

多尺度深度可分离的施工现场安全帽检测

刘金智, 方 睿, 王俊霖, 王畅阳

(成都信息工程大学 计算机学院, 四川 成都 610225)

摘 要: 自动化安全帽佩戴监测能够显著提升监管效率, 减少人为疏漏, 从而降低安全事故的发生率。针对现有安全帽检测算法在复杂背景环境及目标尺度多变的情况下难以实现精确检测的问题, 提出一种基于YOLOv5s的多尺度深度可分离安全帽检测算法—MDSDet。首先, 针对背景环境复杂的问题, 设计了一种由全局可分离特征融合模块与空间通道融合模块组成的可分离特征融合主干网络, 该主干网络能够有效过滤背景信息与冗余特征, 增强模型对全局上下文信息的捕捉能力; 其次, 针对目标尺度多变的问题, 设计了多尺度深度可分离提取模块, 该模块使用不同大小的卷积组合, 增强了模型对不同尺度安全帽的适应性, 显著提升了模型的检测精度。在对比实验中, 所提出的算法在各项指标上均表现优异, mAP达到87.9%, mAP_{head}较YOLOv5s提升了5.1%。实验结果表明, MDSDet算法在复杂环境下能够有效提升对多尺度目标的检测精度。

关键词: 安全帽; 目标检测; YOLOv5s; 深度可分离; 特征融合; 多尺度融合

DOI: 10.11907/rjtk.241775

扫描二维码阅读全文:

中图分类号: TP391

文献标识码: A

文章编号: 1672-7800(2025)010-0164-09



Multi-scale Depth Separable Safety Helmet Detection on Construction Site

LIU Jinzhi, FANG Rui, WANG Junlin, WANG Changyang

(School of Computer Science, Chengdu University of Information Technology, Chengdu 610225, China)

Abstract: Automated helmet monitoring significantly improves regulatory efficiency and reduces human negligence, thereby lowering the occurrence of safety accidents. To address the challenge that existing safety helmet detection algorithms struggle to achieve accurate detection in complex backgrounds and with varying target scales, this paper proposes a multi-scale depth-separable safety helmet detection algorithm—MDSDet based on YOLOv5s. A separable feature fusion backbone network is designed, incorporating the global separable feature fusion module and the spatial channel fusion module, to handle complex background environments. This backbone network effectively filters background noise and redundant features, enhancing the model's ability to capture global contextual information. Additionally, to address the issue of varying target scales, a multi-scale depth-separable extraction module is introduced. This module utilizes convolution combinations of varying sizes to improve the model's adaptability to helmets of different scales, significantly enhancing detection accuracy. Comparative experiments demonstrate that the proposed algorithm achieves excellent performance across various metrics, with an mAP of 87.9%, and an mAP_{head} increase of 5.1% compared to YOLOv5s. The results indicate that the MDSDet algorithm effectively improves the detection accuracy of multi-scale targets in complex environments.

Key Words: safety helmet; object detection; YOLOv5s; depth-separable; feature fusion; multi-scale fusion

0 引言

在全球范围内, 施工安全一直是一个亟待解决的重要问题, 其中安全帽作为预防高空坠物、飞溅物等意外事故

中至关重要的防护设备, 能够有效减少工人头部伤害^[1-2]。然而, 由于监管不完善和人力监督不足, 工人未佩戴安全帽的现象时有发生, 特别是在高风险行业, 如建筑业和制造业, 这类忽视往往导致了大量可避免的伤亡^[3]。早期, 安全帽佩戴的监测主要依赖于人工巡查, 这一方式不仅耗

收稿日期: 2024-09-09

基金项目: 国家重点研发计划项目(2020YFA0608000)

作者简介: 刘金智(1999-), 男, 成都信息工程大学计算机学院硕士研究生, 研究方向为目标检测; 方睿(1974-), 男, 博士, 成都信息工程大学计算机学院教授、硕士生导师, 研究方向为基于网络的计算机应用; 王俊霖(1999-), 男, 成都信息工程大学计算机学院硕士研究生, 研究方向为目标检测; 王畅阳(1999-), 男, 成都信息工程大学计算机学院硕士研究生, 研究方向为目标检测。本文通讯作者: 方睿。

时耗力,而且效率低下,难以全面覆盖所有作业区域和时段。随着技术的进步,安全帽佩戴的监测方法逐渐从传统的人工巡查过渡到基于传感器的直接监测和基于计算机视觉的图像识别技术。基于传感器的方法利用物理传感器(如加速度计、压力传感器等)检测佩戴者的动作或安全帽佩戴状态,具有较高的实时性和准确性,但其部署和维护成本较高,且会限制佩戴者的活动自由。相比之下,基于计算机视觉的方法^[4-5]通过图像处理与模式识别技术检测工人佩戴情况,灵活性更高,但在光照条件、成像角度变化以及背景复杂的情况下,检测效果依然不够稳定,并且由于需要处理大量数据,对计算资源的要求较高。

此外,在实际施工环境中,安全帽检测还面临着更多的挑战,如光照变化、背景干扰以及遮挡物的影响,这些因素进一步导致了漏检和误检现象。同时,目标距离的变化也是影响检测精度的重要因素,特别是在工人远近移动时,头部的成像效果会随之波动,增加了检测的难度。近年来,深度学习技术的迅猛发展为应对这些问题提供了新的路径。特别是基于卷积神经网络(CNN)的目标检测算法^[6],不仅在提高小目标检测精度方面展现了巨大的潜力,还有效降低了误检和漏检的发生率^[7-9],从根本上为安全帽的自动化监测^[10]带来了更多的可能性。例如,邓开发等^[11]结合Faster R-CNN和深度特征流算法,实现了对安全帽的图像与视频检测。Li等^[12]通过增强路径聚合网络提升了YOLOv4算法对小目标的检测性能,并对空间金字塔池化模块和VoVNet结构进行了轻量化改进,进一步优化了模型的效率和精度。Zhao等^[13]在YOLOv5基础上引入双向特征金字塔网络和卷积块注意模块,增加检测层,有效提升网络在复杂场景下的检测能力,且降低了误检率。An等^[14]通过引入全局注意力机制和CBAM模块,加强了YOLOv5s对小目标的特征提取能力,并采用知识蒸馏实现模型的轻量化。Wang等^[15]在YOLOv7-tiny的基础上融合了高效多尺度注意力和自注意力动态检测头,以实现轻量化模型更强的特征表示。Song等^[16]将多尺度空洞卷积与ASPP池化金字塔和标准化高斯瓦瑟斯坦距离相结合,改善了YOLOv8因距离等原因导致安全帽难以检测的问题。然而,这些方法只是将功能模块进行简单串联拼凑,缺乏多尺度信息的交互作用,在面对实际施工环境背景复杂和安全帽目标尺寸不一的问题时,仍需不断探索与优化。

因此,本文提出一种多尺度深度可分离安全帽检测算法(Multi-scale Depthwise Separable Safety Helmet Detection Algorithm, MDSDet)。具体而言,本文提出一种可分离特征融合主干网络(Separable Feature Fusion Backbone, SFFB),旨在有效提取并融合安全帽的关键特征。针对传统方法在复杂场景中难以精准捕捉目标特征的问题,SFFB主干通过全局可分离特征融合模块(Globally Separable Feature Fusion Module, GSFFM)减少了目标区域冗余信息

对特征提取的干扰,同时增强了模型对全局上下文信息的理解能力。为了进一步优化特征质量,将空间通道融合模块(Space Channel Fusion Module, SCFM)与GSFFM相结合,通过对空间和通道信息的学习来过滤背景噪声,有效减少了目标检测过程中背景信息的干扰,从而提高了特征图的纯净度和检测精度。此外,为了解决安全帽检测中常见的多尺度问题,设计并实现了多尺度深度可分离提取模块(Multi-scale Depthwise Separable Extraction Module, MDSEM)。该模块结合了坐标注意力机制、深度可分离卷积和C3模块,不仅增强了模型对不同尺寸安全帽的适应性,还提高了模型对各种尺度目标的检测性能^[17]。实验结果表明,MDSDet算法在安全帽数据集上的表现优于许多主流目标检测方法,其综合指标最佳,其中对未佩戴安全帽类别的检测精度达到80.3%,相较基准模型提高了5.1%,能够更精确地识别施工现场中的各类安全帽。

本文主要贡献总结如下:①为了增强模型在复杂背景下提取目标特征的能力,设计了全局可分离特征融合模块(GSFFM);②针对安全帽检测过程中背景信息冗余的问题,创建了空间通道融合模块(SCFM),以增强模型对安全帽特征的关注,并减少干扰信息;③为解决安全帽尺寸变化的问题,提出了多尺度深度可分离提取模块(MDSEM),通过融合多层次的语义信息来优化模型对不同尺寸安全帽的适应性。

1 基准模型选取及其算法原理

YOLO系列算法因其在速度与精度之间的良好平衡,近年来在目标检测领域受到广泛关注^[18-22]。在众多YOLO版本中,YOLOv5s凭借其适中的模型大小、高效的计算效率和稳定的检测精度,成为本文实验研究的首选基准模型。相较于早期的YOLOv3和YOLOv4,YOLOv5s在检测能力和模型压缩方面取得了显著进步,使其在处理安全帽检测等具体任务时更加高效。同时,与更高版本的YOLOv7-tiny和YOLOv8s相比,YOLOv5s在保证相近精度的同时,具有更小的模型体积和更低的计算需求,这对算力资源有限的施工现场尤为关键。此外,在本文的安全帽数据集上,其优势表现得更加显著。

YOLOv5s的网络结构分为4个部分:输入端、主干、颈部和检测头。输入端主要负责数据预处理,采用Mosaic数据增强技术来增加数据的多样性和复杂性,从而提升模型的泛化能力。主干部分负责特征提取,主要由Conv模块、C3模块、SPPF和Focus结构组成。C3模块增强了模型主干网络的学习能力,减少了梯度信息的重复,提高了特征提取能力,并用于调整模型的深度和宽度。Focus结构通过对输入特征图进行切片操作,减少了模型的参数和计算量,从而提高了检测速度。SPPF通过全局特征与局部特征的加权融合,获取了更全面的特征信息。在颈部部分,

通过上采样和下采样操作融合了不同尺度的特征图,并将这些特征信息传递到检测头。检测头则负责最终输出每个目标的类别及其对应的坐标信息。

2 施工现场安全帽检测算法

在错综复杂的施工场景中,安全帽检测任务面临着前所未有的挑战。这些挑战不仅来源于目标物体本身,如尺寸差异显著、部分或完全遮挡等,还涉及到环境背景的复杂性。特别是当安全帽的颜色、纹理等特征与周围环境高度相似时,传统的检测方法往往难以有效区分。

2.1 MDSDet 网络模型

本文提出的网络结构框架如图 1 所示。在主干网络 SFF 中,加入了 SCFM 和 GSFFM。其中,SCFM 用于进行背景过滤,使主干网络能够更专注于目标信息的学习。随后,通过 GSFF 获得不同尺寸的上下文信息,以增强目标特征的表达。在颈部,将下采样过程中的 C3 模块替换为新的特征提取模块 MDEM,为 Head 的 3 个检测头提供更加丰富的特征融合信息。

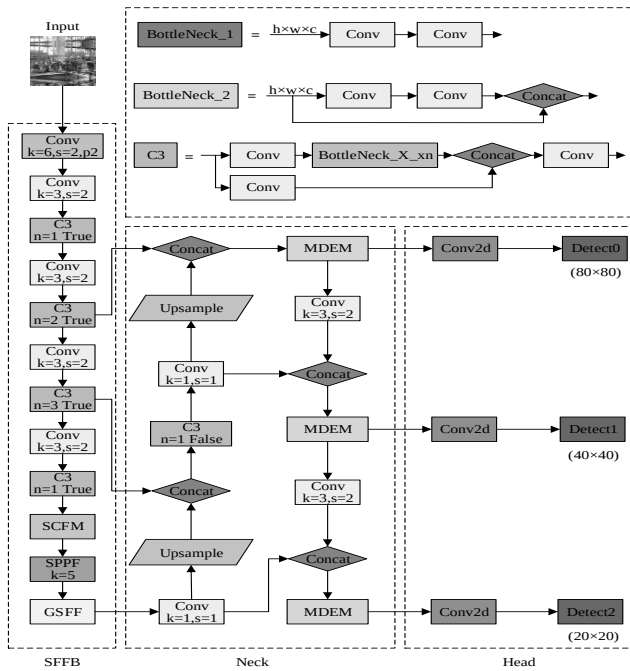


Fig. 1 MDSDet network structure diagram

图 1 MDSDet 网络结构图

2.2 可分离特征融合主干

在现实生活中,检测任务受环境光线、背景杂乱和物体遮挡等因素的影响,使得安全帽与背景之间的信息难以区分。为应对这些挑战,本文设计了 GSFFM 以增强网络的全局特征表达能力并有效融合不同维度的信息。此外,还引入 SCFM 以减少背景干扰,帮助网络更加专注于学习安全帽的特征。

2.2.1 GSFFM 全局可分离特征融合模块

在 YOLOv5s 的设计中,C3 模块在高效特征提取与融

合方面具有显著优势,但在处理复杂背景和精细特征时表现出一定局限,特别是在整合全局上下文信息方面的不足。受全局上下文模块的启发,GSFFM 旨在突破传统卷积操作对局部视野的依赖,赋予网络更深层次的全局感知能力^[23]。如图 2 所示,GSFFM 由 4 个分支组成,巧妙地集成了 3×3 深度可分离卷积、1×1 卷积、原始输入信息以及全局上下文模块。

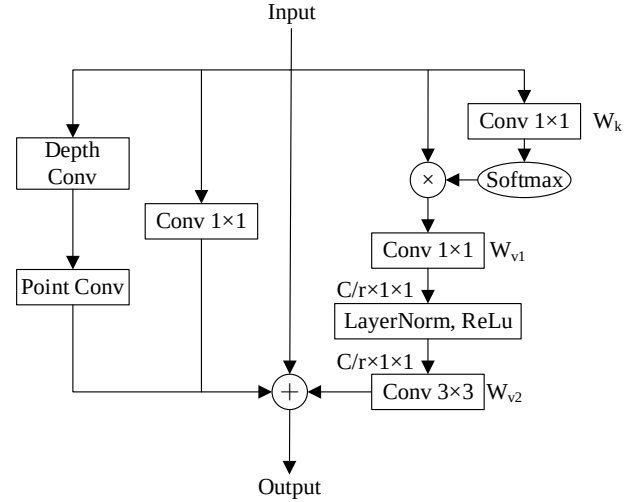


Fig. 2 GSFFM structure diagram

图 2 GSFFM 结构图

具体而言,3×3 深度可分离卷积、1×1 卷积以及原始输入 3 条分支共同实现了对不同尺寸特征信息的提取。通过引入更多的非线性信息来提高网络的泛化能力,同时提取不同尺度的特征信息。第 4 条分支通过全局上下文模块,将提炼出的全局上下文特征融入到其他 3 条分支的特征图中,从而增强了特征图的表示能力。其定义如式(1)所示。

$$Z_i = DSConv(x_i) + Conv(x_i) + x_i + \delta(\cdot) \quad (1)$$

式中: $DSConv(\cdot)$ 表示卷积核大小为 3×3 的深度可分离卷积; $Conv(\cdot)$ 表示 1×1 的普通卷积; x_i 表示特征图的输入; $\delta(\cdot)$ 是全局上下文模块。具体如式(2)所示。

$$\delta(\cdot) = W_{t2} ReLu \left(LN \left(W_{v1} \sum_{j=1}^{N_p} \frac{e^{W_{kx_j}}}{e^{W_{kx_m}}} x_j \right) \right) \quad (2)$$

式中: $\sum_{j=1}^{N_p} \frac{e^{W_{kx_j}}}{e^{W_{kx_m}}} x_j$ 表示全局注意力池化的权重; $W_{t2} ReLu(LN(W_{v1}(\cdot)))$ 表示瓶颈变换; W 表示线性变换矩阵; LN 和 $ReLu$ 分别表示归一化处理 and 激活函数。具体而言,全局上下文模块首先利用 1×1 卷积和 Softmax 函数构建一个注意力权重矩阵。该矩阵能够引导模型聚焦于输入特征图中的重要区域,从而实现全局注意力池化。随后,通过瓶颈变换(一种高效的维度变换策略)来捕捉通道间的复杂依赖关系。

2.2.2 SCFM 空间通道融合模块

由于施工现场环境的复杂性,输入图像中往往包含大

量无用的背景信息,这些信息不仅来源于背景的杂乱,还包括光线和遮挡目标对检测的干扰^[24]。为此,本文在主干的SPPF特征融合之前添加SCFM,以增强模型对安全帽特征的关注度并减少干扰信息。SCFM结构图如图3所示。通过不同卷积核大小的非线性操作和残差连接的组合,增强所有通道之间的关联性,从而有效提升网络在复杂环境下对目标特征的注意力和捕捉能力。

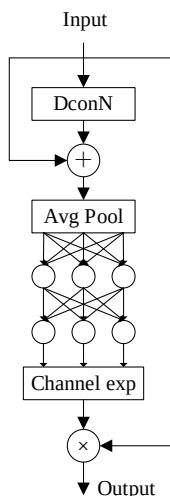


Fig. 3 SCFM structure diagram

图3 SCFM结构图

为获得丰富的特征信息,首先,利用不同卷积核大小的卷积提取目标特征;其次,将串联后的特征信息与浅层输入特征相融合;最后,获得更加全面的语义信息。具体如式(3)所示。

$$D_{out} = x_i + DconN(x_i) \quad (3)$$

式中: $DconN(x_i)$ 是通过 3×3 卷积和 1×1 卷积串联组合实现的。在每次卷积操作后紧跟GELU激活函数和BN批量归一化,如图4所示。

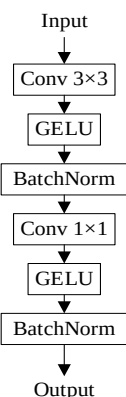


Fig. 4 DconN structure diagram

图4 DconN结构图

在信息融合后,首先,使用全局平均池化层来获取全局信道信息;其次,通过2个全连接层自适应地获得特征映射的注意力;再次,通过通道扩展操作将通道数与输入特征图保持一致;最后,通过元素级乘法将增强的特征信

息整合到特征图中。通过将学习到的空间和通道维度间的相关性融入浅层和深层特征,可以防止所需信息在信息融合过程中丢失,从而提高网络对冗余信息的抗干扰能力。

2.3 MDSEM多尺度深度可分离提取模块

在安全帽检测任务中,目标尺寸的多变性是一个普遍存在的问题。在目标尺寸多样的场景中,检测模型需要充分融合高级和低级特征信息。然而,传统模型中的特征融合网络通常难以有效融合不同目标尺寸的信息,这使得多尺度物体检测变得具有挑战性。为了解决这一问题,本文设计了多尺度深度可分离提取模块,以融合具有不同尺度和语义信息的特征图。MDSEM结构如图5所示。

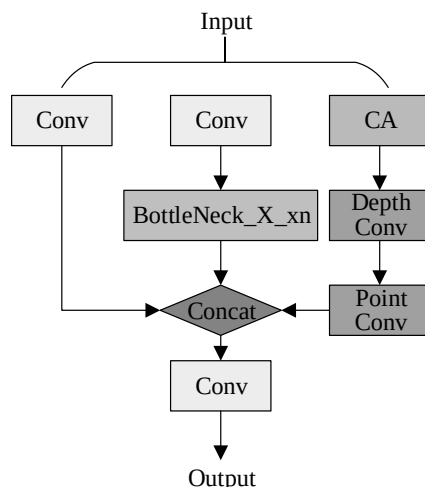


Fig. 5 MDSEM structure diagram

图5 MDSEM结构图

MDSEM由3条支路(B1、B2、B3)构成。B1通过普通 1×1 卷积学习到更加复杂的非线性特征表示。B2使用 1×1 和 3×3 卷积来获得不同感受野大小的特征信息,以提高网络对不同尺度目标的检测能力。具体如式(4)一式(5)所示。

$$F_{out} = Conv_1(Cat(B1, B2, B3)) \quad (4)$$

式中: $Cat(\cdot)$ 表示按通道维度进行拼接。

$$B2 = Conv_1(Conv_1(Conv_3(x_i))) \quad (5)$$

式中: $Conv_1$ 是 1×1 卷积, $Conv_3$ 是 3×3 卷积。

首先,B3采用坐标注意力机制将位置信息嵌入到通道注意力中,使模型不仅能够捕获跨通道的信息,还包含了方向感知和位置感知的信息;其次,将增强后的特征输入到 3×3 的深度可分离卷积中,使模型在处理多尺度目标时能够更准确地定位并识别目标区域,从而使得模型能够更好地适应不同尺度的目标;最后,将3条支路处理后的特征图按照通道维度进行拼接,并使用 1×1 卷积来调整输出特征的通道数。具体如式(6)所示。

$$B3 = DSConv(CA(x_i)) \quad (6)$$

式中: $CA(\cdot)$ 表示坐标注意力机制; $DSConv(\cdot)$ 表示卷积核大小为 3×3 的深度可分离卷积。

3 实验与分析

3.1 数据集

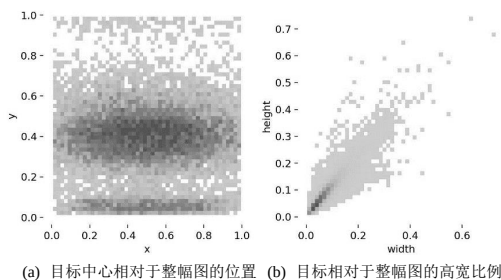
本实验使用的安全帽佩戴检测数据集来源于 Kaggle 机器学习竞赛平台。该数据集包含了 5 000 张图片,每张图片的大小为 416×416 ,并配有 2 501 个标签,采用了 PASCAL VOC 格式进行边界框标注。但该数据集的标签格式不符合本研究的实验需求,并存在边界框标注不准确以及大部分图像未标注的问题。因此,利用 Labelme 工具对数据集进行了重新标注,将标注类别限定为佩戴安全帽和未佩戴安全帽。最终,使用 Python 脚本将标注信息转换为 YOLOv5 模型所能读取的 txt 格式。经过处理后,本文实

验使用的数据集包含 4 490 张图片,标签总数扩展至 18 728 个,其中佩戴安全帽与未佩戴安全帽的标签数量比例约为 3:1。为了进行模型的训练、验证和测试,本文按照 8:1:1 的比例将数据集划分为训练集、验证集和测试集。图 6 和图 7 分别展示了经过图像增强处理的数据集样本及数据处理的可视化结果。从数据集样本展示图可以看出,本研究使用的安全帽数据集涵盖了室内、室外、背景复杂及人员密集等多种场景。通过数据处理的可视化结果还可以看出,绝大多数目标区域较小且分布集中。因为在现实场景中,无论是人眼观察还是监控设备的视野范围内,安全帽和头部区域的尺寸通常相对较小。相较于广阔的视野范围和周围环境中的大型物体,这些安全目标的尺寸特征确实较小,这也符合实际情况。



Fig. 6 Dataset sample

图6 数据集样本



(a) 目标中心相对于整幅图的位置 (b) 目标相对于整幅图的高宽比例

Fig. 7 Target box visual data diagram

图7 目标框可视化数据图

3.2 评价指标

本文采用精确度(Precision)、召回率(Recall)、平均均值精度(mean Average Precision, mAP)和计算量(FLOPs)4个评价指标来评估改进后的网络模型性能。

平均精度 AP 表示单个类别的检测精度。平均均值精度代表所有类别的平均精度的平均值。具体如式(7)所示。

$$mAP = \frac{\sum_{i=1}^C AP_i}{C} \quad (7)$$

式中: C 为样本的类别总数; AP_i 为类别 i 的平均精度。其中, AP_{helmet} 表示为佩戴安全帽目标类别的平均精度, AP_{no_helmet} 表示未佩戴安全帽类别的平均精度。实验结果表明,较低的IoU 阈值意味着对预测结果有较高的容忍度,能够包含更多“正确”的检测结果。然而,过高的IoU 阈值可能导致检测性能下降,因为这将排除许多虽然部分重叠但仍被视为有效预测的边界框。因此,本文默认在IoU 阈值为0.5的情况下进行精度计算,即 $mAP@0.5$ 。此外,为了使实验结果更具说服力,本研究还计算了在不同 IoU 阈值($IoU=[0.5, 0.95]$, 区间间隔为 0.05)下的 mAP,并将其平均值作为评价指标,在文中以 mAP 表示。

3.3 实验环境及参数设置

本文的实验平台基于 Ubuntu 18.04.6 LTS 操作系统,配备了 AMD Ryzen 9 3900X 12-Core Processor 处理器以及 NVIDIA GeForce RTX 3080 显卡,显存为 10GB。模型实验

使用的 CUDA 版本为 11.4, Python 版本为 3.8.5, Pytorch 版本为 1.12.1。网络训练超参数的具体设置如表 1 所示。

Table 1 Hyperparameter setting
表 1 超参数设置

参数	参数值
输入图像尺寸大小	640×640
训练迭代次数	150
批系数	16
优化器	SGD
动量参数	0.937
初始学习率	0.01
权重衰减系数	0.000 5

3.4 模型对比实验

在确保所有硬件条件及参数配置完全一致的前提下,本研究选取了当前主流的目标检测算法和该领域最新的同类算法与 MDSDet 模型进行比较^[14,25]。在安全帽数据集上的对比实验结果如表 2 所示,直观展示了各模型之间的性能差异。

Table 2 Performance comparison of different models
表 2 不同模型性能对比

模型	P/%	R/%	AP _{helmet} /%	AP _{head} /%	mAP@0.5/%	mAP/%	FLOPs/G
SSD	82.3	78.5	90.2	70.0	80.1	53.1	62.7
Retina-Net	82.7	77.9	89.7	69.5	79.6	52.7	81.1
YOLOv5s	89.7	78.7	95.5	75.2	85.4	54.6	15.8
YOLOv6s	84.7	75.7	94.3	70.9	82.6	53.2	44.0
YOLOv7-tiny	89.7	75.6	95.6	75.2	85.4	54.2	13.2
YOLOv7	89.3	83.4	96.1	79.1	87.6	56.2	105.1
YOLOv8s	87.9	81.0	95.0	79.1	87	56.1	28.4
文献 14	90.1	80.7	95.1	77.7	86.4	55.4	12.5
文献 25	88.2	81.3	94.8	79.6	87.2	55.7	28.7
MDSDet	90.9	81.9	95.5	80.3	87.9	56.8	18.5

本文所提出的 MDSDet 模型在检测精确度上达到了 90.9%, 明显优于 YOLOv5s (89.7%)、YOLOv7 (89.3%) 和 YOLOv8s (87.9%), 且超过了文献[14]中提到的最新算法 (90.1%), 彰显了其卓越的检测准确性。而在安全帽检测任务中, 特别是在对未戴安全帽人员的专项检测中, 结果表明在未佩戴安全帽的平均精度 (AP_{head}/%) 上, MDSDet 模型达到了 80.3%, 高于其他所有对比模型, 验证了 MDSDet 模型在检测头部时具有更高的准确性, 这对于实际应用中可能存在的复杂背景和目标尺寸不一的情况尤为重要。在 mAP@0.5 这一关键指标上, MDSDet 模型同样表现优异, 达到了 87.9%, 超越了所有参与对比的模型, 同时 mAP 也高于所有对比模型, 进一步证明了其在安全帽检测任务时, 能够保持较高的整体检测精度。尽管召回率、模型计算量和佩戴安全帽的平均精度并非在所有模型中最高, 但 MDSDet 依然保持了较高的水平, 召回率和佩戴安全帽类别的平均精度仅次于 YOLOv7, 而 YOLOv7 的计算量是 MDSDet 的 6 倍左右, 彰显 MDSDet 模型在效率与效能之间

的平衡。

综上, 相较于基准模型 YOLOv5s、召回率最高的 YOLOv7-tiny 以及同领域最新的改进算法, 本文提出的模型在安全帽检测任务中展现出了更为优异的综合性能, 尤其是对未佩戴安全帽类别的检测, AP_{head} 比 YOLOv5s 提高了 5.1%。实验结果表明, MDSDet 模型在整体性能上优于其他对比模型, 为施工现场安全帽检测提供了更为高效、可靠的解决方案。

3.5 对比主干改进的实验与分析

为了深入探究 SFFB 对于模型在复杂场景下的性能提升, 特别是针对安全帽佩戴检测任务中的有效性, 本文通过对比不同主干改进方法进行了实验与分析, 旨在揭示 SFFB 在特征提取、目标召回及整体检测性能上的独特优势。其他改进方法与 SFFB 改进类似, 在对应位置添加相应改进模块进行对比实验。实验结果如表 3 所示。

Table 3 Comparative experiment of backbone improvement
表 3 主干改进对比实验

改进策略	P/%	AP _{helmet} /%	AP _{head} /%	mAP@0.5/%
+SE	89.1	95.3	78.7	87
+CA	89.5	95.5	78	86.8
+SA	89.6	95.4	77.5	86.4
+EMA	90.4	95.1	78.5	86.8
+SimAM	90.1	95.5	78.2	86.9
+SFFB	90.7	95.5	79.3	87.4

由表 3 可知, SFFB 改进策略在多个关键指标上表现出色, 尤其是在精确度、未佩戴安全帽的平均精度以及总体平均精度方面均领先于其他改进方法。SFFB 的精确度达到了 90.7%, 明显高于其他策略, 同时未佩戴安全帽的平均精度为 79.3%, mAP@0.5 达到了 87.4%, 均为最高值。这一优势主要得益于 GSFFM 和 SCFM 的巧妙融合。GSFFM 模块通过增强全局上下文感知能力, 克服了传统卷积操作局限于局部视野、难以有效整合全局信息的缺点。GSFFM 包含 4 条分支, 其中 3×3 深度可分离卷积和 1×1 卷积的结合, 能够高效提取多尺度特征; 全局上下文模块通过注意力机制, 帮助模型聚焦于输入特征图中的关键区域。同时, SCFM 模块的作用在于减少背景噪声干扰, 强化网络对目标信息的关注度。在施工现场等复杂环境下, 光线变化、目标遮挡等因素容易干扰检测精度, 而 SCFM 通过多尺度卷积核组合和残差连接结构, 增强了通道间的关联性, 有效地过滤掉了无用的背景信息。在特征融合前, SCFM 通过多尺度特征提取及全局平均池化的方式进一步提升了模型对安全帽相关特征的捕捉能力。

由图 8 可视化对比结果可知, GSFFM 和 SCFM 模块的引入为 SFFB 策略的性能提升提供了核心支持, 不仅提高了精确度和 mAP@0.5, 还在复杂环境中展现了对特定目标的强大识别能力, 进一步验证了这一改进方案在实际应用中的有效性。



Fig. 8 Backbone improvement comparison diagram

图8 主干改进对比图

3.6 对比颈部融合网络改进的实验与分析

本文对不同特征融合提取模块进行了全面评估,以验证 MDSEM 在模型中的实用性,实验结果如表 4 所示。

由表 4 可知,MDSEM 在多个评估指标上均优于其他特

Table 4 Comparative experiment of different feature fusion extraction modules

表 4 不同特征融合提取模块对比实验

改进策略	P/%	AP _{helmet} /%	AP _{head} /%	mAP@0.5/%
+EMA	89.5	95.1	78.1	86.6
+ RFACnv	89.4	95	78.1	86.6
+DBB	89.7	95.7	77	86.4
+C2f	89.8	95.7	78.3	87
+MDSEM	90.1	95.8	78.4	87.1

征融合策略(如 EMA、RFACnv、DBB 和 C2f)。具体而言,MDSEM 的精确度达到了 90.1%,安全帽类别的平均精度为 95.8%,而未佩戴安全帽类别的平均精度为 78.4%。这一系列指标的提升验证了 MDSEM 通过多尺度处理的有效性,使模型在融合具有不同尺度和语义信息的特征图时表现得更加出色。这得益于 MDSEM 模块的设计不仅增强了对多尺度目标的检测能力,还有效整合了来自不同卷积核的特征信息。通过引入坐标注意力机制,MDSEM 能够在通道注意力中嵌入位置信息,进一步提升了模型对目标区域的精确定位能力。这一优势在图 9 的对比图中得到了体现,MDSEM 在复杂环境中能够更准确地识别并定位目标,确保了其在实际应用中的有效性和可靠性。



Fig. 9 Neck improvement comparison diagram

图9 颈部改进对比图

3.7 模型消融实验

为验证本文所提出的各个改进方法对 MDSDet 算法的提升效果,以 YOLOv5s 为基准模型在安全帽数据集上进行消融实验,实验结果如表 5 所示。“√”表示使用了该改进方法,“×”表示未使用该改进方法。

由表 5 可知,在基准模型 YOLOv5s 上分别添加改进策略时,本文所提出的 MDSDet 算法有了明显的改进。具体

而言,SCFM 突出了目标区域的特征表达,抑制了非目标区域的信息干扰,与基准模型相比,mAP 提升了 1.5%。GS-FFM 的引入增强了模型对全局上下文信息以及目标融合信息的感知能力,将 mAP 提高至 87.2%,有效提高了模型的检测精度。由于多尺度的处理方式使得 MDSEM 能够更有效地融合具有不同尺度和语义信息的特征图,从而提高模型对多尺度目标的检测能力,模型精确度提升至 90.1%,

Table 5 Ablation experiment results
表 5 消融实验结果

SCFM	GSFFM	MDSEM	P/%	mAP@0.5/%
×	×	×	89.7	85.4
√	×	×	89.7	86.9
×	√	×	89.7	87.2
×	×	√	90.1	87.1
√	√	×	90.7	87.4
√	×	√	90.1	87.3
×	√	√	90.6	87.6
√	√	√	90.9	87.9

mAP 提高了 1.7%。实验结果表明,SCFM、GSFFM 和 MDSEM 这 3 种改进方法均对模型性能有显著的提升作用,且

它们之间存在互补性和协同性。当 3 种改进方法被组合使用时,检测精确度从 89.7% 提升至 90.9%,mAP 从 85.4% 提高到 87.9%,模型的整体性能达到了最优,进而证明了每种改进方法的有效性。

3.8 结果可视化分析

为了直观验证本文改进模型在复杂施工环境下以及对多尺度目标的检测效果,挑选了一系列具有代表性的施工现场图片作为测试样本。这些图片涵盖了环境背景错综复杂、人员高度密集、目标尺度多变(包括远近目标)等多种挑战场景。随后,将基准模型与本文模型进行了详尽的实验对比,并将检测结果以对比图的形式呈现,如图 10 所示。

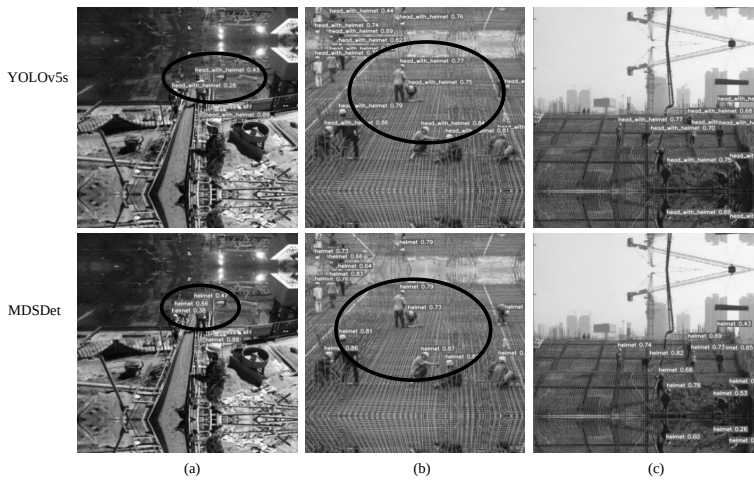


Fig. 10 Comparison of detection results
图 10 检测结果对比图

在图 10(a)中展示了不同远距离造成安全帽图像大小不一的情况。在检测近距离目标时,YOLOv5s 和 MDS-Det 的检测效果相当,但对于远处小目标的检测,本文提出的算法具有更高的检测精度以及查全率。图 10(b)呈现了复杂环境中中距离目标的情况,在图中圈出区域突出了 MDSDet 更高的检测精度。图 10(c)展示的远距离小目标复杂背景,相比之下,本文的改进算法在漏检方面更加优秀,能够更加有效查全安全帽目标。通过可视化结果进一步验证了 MDSDet 算法的卓越适应性,其能够稳定且准确地检测出不同环境下的安全帽,从而有效解决了在复杂场景下检测对不同大小目标的局限性。

4 结语

本文针对实际施工场景中安全帽目标检测所面临的环境背景复杂性和尺度多变性的挑战,设计了一个基于 YOLOv5s 的改进算法 MDSDet。为获得更易区分前景后景的信息和减少冗余信息,设计了一个 GSFFM 和 SCFM 相结合

的算法,充分融合不同感受野下的特征信息,减少更小目标特征信息的丢失。实验结果充分验证了本文所提出改进算法的良好检测性能,mAP 指标比 YOLOv5s 的提高了 2.5%。尤其是对未佩戴安全帽的检测精度由 75.2% 提升至 80.3%,有效提高了安全帽检测的精度。在包含复杂场景和不同尺度目标的施工现场图片测试中,进一步验证了 MDSDet 算法对目标尺寸显著变化和面对复杂施工环境下安全帽检测的有效性。

然而,尽管本研究在大量日间施工场景下取得了显著成果,但对于夜间环境对模型检测性能的影响尚未进行详尽探讨。鉴于夜间施工环境的特殊性,如光照不足和阴影干扰等因素,这些都可能对模型检测精度构成一定挑战。因此,未来研究将致力于构建包含丰富夜间施工场景的数据集,并通过增加夜间条件下的实验研究,为实际施工场景下的安全帽检测提供更加全面和可靠的解决方案。

参考文献:

[1] FENG S N. Application research on predicting and preventing residential construction safety accidents based on artificial intelligence [J]. Jushe, 2024(17): 137-140.

- 冯淑楠. 基于人工智能预测与防范住宅施工安全事故的应用研究[J]. 居舍, 2024(17): 137-140.
- [2] GAO S Q, SUN T. Safety management and accident prevention of construction engineering[J]. Urban Construction Theory Research (Electronic Edition), 2024(16): 48-50.
- 高圣强, 孙涛. 建筑工程安全生产管理及事故预防[J]. 城市建设理论研究(电子版), 2024(16): 48-50.
- [3] FANG W L, DING L Y. Research on intelligent recognition and correction of workers' unsafe behavior[J]. Journal of Huazhong University of Science and Technology (Natural Science Edition), 2022, 50(8): 131-135.
- 方伟立, 丁烈云. 工人不安全行为智能识别与矫正研究[J]. 华中科技大学学报(自然科学版), 2022, 50(8): 131-135.
- [4] YUE S, ZHANG Q, SHAO D, et al. Safety helmet wearing status detection based on improved boosted random ferns[J]. Multimedia Tools and Applications, 2022, 81(12): 16783-16796.
- [5] LI J, LIU H, WANG T, et al. Safety helmet wearing detection based on image processing and machine learning[C]//2017 Ninth International Conference on Advanced Computational Intelligence (ICACI), 2017: 201-205.
- [6] ZENG W B, LI J. Overview of object detection algorithms based on deep learning[J]. Automotive Engineer, 2024(1): 1-11.
- 曾文炳, 李军. 基于深度学习的目标检测算法综述[J]. 汽车工程师, 2024(1): 1-11.
- [7] BAI P R, WANG R, LIU Q Y, et al. DS-YOLOv5: a real-time helmet wear detection and recognition model[J]. Chinese Journal of Engineering Sciences, 2023, 45(12): 2108-2117.
- 白培瑞, 王瑞, 刘庆一, 等. DS-YOLOv5: 一种实时的安全帽佩戴检测与识别模型[J]. 工程科学学报, 2023, 45(12): 2108-2117.
- [8] HOU G Y, CHEN Q H, YANG Z H, et al. Helmet detection algorithm based on improved YOLOv5 [J]. Chinese Journal of Engineering Sciences, 2024, 46(2): 329-342.
- 侯公羽, 陈钦煌, 杨振华, 等. 基于改进 YOLOv5 的安全帽检测算法[J]. 工程科学学报, 2024, 46(2): 329-342.
- [9] LI S X, JIANG H Y, XIONG L, et al. Helmet wearing detection algorithm for hydraulic construction personnel based on STA-YOLOv5 [J]. Journal of Chongqing University, 2023, 46(9): 142-152.
- 李顺祥, 蒋海洋, 熊伶, 等. 基于 STA-YOLOv5 的水利建造人员安全帽佩戴检测算法[J]. 重庆大学学报, 2023, 46(9): 142-152.
- [10] LI G P, JIANG Y. Intelligent site construction and exploration of construction enterprises [J]. Informatization of China Construction, 2024(10): 36-41.
- 李国鹏, 蒋艺. 建筑业企业智慧工地建设与探索[J]. 中国建设信息化, 2024(10): 36-41.
- [11] DENG K F, ZOU Z Y. Implementation and analysis of helmet wearing detection based on deep learning [J]. Computer Era, 2020(7): 12-15, 21.
- 邓开发, 邹振宇. 基于深度学习的安全帽佩戴检测实现与分析[J]. 计算机时代, 2020(7): 12-15, 21.
- [12] LI H, WU D, ZHANG W, et al. YOLO-PL: helmet wearing detection algorithm based on improved YOLOv4 [J]. Digital Signal Processing, 2024, 144: 104283.
- [13] ZHAO L, TOHTI T, HAMDULLA A. BDC-YOLOv5: a helmet detection model employs improved YOLOv5[J]. Signal, Image and Video Processing, 2023, 17(8): 4435-4445.
- [14] AN Q, XU Y, YU J, et al. Research on safety helmet detection algorithm based on improved YOLOv5s[J]. Sensors, 2023, 23(13): 5824.
- [15] WANG S, WU P, WU Q. Safety helmet detection based on improved YOLOv7-tiny with multiple feature enhancement [J]. Journal of Real-Time Image Processing, 2024, 21(4): 120.
- [16] SONG X, ZHANG T, YI W. An improved YOLOv8 safety helmet wearing detection network[J]. Scientific Reports, 14(1): 1-14.
- [17] HOU Q, ZHOU D, FENG J. Coordinate attention for efficient mobile network design [C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2021: 13713-13722.
- [18] REDMON J, FARHADI A. Yolov3: an incremental improvement [DB/OL]. <https://arxiv.org/abs/1804.02767>.
- [19] BOCHKOVSKIY A, WANG C Y, LIAO H Y M. YOLOV4: optimal speed and accuracy of object detection [DB/OL]. <https://arxiv.org/abs/2004.10934>.
- [20] ZHU X, LYU S, WANG X, et al. TPH-YOLOv5: improved YOLOv5 based on transformer prediction head for object detection on drone-captured scenarios [C]//Proceedings of the IEEE/CVF International Conference on Computer Vision, 2021: 2778-2788.
- [21] LI C, LI L, JIANG H, et al. YOLOv6: a single-stage object detection framework for industrial applications [DB/OL]. <https://arxiv.org/abs/2209.02976>.
- [22] WANG C Y, BOCHKOVSKIY A, LIAO H Y M. YOLOv7: trainable bag-of-freebies sets new state-of-the-art for real-time object detectors [C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2023: 7464-7475.
- [23] CAO Y, XU J, LIN S, et al. GCNET: non-local networks meet squeeze-excitation networks and beyond [C]//Proceedings of the IEEE/CVF International Conference on Computer Vision Workshops, 2019: 1971-1980.
- [24] WANG Y, ZHANG J, KAN M, et al. Self-supervised equivariant attention mechanism for weakly supervised semantic segmentation [C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2020: 12275-12284.
- [25] DENG L, ZHOU J, LIU Q. Helmet net: an improved YOLOv8 algorithm for helmet wearing detection[J]. International Journal of Networked and Distributed Computing, 2024, 12: 329-343.

(责任编辑: 陈维捷)