

## 复杂场景下的改进 YOLOv4 安全帽检测算法

谢国波<sup>1</sup>, 唐晶晶<sup>1</sup>, 林志毅<sup>1\*</sup>, 郑晓锋<sup>1</sup>, 方明<sup>2</sup><sup>1</sup>广东工业大学计算机学院, 广东 广州 510006;<sup>2</sup>云南电网有限责任公司输电分公司, 云南 昆明 650011

**摘要** 为了有效地检测复杂场景下施工人员的安全帽佩戴情况,减少安全隐患,提出一种改进的YOLOv4安全帽检测算法(SMD-YOLOv4)。首先采用SE-Net注意力模块强化模型主干网络提取有效特征的能力;然后使用密集空洞空间金字塔池化(DenseASPP)代替网络中的空间金字塔池化(SPP)以减少信息丢失,优化全局上下文信息的提取;最后在PANet部分增加特征融合的尺度并引入深度可分离卷积,使网络在获得复杂背景下小目标细节信息的同时不降低网络推理速度。实验结果表明:在自建实验数据集下,SMD-YOLOv4算法的平均精度均值(mAP)达97.34%,较目前具有代表性的Faster R-CNN、SSD、YOLOv5、YOLOx和YOLOv4算法,分别高出了26.41个百分点、6.44个百分点、3.25个百分点、1.49个百分点和3.19个百分点,能满足实时检测的需要。

**关键词** 目标检测; YOLOv4; 安全帽检测; 注意力机制; 多尺度特征融合; DenseASPP

中图分类号 TP391.4

文献标志码 A

DOI: 10.3788/LOP221388

## Improved YOLOv4 Helmet Detection Algorithm Under Complex Scenarios

Xie Guobo<sup>1</sup>, Tang Jingjing<sup>1</sup>, Lin Zhiyi<sup>1\*</sup>, Zheng Xiaofeng<sup>1</sup>, Fang Ming<sup>2</sup><sup>1</sup>School of Computer Science and Technology, Guangdong University of Technology, Guangzhou 510006, Guangdong, China;<sup>2</sup>Transmission Branch of Yunnan Power Grid Co., Ltd., Kunming 650011, Yunnan, China

**Abstract** An improved helmet detection algorithm for YOLOv4 (SMD-YOLOv4) is proposed to effectively detect whether construction workers are wearing helmets in complex scenes and reduce safety hazards. First, the SE-Net attention module is used to improve the ability of the model backbone network to extract effective features. Next, a dense atrous space pyramid pooling (DenseASPP) is used instead of spatial pyramid pooling (SPP) in the network to reduce information loss and optimize the extraction of global contextual information. Finally, the scale of feature fusion is increased in the PANet part and deep separable convolution is introduced to obtain detailed information about small targets in complex contexts without slowing down the network inference speed. The experimental results show that the mean average precision (mAP) of SMD-YOLOv4 algorithm reaches 97.34% on the self-built experimental dataset, which is 26.41 percentage points, 6.44 percentage points, 3.25 percentage points, 1.49 percentage points, and 3.19 percentage points higher than that of the current representative Faster R-CNN, SSD, YOLOv5, YOLOx, and original YOLOv4 algorithms, respectively, and can meet the real-time detection requirements.

**Key words** object detection; YOLOv4; helmet detection; attentional mechanism; multi-scale feature fusion; DenseASPP

## 1 引言

在施工现场的安全管理中,监督施工人员佩戴安全帽是必不可少的,安全帽可以缓冲和分散瞬间的冲击力<sup>[1]</sup>,避免或减少高空坠物对头部的直接伤害,是施工人员最基本的个体防护用具。在施工现场安全帽佩戴的早期监管中,通常采用人工监督的方式判断工作人员是否佩戴安全帽,但由于施工人员流动性大,巡检

人员很难全面了解各单位的安全施工情况,既增加额外的人力成本又容易出现安全隐患<sup>[2]</sup>。因此,研究施工现场工作人员安全帽的佩戴检测具有重要意义。

随着计算机技术的发展,机器学习技术被广泛应用到安全帽佩戴检测中。目前针对安全帽佩戴的检测算法主要分为基于传统机器学习的目标检测算法和基于深度学习的目标检测算法。基于传统机器学习的检测方法主要是通过安全帽的形状和颜色特征进行识别

收稿日期: 2022-04-22; 修回日期: 2022-05-30; 录用日期: 2022-06-16; 网络首发日期: 2022-06-26

基金项目: 国家自然科学基金(61802072)、广州市科技计划项目(201902020012)

通信作者: \*lzy291@gdut.edu.cn

的。刘晓慧等<sup>[3]</sup>使用皮肤颜色检测方法对人脸进行定位,并使用支持向量机(SVM)<sup>[4]</sup>实现对安全帽的检测。Rubaiyat等<sup>[5]</sup>使用梯度方向直方图(HOG)特征检测施工人员,将基于颜色的特征提取技术和随机圆Hough变换算法结合,用于安全帽佩戴检测。虽然类似于上述传统机器学习的安全帽检测算法检测速度较快,但需根据特定检测对象去人为设计特征和训练分类器,同时由于特征单一、泛化能力较差,无法对复杂施工环境下的目标进行有效检测,检测准确率较低。

基于深度学习的目标检测算法分为“双阶段”和“单阶段”的目标检测算法。Fan等<sup>[6]</sup>集成Tiny Face和经典的双阶段检测算法Faster R-CNN<sup>[7]</sup>,对复杂场景下的安全帽佩戴情况进行检测,虽然检测精度有所提升但检测速度依旧较慢。单阶段检测算法则在检测速度方面有较大的提升,主要的单阶段检测算法为single shot MultiBox detector(SSD)<sup>[8]</sup>和you only look once(YOLO)<sup>[9-12]</sup>系列模型,因此许多学者在单阶段模型的基础上对安全帽检测问题进行了进一步研究。许凯等<sup>[13]</sup>通过对YOLOv3的损失函数进行改进,减小了模型检测安全帽时由于正负样本不均衡带来的误差。Zhou等<sup>[14]</sup>提出了一种轻量级的YOLOv4算法,提高了检测安全帽效率。Deng等<sup>[15]</sup>使用多尺度图像训练YOLOv4模型,该模型能够适应不同分辨率的图像,在小尺寸安全帽检测方面取得了较好的效果。杨永波等<sup>[16]</sup>通过替换YOLOv5的主干网络并替换模型中非

极大值抑制的方法实现了模型的轻量化,提高了模型在复杂场景下对被遮挡安全帽的辨识度。曹捷等<sup>[17]</sup>利用数据增扩方式丰富数据集中的场景,并结合实例归一化策略提高了模型应对复杂作业环境的鲁棒性。

上述单阶段安全帽检测算法虽然从不同方面提升了检测性能,但都难以在复杂的场景下保持较好的检测效果,原因在于视频监控场景中的安全帽目标多为中小目标,使得模型在复杂背景下难以正确区分目标与背景。因此,本文从小目标细节信息和图像整体深层语义信息着手,使用单阶段的YOLOv4模型并对模型进行如下改进:使用 Squeeze-and-Excitation networks(SE-Net)<sup>[18]</sup>注意力模块增强主干网络的特征提取能力;融合多尺度信息,在加强浅层特征与深层特征融合的同时引入深度可分离卷积,提升算法对小目标的检测能力;将模型中的 spatial pyramid pooling networks(SPP-Net)部分替换为密集空洞空间金字塔池化(DenseASPP)<sup>[19]</sup>以获取多尺度的信息增益,避免深层网络信息的丢失。实验结果表明,SMD-YOLOv4算法可以提高在复杂环境下对安全帽的检测能力。

## 2 YOLOv4算法

YOLOv4算法由输入端(Input)、主干特征提取网络(Backbone)、颈部(Neck)和检测头(Head)组成,网络结构如图1所示。Backbone部分为CSPDarknet53网络,包含5个残差块。其中,CBM模块由卷积运算

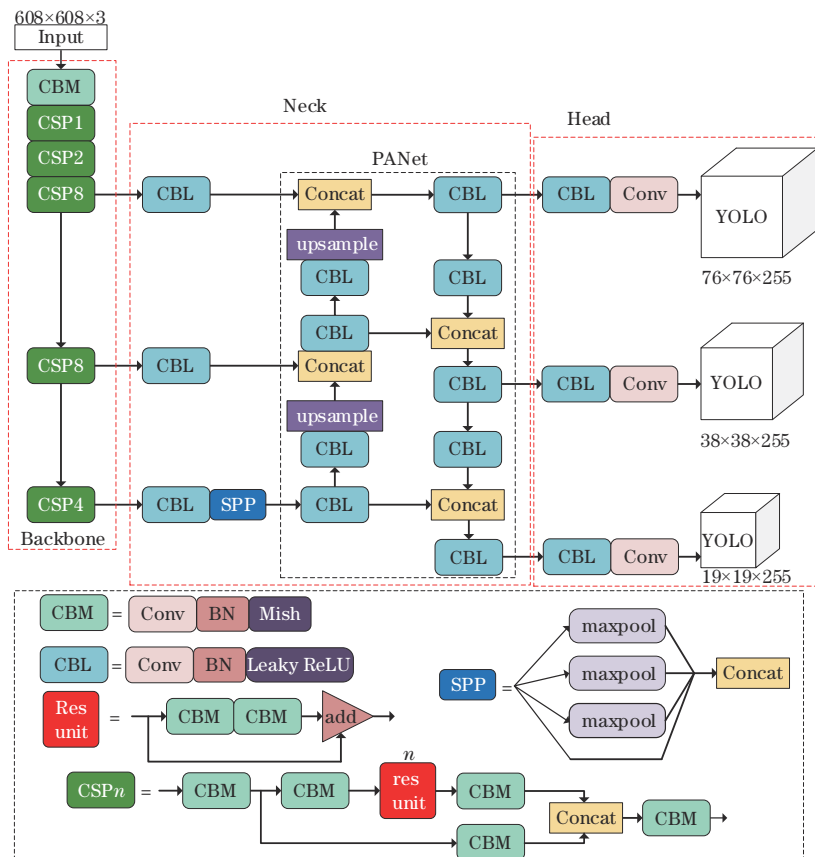


图1 YOLOv4网络结构

Fig. 1 YOLOv4 network structure

(Conv)、批量归一化(BN)和 Mish 激活函数组成; CSP<sub>n</sub>为将 CSP 残差模块重复  $n$  次的大残差模块。

Neck 部分包括 SPP-Net 和路径聚合网络(PANet),其中 CBL 模块由 Conv、BN 和 Leaky ReLU 激活函数组成。SPP 模块用于增加高层网络的感受野,PANet 则在自顶向下和自底向上两个路径方向上对主干特征提取网络生成的三个尺度特征图进行信息融合,得到三种不同尺度的特征图。

Head 部分利用预测值与真实值的交并比(IoU)进行非极大值抑制,筛除无效预测边界框,返回优化结果。

### 3 SMD-YOLOv4 安全帽佩戴检测模型

由于复杂背景的干扰会影响 YOLOv4 检测中小目标安全帽的效果,对此,本文从全局多尺度上下文信息和细节信息方面改进模型,增强模型对复杂背景下样本的区分能力,改善对小目标安全帽的检测效果。图 2 为所提 SMD-YOLOv4 的结构,其中 DBM 模块由深度可分离卷积(Dw)、批量归一化和 Mish 激活函数组成;DBL 模块由深度可分离卷积、批量归一化和 Leaky ReLU 激活函数组成。

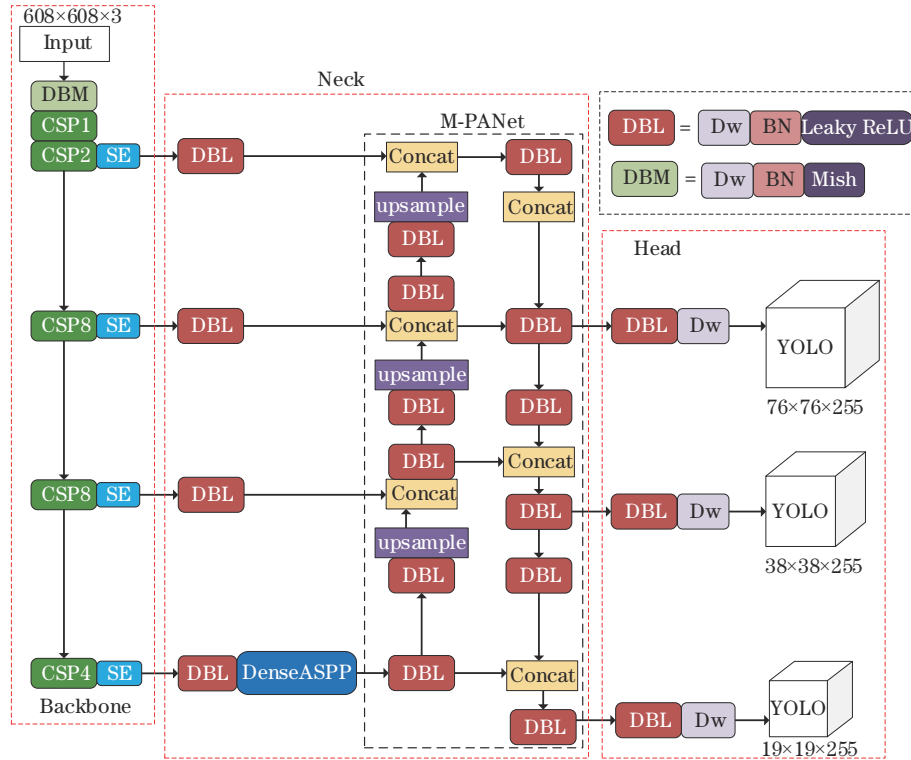


图 2 SMD-YOLOv4 网络结构  
Fig. 2 SMD-YOLOv4 network structure

总体改进思路如下:首先在主干网络末端引入 SE-Net 注意力模块,提高小目标特征的权重占比,用于后续的训练;其次,为弥补高层特征图分辨率降低造成的空间信息丢失,将密集空洞空间金字塔池化用于提取深层语义信息,获取大量有效特征;最后在原有网络基础上增加一个特征融合尺度,形成新的多尺度信息融合检测网络 M-PANet,使网络学习到更多边缘轮廓等细节信息,提高算法在复杂背景下对小目标的检测能力,同时引入深度可分离卷积减少网络中的卷积运算量,避免模型检测速度降低。

#### 3.1 SE-Net 注意力模块

人类可以通过快速扫描全局图像获得需要关注的目标区域,获得关注焦点后,在该区域投入更多的资源从而获得目标的更多细节信息,抑制不太有用的特征,而注意力机制与此相似。在所进行的安全帽佩戴检测

中,注意力机制可以帮助算法更有效地提取重要的特征信息,减少对无效特征信息的关注。因此,本文将 SE-Net 嵌入到 YOLOv4 主干网络末端,增强主干网络对输出特征图的信息整合能力。SE-Net 结构如图 3 所示,主要包括 Squeeze 和 Excitation 部分。

Squeeze 部分通过对特征图进行全局平均池化得到当前特征图各个通道上的全局压缩特征向量;

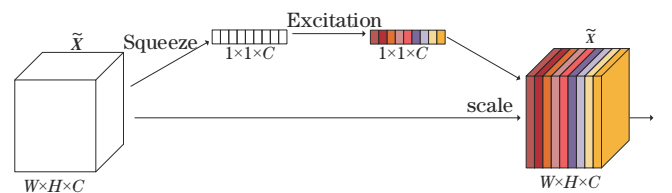


图 3 SE-Net 结构  
Fig. 3 SE-Net structure

Excitation 部分包含两个全连接层,对全局压缩后的特征量进行一次全连接得到低维向量;经过 ReLU 激活函数后再进行一次全连接,将特征向量扩张到原本的维度;随后进行 Sigmoid 激活操作,得到权值在 (0, 1) 之间的权重矩阵;最后进行 scale 操作,即使用权重矩阵对原特征图进行加权,其结果成为网络下一层的输入。SE-Net 通过此方式增加重要特征通道的权重,可提高目标检测网络对细节信息的获取能力,相关数学表达式分别为

$$\mathbf{Z}_c = F_{sq}(\mathbf{X}_c) = \frac{1}{H \times W} \sum_{i=1}^H \sum_{j=1}^W X_c(i, j), \quad (1)$$

$$\mathbf{s} = F_{ex}(\mathbf{z}, \mathbf{W}') = \sigma[g(\mathbf{z}, \mathbf{W}')] = \sigma[W_2 \delta(W_1 \mathbf{z})], \quad (2)$$

$$\tilde{\mathbf{X}}_c = F_{scale}(\mathbf{X}_c, \mathbf{s}_c) = \mathbf{X}_c \cdot \mathbf{s}_c, \quad (3)$$

式中:  $W$  和  $H$  分别为特征图的宽和高;  $\mathbf{Z}_c$  为  $c$  通道上的全局压缩特征向量;  $\mathbf{X}$  为输入特征图;  $F_{sq}$  为 Squeeze 函数;  $F_{ex}$  为 Excitation 函数;  $\sigma$  为 Sigmoid 函数;  $g$  为进行两次全连接操作的函数;  $\delta$  为 ReLU 函数;  $\mathbf{s}$  为得到的权重矩阵;  $F_{scale}$  为通道加权函数;  $\tilde{\mathbf{X}}$  为输出特征图。

### 3.2 特征融合改进

在 YOLOv4 中,输入图像被处理为尺寸为 608×608 的图像。其中 PANet 部分对从主干特征提取网络

中提取的 3 个尺度的特征图进行自上而下和自下而上的反复特征融合,返回 3 个尺度 (19×19、38×38、76×76) 的融合特征图。然而对施工场景下的安全帽佩戴检测而言,检测时有相当一部分目标是小目标, PANet 的融合特征缺乏对浅层特征表示,不利于对小目标的定位和上下文信息的学习。针对这一问题,在 PANet 中融入了浅层特征图的信息。随着特征融合尺度的增加,模型可以对更细粒度的信息进行学习,以提高模型对小目标安全帽的检测能力。虽然这种方式使网络模型得到扩张,增强了网络对细节特征的提取能力,但同时网络的计算量也相应增加,影响了模型的检测速度。为了适当提高模型检测速度,摒弃网络中的普通卷积运算,引入深度可分离卷积。深度可分离卷积由两步卷积运算组成,先进行逐通道卷积(深度卷积)再进行逐点卷积(1×1 卷积),可以减少卷积计算量,提高训练和检测的速度。最终,形成针对施工现场检测安全帽的新多尺度信息融合检测网络(M-PANet),如图 4 所示。其中 P1、P2、P3 分别为原 PANet 所融合的 3 个不同尺度的特征图; P4 为本文融入的浅层特征图,大小为 152×152,包含更多的边缘轮廓信息。可以看到 P4 特征图的融入使网络的三个检测层具有更丰富的细节信息,有利于对小目标的定位。

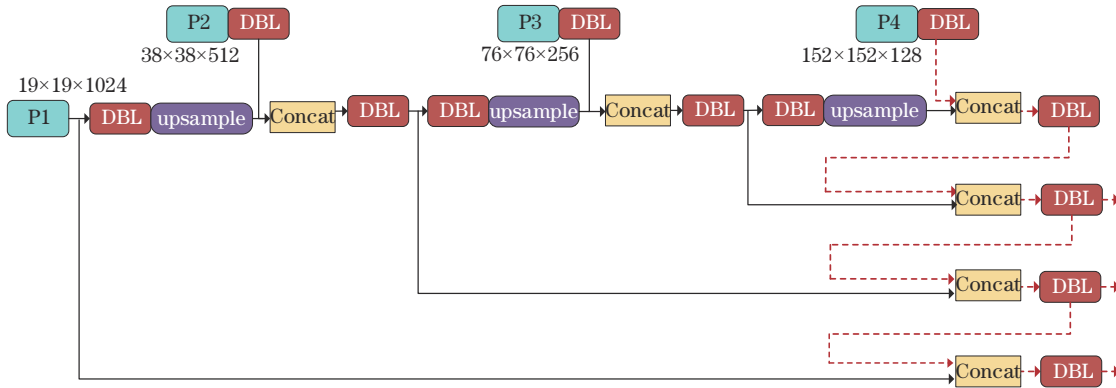


图 4 多尺度信息融合检测网络

Fig. 4 Multi-scale information fusion detection network

### 3.3 空间金字塔改进

YOLOv4 中的空间金字塔池化部分主要是通过池化操作进行特征融合的,进而增加感受野,但是每次的池化操作都会牺牲一定的空间分辨率,多次池化后可能造成大量特征丢失,影响语义信息的提取效果,再对得到的金字塔特征图进行后续的特征融合时,不利于对复杂场景下的目标进行识别和分类。空洞卷积的出现使网络可以在不损失信息的情况下提升感受野;空洞空间金字塔池化(ASPP)<sup>[20]</sup>网络则将不同扩张率

量的像素点,采样不密集导致大量信息丢失,并且当空洞率增加到一定程度时,空洞卷积将变得收效甚微,因此将 DenseASPP 网络引入到模型中。DenseASPP 网络使用不同扩张率 ( $d=3, 6, 12, 18, 24$ ) 的空洞卷积,而且在特征联接方式上不同于 ASPP,它融合了 DenseNet 的思想,将所有的空洞卷积层堆叠并紧密地连接以获得更大的感受野和更多尺度的特征图,形成了一个更密集的特征金字塔,可从特征图中有效获取多尺度的空间上下文信息。图 5 为 DenseASPP 的网络结构,其中  $c$  为 contact 联接操作。

DenseASPP 网络将 6 个不同尺度的特征层并联,首先将输入的特征图直接传到输出位置,然后依次进行 5 种不同扩张率的空洞卷积,而每一个空洞卷积层的输入都是前面所有卷积层的输出和初始输入特征图

空洞卷积的感受野数学表达式为

$$S = (d - 1) \times (k - 1) + k, \quad (4)$$

式中:  $S$  为最终感受野大小;  $d$  为扩张率;  $k$  为卷积核大小。但由于 ASPP 中的空洞卷积每次计算时只选取少

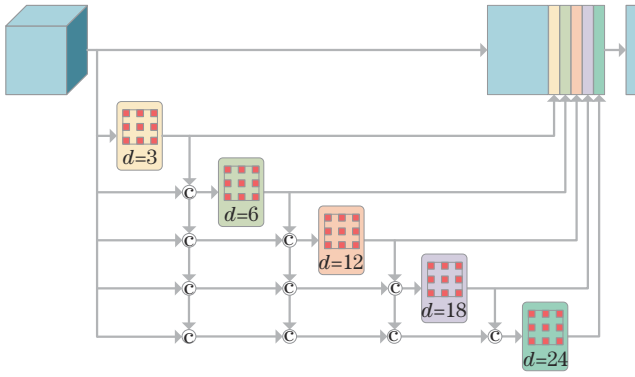


图 5 DenseASPP 网络结构

Fig. 5 DenseASPP network structure

的联接,这种密集连接的方式避免了扩张率过大导致的卷积退化,最终得到大量有效特征。其中,当两个空洞卷积层相连接时,得到的感受野为

$$S = S_1 + S_2 - 1, \quad (5)$$

式中: $S_1$ 、 $S_2$ 分别为两个空洞卷积层的感受野大小; $S$ 为卷积层连接后的感受野大小。

## 4 实验结果与分析

### 4.1 数据集与测评指标

使用 PyTorch 深度学习框架,在服务器上搭建所提安全帽佩戴检测模型。实验环境为:GPU (GeForce\_RTX\_3090, 24G 显存)、Linux 操作系统、Python3.8、CUDA11.1、Ubuntu16.04、内存 125 GB。

为评价所提安全帽佩戴检测方法的有效性,以精确率( $P$ )、召回率( $R$ )和平均精度均值(mAP)为模型的测评指标,具体计算公式为

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

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

$$P_{mAP} = \frac{\sum_{i=1}^k P_{A_i}}{k}, \quad (8)$$

$$P_A = \int_0^1 P(r) dr, \quad (9)$$

式中: $N_{TP}$ 为被模型正确划分为正样本的数量; $N_{FP}$ 为被错误地划分为正样本的数量; $N_{FN}$ 为被错误地划分为负样本的数量; $P_A$ 为单个类别的平均精度,使用积分法计算精确率和召回率的 P-R 曲线与坐标轴所围面积, $P_{mAP}$ 即为所有类别  $P_A$  的均值; $k$ 为检测的总类别数。

所用数据集采集于施工作业现场录像视频截图文件和互联网爬虫数据,总共 6890 张图像。使用 labelImg 对目标进行标注,标注信息分别为正确佩戴安全帽的“hat”和未佩戴安全帽的“person”,并将数据转换为 VOC2007 数据集的格式,以 7.5:1.5:1 的比例将数据集随机划分为训练集、验证集和测试集,并在实验过程中通过模型训练和验证过程中的损失值和 mAP 的变化来判断模型是否收敛,观察模型的训练情况。其中 train loss 为模型训练时的损失值, val loss 为模型验证时的损失值; train mAP 为模型训练时的 mAP 值, val mAP 为模型验证时的 mAP 值。

根据迁移学习<sup>[21]</sup>的思想,主干网络提取的特征是通用的。因此,为加快网络拟合速度,在 SMD-YOLOv4 中加载已经在 ImageNet 数据集上训练好的预训练模型,使用预训练模型的权值信息对 SMD-YOLOv4 的参数进行初始化,提高模型训练效果。随后将训练分为两个阶段以平衡训练资源与训练时间,分别为主干网络冻结训练阶段和全局训练阶段。在冻结训练阶段将更多资源放在特征融合部分的网络参数,从而加快网络训练速度,训练 70 个 epoch 之后将主干网络解冻,对全局网络进行微调,训练 60 个 epoch。学习率的动态调整策略如下:两个训练阶段的初始学习率均为 0.001,衰减系数为 0.92,即每训练 1 轮便将学习率衰减为上一个学习率的 0.92。

经过 130 个 epoch 的迭代, SMD-YOLOv4 模型的训练情况如图 6 所示,图 6(a)为损失值变化情况,图 6(b)为 mAP 值变化情况。可以看出:图 6(a)中模型的损失值在前 20 个 epoch 迅速下降,训练损失值在主干网络解冻时有所波动,在 100 个 epoch 后训练损失与验证损失逐渐趋于稳定,分别稳定在 1.24 和 1.07,模型达到收敛状态;从图 6(b)可以看出,在冻结训练

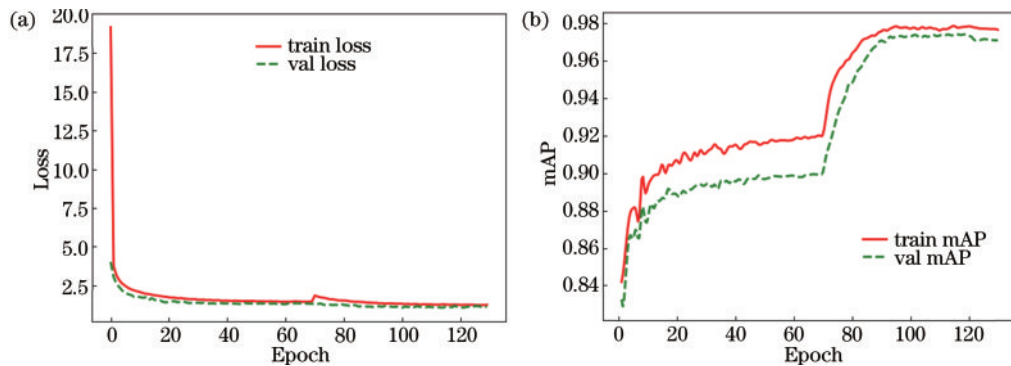


图 6 SMD-YOLOv4 模型训练情况。(a)损失值;(b)mAP 值

Fig. 6 SMD-YOLOv4 model training situation. (a) Loss; (b) mAP

阶段模型训练和验证的 mAP 值差距较大,在 70 个 epoch 时趋于平缓,主干网络解冻后 mAP 值二次大幅提高,在 90 个 epoch 后趋于平缓且模型训练和验证的 mAP 值的差距开始缩小,在 117 个 epoch 时达到了最佳的训练效果,此时 train mAP 为 0.9878, val mAP 为 0.9742,训练结果较为理想。

表 1 消融实验结果对比

Table 1 Comparison of the results of ablation experiments

YOLOv4	SE	Multi-scale	Dw	DenseASPP	P / %	R / %	mAP / %	T / ms
✓					92.30	89.76	94.15	24.7
✓	✓				93.72	88.78	94.74	34.6
✓	✓	✓			94.80	89.60	95.41	51.4
✓	✓	✓	✓		94.51	89.40	95.23	28.5
✓	✓	✓	✓	✓	95.19	93.00	97.34	47.1

由表 1 可以看出:原始 YOLOv4 的精确率、召回率以及平均精度均值分别为 92.30%、89.76% 和 94.15%;以此为基准,基本每个模块的改进都会使测评结果有一定的提高;注意力机制 SE 与多尺度特征融合 Multi-scale 的组合稍微降低了原模型的召回率,但精确率和平均精度均值均得到提高;随着各模块的引入,模型的检测时间也随之增加,深度可分离卷积 Dw 的加入使模型在检测效果稍为降低的情况下大幅减少了检测时间,以平衡模型的检测效果和效率;与原始 YOLOv4 相比,同时引入注意力机制、多尺度特征融合、深度可分离卷积和 DenseASPP 的网络的检测时间稍为增加,但在满足实时检测的条件下模型达到了最优的 95.19% 精确率、93.00% 召回率和 97.34% 平均精度均值,分别比基准模型提高了 2.89 个百分点、3.24 个百分点和 3.19 个百分点,其中召回率有较大提高,说明模型正确检测出目标的能力有了提升。消融实验结果表明,SMD-YOLOv4 能够有效地提高施工现场下安全帽佩戴检测的效果,具有可行性。

## 4.2 消融实验

为验证各项改进对模型检测性能的影响,以原始 YOLOv4 为基准进行消融实验,实验结果如表 1 所示,其中“✓”为使用该部分内容,“T”为每检测一张图片需要的时间。

## 4.3 检测效果对比

为了验证 SMD-YOLOv4 模型的有效性,使用实验数据集中的测试集对模型进行测试,比较 YOLOv4 和 SMD-YOLOv4 在相同图像数据中的检测效果,检测效果对比如图 7 所示。其中,图 7(a)为 YOLOv4 算法的检测效果,图 7(b)为 SMD-YOLOv4 算法的检测效果,检测类别和置信度信息在边界框显示,“h”代表佩戴安全帽的“hat”标签,“p”代表未佩戴安全帽的“person”标签。

相比于图 7(a),图 7(b)中检测出了更多的遮挡目标和较小的目标,且检测到的目标置信度也更高。在第一张密集目标的图片中,YOLOv4 模型漏检了中间靠下方的一个被遮挡的目标;在第二张图片的弱光照环境中,SMD-YOLOv4 则更有效地克服了 YOLOv4 模型的漏检情况;由于拍摄原因,第三张图片中存在不完全拍摄的工人头部图像,SMD-YOLOv4 也能很好地判断出来其是否佩戴安全帽。实验结果表明,SMD-YOLOv4 能够有效地改善在复杂环境下的检测效果。

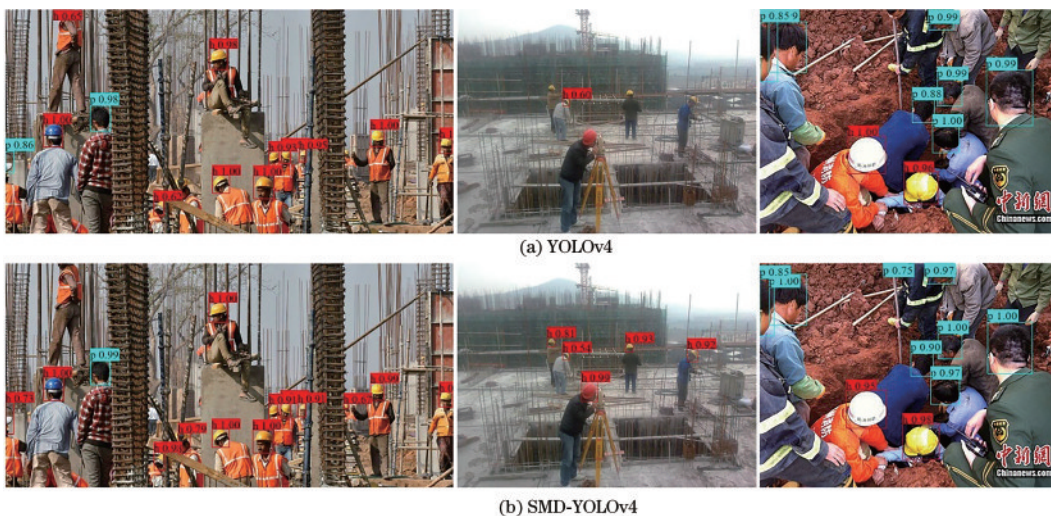


图 7 不同模型的检测效果对比

Fig. 7 Detection effect comparison of different models

#### 4.4 深度学习目标检测算法对比实验

为进一步验证 SMD-YOLOv4 在安全帽佩戴检测方面的有效性,选取了一些有代表性的深度学习目标检测算法进行对比实验。在同样的实验环境下,使用相同的数据集、数据划分策略、模型训练策略,对 SMD-YOLOv4 与 Faster R-CNN、SSD、YOLOv5、YOLOx 目标检测算法进行对比,实验结果如表 2 所示。

表 2 与其他深度学习目标检测算法性能对比

Table 2 Performance comparison with other deep learning target detection algorithms

Model	P / %	R / %	mAP / %	T / ms
Faster R-CNN	57.31	75.31	70.93	40.1
SSD	93.79	76.67	90.90	12.9
YOLOv5	94.81	86.16	94.09	23.7
YOLOx	94.51	91.33	95.85	24.6
SMD-YOLOv4	95.19	93.00	97.34	47.1

从表 2 可以看出:由于本实验数据集的目标中有相当一部分为小目标和密集目标,Faster R-CNN 算法在该数据集上的表现欠佳;YOLOv5 算法与 SSD 算法类似,检测速度较快,且有较高的精确度,但召回率较低,会影响模型在复杂场景下对目标的正确识别,性能上 YOLOv5 比 SSD 算法更佳;YOLOx 很好地平衡了精确度和召回率,达到了较高的 mAP 值;SMD-YOLOv4 在不牺牲精确率的情况下使得召回率得到较大提高,性能方面较 YOLOv5 分别提升了 0.38 个百分点、6.84 个百分点、3.25 个百分点,较 YOLOx 分别提升了 0.68 个百分点、1.67 个百分点、1.49 个百分点。对比可知,虽然 SMD-YOLOv4 的检测时间较其他模型有略微的增加,但仍能满足实时检测的需要,因此 SMD-YOLOv4 能在满足实时检测的要求下有更好的检测效果。

图 8 为各模型在不同场景下的检测效果对比图。

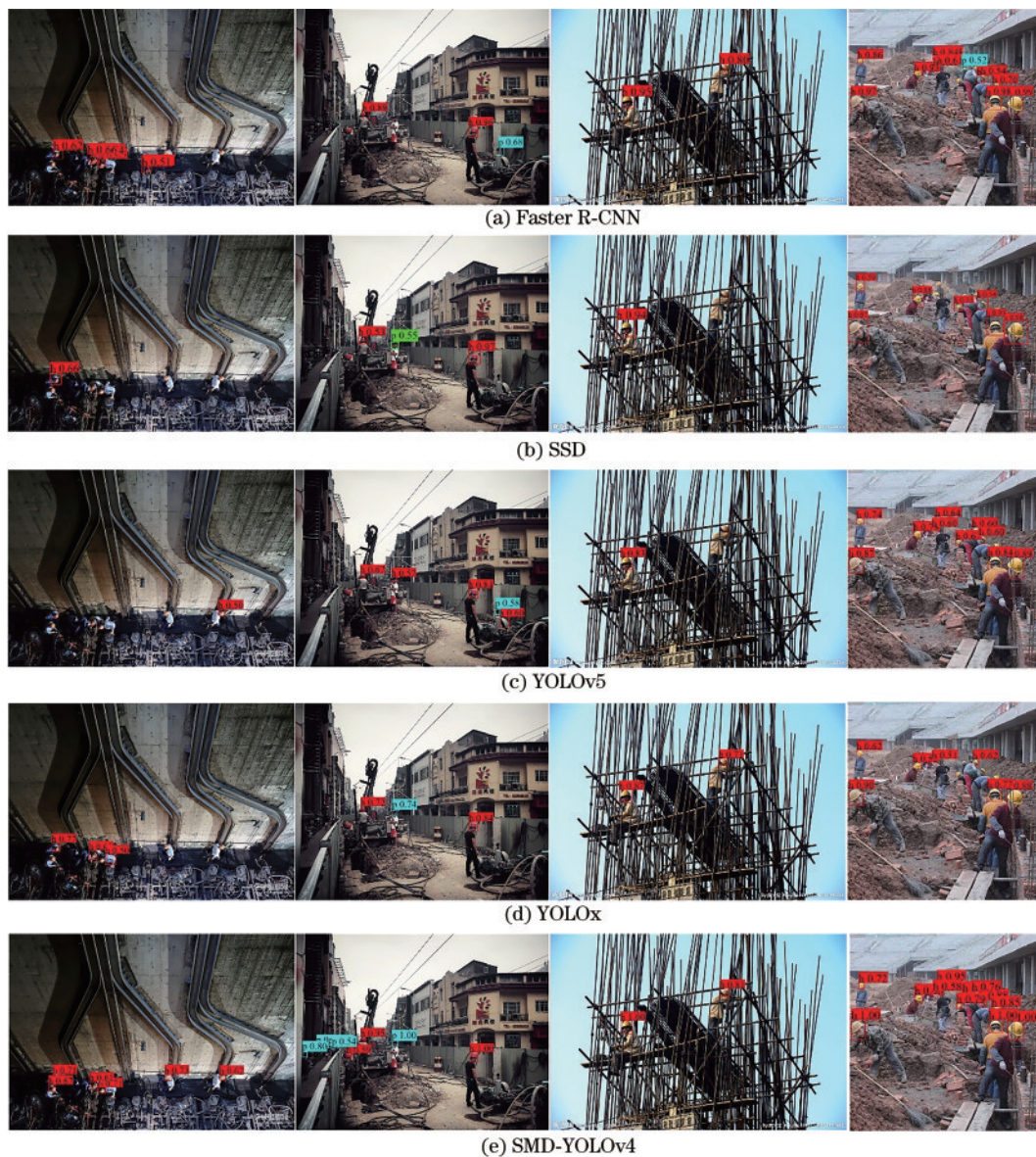


图 8 与各深度学习目标检测算法检测效果对比

Fig. 8 Comparison of detection effect with other deep learning target detection algorithms

第一张为俯视角度的施工图像,可以看到 Faster R-CNN 有很多重叠的检测框,即有较多的错检目标;SSD 和 YOLOv5 在此视角下只正确检测到一个目标;YOLOx 检测到了较多的目标,但是也存在漏检;SMD-YOLOv4 则较好地检测出了图片中的 6 个目标。第二张图的场景为施工街区,SMD-YOLOv4 不仅准确检测到施工区人员的安全帽佩戴情况,也检测到了左边行人的情况,其他模型皆因图片背景复杂且待检目标过小,漏检了目标,且 Faster R-CNN 和 YOLOv5 有明显错检的目标。第三张是工人在脚手架上施工的图像,Faster R-CNN、YOLOx 和 SMD-YOLOv4 在该遮挡场景下有较好的表现,SMD-YOLOv4 以更高的置信值得到了检测结果。第四张图中的目标距离远、密集、尺寸小,SMD-YOLOv4 在该场景下基本检测出了全部目标。总体来说,Faster R-CNN 模型存在较为明显的错检问题,SSD、YOLOv5 和 YOLOx 模型存在一定程度的小目标漏检情况,SMD-YOLOv4 在复杂环境下则具有较好的鲁棒性。

## 5 结 论

为了提升复杂场景下施工作业的安全帽佩戴检测效果,构建了一种 SMD-YOLOv4 目标检测框架。首先在主干网络上引入 SE-Net 注意力模块以提取更多的小目标特征,其次使用 DenseASPP 优化网络高层语义信息的提取,最后在特征融合网络中融入大量细节信息,从而将高层特征图丰富的语义信息与低层特征图的更多细节信息融合,使模型在复杂的环境下有良好的表现,同时以深度可分离卷积替代普通卷积运算,提高网络检测速度。实验结果表明:SMD-YOLOv4 的 mAP 较 YOLOv4 提升了 3.19 个百分点,达 97.34%,召回率提升了 3.24 个百分点,能够有效地对安全帽的佩戴情况进行识别;与 Faster R-CNN、SSD、YOLOv5 和 YOLOx 算法相比,SMD-YOLOv4 检测每张图片的时间比其中检测速度最快的算法 SSD 增加了 34.2 ms,但 mAP 和召回率比其中检测效果最好的算法 YOLOx 分别提高了 1.67 个百分点和 1.49 个百分点,在得到最佳检测效果的同时仍满足实时检测的需求。未来将继续使用单阶段目标检测算法,在保持良好检测效果的情况下对模型检测效率进行进一步优化,为施工作业现场的安全监管提供更多技术参考。

## 参 考 文 献

- [1] 李启月,王靖博,王宏伟,等.工业安全帽抗冲击性能研究[J].中国安全生产科学技术,2021,17(3):182-186.  
Li Q Y, Wang J B, Wang H W, et al. Study on impact resistance of industrial safety helmet[J]. Journal of Safety Science and Technology, 2021, 17(3): 182-186.
- [2] 王宇向,王贞,吴斌,等.智慧工地的安全帽佩戴检测算法研究综述[J].武汉理工大学学报,2021,43(10):56-62.
- [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(3):365-370.  
Liu X H, Ye X N. Skin color detection and Hu moments in helmet recognition research[J]. Journal of East China University of Science and Technology (Natural Science Edition), 2014, 40(3): 365-370.
- [4] 王福斌,刘梦竹, Tu P. 基于级联支持向量机的飞秒激光烧蚀光斑分类[J].中国激光,2021,48(6):0602108.  
Wang F B, Liu M Z, Tu P. Spot ablated by femtosecond laser classification based on cascaded support vector machine[J]. Chinese Journal of Lasers, 2021, 48(6): 0602108.
- [5] Rubaiyat A H M, Toma T T, Kalantari-Khandani M, et al. Automatic detection of helmet uses for construction safety [C]//2016 IEEE/WIC/ACM International Conference on Web Intelligence Workshops (WIW), October 13-16, 2016, Omaha, NE, USA. New York: IEEE Press, 2016: 135-142.
- [6] Fan Z M, Peng C B, Dai L C, et al. A deep learning-based ensemble method for helmet-wearing detection[J]. PeerJ Computer Science, 2020, 6: e311.
- [7] Ren S Q, He K M, 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.
- [8] Liu W, Anguelov D, Erhan D, et al. SSD: single shot MultiBox detector[M]//Leibe B, Matas J, Sebe N, et al. Computer vision-ECCV 2016. Lecture notes in computer science. Cham: Springer, 2016, 9905: 21-37.
- [9] Redmon J, Divvala S, Girshick R, et al. You only look once: unified, real-time object detection[C]//2016 IEEE Conference on Computer Vision and Pattern Recognition, June 27-30, 2016, Las Vegas, NV, USA. New York: IEEE Press, 2016: 779-788.
- [10] Redmon J, Farhadi A. YOLO9000: better, faster, stronger[C]//2017 IEEE Conference on Computer Vision and Pattern Recognition, July 21-26, 2017, Honolulu, HI, USA. New York: IEEE Press, 2017: 6517-6525.
- [11] Redmon J, Farhadi A. YOLOv3: an incremental improvement[EB/OL]. (2018-04-08)[2020-07-24]. <https://arxiv.org/abs/1804.02767>.
- [12] Bochkovskiy A, Wang C Y, Liao H Y M. YOLOv4: optimal speed and accuracy of object detection[EB/OL]. (2020-04-23)[2021-02-04]. <https://arxiv.org/abs/2004.10934>.
- [13] 许凯,邓超.基于改进YOLOv3的安全帽佩戴识别算法[J].激光与光电子学进展,2021,58(6):0615002.  
Xu K, Deng C. Research on helmet wear identification based on improved YOLOv3[J]. Laser & Optoelectronics Progress, 2021, 58(6): 0615002.
- [14] Zhou M D, Fang Z L, Zhao B, et al. Safety helmet wearing detection and recognition based on YOLOv4 [C]//2021 3rd International Academic Exchange Conference on Science and Technology Innovation (IAECST),



- December 10-12, 2021, Guangzhou, China. New York: IEEE Press, 2021: 798-802.
- [15] Deng B Y, Lei X C, Ye M. Safety helmet detection method based on YOLO v4[C]//2020 16th International Conference on Computational Intelligence and Security (CIS), November 27-30, 2020, Guangxi, China. New York: IEEE Press, 2020: 155-158.
- [16] 杨永波, 李栋. 改进 YOLOv5 的轻量级安全帽佩戴检测算法[J]. 计算机工程与应用, 2022, 58(9): 201-207.  
Yang Y B, Li D. Lightweight helmet wearing detection algorithm of improved YOLOv5[J]. Computer Engineering and Applications, 2022, 58(9): 201-207.
- [17] 曹捷, 郭志彬, 潘立志, 等. 高空作业场景下的安全带穿戴检测[J]. 湖南科技大学学报(自然科学版), 2022, 37(1): 92-99.  
Cao J, Guo Z B, Pan L Z, et al. Detection of safety belt wearing in aerial work[J]. Journal of Hunan University of Science and Technology (Natural Science Edition), 2022, 37(1): 92-99.
- [18] Hu J, Shen L, Sun G. Squeeze-and-excitation networks [C]//2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, June 18-23, 2018, Salt Lake City, UT, USA. New York: IEEE Press, 2018: 7132-7141.
- [19] Yang M K, Yu K, Zhang C, et al. DenseASPP for semantic segmentation in street scenes[C]//2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, June 18-23, 2018, Salt Lake City, UT, USA. New York: IEEE Press, 2018: 3684-3692.
- [20] Chen L C, Papandreou G, Kokkinos I, et al. DeepLab: semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected CRFs[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2018, 40(4): 834-848.
- [21] 雷相达, 王宏涛, 赵宗泽. 整合迁移学习与全卷积网络的小样本机载激光雷达点云分类[J]. 中国激光, 2021, 48(16): 1610001.  
Lei X D, Wang H T, Zhao Z Z. Small-sample airborne LiDAR point cloud classification based on transfer learning and fully convolutional network[J]. Chinese Journal of Lasers, 2021, 48(16): 1610001.