

doi:10.11835/j.issn.1000.582X.2023.12.005

基于改进 YOLOv4 颈部优化网络的安全帽佩戴检测方法

徐先峰, 王 轲, 马志雄, 姚景杰, 赵万福

(长安大学能源与电气工程学院, 西安 710016)

摘要:针对安全帽佩戴检测时易受复杂背景干扰, 解决 YOLOv4 网络检测速度慢、内存消耗大、计算复杂度高、对硬件性能要求较高等问题, 引入改进 YOLOv4 算法优化安全帽佩戴检测方法。引入 MobileNet 网络轻量化 YOLOv4、跨越模块特征融合, 实现高层语义特征和低层语义特征有效融合。针对图像中小目标分辨率低, 信息特征少, 多尺度并存, 导致在连续卷积过程中易丢失特征信息等问题, 采用改进特征金字塔 FPN 和注意力机制等颈部优化策略聚焦目标信息, 弱化安全帽检测时背景信息的干扰。仿真结果表明, 基于改进的 YOLOv4 颈部优化网络安全帽佩戴检测算法在 CPU 平台下的检测速度为 34.28 FPS, 是基础 YOLOv4 网络的 16 倍, 检测精度提升了 4.21%, 检测速度与检测精度达到平衡。

关键词: 安全帽佩戴检测; YOLOv4 算法; 检测精度; 检测速度

中图分类号: TP391

文献标志码: A

文章编号: 1000-582X(2023)12-043-12

Helmet wearing detection method based on improved YOLOv4 neck optimized network

XU Xianfeng, WANG Ke, MA Zhixiong, YAO Jingjie, ZHAO Wanfu

(School of Energy & Electronic Engineering, Chang'an University, Xi'an 710016, P. R. China)

Abstract: Helmet wearing detection is susceptible to interference from complex background. To address this challenge, the YOLOv4 algorithm is employed. However, YOLOv4 faces issues such as slow detection speed, high memory consumption, computational complexity and demanding hardware performance requirements. This study integrates the MobileNet network to alleviate these challenges in YOLOv4. Additionally, cross-module feature fusion is introduced in MobileNet network, enabling effective fusion of high-level semantic features and low-level semantic features. Despite these advancements, small targets in images poses problems such as low resolution, limited informative features, coexistence of multiple scales, and potential loss of feature information during continuous convolution. To mitigate these issues, this paper introduces neck optimization strategies, such as improving the feature pyramid FPN and introducing/improving attention mechanism. These strategies focus on

收稿日期: 2022-07-02 网络出版日期: 2023-08-30

基金项目: 陕西省重点研发计划资助项目 (2021GY-098); 长安大学中央高校基本科研业务费专项资助项目 (300102321504, 300102321501, 300102321503)。

Supported by the Key Research and Development Program of Shaanxi Province (2021GY-098) and the Fundamental Research Funds for the Central Universities (300102321504, 300102321501, 300102321503).

作者简介: 徐先峰 (1982—), 男, 副教授, 博士, 主要从事信号处理、深度学习理论及应用、智能电网等方向研究, (E-mail) xxf_chd@163.com。

target information and reduce interference from background information during helmet detection. Simulation results show that the improved YOLOv4 neck-optimized network achieves a detection speed of 34.28FPS on the CPU platform, which is about 16 times faster than YOLOv4 network. Moreover, its detection accuracy is 4.21% higher than that of the YOLOv4 algorithm. This optimized algorithm strikes a balance between speed and accuracy.

Keywords: safety helmet wear detection; detection precision; YOLOv4; detection speed

建筑施工安全是建筑施工现场作业的生命线^[1]。为了能够对施工人员安全帽佩戴进行全天候、无死角监督^[2-3]；高准确性、高实时性满足复杂背景安全帽检测算法至关重要。

为实现安全帽佩戴自动检测, 研究人员进行大量研究, 但基于传统安全帽佩戴检测主要通过捕获边缘、颜色等特征^[4-5], 对于复杂施工环境下检测效果一般, 且效率低、耗时长, 无法真正应用到实际施工现场, 这使准确、高效实现安全帽佩戴检测成为一项艰巨任务。随着人工智能技术发展, 基于深度学习方法逐渐成为目标检测的主流方法^[6-7], YOLOv4^[8]作为经典基于深度学习的目标检测算法之一, 在处理复杂背景下安全帽检测时表现出优异的检测效果, 但存在训练步骤繁琐、测试训练速度慢、占用计算机内存空间大等缺点。轻量级网络 MobileNet-SSD^[9]充分利用密集块中卷积层产生的输出特征图, 重复使用这些特征, 降低了参数和计算成本, 然而其易出现漏检、错检现象。

研究的主要贡献为: ①充分结合深度学习方法和轻量级网络 MobileNet-SSD 两者的优点, 在 YOLOv4 网络基础上对其改进, 高效准确实现安全帽的检测与识别; ②引入轻量级 MobileNet 网络加快安全帽的佩戴检测速度, 解决 YOLOv4 算法参数量巨大、检测速度缓慢等问题。进一步改进 MobileNet 网络结构, 将浅层语义特征进行加强; ③通过改进特征金字塔和引入并改进注意力机制等颈部优化策略聚焦目标信息, 解决图像中小目标分辨率低, 信息特征少, 多尺度并存, 小目标在连续卷积过程中容易丢失特征信息等问题; ④基于改进 YOLOv4 算法的安全帽佩戴监测方法实现了在极端可视化条件和极端小目标情况下均对安全帽佩戴的快速准确检测, 并且检测速度与检测精度达到平衡。

1 安全帽佩戴检测算法

1.1 YOLOv4 模型算法

YOLOv4 是经典的目标检测算法, 主要是对分帧标注后的图片进行特征提取, 利用回归思想对目标进行分类和位置定位, 最终得到检验框位置、类别及置信度。YOLOv4 在复杂背景下具有优异的检测效果, 适用于安全帽佩戴场景检测^[10]。

考虑到实际应用中的视频监控设备多使用 CPU 平台, 不具备 GPU 的并行加速计算能力, 在 CPU 平台下 YOLOv4 算法检测速度缓慢, 时效性差, 难以满足实时性要求, 在 YOLOv4 算法基础上引入 MobileNet 轻量级网络并对其进行改进。轻量化 YOLOv4 网络可有效加快检测速度, 但对于监控视频中含有多夜间极端条件和极端小目标时易造成安全帽佩戴的漏检、错检等现象。因此在轻量化网络基础上, YOLOv4 颈部网络中改进特征金字塔 FPN 设计了信息连接层, 小目标特征信息在提取过程中保持信息完整, 获得更好鲁棒性的语义信息, 改进注意力机制聚焦目标信息, 弱化安全帽检测时背景信息的干扰。

1.2 YOLOv4 模型算法

按照改进顺序依次对 MobileNet 网络、新型特征金字塔和轻量双注意力机制进行介绍。将其称为基于改进的 YOLOv4 颈部优化网络算法, 其网络结构如图 1 所示。

1.2.1 MobileNet 轻量化分析

MobileNet 是一种兼备检测精度和检测速度的轻量级神经网络, 通过构建深度可分离卷积改变网络计算方式, 降低网络参数量、模型复杂度, 提高模型的检测速度。深度可分离卷积是实现轻量级神经网络最关键一步。深度可分离卷积将标准的卷积分解成逐点卷积和深度卷积^[10-11]。

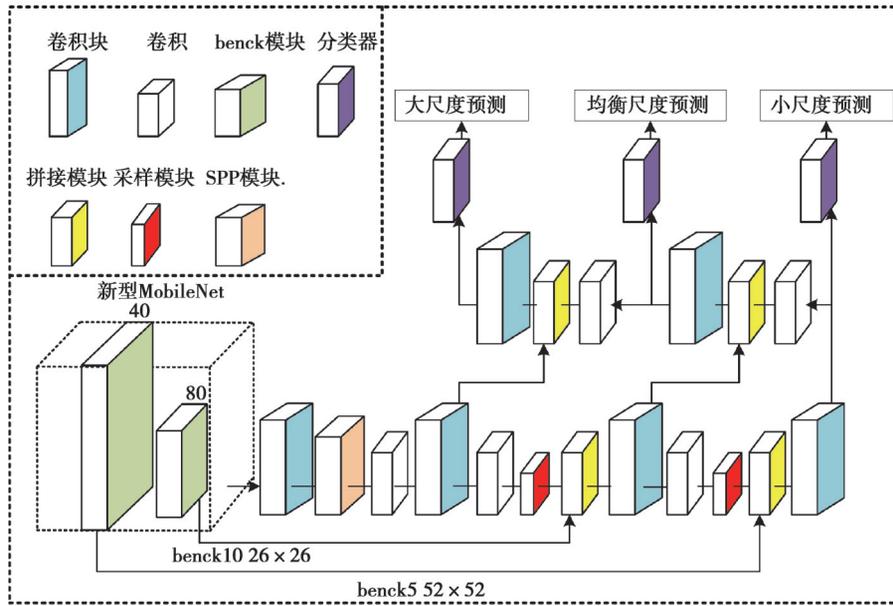


图 1 基于改进的 YOLOv4 颈部优化算法网络结构

Fig. 1 Network structure based on improved YOLOv4 neck optimization algorithm

假设输入特征图大小为 $D_F \times D_F \times M$, 其输出特征图大小为 $D_F \times D_F \times N$, 其中: D_F 表示特征图的高和宽, M 和 N 表示特征图的通道数。

$$S_a = D_K \times D_K \times M \times N \times D_F \times D_F, \tag{1}$$

$$S_c = D_K \times D_K \times M \times D_F \times D_F, \tag{2}$$

$$S_d = D_F \times D_F \times N \times M, \tag{3}$$

$$S_b = S_c + S_d = D_F \times D_F \times N \times M + D_K \times D_K \times M \times D_F \times D_F, \tag{4}$$

$$\frac{S_b}{S_a} = \frac{D_K \times D_K \times M \times D_F \times D_F + D_F \times D_F \times N \times M}{D_K \times D_K \times M \times N \times D_F \times D_F} = \frac{1}{N} + \frac{1}{D_K^2}, \tag{5}$$

其中: S_a 为传统卷积参数的总数; 深度可分离卷积的总计算量 S_b 为深度卷积计算量 S_c 与逐个卷积计算量 S_d 之和; S_c 为深度卷积计算量; S_d 为卷积计算量; S_b/S_a 为深度可分离卷积总计算量与传统卷积计算量之比。明显深度可分离卷积的计算量比传统卷积核小很多, 卷积核个数 N 越大, 计算量更大。

YOLOv4 网络中的基础网络 CSPDarknet53 是传统卷积网络, 其一层的计算量经公式 (1) 得到, 经计算 CSPDarknet53 的参数量为 27.6×10^7 。对于 MobileNet 网络中采用深度可分离卷积, 每一层对应的计算量由公式 (4) 计算, 经计算 MobileNet 网络的参数量为 4.2×10^7 。利用 MobileNet 网络替换 YOLOv4 的 CSPDarknet53 网络, 参数量下降为原来的 $\frac{7}{46}$ 。YOLOv4 网络更加轻量化, 有效减少内存消耗, 为 YOLOv4 安全帽佩戴检测算法在 CPU 平台应用提供了可能。

在目标检测领域, 一张图片可以产生数以千计的预测样本, 其中大部分预测样本与目标检测的背景有关, 称其为负样本, 只有一小部分样本与检测目标有关为正样本, 为减少随机预测框数量, 在 MobileNet 网络基础上创新性引入跨越模块 (crossing block), 跨越模块由卷积层和池化层组成, 如图 2 所示, 其主要通过增加神经网络中独特描述特征的数量, 降低计算复杂度。在跨越模块中为了保留目标的显著特征和减小空间大小, 首先对特征图进行全局平均池化, 然后经批量归一化、Relu 激活函数、卷积层等操作。批量处理归一化 (BN) 和 ReLU 非线性应用在每个卷积层之后, 除了最后一个没有使用 ReLU 的卷积层。选择全局平均池化是因为直接实现了降维, 极大减少网络参数量, 并对局部信息进行整合。

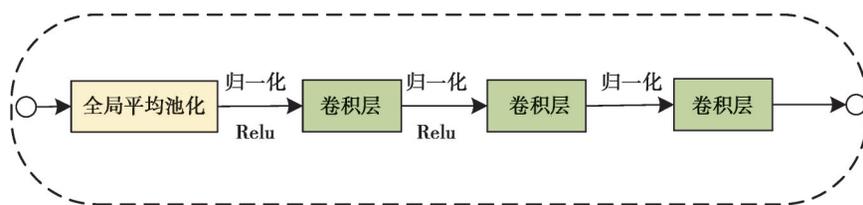


图 2 跨越模块

Fig. 2 crossing block

在 MobileNet 网络的结构基础上引入跨越模块称其为改进的 MobileNet 网络,其网络结构如图 3 所示。在第一个卷积块之后插入 2 个跨越模块,直接将浅层语义信息直接传递到第 6 个 benck 模块和第 10 个 benck 模块,对浅层语义特征进行加强。跨越模块将浅层语义特征直接从第一层传递到网络最后一层。在神经网络中,层空间大小逐渐减小,无法直接进行连接。利用 add 方式将不同空间大小的层连接起来,利用从第一层提取的语义特征丰富最后一层。add 方式在保证维度不变情况下,每一维的信息量增多,增强目标检测的分类效果。

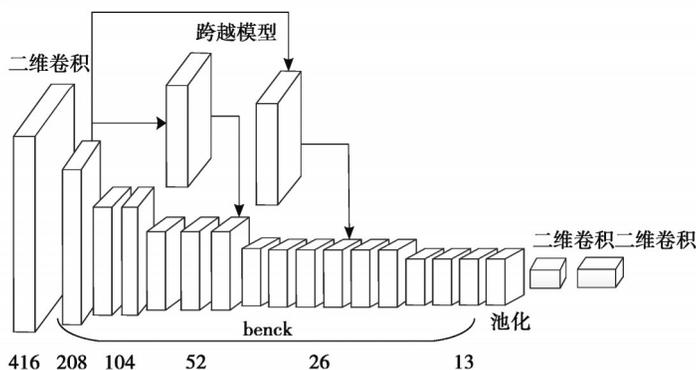


图 3 改进的 MobileNet 网络结构

Fig. 3 Improved MobileNet network architecture

1.2.2 新型特征金字塔

特征金字塔 FPN 提出使多尺度特征融合技术在计算机视觉领域得到广泛应用,特征提取能力显著加强。研究人员基于 FPN 特征网络结构进行跨尺度特征融合的大量研究。FPN 特征金字塔^[12]利用浅层语义特征区分简单目标,深层语义特征区分复杂目标,PANet^[13]在特征金字塔 FPN 的基础上增加了额外路径聚合层,其双向融合方式较为简单。引入可学习权重对不同特征层的特征信息进行加强,重复利用高层语义特征和低层位置特征,形成双向特征金字塔 BiFPN^[14],如图 4(a)所示。

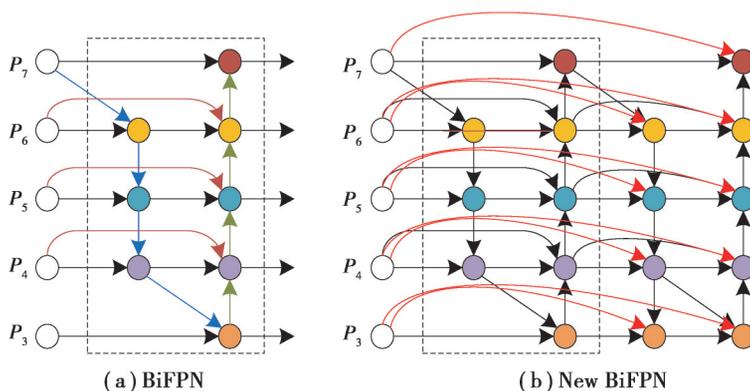


图 4 特征金字塔

Fig. 4 Feature Pyramid Networks

安全帽佩戴检测过程中需要利用摄像头拍摄到的监控画面,摄像头距离施工人员的远近导致拍摄目标大小形状往往各不相同。为了解决多尺度以及在特征提取时浅层语义特征信息丢失问题,笔者研究了一种新的基于双向尺度连接特征融合的新型特征金字塔,如图 4(b)所示该网络具有信息直接连接层和浅层信息融合层。针对小目标在连续卷积过程中特征信息丢失问题,在特征融合网络中设计了信息连接层,直接在双向跨度特征金字塔的每个节点迭代输出与初始输入特征映射进行特征融合,使小目标特征信息在提取过程中保持信息完整。

使用多尺度目的是为了融合不同分辨率下特征图信息。给定一组多尺度输入 $P^m = (P_1^m, P_2^m, \dots, P_n^m)$, 其中 P_i^m 表示第 i 层的特征信息。多尺度融合方法其公式为

$$P_3^{\text{out}} = \text{conv}(P_3^{\text{in}}), \quad (6)$$

$$P_2^{\text{out}} = \text{conv}(P_2^{\text{in}} + \text{Resize}(P_3^{\text{out}})), \quad (7)$$

$$P_1^{\text{out}} = \text{conv}(P_1^{\text{in}} + \text{Resize}(P_2^{\text{out}})), \quad (8)$$

其中:conv 表示卷积操作;Resize 表示为了匹配特征尺度进行的上采样或下采样操作。

使用特征加权融合方法对不同特征尺度的特征层进行融合,对每个输入添加额外的权重,使得网络对不同特征输入进行调整与融合,其融合公式如(9)所示

$$O = \sum_i \frac{\omega_i}{\varepsilon + \sum_j \omega_j} \cdot I_i, \quad (9)$$

其中: ω_i 是一个可学习的权重,其值位于 0 和 1 之间。通过添加一个很小的 ε 来保证数值的稳定性; I_i 表示第 i 个特征输入量。

如图 4(a)所示,其中虚线表示特征金字塔的基本单元。对其中 P_4 层为例进行迭代,其中包括 2 层,中间层和输出层,如公式(10)、(11)所示

$$P_4^{\text{id}} = \text{conv}\left(\frac{\omega_1 \cdot P_4^{\text{in}} + \omega_2 \cdot \text{Resize}(P_3^{\text{in}})}{\omega_1 + \omega_2 + \varepsilon}\right), \quad (10)$$

$$P_4^{\text{out}} = \text{conv}\left(\frac{\omega_1^! \cdot P_4^{\text{in}} + \omega_2^! \cdot P_4^{\text{id}} + \omega_3^! \cdot \text{Resize}(P_3^{\text{out}})}{\omega_1^! + \omega_2^! + \omega_3^! + \varepsilon}\right), \quad (11)$$

其中: P_4^{id} 表示 P_4 层中间层的输出; ω_i 和 Resize 跟前文一致; P_4^{out} 表示 P_4 输出层的输出。

新型特征金字塔是在 BiFPN 特征金字塔基础上设计了信息直接连接层,如图 4(b)中红线所示将每个节点的迭代输出与初始输入特征映射进行特征融合。利用信息直接连接层在双向跨尺度连接特征融合金字塔的每个迭代输出节点上与初始输入特征映射进行特征融合,使小目标特征信息在特征提取过程中保持完整。计算方法如公式(12)所示

$$P_4^{\text{out}} = \text{conv}\left(\frac{\omega_1^! \cdot P_4^{\text{in}} + \omega_2^! \cdot P_4^{\text{id}} + \omega_3^! \cdot \text{Resize}(P_3^{\text{out}}) + \omega_4^! \cdot P_4^{\text{in}}}{\omega_1^! + \omega_2^! + \omega_3^! + \omega_4^! + \varepsilon}\right). \quad (12)$$

1.2.3 轻量双注意力机制

注意力机制在目标检测领域得到广泛应用,其能够聚焦目标信息,降低背景信息对检测效果的影响,有效提高神经网络对于目标的识别能力。为充分考虑安全帽检测的准确性和实时速度,利用通道注意力和空间注意力机制模型,笔者研究了一种轻量双注意力机制模型 LDAM(lightweight dual attention module),如图 5 所示。LDAM 借鉴了 DAM^[15]和 CBAM^[16]的优势,在有效性跟效率性之间取得平衡。

轻量双注意力机制依次通过通道注意力机制和空间注意力机制,在进行聚焦目标信息的基础上,没有产生额外的权重和计算成本。图 7 红色虚线框是一种注意力机制全局池化模型。其主要应用有两方面:用于全局空间上下文池化;用于全局通道上下文池化。

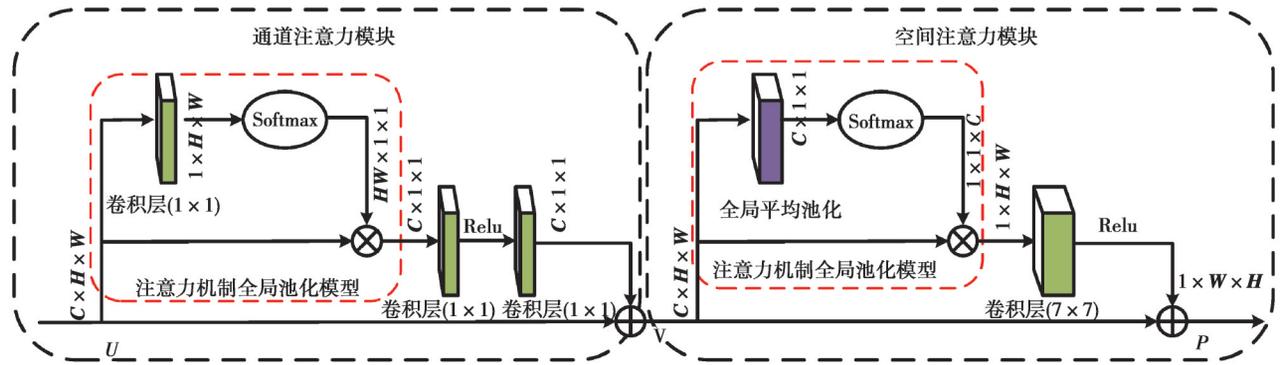


图5 LDAM模型

Fig. 5 Lightweight Dual Attention Module

给定输入特征图 $U \in R^{C \times W \times H}$, 注意力机制全局池化模型通过 Softmax 计算得到自注意力矩阵 M_1 , 如式(13)

$$M_1 = \text{Softmax}(W_1, U), \quad (13)$$

其中: W_1 代表权重向量; Softmax 表示归一化函数。

最后自注意力矩阵 M_1 与输入特征向量 U 进行相乘, 得到向量 a

$$a = U \cdot M_1, \quad (14)$$

其中: a 代表每个通道的全局信息; 通道注意力输出特征图 $V \in R^{C \times W \times H}$ 经过全局平均池化网络, 后经过 Softmax 计算得到每个空间的全局信息 β 。

轻量双注意力机制模型依次通过通道注意力机制和空间注意力机制可将数据量从 $O(CWH)$ 减少到 $O(WH)$, 有效提高目标检测速度。

2 实验结果与分析

2.1 数据集构建

实际工程应用信号运行环境更加复杂, 所采集振动信号噪声干扰更强, 针对目前安全帽数据集规模较小, 为了正负样本保持平衡, 保证数据样本多样性, 依托互联网平台大量采集安全帽图像, 利用视频截图软件 YoloMark 对人员流动性大、建筑物遮挡、夜间可视化条件差等复杂施工背景下的视频进行分帧截取图片, 截图的图片覆盖了复杂背景下的各种施工人员佩戴安全帽情形。数据集中包含图像 10 032 张, 其中包括安全帽目标 13 909 个, 未佩戴安全帽人脸目标 112 728 个。

将数据集随机划分为 2 部分: 训练集和测试集。其中训练集为 4 636 张安全帽图像和 3 396 张人脸图像, 测试集为 1 000 张安全帽图像和 1 000 张人脸图像。



图6 真实场景下的检测图像

Fig. 6 Detection image in real scene

2.2 模型训练

为验证基于改进的 MobileNet-YOLOv4 安全帽佩戴检算法,获得最优模型,笔者迭代轮次设置为 100 步, IOU 阈值设置为 0.5。模型训练 100 个迭代轮次(Epoch),前 50 个迭代轮次引入冻结训练来加快训练速度,防止训练初期权值被破坏。刚开始训练时,为使损失函数快速下降,批量大小(Batch size)设置为 16,一个迭代轮次迭代次数为 383。后 50 个轮次解除冻结进行全网络训练。为防止错过最优点,批量大小为 8,一个迭代轮次迭代次数为 767。参数设置如表 1 所示。

表 1 基础参数表

Table 1 Base parameter table

参数	Value
批量大小	16.000 00
输入大小	416×416
迭代轮次	100.000 00
学习率	0.005 89
衰减系数	0.000 44
动量系数	0.903 00
IOU 阈值	0.500 00

2.3 评价指标

通常衡量目标检测算法性能指标主要围绕检测速度和检测精度 2 个方面。检测任务的精度用平均精确度(mAP)及精确率与召回率的调和平均数(F1)等评价指标来衡量;1 s 处理图片的数量(FPS),其用于评价算法的检测速度。主要依据这 4 个指标对安全帽佩戴检测算法进行评价。

1) mAP

mAP 表示平均准确精度,通常需要以下 3 个指标进行综合评价:召回率 Recall、准确率 Precision、平均精度 Average precision,计算公式如(15)、(16)所示。

$$\text{Recall} = \frac{TP}{TP + FN}, \quad (15)$$

$$\text{Precision} = \frac{TP}{TP + FP}, \quad (16)$$

其中:TP(true positive)即模型预测正确的正样本;FP(false positive)代表模型预测错误的正样本;FN(false negative)代表模型预测错误的负样本;TN(true negative)代表模型预测正确的负样本。召回率表示在所有标注中正样本所占的比重,准确率表示模型预测为正样本占总体正样本的比重。Precision-recall 曲线(PR 曲线),是以 Recall 为横轴、Precision 为纵轴,反映了分类器对正样本的识别准确度和对正样本的覆盖能力之间的权衡。AP 为 PR 曲线与 X 轴围成图形的面积。

对于连续的 PR 曲线,如公式(17)所示

$$AP = \int_0^1 PR dr. \quad (17)$$

对于离散的 PR 曲线

$$AP = \sum_{k=1}^n P(k) \Delta r(k). \quad (18)$$

每种类型的目标都对应着一个 AP, mAP 是所有 AP 的平均值

$$\text{mAP} = \frac{\sum_{i=1}^n AP_i}{n}. \quad (19)$$

2) $F1$

$F1$ 是准确率和召回率的调和平均,如公式(20)所示。

$$F1 = 2 \cdot \frac{\text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}} \quad (20)$$

$F1$ 的取值范围在 0 和 1 之间,取值越大表明检测精度越理想。

3) FPS

FPS(frame per second)即 1 s 可以处理图片的数量。不同的检测平台其性能各不相同,因此评估 FPS 参数时必须要在同一设备上进行测试。可以对单位时间内处理的图片数量进行统计,数量越多,表示速度越快,也可以对处理单个图片所需的时间来测量检测速度,时间越短,表示速度越快。

2.4 结果分析

为验证研究算法在复杂工地背景下检测性能的优越性,对测试集进行测试,统计 2 类目标安全帽和未佩戴安全帽的人脸检测精度 AP ,其中以召回率 Recall 为横轴,准确率 Precision 为纵轴,测试结果如图 7 所示。

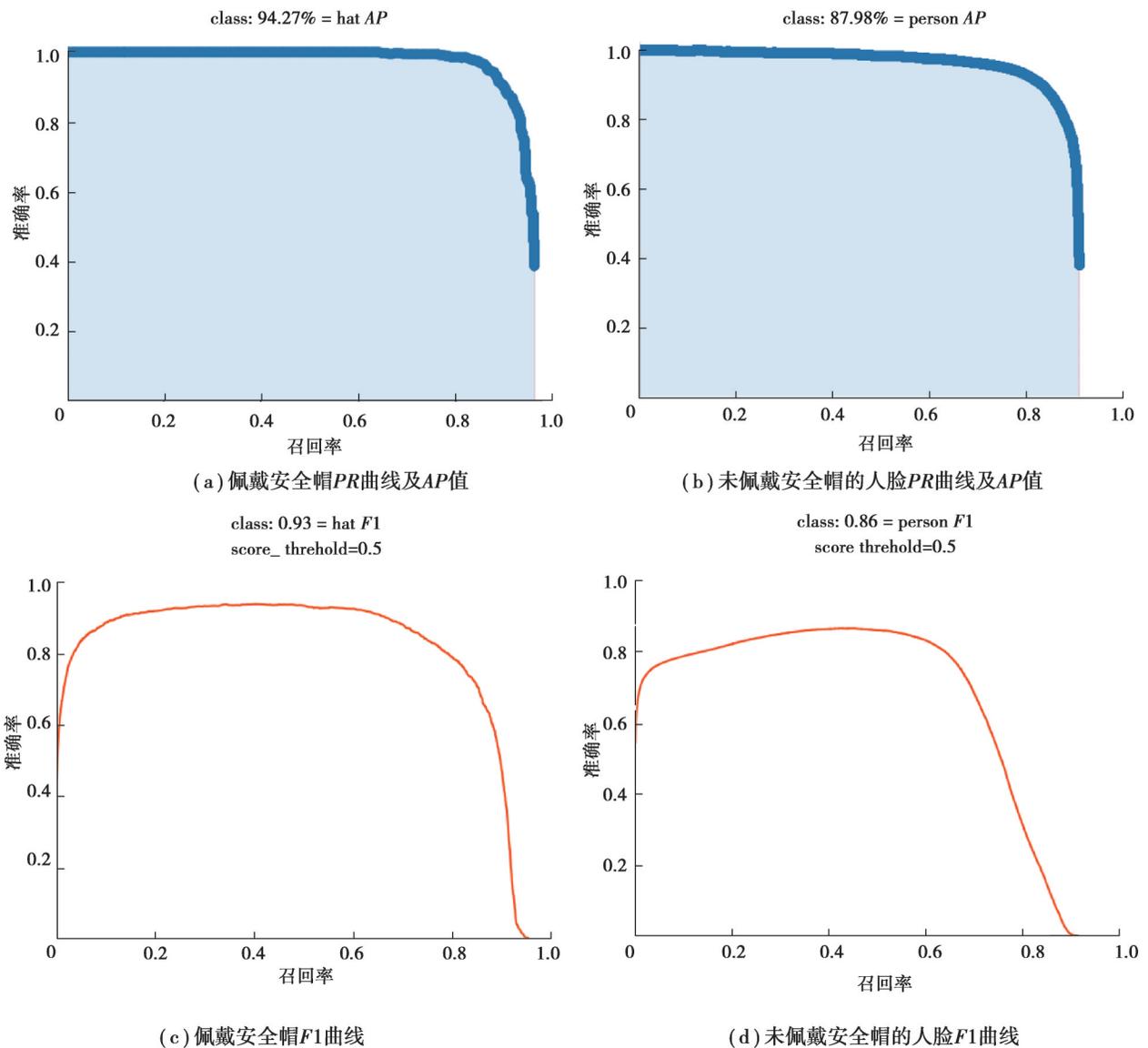


图 7 两类目标 PR 曲线、 $F1$ 曲线及 AP 值

Fig. 7 Two kinds of target PR curve, $F1$ curve and AP value

由图 7 可知,安全帽的 AP 为 94.27%,其 AP 大于 90%,精度较高。未佩戴安全帽的人脸的 AP 为 88.98%,整体模型的精度为 91.12%,佩戴安全帽的 $F1$ 值为 0.93,未佩戴安全帽的人脸 $F1$ 值为 0.86。

除了对检测精度 AP 进行检测,还需对检测速度 FPS 进行探究。使用 GPU 型号为 GTX 1080Ti,有效提高计算机计算性能,在实际应用场景中,视频监控一般使用普通的 CPU 设备,不具备 GPU 的并行加速计算能力。在其他硬件平台相同下,分别在使用 GPU 和 CPU 设备下对最终模型的检测速度进行统计,基于改进的 YOLOv4 颈部优化网络算法在不同平台下的检测速度如表 2 所示。

表 2 改进的 YOLOv4 颈部优化网络在不同平台下检测速度比较
Table 2 Improved YOLOv4 neck optimization network compares the detection speed under different platforms

平台	FPS
CPU	34.28
GPU	85.86

与 YOLOv4 在 CPU 平台下的速度(2.14 FPS)对比,基于改进的 YOLOv4 颈部优化网络的检测速度是其 16 倍左右,同时检测精度相比于 YOLOv4 算法检测精度提升了 4.21%,基于改进的 YOLOv4 颈部优化网络无论在 CPU 还是 GPU 平台下,检测速度都超过 24 FPS,满足实时性要求。因此,基于改进的 YOLOv4 颈部优化网络在保证检测精度的同时,在 CPU 平台下检测速度达到 34.28 FPS,使基于改进的 YOLOv4 颈部优化网络在 CPU 平台上顺利实现,在检测精度和检测速度 2 个方面均表现出良好检测性能。

为进一步验证提出算法具有较高检测精度和检测速度,使用相同的测试数据集在 Faster-RCNN、SSD 等目标检测的经典算法进行对比,其中 YOLOv4、Faster R-CNN、基于 YOLOv4 的颈部优化安全帽佩戴检测等算法的损失随迭代步长的变化曲线如图 8 所示,其训练结果如表 3 所示。

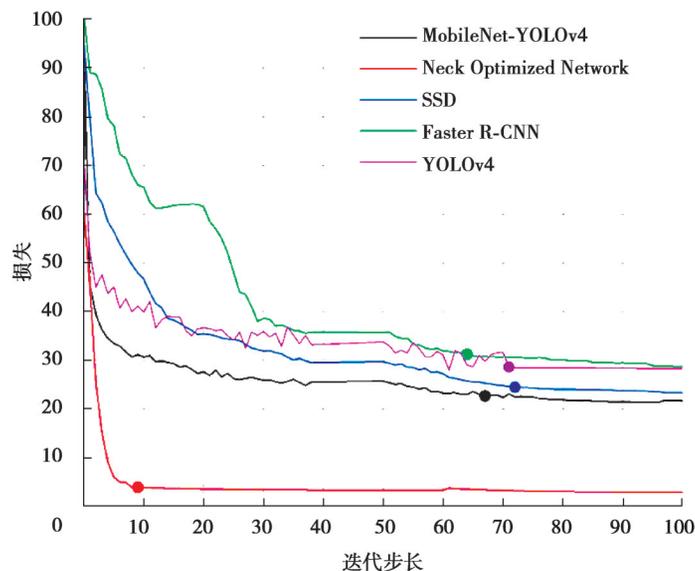


图 8 各模型训练损失函数曲线图

Fig. 8 Curves of training loss function for each model

分析图 8,在整个训练过程中所有模型均收敛,其中基于 YOLOv4 的颈部优化网络新模型具有更低的损失值,且迭代到 10 步左右时,loss 均不再震荡且趋于稳定,相比于 YOLOv4 模型、MobileNet-YOLOv4 等模型其收敛速度更迅速、平稳,同时基于 YOLOv4 颈部优化网络新模型收敛到更低的损失值。

表3 不同算法在同一平台下的检测速度、平均准确率

Table 3 Detection speed and average accuracy of different algorithms under the same platform

模型	模型大小/MB	FPS/GPU	mAP/%	F1	
				安全帽	人
Neck Optimized Network	42.36	85.36	91.12	0.93	0.86
MobileNet-YOLOv4	48.42	78.14	86.58	0.91	0.78
YOLOv4	245.78	13.09	87.43	0.92	0.77
SSD	91.09	15.31	80.25	0.82	0.78
Faster-RCNN	932.98	0.65	93.32	0.94	0.86

观察表3, Faster-RCNN具有较高检测精度,但其检测速度缓慢,无法实时安全帽佩戴检测;YOLOv4和SSD算法检测速度相当,但YOLOv4检测精度高于SSD的检测精度;基于改进的YOLOv4颈部优化网络新模型其模型参数量更小且平均检测精度为91.12%,检测速度达到85.36 FPS,调和平均更接近于1,该模型在检测精度和检测速度2个方面均表现出良好的检测性能,因此选择基于颈部优化的网络新模型作为最终模型,极大提高安全帽的检测速度,使安全帽佩戴检测能够实时实现。

为了更直观展现改进的YOLOv4颈部优化网络新模型的检测效果,选取单目标和多目标情形、小目标情形和极端小目标情形、复杂背景下的测试图像序列进行检测效果验证。图9为单目标安全帽佩戴测试结果,图10为多目标安全帽佩戴测试结果。分别选取在施工状态下不同姿态视角下的单目标场景以及多目标场景,对佩戴安全帽的施工人员使用蓝色框对其位置进行标注,对未佩戴安全帽的施工人员使用红色框进行标注。



图9 单目标下基于改进的YOLOv4颈部优化网络的检测结果

Fig. 9 Display of detection results based on improved YOLOv4 neck optimization network under single objective



图10 多目标下基于改进的YOLOv4颈部优化网络的检测结果

Fig. 10 Display of detection results based on improved YOLOv4 neck optimization network under multi-objective

图9、10分别选取了在施工状态下不同姿态视角下的单目标场景、多目标场景。通常单目标图像以及无遮挡的多目标检测过程较为容易,因此基于改进的YOLOv4颈部优化网络对单目标图像、多目标图像中均能做出正确检测,检测结果基本满足需求。

对于在夜间极端条件和含有极端小目标的情形,基于改进的 YOLOv4 颈部优化网络算法与 YOLOv4 算法两者检测效果差异较大,针对人员流动性大、不同光照条件、人员之间严重遮挡以及存在密集检测目标环境下的目标检测。图 11-12 为 2 种不同算法在极端条件下的检测效果对比图。



(a) 人员流动性大环境

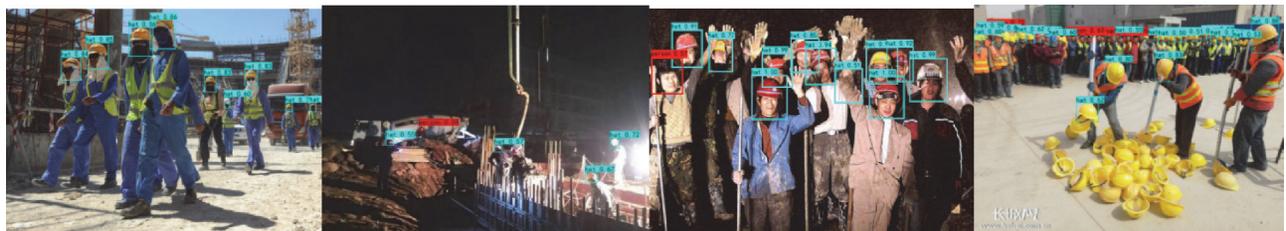
(b) 夜间极端条件环境

(c) 人员之间严重遮挡环境

(d) 密集检测目标环境

图 11 YOLOv4 的检测结果

Fig. 11 YOLOv4 test results



(a) 人员流动性大环境

(b) 夜间极端条件环境

(c) 人员之间严重遮挡环境

(d) 密集检测目标环境

Fig. 12 Detection results based on the improved YOLOv4 neck optimization network

由图 12(a)可知,在可视化条件良好,但人员流动性较大、建筑设施严重干扰的施工场所,基于颈部优化的安全帽佩戴检测算法相对于 YOLOv4 检测算法,检测出了所有检测目标,未发生漏检或错检现象;如图 12(b)所示,在夜间可视化条件差以及强光照射条件下,基于颈部优化下的安全帽佩戴检测算法能够检测出绝大多数目标,未发生漏检等现象;如图 12(c)所示,在光照条件严重不足且各个目标之间存在严重遮挡的情况下,基于改进的 YOLOv4 颈部优化网络的安全帽检测算法也能够正确检测;图 12(d)可知,对于存在大量极端小目标且各个小目标之间存在严重遮挡,基于改进的 YOLOv4 颈部优化网络的安全帽检测算法能够检测出绝大多数目标,表现出了优良的小目标检测性能。

3 结 语

针对安全帽佩戴检测时易受复杂背景的干扰等问题,采用 YOLOv4 算法进行安全帽佩戴检测,检测精度达到 86.91%,在 GPU 平台下检测速度达到了 13.09 FPS。但 YOLOv4 算法在 CPU 平台下检测速度缓慢、内存消耗大、小目标分辨率低,信息特征少等问题,引入 MobileNet 网络轻量化 YOLOv4、跨越模块,采用改进特征金字塔 FPN 和改进注意力机制等颈部优化策略聚焦安全帽目标信息,优化网络模型在 CPU 平台下检测速度达到 34.28 FPS,是 YOLOv4 网络的 16 倍左右,同时其检测精度相比于 YOLOv4 算法检测精度提升了 4.21%,达到 91.12%。在夜间极端条件和含有极端小目标等特殊情况下均能实现对安全帽佩戴的快速准确检测,检测性能优秀,极大提升施工作业人员佩戴安全帽的监管效率。

参考文献

- [1] Achilov S S. Expert method of quality management of road construction project[J]. Middle European Scientific Bulletin, 2021, 15:692.
- [2] Peng H, Zhang Z W. Helmet wearing recognition of construction workers using convolutional neural network[J]. Wireless Communications and Mobile Computing, 2022: 1-8.
- [3] 邵延华,张铎,楚红雨,等.基于深度学习的 YOLO 目标检测综述[J].电子与信息学报,2022,44(10):3697-3708.

- Shao Y H, Zhang D, Chu H Y, et al. A review of YOLO object detection based on deep learning[J]. Journal of Electronics & Information Technology, 2022, 44(10): 3697-3708.(in Chinese)
- [4] Shen J E, Xiong X, Li Y, et al. Detecting safety helmet wearing on construction sites with bounding-box regression and deep transfer learning[J]. Computer-Aided Civil and Infrastructure Engineering, 2021, 36(2): 180-196.
- [5] Wu H, Zhao J S. An intelligent vision-based approach for helmet identification for work safety[J]. Computers in Industry, 2018, 100: 267-277.
- [6] He D Q, Qiu Y F, Miao J, et al. Improved Mask R-CNN for obstacle detection of rail transit[J]. Measurement, 2022, 190: 110728.
- [7] Sun X D, Wu P C, Hoi S C H. Face detection using deep learning: an improved faster RCNN approach[J]. Neurocomputing, 2018, 299: 42-50.
- [8] Bochkovskiy A, Wang C Y, Liao H Y M. YOLOv4: optimal speed and accuracy of object detection[EB/OL]. 2020: arXiv: 2004.10934. <https://arxiv.org/abs/2004.10934>
- [9] Biswas D, Su H B, Wang C Y, et al. An automatic traffic density estimation using Single Shot Detection (SSD) and MobileNet-SSD[J]. Physics and Chemistry of the Earth, 2019, 110: 176-184.
- [10] Liu Y Y, Jiang W R. Detection of wearing safety helmet for workers based on YOLOv4[C]//2021 International Conference on Computer Engineering and Artificial Intelligence (ICCEAI). August 27-29, 2021, Shanghai, China: IEEE, 2021: 83-87.
- [11] Tan M X, Le Q V. MixConv: mixed depthwise convolutional kernels[EB/OL]. 2019: arXiv: 1907.09595. <https://arxiv.org/abs/1907.09595>
- [12] Peng F, Miao Z, Li F, et al. S-FPN: a shortcut feature pyramid network for sea cucumber detection in underwater images[J]. Expert Systems With Applications, 2021, 182: 115306.
- [13] Li C F, Wang B P. A YOLOv4 model with FPN for service plates detection[J]. Journal of Electrical Engineering & Technology, 2022, 17(4): 2469-2479.
- [14] Zheng C Y, Zhang J H, Wang J N, et al. Double-branch dehazing network based on self-calibrated attentional convolution[J]. Knowledge-Based Systems, 2022, 240: 108148.
- [15] Wang S H, Zhou Q, Yang M, et al. ADVIAN: Alzheimer's disease VGG-inspired attention network based on convolutional block attention module and multiple way data augmentation[J]. Frontiers in Aging Neuroscience, 2021, 13: 687456.
- [16] Safety-Helmet-Wearing-Dataset[DB/OL].<https://github.com/njvisionpower/Safety-Helmet-Wearing-Dataset>, 2019-12-17.

(编辑 侯 湘)