output to form a residual module; at the same time, the classification loss function was improved to balance the influence of positive and negative samples and difficult and easy samples on the model. In order to verify the effectiveness of the method, it was verified on the NVIDIA GTX1660Ti. The experimental results show that the *mAP* of the improved YOLOv3 helmet wearing detection algorithm increase by 4.84%, the detection ability of occluded targets and small targets is improved, and the model has strong robustness.

Key words: helmet; YOLOv3; feature fusion; loss function

安全帽是建筑施工人员必须佩戴的防护用具,它可以在发生高空坠物和头部撞击等危害时,有效保护甚至挽救人的生命^[1]。但目前对建筑工人是否佩戴安全帽的检测多是通过人来观察监控的视频^[2],这种方法不仅耗时且效率低,因此针对安全帽佩戴的检测成为研究热点。

现有的安全帽佩戴检测方法有两类。一类是基于传统机器学习的算法,如刘晓慧等^[3]通过肤色定位到人脸,用 SVM 检测 Hu 特征,在人脸以上区域检测安全帽;何慧敏等^[4]使用卷积神经网络检测行人,对行人部分做 HSV 空间变换,得到安全

帽颜色的区域,再对图像做圆 Hough 变换,检测到的圆形物体即为安全帽;贾峻苏等^[5]将方向梯度直方图、颜色特征和基于块的局部二值模式直方图结合成特征向量,再使用 SVM 检测安全帽等。传统算法依赖较强的先验知识,需要耗费大量的时间设计特征,检测精度低、耗时过长。另一类是基于深度学习的算法,分为两种。一种是基于区域建议机制的 Two Stage 方法,如 Girshick R 等^[6-9]使用改进 Faster R-CNN 检测安全帽佩戴情况,该类方法检测精度较高但是耗时较长,且鲁棒性较差。另一种是基于回归模型的 One Stage 方

收稿日期:2020-08-11

基金项目:国家自然科学基金资助项目(61472282);安徽高校自然科学研究重点项目(KJ2019A0065);安徽省教育厅高校科学研究重大项目(KJ2019ZD05)

作者简介:马小陆(1979-),男,安徽芜湖人,博士,博士后,副教授,主要研究方向为嵌入式、车联网和服务机器人。

法^[10-13],如 SSD、YOLO 系列等,如王秋余^[14]对 YOLO 网络结构进行改进,提出一种基于半监督学习的安全帽佩戴识别算法;方明等^[15]在 YOLOv2 的基础上加入密集块用以快速检测安全帽等,该方法检测速度较快但是精度较低,容易漏检小目标。

综上所述可以看出,传统算法的精度和速度均达不到安全帽佩戴检测的实际需要;Two Stage 方法的检测时间过长,且对不同场景的鲁棒性较差,也不适用于复杂的施工现场;One Stage 方法的速度较快,但是对小目标检测效果较差,容易造成漏检和误检,模型精度不高。

针对现有的安全帽佩戴检测算法存在的检测精度较低、鲁棒性较差等问题,本文综合了 One Stage 和 Two Stage 方法的特点及性能后选择 YOLOv3 作为主体算法,YOLOv3 使用了 One Stage 思想,在特征图上直接回归目标的类别和位置,检测速度较快,同时吸收了残差网络和特征金字塔等算法优点,提高了模型的精度。本文在 YOLOv3 的基础上,针对检测对象的特点提出一种改进 YOLOv3 的安全帽佩戴检测算法。采用数据增强手段进行数据集扩充,以提升模型的鲁棒性;删除大尺度输出的同时增加一个特小尺度输出以检测特小目标,在输出端引入了跳跃连接构成残差模块,同时改进分类损失,以提升模型的检测精度。

1 YOLOv3 简介

1.1 网络结构

YOLOv3 由特征提取网络 Darknet-53 和检测层两部分组成,如图 1 所示。首先将输入图片缩放到 416×416 大小,进入特征提取网络 Darknet-53,通过卷积操作提取不同程度的特征,其中使用了残差结构 Residual Block^[16],有助于深层网络训练时仍能保持收敛,防止过拟合,经过 5 次下采样后变成 13×13 大小的特征图,经过卷积后得到第一层输出 13×13。同时借鉴了特征金字塔 (Feature Pyramid Network, FPN)^[17]原理,将第一层输出上采样,并与输入图片下采样得到的特征图进行张量拼接,重复操作后得到三层不同尺度的输出,分别为 13×13、26×26 和 52×52,三种输出可以检测不同大小的目标,提高了模型的精度。

检测时,将输入图像分成 $S \times S$ 个网格 (grid cells),如果目标的中心点在某个 grid cell 中,那么由该 grid cell 负责目标的预测,每个 grid cell 会预测出 3 个边界框 (bounding box, bbox),网络输出的维度如式 (1) 所示:

$$S \times S \times B \times (4 + 1 + C) \quad (1)$$

式中, B 表示每个 grid cell 预测的 bbox 数量,默认为 3;4 表示 bbox 的 4 个参数,分别为中心坐标以及宽高 (x, y, w, h);1 表示该类别的置信度

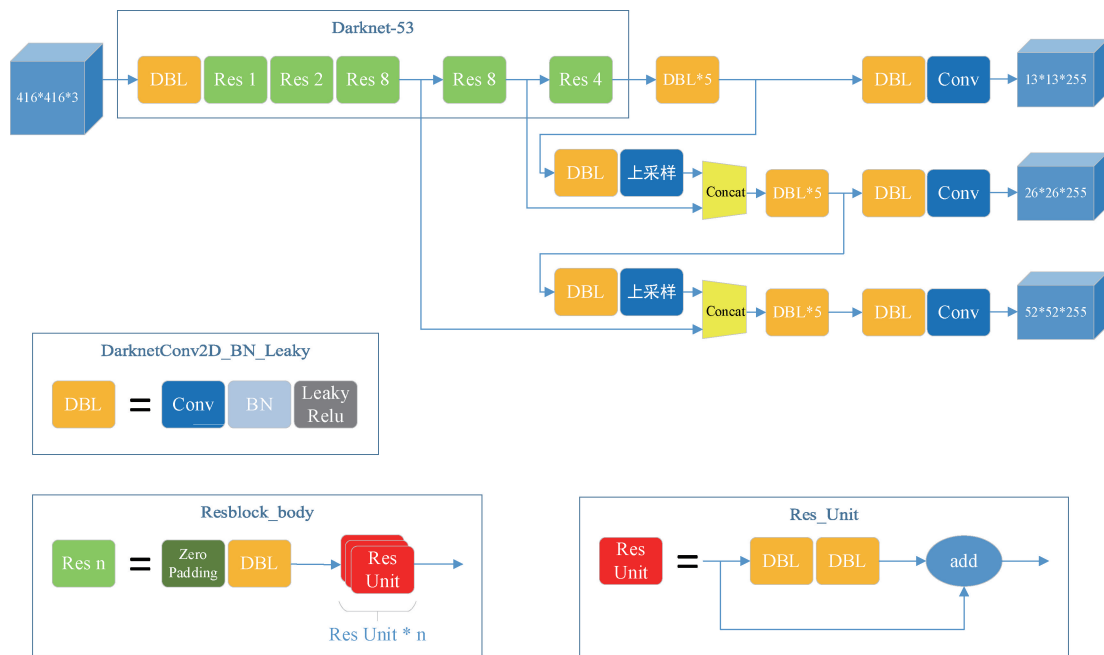


图 1 YOLOv3 网络结构

Fig. 1 Network structure of YOLOv3

(confidence); C 表示类别的总数。预测的类别置信度如式(2)所示:

$$P_r(\text{class}_i | \text{object}) \times P_r(\text{object}) \times \text{IoU}_{\text{pred}}^{\text{truth}} \quad (2)$$

式中,若 grid cell 中有目标,则 $P_r(\text{object}) = 1$,反之 $P_r(\text{object}) = 0$; $P_r(\text{class}_i | \text{object})$ 表示第 i 类目标的置信度; $\text{IoU}_{\text{pred}}^{\text{truth}}$ 表示预测框 bbox 与真实框 (Ground Truth, GT) 的交并比 (Intersection over Union, IoU),可以用来衡量预测的精准性;对所有结果使用 NMS 算法进行筛选得到最终输出。

1.2 边框预测

YOLOv3 使用了先验框 anchor box 机制, anchor box 最初由 Faster R-CNN 提出,使用统计学中的先验数据或人类的经验作为约束条件。比如需要检测的对象为行人,那么目标一定是瘦高的,只需要在目标中心点周围寻找瘦高的 bbox 即可,这样可以缩小检测的范围,有助于模型的收敛。

原始 YOLOv3 对 COCO 数据集聚类出 9 个 anchor box: (10×13) 、 (16×30) 、 (33×23) 、 (30×61) 、 (62×45) 、 (59×119) 、 (116×90) 、 (156×198) 、 (373×326) ,并平均分配给 3 个尺度的输出。其中 13×13 的输出有最大的感受野,使用最大的三个 anchor box 检测大型目标; 26×26 的输出感受野中等,使用中间三个 anchor box 检测中等目标; 52×52 的输出的感受野最小,使用最小的三个 anchor box 检测小目标。grid cell 预测目标时,在预设的 anchor box 基础上,通过平移变换加缩放即可得到 bbox,如图 2 所示。

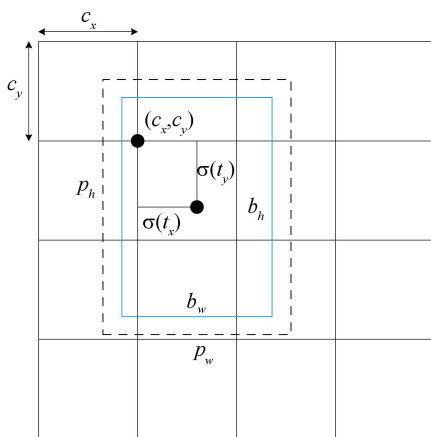


图 2 边框预测

Fig. 2 Bounding box prediction

图 2 中,虚线框表示预设的先验框 anchor box,实线框表示网络的预测框 bbox, bbox 的中心点坐标以及宽高 (b_x, b_y, b_w, b_h) 的计算如式(3)~(6)所示:

$$b_x = \sigma(t_x) + c_x \quad (3)$$

$$b_y = \sigma(t_y) + c_y \quad (4)$$

$$b_w = p_w e^{t_w} \quad (5)$$

$$b_h = p_h e^{t_h} \quad (6)$$

式(3)~(6)中, (c_x, c_y) 表示目标中心所在的 grid cell 左上角的坐标; p_w, p_h 分别表示预设的 anchor box 映射到特征图中的宽和高; (t_x, t_y, t_w, t_h) 则是网络需要训练求解的参数; $\sigma(t_x), \sigma(t_y)$ 表示使用 Sigmoid 函数将 t_x, t_y 压缩到区间 $(0, 1)$ 上,否则模型无法收敛。

1.3 损失函数

YOLOv3 的损失函数分为四个部分,如式(7)所示,分别为 bbox 坐标损失 Loss_{xy} 、bbox 宽高损失 Loss_{wh} 、置信度损失 $\text{Loss}_{\text{conf}}$ 以及类别损失 $\text{Loss}_{\text{class}}$,具体计算如式(8)~(11)所示:

$$\text{Loss}(\text{object}) = \text{Loss}_{xy} + \text{Loss}_{wh} + \text{Loss}_{\text{conf}} + \text{Loss}_{\text{class}} \quad (7)$$

$$\text{Loss}_{xy} = \lambda_{\text{coord}} \sum_{i=0}^{S^2} \sum_{j=0}^B I_{ij}^{\text{obj}} (2 - w_i \times h_i) [(x_i - \hat{x}_i)^2 + (y_i - \hat{y}_i)^2] \quad (8)$$

$$\text{Loss}_{wh} = \lambda_{\text{coord}} \sum_{i=0}^{S^2} \sum_{j=0}^B I_{ij}^{\text{obj}} (2 - w_i \times h_i) [(w_i - \hat{w}_i)^2 + (h_i - \hat{h}_i)^2] \quad (9)$$

$$\text{Loss}_{\text{conf}} = - \sum_{i=0}^{S^2} \sum_{j=0}^B I_{ij}^{\text{obj}} [C_i \log(\hat{C}_i) + [(1 - C_i) \log(1 - \hat{C}_i)] - \lambda_{\text{noobj}} \sum_{i=0}^{S^2} \sum_{j=0}^B I_{ij}^{\text{noobj}} [C_i \log(\hat{C}_i) + (1 - C_i) \log(1 - \hat{C}_i)] \quad (10)$$

$$\text{Loss}_{\text{class}} = - \sum_{i=0}^{S^2} I_{ij}^{\text{obj}} \sum_{c \in \text{classes}} \{p_i(c) \log[\hat{p}_i(c)] + [1 - p_i(c)] \log[1 - \hat{p}_i(c)]\} \quad (11)$$

式(8)~(11)中, i 表示 grid cell 序号; j 表示每个 grid cell 预测的 bbox 序号; (x_i, y_i) 表示第 i 个 grid cell 预测的 bbox 中心点坐标; w_i, h_i 分别表示 bbox 的宽和高; $p(c)$ 表示目标属于第 c 个类别的概率; λ_{coord} 为权重系数; λ_{noobj} 为惩罚权重系数; I_{ij}^{obj} 表示第 i 个 grid cell 的第 j 个 bbox 是否负责预测该目标,取值 0 或 1。

2 YOLOv3 改进

2.1 输出尺度改进

YOLOv3 有大、中、小三个不同尺度的输出,每层输出上预设三个 anchor box,在预设的 anchor box 基础上通过平移和缩放变换即可得到预测框 bbox,所以 anchor box 的设定对预测的精准性有很

大影响,并且其大小与 YOLOv3 的输出尺度相对应,大尺度的输出上设定 3 个较大的 anchor box,以此类推,因此对输出的尺度改进取决于 anchor box 的大小。anchor box 的值是对数据集的标注框聚类得到的,使用 K-means 聚类算法对“安全帽”数据集重新聚类生成新的 anchor box,令 K 取不同的值即选择不同数量的聚类中心得到的 anchor box 大小,如表 1 所示。

表 1 不同 K 值对应的 anchor box

Tab. 1 Anchor box corresponding to different K values

$K=3$	$K=6$	$K=9$	$K=12$
(7, 14)	(6, 12)	(5, 10)	(5, 10)
(13, 24)	(8, 16)	(6, 13)	(6, 13)
(36, 61)	(11, 22)	(8, 16)	(7, 14)
—	(18, 31)	(10, 20)	(8, 18)
—	(30, 51)	(13, 25)	(9, 16)
—	(62, 107)	(19, 33)	(10, 21)
—	—	(28, 47)	(14, 25)
—	—	(44, 75)	(18, 31)
—	—	(82, 140)	(24, 42)
—	—	—	(34, 58)
—	—	—	(51, 88)
—	—	—	(92, 157)

由表 1 数据可知,当 K 值继续增大时,聚类中心相距较近,为了避免增加计算量产生冗余,选取 $K=6$ 时的 anchor box 大小。通过比较重新聚类出的 6 个 anchor box 与原来的 9 个 anchor box 大小可知,“安全帽”数据集的目标中没有特别大的尺寸,多了更小的尺寸。针对 anchor box 与输出尺度一

一对应,将原有大尺度输出删除,并且增加一个特小尺度输出以检测特小目标。

在深度学习中,加深网络层数多次进行卷积和下采样操作可以提取更强的语义特征,但与此同时,小像素目标也会更容易丢失特征信息。因此将原始的输入尺寸从 416×416 增大至 608×608 ,以确保多次下采样之后,小像素目标的特征不会丢失,经过 5 次下采样操作得到的特征图尺寸依次为: 304×304 、 152×152 、 76×76 、 38×38 、 19×19 ,此时 YOLOv3 的三层输出分别为: 19×19 、 38×38 、 76×76 。根据 K-means 聚类得到的 anchor box 的尺寸情况,没有特别大的 anchor box,将检测大目标的输出 19×19 删除,针对最小 anchor box 为 (6, 12),增加一个特小尺度输出以检测特小目标:将 76×76 的输出进行上采样至 152×152 ,同时与输入图片第二次下采样得到的 152×152 特征图进行张量拼接,这样同时保留了浅层特征的高分辨率信息和深层特征的高语义信息,得到了第三层输出 152×152 。

此时新设定的三层输出以及对应的 anchor box 大小分配如表 2 所示。

表 2 不同输出分配的 anchor box

Tab. 2 Anchor box for different output scales

输出	38×38	76×76	152×152
anchor box	(30, 51)	(11, 22)	(6, 12)
	(62, 107)	(18, 31)	(8, 16)
检测对象	中等目标	小目标	特小目标

改变输出尺度之后的 YOLOv3 网络结构如图 3 所示。

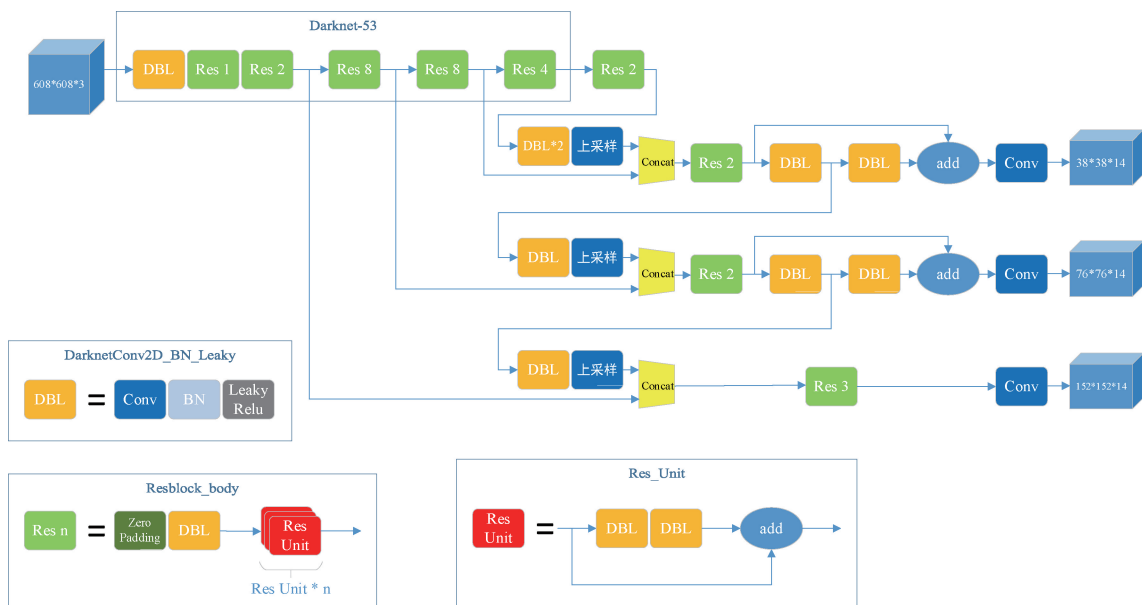


图 3 改进 YOLOv3 网络结构

Fig. 3 Improved YOLOv3 network structure

2.2 引入残差模块

深层网络中小目标所包含的特征信息较少,为了使检测层可以检测到深层特征图中包含的信息,在输出端的卷积层之间引入跳跃连接构成残差模块实现网络间的信息传递,这样在网络加深的同时可以提高对深层特征的检测能力,保证了模型检测的准确度。残差单元结构如图4所示。

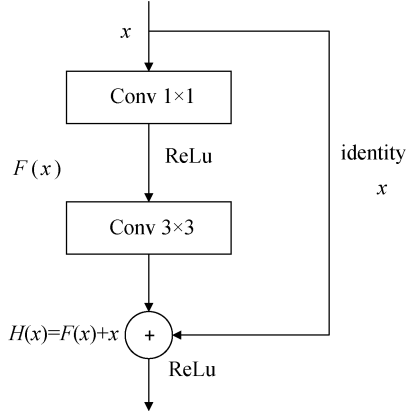


图4 残差单元结构

Fig. 4 The structure of the Residual Block

当输入为 x 时,学习到的特征记为 $H(x)$,此时的残差为 $F(x) = H(x) - x$,因为残差 $F(x)$ 一般比较小,即需要学习的内容比较少,因此残差学习比原始特征直接学习更容易。

残差单元的计算如式(12)和式(13)所示:

$$y_l = h(x_l) + F(x_l, W_l) \quad (12)$$

$$x_{l+1} = f(y_l) \quad (13)$$

式(12)和式(13)中, x_l, x_{l+1} 分别表示第 l 个残差单元的输入和输出; F 是残差函数,表示学习到的残差; $h(x_l) = x_l$ 表示恒等映射; f 表示 ReLu 激活函数。通过递归可以得到浅层单元 l 到深层单元 L 的特征表达式,如式(14)所示:

$$x_L = x_l + \sum_{i=l}^{L-1} F(x_i, W_i) \quad (14)$$

传统的卷积神经网络相当于一层层地做仿射变换以及非线性变换,仿射变换主要依靠矩阵的乘法,所以传统网络总体为乘法性质的计算,而加入了跳跃连接构成残差模块之后,计算性质由乘法变为加法,由式(14)可以看出,任意深度单元 L 的特征 x_L 可以表达为浅层单元 l 的特征 x_l 加上一个形如 $\sum_{i=l}^{L-1} F$ 的残差函数,表明了任何单元 L 和 l 之间都具有残差特性。

因此残差模块可以保证在深层的网络结构中提取到浅层网络的高分辨率信息,同时保留深层网络的高语义信息,提升了对小目标特征的检测

能力。

2.3 损失函数改进

One Stage 算法精度要低于 Two Stage 算法,因为 One Stage 算法没有在检测之前先使用区域建议网络提取图片中待检测目标所处的部分区域。一幅图片中只有少部分区域是包含待检测目标的,将图片中含有待检测目标“安全帽”的部分区域称为正样本;将图片中不含有待检测目标的其它区域称为负样本,如图5所示。

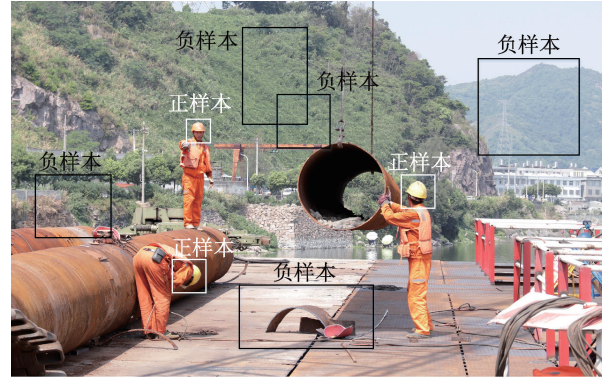


图5 正负样本失衡

Fig. 5 Positive and Negative samples imbalance

图5中,白色框中包含待检测目标“安全帽”,为正样本;黑色框中均不含有待检测目标,为负样本;可知负样本数量远大于正样本,造成了正负样本比例失衡现象,导致了总体损失 Loss 负样本占大部分,影响了模型的收敛。较难分类的样本称为困难样本;容易分类的样本称为简单样本。又因为正样本中待检测目标的类别大多较难分类,所以正样本多数为困难正样本;负样本大多容易分类,所以负样本多数为简单负样本。由式(11)可知,YOLOv3 以交叉熵作为分类损失,可以简化成式(15)的形式:

$$\begin{aligned} \text{Loss}_{\text{class}} &= -y \log \hat{y} - (1-y) \log(1-\hat{y}) \\ \log(1-\hat{y}) &= \begin{cases} -\log \hat{y}, & y=1 \\ -\log(1-\hat{y}), & y=0 \end{cases} \quad (15) \end{aligned}$$

式(15)中, \hat{y} 是经过激活函数之后的输出,取值 $(0, 1)$ 之间,对于正样本,输出的概率越大则损失 Loss 越小,对于负样本,输出的概率越小则损失 Loss 越小。由于简单负样本在总体训练样本中的占比过大,模型在迭代过程中收敛缓慢,且可能无法优化至最好。

为了平衡不同样本对模型的贡献,本文引入焦点损失函数(Focal Loss, FL)^[18]思想,在原有函数加入调节因子以降低简单易区分样本在训练中

所占的比重,使模型更专注于困难样本的分类。改进后的分类损失简化成式(16)的形式:

$$Loss_{class} = \begin{cases} -\alpha(1-\hat{y})^\gamma \log \hat{y}, & y = 1 \\ -(1-\alpha)\hat{y}^\gamma \log(1-\hat{y}), & y = 0 \end{cases} \quad (16)$$

式(16)中, $\gamma > 0$,以正样本为例,若分类概率 \hat{y} 较大则表示较容易分类,即为简单样本,此时 $(1-\hat{y})^\gamma$ 的值就会很小,而分类概率较低的困难样本,损失值相对就比较大,负样本同理。这样就降低了简单样本对损失的贡献,使模型更加关注难以区分的样本,同时加入平衡因子 α ,用来平衡正负样本比例失衡现象。

3 实验结果与分析

实验平台的软硬件配置如表 3 所示。

表 3 实验平台配置

Tab. 3 Experimental platform configuration	
实验环境	配置
操作系统	Ubuntu 18.04 LTS
CPU	i7-9750H 2.6 GHz
GPU	NVIDIA GTX1660Ti
显卡驱动	NVIDIA 440.100
GPU 加速库	CUDA10.0, cudnn7.4
数据处理	Python3.6, OpenCV3.4.1
编译器	gcc4.9, g++4.9
框架	Darknet

3.1 训练

由于针对安全帽佩戴检测没有官方的数据集,为了进行本次研究,从互联网上搜集“佩戴安全帽的人”以及“没有佩戴安全帽的人”图片共 7 000 余张。使用标注工具 LabelImg 对图片内目标进行标注,将“佩戴安全帽的人”归于“hat”类,将“没有佩戴安全帽的人”归于“person”类,标注过程如图 6 所示。

标注完成后,会得到相应的 xml 文件,xml 文件中存放了每一个目标 GT 框的标签信息,包括输入图像的大小以及目标的类别和位置信息,如图 7 所示。

所有的图片标注完后即完成“安全帽数据集”建立,将数据集按照 7:1:2 的比例划分成训练集 train、验证集 val 以及测试集 test。train 是参与训练的主要样本,在训练过程中对误差进行梯度下降,训练权重参数;val 是训练过程中单独留出的样本,每训练一段时间验证一下模型的检测效果,并调整参数优化模型;test 是用来评估最终模型的能

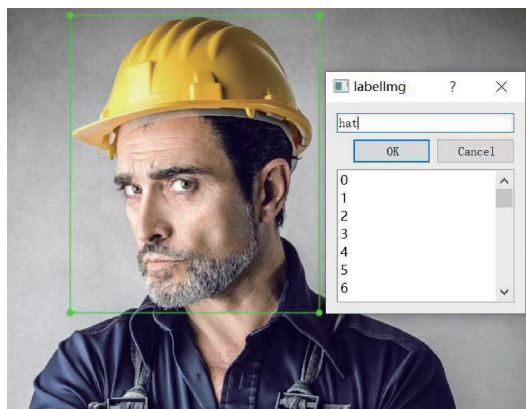


图 6 数据集的标注

Fig. 6 Data set annotation

```

<annotation>
  <folder>data_set</folder>
  <filename>0001.jpg</filename>
  <path>/home/alien/桌面/data_set/0001.jpg</path>
  <source>
    <database>Unknown</database>
  </source>
  <size>
    <width>1000</width>
    <height>762</height>
    <depth>3</depth>
  </size>
  <segmented>0</segmented>
  <object>
    <name>hat</name>
    <pose>Unspecified</pose>
    <truncated>0</truncated>
    <difficult>0</difficult>
    <bndbox>
      <xmin>334</xmin>
      <ymin>15</ymin>
      <xmax>630</xmax>
      <ymax>370</ymax>
    </bndbox>
  </object>
</annotation>

```

图 7 标签信息

Fig. 7 Label information

力,不参与训练。为了提高模型的泛化能力,使用数据增强的手段对数据集进行扩充,具体操作为:改变图像的饱和度、亮度以及色调等,如图 8 所示。

图 8 中,增加亮度及减少亮度可以模拟光线不同时场景的变化情况,改变色调可以模拟不同颜色的安全帽,这样可以增加样本的数量以提高模型对不同场景的鲁棒性,并且采用多尺度训练的方式,在将样本送入网络训练前随机改变样本的大小,以提高模型对不同尺度目标的检测能力。

将训练批次数(Batches)设置为 50 200 次,分别将数据集送至 YOLOv3 和改进 YOLOv3 中进行训练,训练过程中模型的平均损失值(Average Loss)由式(7)可以计算出,损失值越小表示训练出的模型与数据集的真实模型越接近;训练期间 YOLOv3 和改进 YOLOv3 的损失函数曲线(Loss Curve)如图 9 所示。



图8 数据增强示例图

Fig. 8 Data augmentation example

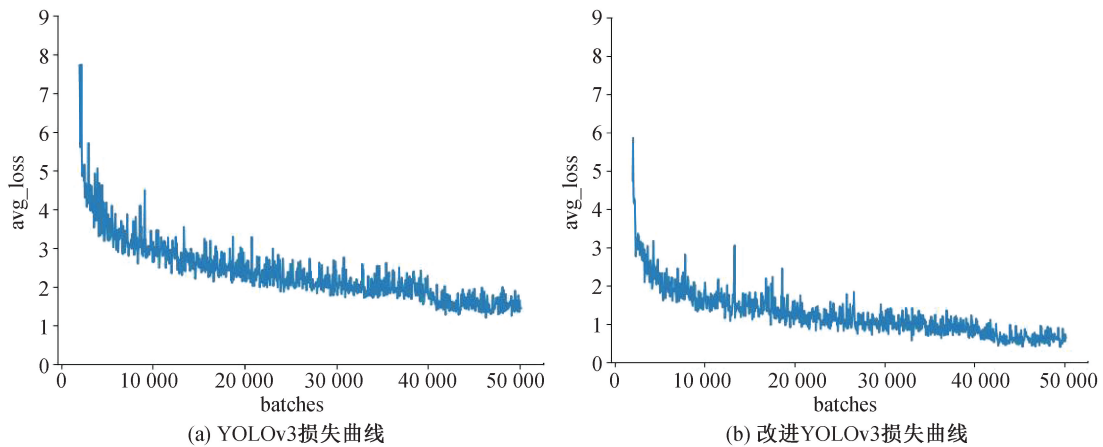


图9 损失曲线

Fig. 9 The Loss curve

图9中,横坐标 batches 表示训练的批次数,纵坐标 avg_loss 表示平均损失值。由图9(a)和图9(b)可以看出,两次训练过程网络损失均保持稳定下降趋势并在迭代 50 200 次左右收敛,说明训练基本顺利。同时,与图9(a)的曲线相比,图9(b)的曲线收敛速度更快,并且网络波动幅度变小,收敛于一个更小的值,说明改进后的模型更接近数据集的真实模型。

3.2 测试

在目标检测领域,一般使用平均精度(mean Average Precision, mAP)评估模型的效果,AP 表示单个类别的平均准确率,mAP 表示多个类别的 AP 的均值^[19]。mAP 的值越大表示模型的整体识别准确率越高。AP 可以由模型分类结果的混淆矩阵求得,混淆矩阵的定义如表4所示。

表4 混淆矩阵

Tab. 4 Confusion matrix

真实情况	预测结果	
	正例	反例
正例	TP(真正例)	FN(假反例)
反例	FP(假正例)	TN(真反例)

精确率 Precision = $\frac{TP}{TP + FP}$, 即预测为正例的样本中,实际是正例的占多少。

召回率 Recall = $\frac{TP}{TP + FN}$, 即所有的正样本,被正确预测的占多少。

以 Recall 为横轴,Precision 为纵轴绘成的曲线图称为 PR 曲线,PR 曲线综合了精确率与召回率的信息,曲线与两坐标轴围成的面积即为 AP 的值,如式(17)所示。

$$AP = \int_{0P}^1 (R) dR \quad (17)$$

测试集上测得改进前及改进后模型的 PR 曲线如图 10 所示。由图计算得出 YOLOv3 的 $AP_{hat} = 88.97\%$, $AP_{person} = 85.23\%$, $mAP = 87.10\%$;改进 YOLOv3 的 $AP_{hat} = 90.59\%$, $AP_{person} = 93.30\%$, $mAP = 91.94\%$ 。

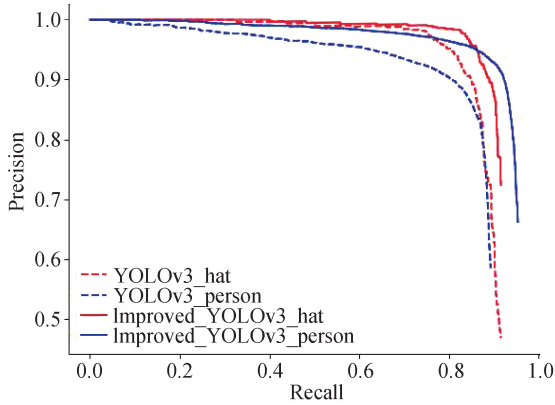


图 10 PR 曲线
Fig. 10 The PR curve

图 10 中,横坐标 Recall 表示模型的召回率;纵坐标 Precision 表示模型的精确率;红色和蓝色的虚线分别表示原始 YOLOv3 的“hat”和“person”的 PR 曲线,红色和蓝色的实线分别表示改进 YOLOv3 的“hat”和“person”的 PR 曲线,由图 10

可以看出改进 YOLOv3 的两个类别 PR 曲线坡度更加平滑,精确率 Precision 和召回率 Recall 均得到了提升,曲线与坐标轴围成的面积也更大,即平均准确率 AP 得到了提高。可见改进的 YOLOv3 模型对两类目标检测能力均有不同幅度地提升,模型的整体检测精度更高。

实验中,为了验证两种模型的精度,在测试集中挑选部分样本进行测试,两种模型的检测效果对比如图 11 所示,图 11(a)为原图,图 11(b)为 YOLOv3 的检测图,图 11(c)为改进 YOLOv3 检测图。

由图 11 可知,两种模型均可以检测到佩戴安全帽的目标“hat”以及未佩戴安全帽的目标“person”。同时,由第一组检测结果可知,原始 YOLOv3 漏检了部分被遮挡目标和重叠目标,而在改进 YOLOv3 中可以成功检测到;由第二组检测结果可知,原始 YOLOv3 无法判断中间的目标为“hat”还是“person”,属于困难样本,而改进后的 YOLOv3 可以正确的区分;由第三组检测结果可知,原始 YOLOv3 无法检测到特小目标,而改进后的 YOLOv3 可以检测更多的小目标。对比检测结果可知,改进后的 YOLOv3 提升了对被遮挡的目标、困难样本以及小目标的检测能力,检测精度有所提高。

同时,为了测试改进后的 YOLOv3 模型的泛化



(a) 原图 (b) YOLOv3 (c) 改进YOLOv3

图 11 检测对比图

Fig. 11 Comparison of detection results

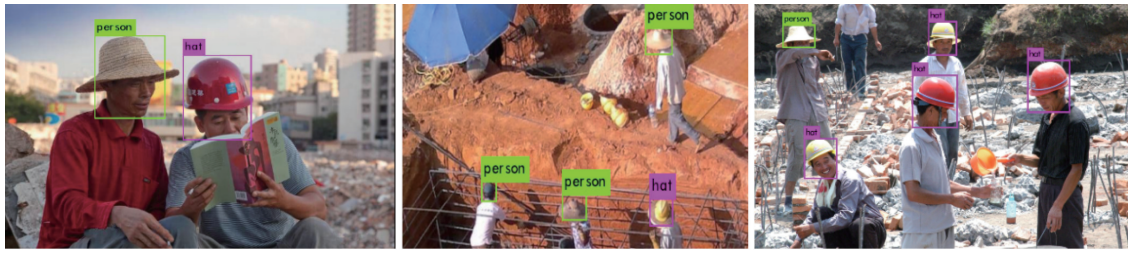


图 12 不同场景测试图

Fig. 12 Detection results of different scenes

能力和鲁棒性,挑选不同的场景以及不是佩戴“安全帽”的目标进行测试,测试结果如图 12 所示。

由图 12 可以看出,改进后的 YOLOv3 模型在不同场景下均有较强的检测能力,同时模型可以判断出工人所佩戴的是否为安全帽,若佩戴的不是安全帽,则会将其归类于“person”。由此可知,改进后的模型泛化能力和鲁棒性较强,受场景、光照及色彩等变换的影响较小,整体的检测效果较为理想。

4 结论

本文提出了一种安全帽佩戴检测算法,以 YOLOv3 算法为主体,聚类出更符合“安全帽”的 anchor box;并根据 anchor box 大小修改了网络的输出尺度,降低了输出维度并减小了模型大小;引入跳跃连接在输出端构建残差模块,更有效地对特征进行表示;同时改进分类损失,平衡了正负样本失衡现象以及难易样本对模型的影响。实验结果表明此方法提升了 YOLOv3 的模型性能,增强了鲁棒性,针对小目标的检测能力得到了提高,具有较高的精度,达到了安全帽佩戴检测的实际需要。

参考文献:

- [1]常欣,刘鑫萌. 建筑施工人员不合理佩戴安全帽事故树分析[J]. 吉林建筑大学学报, 2018, 35(06): 65-69.
- [2]狄春杰,胡瑞卿. 行为安全研究在建筑施工安全管理中的应用与分析[J]. 建筑安全, 2020, 35(01): 48-51.
- [3]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(自然科学版), 2014, 40(03): 365-370.
- [4]何慧敏,林志贤,郭太良. 基于卷积神经网络的行人安全帽自动识别[J]. 有线电视技术, 2018(03): 104-108.
- [5]贾峻苏,鲍庆洁,唐慧明. 基于可变形部件模型的安全头盔佩戴检测[J]. 计算机应用研究, 2016, 33(03): 953-956.
- [6]GIRSHICK R, DONAHUE J, DARRELL T, et al. Re-

gion-based Convolutional Networks for Accurate Object Detection and Segmentation[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence. IEEE: Piscataway, NJ, 2015, 38(1): 142-158.

- [7]GIRSHICK R. Fast R-CNN[C]//IEEE International Conference on Computer Vision. IEEE: Piscataway, NJ, 2015: 1440-1448.
- [8]REN S Q, HE K M, GIRSHICK R, et al. Faster R-CNN: Towards Real-time Object Detection with Region Proposal Networks[C]//Proceedings of the 28th International Conference on Neural Information Processing Systems. IEEE: Piscataway, NJ, 2015: 91-99.
- [9]王忠玉. 智能视频监控下的安全帽佩戴检测系统的设计与实现[D]. 北京:北京邮电大学, 2018.
- [10]LIU W, ANGUELOV D, ERHAN D, et al. SSD: Single Shot Multi Box Detector[C]//European Conference on Computer Vision. Springer: Berlin, DE, 2016: 21-37.
- [11]REDMON J, DIVVALA S, GIRSHICK R, et al. You Only Look Once: Unified, Real-time Object Detection [C]//IEEE Conference on Computer Vision and Pattern Recognition. IEEE: Piscataway, NJ, 2016: 779-788.
- [12]REDMON J, FARHADI A. YOLO9000: Better, Faster, Stronger[C]//IEEE Conference on Computer Vision and Pattern Recognition. IEEE: Piscataway, NJ, 2017: 6517-6525.
- [13]REDMON J, FARHADI A. YOLOv3: An Incremental Improvement[J]. arXiv: 1804. 02767, 2018.
- [14]王秋余. 基于视频流的施工现场工人安全帽佩戴识别研究[D]. 武汉:华中科技大学, 2018.
- [15]方明,孙腾腾,邵桢. 基于改进 YOLOv2 的快速安全帽佩戴情况检测[J]. 光学精密工程, 2019, 27(05): 1196-1205.
- [16]HE K, ZHANG X, REN S, et al. Deep Residual Learning for Image Recognition[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. IEEE: Piscataway, NJ, 2016: 770-778.
- [17]LIN T, DOLLAR P, GIRSHICK R, et al. Feature Pyramid Networks for Object Detection [C]//IEEE Conference on Computer Vision and Pattern Recognition. IEEE: Piscataway, NJ, 2017: 936-944.
- [18]LIN T Y, GOYAL P, GIRSHICK R, et al. Focal Loss for Dense Object Detection[C]//2017 IEEE International Conference on Computer Vision (ICCV). IEEE, 2017:2999-3007.
- [19]任嘉锋,熊卫华,吴之昊,等. 基于改进 YOLOv3 的火灾检测与识别[J]. 计算机系统应用, 2019, 28(12): 171-176.

(责任编辑 周雪梅)