引用:盛耀萱,李烨,张保存,等. 基于协同注意力和特征融合的多类型个人防护装备检测算法[J]. 聊城大学学报(自然科学版),2026,39(5):633-642.

Citation: Sheng Yaoxuan, Li Ye, Zhang Baocun, Chen Yangjun, Liu Qingyi. A multi-type personal protective equipment detection algorithm based on synergistic attention and feature fusion[J]. Journal of Liaocheng University (Natural Science Edition), 2026, 39(5):633-642.

文章编号 1672-6634(2026)05-0633-10 **DOI** 10.19728/j.issn1672-6634.2025120015

基于协同注意力和特征融合的多类型个人防护装备检测算法

盛耀萱[1],李烨[2],张保存[2],陈洋军[3],刘庆一[1]

(1. 山东科技大学 电子信息工程学院,山东 青岛 266590;2. 国家石油天然气管网集团有限公司山东分公司,山东 济南 250002;3. 深圳市科亿达科技有限公司,广东 深圳 518100)

摘要 为确保工人在生产作业中的安全,个人防护装备的佩戴至关重要。基于深度学习的目标检测技术能够低成本、快速地识别防护装备,但在复杂背景和多类型防护装备场景下,存在识别精度下降的挑战。为此,本文提出了一种高性能检测模型 SCASFF-YOLO,在 YOLOv10s 的基础上融合了两项关键改进:其一,引入 SC-PSA 模块,通过空间与通道协同注意力机制强化特征提取能力;其二,在检测头中结合自适应空间特征融合模块,并加入深度可分离卷积优化多尺度特征融合,在提升多类型防护装备检测性能的同时降低计算量。实验证明,所提出的模型在 CHV 数据集上取得了目前同类检测方法中的更高检测精度,为多类型个人防护装备的智能化检测提供了一个具有广阔应用场景的新解决方案。

关键词 个人防护装备;YOLO;空间与通道协同注意力;自适应空间特征融合

中图分类号 TP391.4 **文献标志码** A **开放科学(资源服务)标识码(OSID)**

A multi-type personal protective equipment detection algorithm based on synergistic attention and feature fusion

SHENG Yaoxuan[1], LI Ye[2], ZHANG Baocun[2], CHEN Yangjun[3], LIU Qingyi[1]

(1. School of Electronic and Information Engineering, Shandong University of Science and Technology, Qingdao 266590, China; 2. Shandong Branch of China National Petroleum and Gas Pipeline Network Corporation, Jinan 250002, China; 3. Shenzhen Keyida Technology Co. Ltd., Shenzhen 518100, China)

Abstract To ensure the safety of workers during production operations, the wearing of personal protective equipment is of vital importance. Deep learning-based object detection technology can identify PPE at low cost and high speed, but it faces the challenge of reduced recognition accuracy in complex backgrounds and scenarios with multiple types of PPE. To address this issue, this paper proposes a high-performance detection model, SCASFF-YOLO, which integrates two key improvements on the basis of YOLOv10s: first, it introduces the SC-PSA (Spatial Channel- Pyramid Squeeze Attention) module, which enhances feature ex-

收稿日期:2025-12-19

基金项目:山东省自然科学基金项目(ZR2024MD120)资助

通信作者:刘庆一,男,汉族,博士,讲师,研究方向:图像处理、目标检测与模式识别,E-mail:lqyraphael@163.com。

traction capabilities through a spatial and channel collaborative attention mechanism; second, it combines an adaptive spatial feature fusion module in the detection head and incorporates depthwise separable convolution to optimize multi-scale feature fusion, thereby improving the detection performance of multiple types of PPE while reducing computational load. Experiments have demonstrated that the proposed model achieves the highest detection accuracy among current similar methods on the CHV dataset, providing a new solution and application prospects for the intelligent detection of multiple types of personal protective equipment.

Keywords personal protective equipment; YOLO; spatial and channel synergistic attention; adaptive structure feature fusion

0 引言

个人防护装备(Personal Protective Equipment,PPE)在保障工人安全方面具有至关重要的作用。根据国家统计局数据显示,2024 年中国共有 19 626 人因生产安全事故死亡,其中机械伤害、物体打击等是主要原因[1]。英国健康与安全执行局的统计亦表明,2024 年英国有 138 名工人死于工伤事故,并发生了超过 60 万起非致命事故,其中多数与被移动物体撞击或机械设备有关[2]。因此,各国卫生与安全管理机构普遍要求工人佩戴安全帽、反光背心、手套等防护装备,以减少事故风险[3]。然而,在实际作业中,仍有相当数量的工人因认为佩戴 PPE 不必要或存在不适感而选择忽视,导致企业只能依赖人工检查防护装备的佩戴情况。传统人工检测方式不仅容易产生视觉疲劳,引发误判与漏检,还耗大量人力与物资,难以满足高风险行业对安全管理的要求[4]。

近年来,深度学习技术的突破性进展为这一问题的解决提供了新的思路,研究者开始探索基于深度学习的自动检测技术,通过目标检测算法检测工人防护装备佩戴情况,具有成本低、效率高、部署便捷等优势,已在多种场景中取得实际应用成果。例如,卷积神经网络[5](Convolutional Neural Network,CNN)能够从收集的具有注释的数据中自主学习有用的规律,而不受某些固定特征的限制,得到了更多学者的青睐。刘欣宜[6]等人为了解决污染场地修复中作业人员的安全监测和管理问题,提出了一种快速的区域卷积神经网络(Fast Region-based Convolutional Network,FasterR-CNN)的作业人员着装规范性检测算法,同时检测安全帽和防毒面具两种防护设备。在最近的研究中,Nath 等人[7]为验证建筑人员安全帽和背心的合规性,选用了均衡速度和准确性的 YOLO 算法进行模型研究。Delhi 等人[8]同样关注的是建筑工人防护安全,在 YOLOv3 算法的基础上结合迁移学习开发了一个检测框架,用于检测安全帽和安全夹克的使用合规性。2023 年,Wang 等人[9]通过在 YOLOv5 的基础上引入 MobileNetv3 作为新的骨干,并在特征融合过程中加入残边,提出了 YOLO-M。2024 年,Li 等人[10]提出一种基于改进 YOLOv8 的轻量级算法,用于构建数字安全帽检测系统。通过整合 VoV-GSCSP 模块,降低了模型复杂度,提高了检测精度。2024 年,Liu 等人[11]提出了一种基于 YOLOv8n 的改进检测模型 ML-YOLOv8n-Light 以检测电力工人的个人防护装备。通过将一种新颖的、轻量级的多尺度分组卷积方案集成到体系结构中,来解决安全保护设备尺寸上的差异。

然而,将深度学习应用于 PPE 检测存在许多挑战。首先在复杂背景下,可能导致模型识别的困难。其次。深度学习已广泛应用于安全帽、口罩等单一类型防护装备的目标检测任务,并取得了较好的检测效果。然而,在际工业与施工场景中,工作人员通常同时佩戴多种防护装备,仅针对单一类别的检测难以满足安全需求。尤其是在“人、背心及多类别头盔”的检测任务中,不同目标在外观特征、语义层级上的显著差异,以及头盔类别间的细粒度区分,使得检测难度进一步增加。然而,现有研究主要集中于单一类型防护装备的检测,对多类型防护装备的多目标检测问题关注不足。当检测任务扩展至多目标时,模型训练复杂度显著上升,容易影响检测精度与效率[12]。为了解决上述问题,本文提出了一种高性能目标检测模型 SCASFF-YOLOv10s。该模型融合了两个关键优化策略,提高了多类型防护装备检测的准确度,为复杂场景下的 PPE 检测提供了新的思路。本文的主要贡献如下。

提出了 SC-PSA 模块,通过将原 PSA 模块中的多头自注意力机制替换为空间通道协同注意力机制

(Spatial Channel Synergistic Attention，SCSA)。不同于以往仅进行简单串联或加权融合的空间与通道注意力机制，SCSA 在共享多语义空间下实现了通道与空间特征的逐级交互建模，即先在共享语义空间中建立跨域关联，再通过渐进式通道自注意力强化关键区域响应，对复杂场景下的目标具有更强的特征提取能力和鲁棒性。改进了 YOLOv10s 的检测头，将自适应空间特征融合(Adaptive Structure Feature Fusion，ASFF)机制与检测头相结合，并采用深度可分离卷积(Depthwise Separable Convolution，DSConv)替代 ASFF 中的传统卷积层。ASFF 通过自适应权重分配机制，实现了多尺度特征的最优融合，提升了模型对不同尺寸个人防护装备的检测能力。改进后的 DSConv-ASFF 结构降低了计算复杂度，使模型在保持高精度的同时具备更快的推理速度。

1　相关工作

YOLO 是目前最常用的边缘目标检测模型，在计算机视觉与智能监测领域被广泛应用，因其端到端结构和高实时性，已成为轻量化检测任务的核心框架之一[13]。Joseph 等人[14]首先提出了 YOLO 算法，为直接回归目标盒实现实时检测奠定了基础。Ultralytics[15]推出了 YOLOv5 模型，该模型采用 C3 轻量化特征提取结构实现了轻量化性能。YOLOv5 也是第一个分为 s、m、l、x 四个版本，以满足不同的需求，使其成为使用最广泛的目标检测模型。然而，基于锚点的生成预测盒的方法导致了许多重复的冗余盒，增加了计算负荷。2023 年，Ultralytics 公司[16]推出了 YOLOv8 模型，该模型使用 C2f 来获取更多的梯度流动信息，并采用无锚的方法来加快检测速度，减少了时间和计算能力要求。但是，后处理操作仍然是需要的。Wang 等人[17]提出的 YOLOv10 模型在此基础上进一步改进，通过一致的双重分配策略实现了端到端检测，彻底消除了对 NMS 的依赖，大幅减少了检测时间并提高了推理效率。

YOLOv10 标志着 YOLO 模型系列的重大进步，提供比其前身改进的精度和计算速度。该模型引入了四种架构变体：包括 YOLOv10n、YOLOv10s、YOLOv10m 和 YOLOv10l，覆盖从轻量化到高精度的多种需求。如图 1 所示，YOLOv10 的网络结构分为骨干网络、颈部网络和检测头三部分。

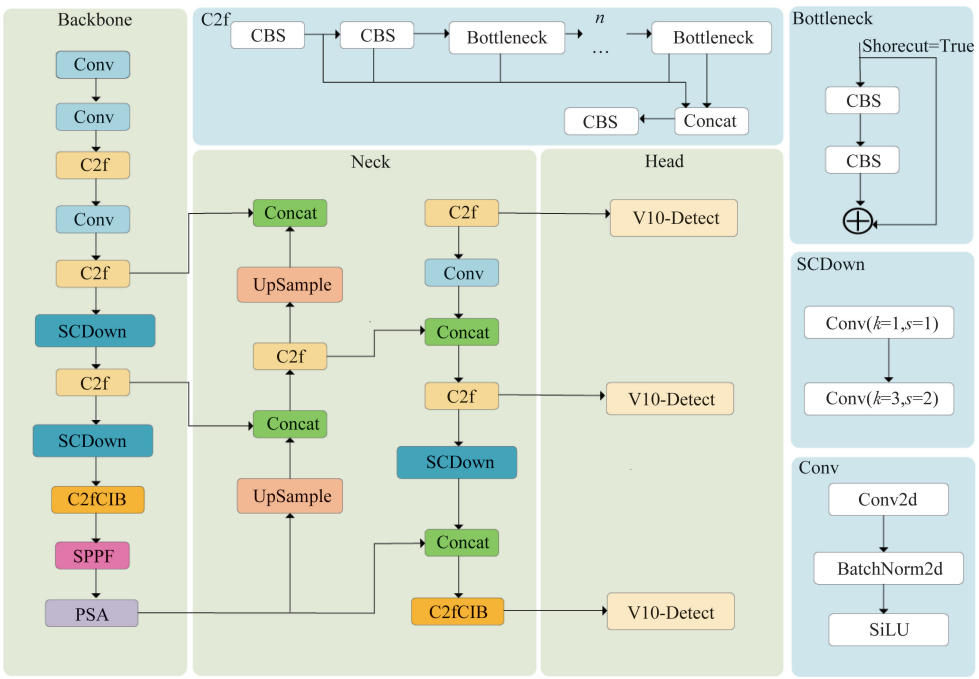

图 1　YOLOv10 的网络结构

Fig. 1　Network architecture of YOLOv10

YOLOv10 的骨干网络采用改进的 CSPNet 结构，高效提取特征并减少计算冗余；颈部网络融合 PANet 与空间通道解耦下采样，实现多尺度特征高效融合[18]；检测头引入一致性双重分配策略与轻量化设计，大幅提升实时性与精度，使结构更简洁、适合工业生产中的目标检测。

2 方法

2.1 改进的 YOLO 模型

本文提出的 SCASFF-YOLO 在 YOLOv10s 的基础上对模型结构进行了两项改进。首先,提出 SC-PSA 模块,用空间与通道协同注意力机制(SCSA)替代原 PSA 模块中的多头自注意力机制(Multi-headed Self-attention,MSHA),通过结合空间与通道维度的信息引导,增强特征表达过程中的语义整合能力。其次,对检测头进行改进,将自适应空间特征融合机制(ASFF)引入检测头,并使用深度可分离卷积(DSConv)替代 ASFF 中的传统卷积结构,从而构建更高效的特征融合模块。

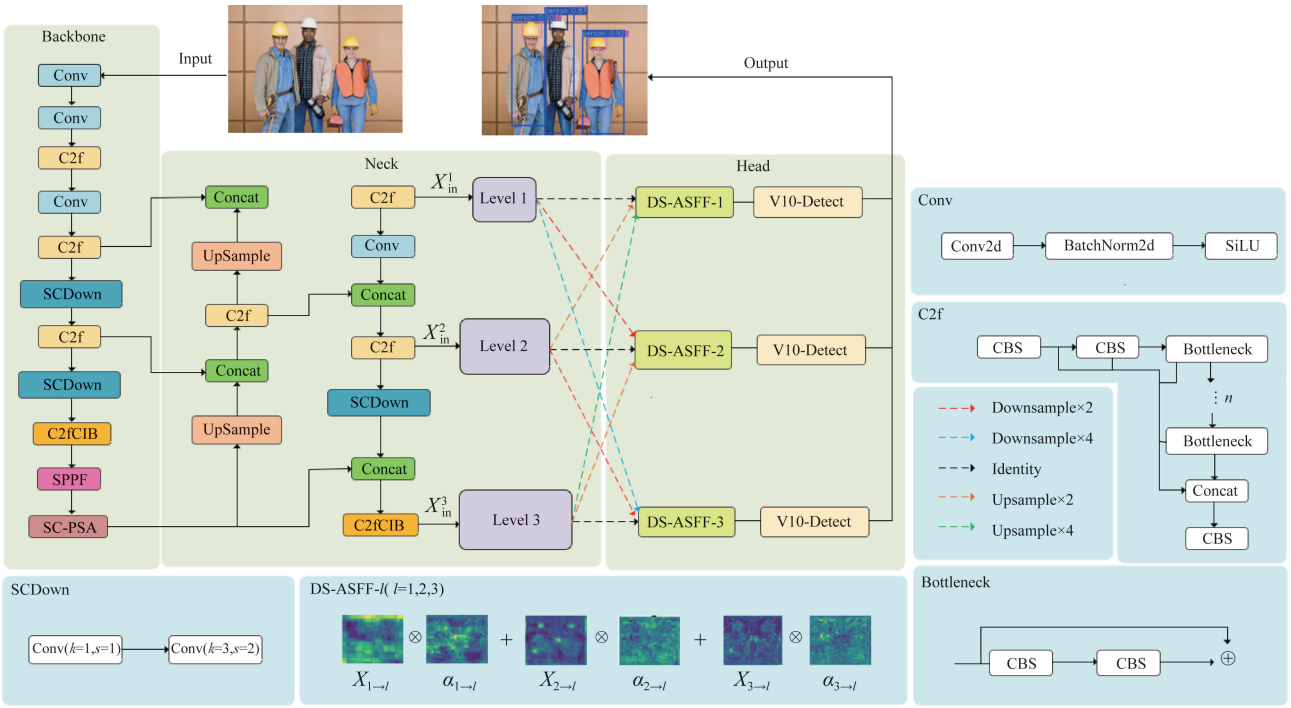

图 2 SCASFF-YOLO 的网络结构

Fig. 2 Network structure of SCASFF-YOLO

整体而言,SCASFF-YOLO 在保持原有结构优势的基础上,通过优化注意力机制与特征融合方式,有效提升了在复杂场景下的检测精度,尤其增强了多目标检测的识别能力。SCASFF-YOLO 的结构图 2 如上。

2.2 SC-PSA 模块

在 SCASFF-YOLO 模型中,本文构建了 SC-PSA 模块,该模块引入了 SCSA 注意机制来取代 PSA 中传统的多头自注意机制。这种改进有效地结合通道和空间注意力的优势,充分利用多语义信息,从而提高视觉任务的表现。SC-PSA 模块的整体结构如图所示。

SCSA 机制通过自适应的权重分配,能够同时关注通道特征和空间位置,从而更加精准地捕捉和理解图像中的信息。传统的 PSA 的 MSHA 机制虽然在捕获全局上下文方面具有出色表现,但其计算成本较高。特别是

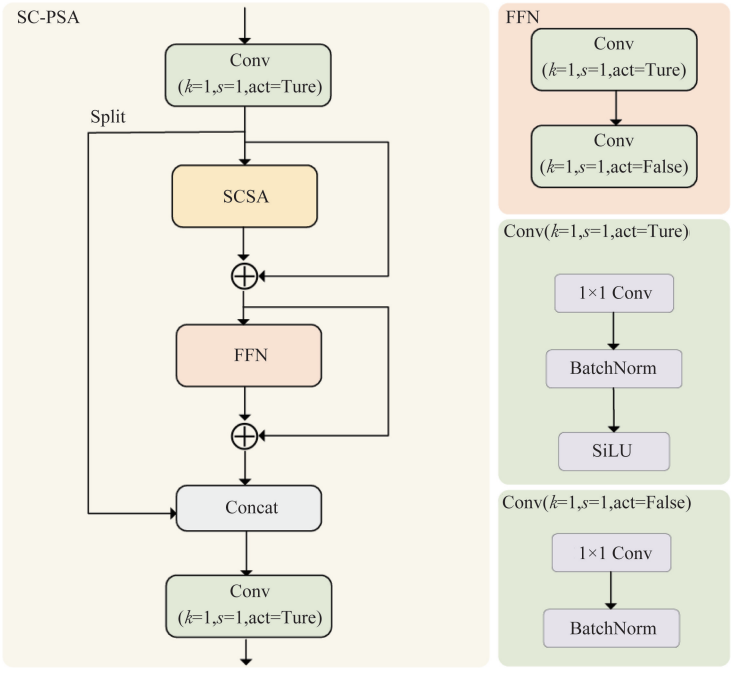

图 3 SC-PSA 的结构图

Fig. 3 Structure diagram of SC-PSA

在处理复杂图像时，其二次复杂度会导致性能瓶颈。与传统的 PSA 的 MSHA 机制相比，SCSA 能够更加高效地提取特征，减少冗余计算，使得模型在处理复杂图像时更加精确且高效。

SCSA 的核心思想是通过协同建模特征间的多语义关系和通道间的渐进依赖，从而实现更精准的特征增强。SCSA 由共享多语义空间注意力（Shared Multi-Semantic Spatial Attention，SMSA）和渐进通道自注意力（Progressive Channel-Wise Self-Attention，PCSA）两个子模块组成。SMSA 与 PCSA 分别用于建模空间维度和通道维度的特征关联。其结构图如图 4。

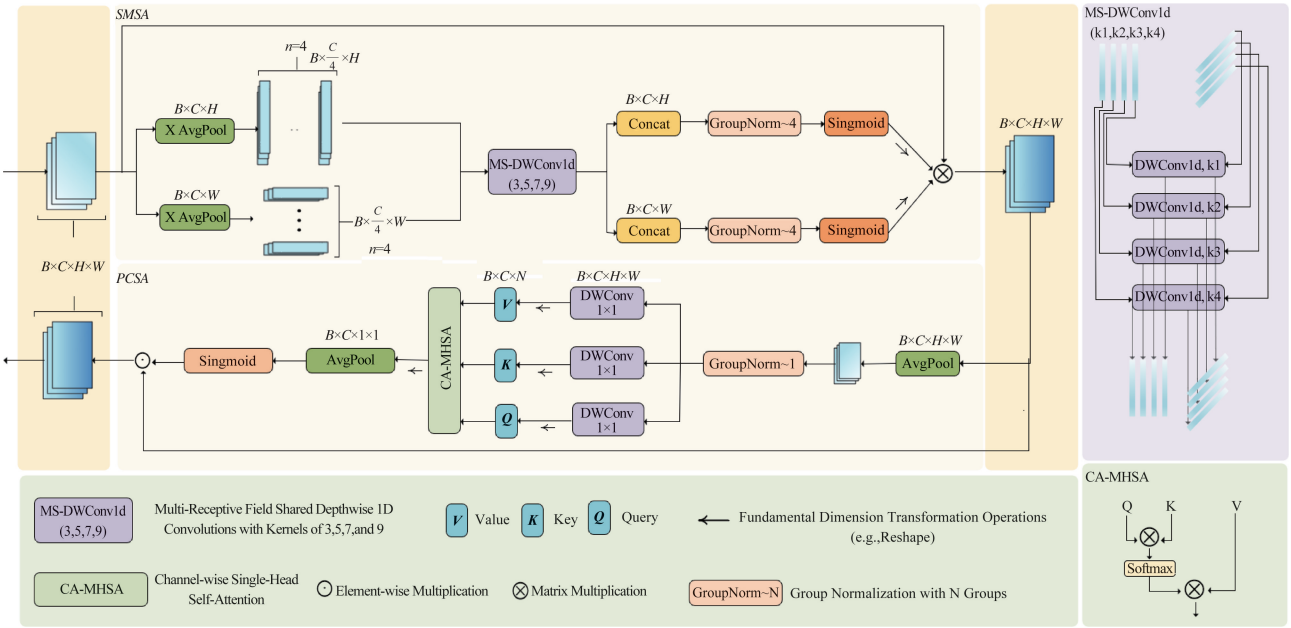

图 4　SCSA 的结构图

Fig. 4　Structure diagram of SCSA

共享多语义空间注意力（SMSA），旨在对输入特征图进行多语义空间的划分，使模型在多个语义子空间中捕捉区域之间的全局依赖关系。首先沿着高度和宽度维度分解给定的输入 $\boldsymbol{X}=\mathbf{R}^{R\times C\times H\times W}$，对每个维度应用全局平均池化，从而创建两个单向 1D 序列结构。将特征集划分为 $\frac{C}{K}$ 个大小相同、独立的子特征 X_H^i 和 X_W^i，每个子特征的通道数为 K，设置默认值 $K=4$。分解成子特征的过程如

$$X_H^i=X_H\left[:,(i-1)\times\frac{C}{K}:i\times\frac{C}{K},:\right], \tag{1}$$

$$X_W^i=X_W\left[:,(i-1)\times\frac{C}{K}:i\times\frac{C}{K},:\right], \tag{2}$$

式中 X^i 表示第 i 个子特征，其中 $i\in(1,K)$。在获得各子特征后，分别对其施加不同尺度的一维深度卷积操作，卷积核尺寸设置为 3、5、7 和 9，以捕获多尺度的空间上下文信息。通过对子特征进行分解并建模不同语义空间特征后，再将不同语义的子特征进行拼接，并采用组规范化（Group Normalization，GN）进行归一化处理，最终得到 SMSA 模块的输出特征。

PCSA 主要用于建模特征通道之间的相互依赖关系。PCSA 采用了基于平均池化的渐进式压缩方法，与使用普通卷积运算建模通道依赖相比，PCSA 表现出更强的输入感知能力，并有效地利用 SMSA 提供的空间先验来深化学习。具体而言，PCSA 的计算流程如下。首先，对 SMSA 输出的特征图 X_s 执行渐进式空间下采样。具体采用两级平均池化操作以逐步压缩特征图尺寸，其计算形式如

$$X_P=\mathrm{Pool}_{(7,7)}^{(H,W)\to(H',W')}(X_s), \tag{3}$$

式中 $\mathrm{Pool}_{(k,k)}^{(H,W)\to(H',W')}(\cdot)$ 表示内核大小 $k\times k$ 的池化操作，依次采用 $k=2$ 与 $k=4$ 的两级池化操作，特征图 X_s 进行逐级压缩，将分辨率从 (H,W) 缩放到 (H',W')。

其次，通过投影函数 $F_{\mathrm{proj}}(\cdot)$ 将下采样后的特征映射为查询（Query，$\boldsymbol{Q}$）、键（Key，$\boldsymbol{K}$）与值（Value，$\boldsymbol{V}$）向

量,$F_{\text{proj}}(\cdot)$定义

$$F_{\text{proj}}=DW\text{Conv1}d_{(1,1)}^{C\to C}\ , \tag{4}$$

$$\boldsymbol{Q}=F_{\text{proj}}^{Q}(\boldsymbol{X}_p),K=F_{\text{proj}}^{K}(\boldsymbol{X}_p),\boldsymbol{V}=F_{\text{proj}}^{V}(\boldsymbol{X}_p), \tag{5}$$

式中 $F_{\text{proj}}(\cdot)$为 1×1、输出通道数均为 $\boldsymbol{C}$ 的逐深度卷积(Depthwise Convolution,DWConv)。此处的参 C 即输入特征图 X_s 的通道数。通过公式(5),得到 $\boldsymbol{Q}$、$\boldsymbol{K}$ 和 $\boldsymbol{V}$ 后,再通过通道自注意力机制对通道间的相关性进行建模,其计算过程如

$$\boldsymbol{X}_{\text{attn}}=\text{Attn}(\boldsymbol{Q},\boldsymbol{K},\boldsymbol{V})=\text{Softmax}\left(\frac{QK^{T}}{\sqrt{C}}\right)\boldsymbol{V}, \tag{6}$$

式中 $\boldsymbol{X}_{\text{attn}}$ 是通过注意力机制 Attn(·)计算得到的中间特征表示,softmax 表示归一化激活函数。$\boldsymbol{QK}^{\text{T}}$ 表示 $\boldsymbol{Q}$ 与 $\boldsymbol{K}$ 的点积。最后 PCSA 模块的整体输出由公式(7)表示,该过程将 $\boldsymbol{X}_{\text{attn}}$ 压缩为 $C\times1\times1$ 的通道权重向量,并通过通道激活函数 $\sigma(\cdot)$ 生成通道注意力权重,随后与输入特征 $\boldsymbol{X}_s$ 进行逐通道相乘,实现通道维度上的自适应特征重标定

$$\text{PCSA}(\boldsymbol{X}_s)=\boldsymbol{X}_c=\boldsymbol{X}_s\times\sigma(\text{Pool}_{(H',W')}^{(H',W')\to(1,1)}(\boldsymbol{X}_{\text{attn}}))\ 。 \tag{7}$$

在此基础上,SCSA 模块将 SMSA 和 PCSA 两部分并联集成,通过融合空间与通道两个维度的注意力机制,实现更全面的信息增强。这样的协同机制使得模型不仅可以感知空间上的多语义特征,还能理解通道间的高阶依赖,提升表示能力。

2.3 基于 ASFF 的 YOLOv10s 检测头优化

为提升模型对不同尺度个体防护装备目标的检测能力,同时兼顾模型推理效率与精度,本文对 YOLOv10s 的检测头进行了结构性改进。具体而言,本文将自适应空间特征融合机制(ASFF)引入 YOLOv10s 的检测头结构中,以更有效地整合来自不同尺度的特征信息;此外,为降低计算复杂度,进一步在 ASFF 模块中用深度可分离卷积(DSConv)替代传统卷积算子,构建轻量级的 DSConv-ASFF 结构。这一优化策略在保障检测精度的基础上,有效提升了推理速度,适用于实际部署场景。

ASFF 旨在通过自适应地融合来自不同尺度的特征图,来提升模型对不同尺寸目标的检测能力。ASFF 通过上采样、下采样或卷积处理不同层级(Level 0、Level 1、Level 2)的特征图,将三个不同尺度的输入特征图 $\{\boldsymbol{X}_{\text{in}}^{1},\boldsymbol{X}_{\text{in}}^{2},\boldsymbol{X}_{\text{in}}^{3}\}$ 映射到第 l 级,映射结果 $\{\boldsymbol{X}_{1\to l},\boldsymbol{X}_{2\to l},\boldsymbol{X}_{3\to l}\}$ 具有共享相同的通道数和维度。在每个级别,映射结果可生成一组可学习的权重 $\{\boldsymbol{\alpha}_{1\to l},\boldsymbol{\alpha}_{2\to l},\boldsymbol{\alpha}_{3\to l}\}$,且满足约束条件 $\sum_{c=1}^{3}\alpha_{c\to l}=1$。在第 l 层 ASFF 的输出结果

$$X_{\text{out}}^{l}=\sum_{c=1}^{3}\alpha_{c\to l}\boldsymbol{X}_{c\to l}\ 。 \tag{8}$$

该过程从多尺度特征图中提取信息,并在特征维度上进行自适应融合。融合后的特征图将通过卷积层进一步处理,以完成边界框定位与类别预测。而本文对卷积层进行了改进,采用了深度可分离卷积(DSConv)代替传统卷积操作。传统卷积操作:输入通道整体一起卷积,每个卷积核大小为,直接得到输出通道。计算量和参数量都比较大。而深度可分离卷积:先做逐通道卷积(Depthwise Convolution)每个输入通道单独卷积,再做逐点卷积(Pointwise Convolution),即使用 1×1 卷积对不同通道的输出进行融合。DSConv 的结构如图 5 所示,深度可分离卷积将传统卷积中空间特征提取与通道信息融合两个过程进行解耦处理。对于输入特征图 X ,其空间尺寸为 $F\times F$,输入通道数为 M 。在第一阶段,逐通道卷积针对每一个输入通道分别采用一个大小为 $K\times K\times1$ 的卷积核进行卷积运算,因此共使用 M 个卷积核,得 M 个中间特征图。该过程仅在各通道内部完成局部空间信息提取,不涉及通道之间的信息交互,从而能够以较低的计算开销实现边缘、纹理等局部特征的建模。

在第二阶段,逐点卷积采用 N 个大小为 $1\times1\times M$ 的卷积核,对前一阶段得到的 M 个中间特征图进行线性组合与通道融合,最终生成通道数为 N 的输出特征图。通过这种分解方式,逐通道卷积首先在各通道内部完成空间特征提取,逐点卷积再对不同通道的特征信息进行融合,从而共同实现完整的特征映射过程。相较于标准卷积,该结构在显著降低计算量和参数量的同时,仍能保持较好的特征提取能力。

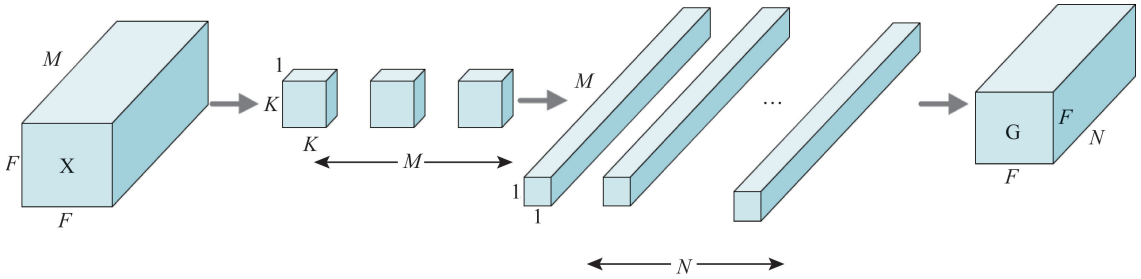

图 5　DSConv 的结构图

Fig. 5　Structure diagram of DSConv

3　实验结果

3.1　数据集与实验平台

本研究在 CHV 通用数据集上对本文所提出的算法进行了测试，CHV 数据集[19]是从约一万张互联网及开放数据源图像中筛选出的 1 330 张高质量图像构成，包含 9 209 个标注实例。数据筛选严格考虑了真实工地场景的多样性、手势变化、多角度拍摄条件以及个人防护装备(PPE)类别的完整性。数据划分为训练集(1 064 幅)、验证集(133 幅)和测试集(133 幅)，标注类别为人员、安全背心、蓝色头盔、红色头盔、白色头盔和黄色头盔。该数据集虽包含少量非工地背景样本，但整体具有较好的场景代表性和检测挑战性，目前已通过 Google Drive 平台公开提供研究使用。

为了确保各种方法之间的公平比较，本研究使用具有 24 GB 内存的 Nvidia RTX 4090GPU 进行训练。实验环境配置是 Cuda12.4 和 Torch2.0.1。在实验过程中，输入图像尺寸统一设置为 640×640，批大小设置为 16，训练轮数设置为 100 epoch。优化器采用 Adam，初始学习率设置为 0.001，动量系数设置为 0.937，权重衰减系数设置为 0.000 5。

3.2　评价指标

为全面评估模型的检测性能，本文采用精确率、召回率、平均精度、平均精度均值、以及模型复杂度相关指标作为评价标准。

为衡量预测框与真实框之间的重叠程度，引入交并比，记为 I，其定义为预测框与真实框交集面积与并集面积之比，计算如

$$I=\frac{S_{\mathrm{i}}}{S_{\mathrm{u}}}, \tag{9}$$

式中 S_{i} 表示预测框与真实框的交集面积，S_{u} 表示预测框与真实框的并集面积。I 的取值范围为[0,1]，其值越大，说明预测框与真实框的匹配程度越高。实际评价过程中，通过设定交并比阈值来判断预测结果是否为正确检测，即当 I 不小于该阈值时，认为预测框与真实框匹配成功，否则视为误检或漏检。

基于上述判定标准，设 T_p 表示该检测到的目标被正确检测出来的数量；F_p 表示本不该检测到却被误检出来的数量；T_n 表示本来就没有目标且未被检测出来的数量；F_n 表示本该检测到但被漏检的数量。在此基础上，精确率记为 P，召回率记为 R，其计算公式分别如

$$P=\frac{T_p}{T_p+F_p}, \tag{10}$$

$$R=\frac{T_p}{T_p+F_n}, \tag{11}$$

式中精确率 P 表示在所有被检测为目标的结果中，真正为目标的比例；召回率 R 表示在所有实际存在的目标中，被成功检测出来的比例。在此基础上，平均精度记为 A，其定义为在不同置信度阈值下生成的精度-召回率曲线(Precision-Recall Curve)所围成的面积，计算公式如

$$A=\int_0^1 P(R)\mathrm{d}R\ , \tag{12}$$

对于包含 n 个类别的目标检测任务，平均精度均值记为 M，表示各类别平均精度的算术平均值，其表达式如

$$M=\frac{1}{n}\sum_{i=1}^{n}A_i, \tag{13}$$

式中 N 表示类别数量，A_i 表示第 i 类的平均精度。M 能够从整体上反映模型在多类别目标检测任务中的综合性能。

在上述基础指标之上，本文进一步采用不同交并比阈值下的平均精度均值对模型进行补充评价。其中，M_{50} 表示在交并比阈值为 0.5 条件下得到的平均精度均值，该指标主要反映模型在较低定位精度要求下的检测能力。M_{50-90} 表示在交并比阈值从 0.5 到 0.9、步长为 0.05 的多个条件下所得平均精度均值的平均结果。总体而言，M_{50} 和 M_{50-90} 的数值越大，表明模型检测性能越好。

此外，本文采用参数量和计算量衡量模型复杂度。其中，参数量记为 C，即模型中可学习参数的总数量，用于表征模型规模；计算量记为 F，即浮点运算总次数，表示模型一次前向推理所需的浮点运算次数，用于表征模型的计算复杂度。一般而言，C 越大，模型规模越大；F 越大，模型计算复杂度越高。

3.3 消融实验

为了客观地评估 SCASFF-YOLO 每项改进的有效性，我们对 CHV 数据集进行了测试。这些实验以原始的 YOLOv10s 模型为基线进行设计，所有输入图像的大小都调整为 640×640。我们引入了 SC-PSA 模块与 DS-ASFF，进行消融实验。实验结果如表 1 所示，主要从参数量 C、计算量 F、精确率 P、召回率 R、以及检测精度(M_{50} 与 M_{50-90})指标进行对比分析。

首先，基线模型 YOLOv10s 在参数量 C 仅为 8.07 M 的情况下，计算量 F 为 24.8 G，M_{50} 为 89.6%，M_{50-90} 为 55.8%，展现了较好的轻量化特性。然而，由于其特征表达能力有限，在复杂场景下的多尺度 PPE 检测精度仍有提升空间。在基线模型的基础上引入 SC-PSA 模块，参数量 C 略微下降至 7.86 M，计算量 F 下降至 24.4 G。这是由于模块中采用了更高效的空间-通道协同注意力机制，对部分计算进行了结构优化。在性能方面，该模型的 M_{50} 提升至 90.6%，相比基线提升了 1.0 个百分点，同时 M_{50-90} 提升至 56.2%，表明在多 IoU 阈值下均具有更稳定的检测表现。

表 1 消融实验

Tab. 1 Ablation experiment

N	YOLOV10s	SC-PSA	DS-ASFF	C /M	F /G	P	R	M_{50}	M_{50-90}
1	√	-	-	8.07	24.8	0.886	0.849	0.896	0.558
2	√	√	-	7.86	24.4	0.908	0.844	0.906	0.562
3	√	√	√	13.7	32.5	0.938	0.841	0.913	0.567

这一结果说明 SCSA 模块能够更充分地挖掘多语义信息，从而增强模型在复杂背景下的特征提取能力。在同时引入 SC-PSA 与 DS-ASFF 结构后，模型参数量 C 增加至 13.7 M，计算量 F 增加至 32.5 G。尽管计算复杂度有所提高，但检测性能也取得了最优结果：M_{50} 达到 91.3%，比基线提升 1.7 个百分点；M_{50-90} 提升至 56.7%，相较于基线提升 0.9 个百分点。这一结果表明，在保持较高效率的同时，ASFF 与 DSConv 的结合能够有效实现多尺度特征的最优融合，提升模型对不同尺寸 PPE 的检测能力，尤其在小目标检测中表现出更高的鲁棒性。

3.4 对比实验及其分析

为验证本文所提出改进模型在目标检测任务中的有效性与优越性，本节将其与主流 YOLO 系列模型进行性能对比。对比模型包括 YOLOv8s、Gold-YOLO、DAMO-YOLO 以及 Mamba-YOLO。实验在相同数据集与训练策略下进行，以确保公平性。各模型在参数量 C、计算量 F、精确率 P、召回率 R、以及检测精度(M_{50} 与 M_{50-90})指标的表现如表 2 所示。

从表中结果可以看出，本文模型在各项指标上均取得最优表现。具体而言，本文模型的 P 值达到 0.938，较 YOLOv8s 提升约 1.8%；R 值提升至 0.841，较 YOLOv8s 增长 0.5%，说明模型在减少漏检方面表现更为出色。在整体检测精度上，M_{50} 达到 0.913，较 YOLOv8s 提升 1.0%，相比 Gold-YOLO 与 DAMO-YOLO 分别提升 3.7%与 5.1%。此外，在更严格的 M_{50-90} 指标下，本文模型取得 0.567 的最高值，相较于现有模型提升明显，充分体现了其在不同 IoU 阈值下的稳定性与泛化能力。

表 2　不同检测模型的性能对比

Tab. 2　Performance comparison of different detection models

Model	P	R	M_{50}	M_{50-90}
Yolov8s	0.921	0.837	0.903	0.561
Gold-YOLO	0.884	0.829	0.876	0.551
DAMO-YOLO	0.871	0.828	0.862	0.537
MambaYOLO	0.896	0.831	0.885	0.545
Ours	0.938	0.841	0.913	0.567

性能提升的主要原因在于结构设计的优化。首先，本文引入 DSConv 取代部分标准卷积层，在显著降低参数量与计算复杂度的同时，保持了高层特征的表达能力。其次，检测头部分加入 ASFF 模块，实现了多尺度特征的自适应融合，使网络能够根据目标特征分布动态调整融合权重，从而提升对小目标与边界模糊目标的检测能力。最后，引入 SCSA 模块增强了通道间的语义交互，使模型在复杂背景下仍能有效聚焦关键区域。与 Mamba-YOLO 等近年来的改进模型相比，本文方法不仅在检测精度上占优，同时兼顾了模型轻量化与推理速度。Mamba-YOLO 虽在特征融合方面具备一定优势，但结构复杂，推理延迟较高；而本文模型通过分层特征融合与轻量化卷积的结合，实现了在保证检测精度的前提下的高效推理，具备良好的工程部署潜力，尤其适用于工业安全检测及边缘计算设备等实时场景。

对比实验的部分可视化结果如图 6 所示，涵盖单人、多人、远景、昏暗及遮挡等典型检测场景，能够较全面地反映模型在不同环境下的适应性与鲁棒性。从整体检测效果来看，SCASFF-YOLO 能够在多种复杂条件下实现高精度检测。无论是在人员间隔较小的场景〔如图 6(b)〕，还是光线不足的昏暗或夜间场景〔如图 6(d)与图 6(e)〕，模型都能准确识别出人员及其佩戴的防护装备，并对背心及不同颜色的头盔进行有效分类。检测框分布均匀、边界清晰，说明模型在多人交互和低光照环境中均能保持良好的特征分离能力和稳定的检测输出。在远景俯视场景中〔如图 6(c)〕，工人目标较小、背景复杂，传统检测模型往往容易出现漏检现象，而 SCASFF-YOLO 依然能准确检测出远距离目标并保持较高置信度，体现出其对小目标和不同视角的良好适应性。此外，在遮挡和复杂背景条件下〔如图 6(f)〕，模型依然能够完整识别被部分遮挡的人员及防护装备，边界框位置准确、类别判定正确。

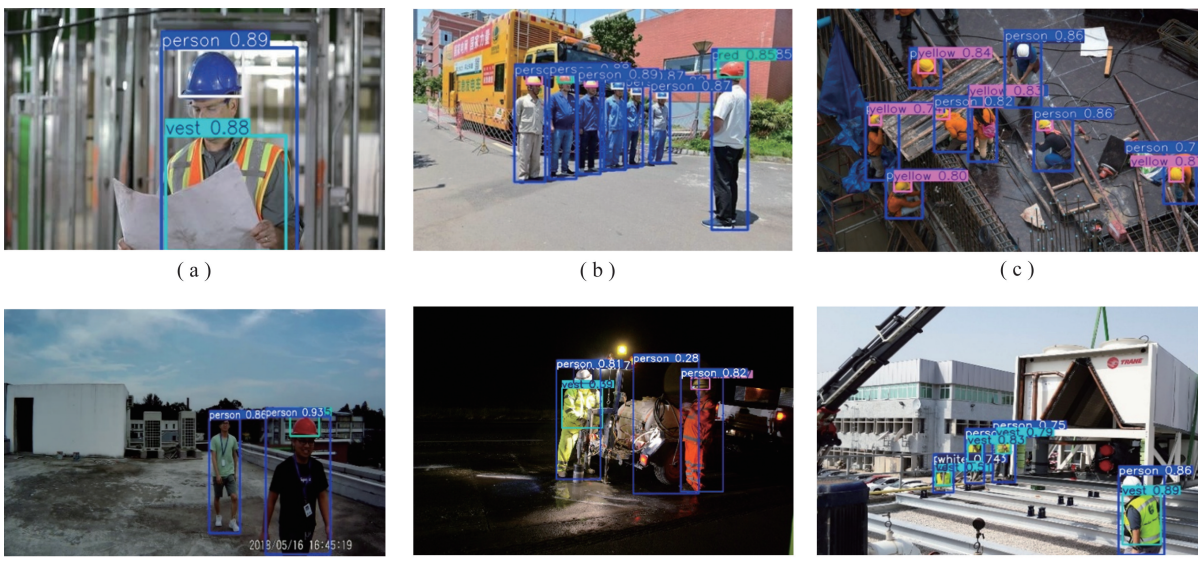

图 6　SCASFF-YOLO 在 CHV 数据集上不同场景下的检测结果可视化

Fig. 6　Visualization of the detection results of SCASFF-YOLO in different scenarios on the CHV dataset

总体来看，SCASFF-YOLO 在 CHV 数据集上的检测结果显示，该模型能够稳定应对多种复杂环境，包括多人密集场景、俯视远景、低光照与遮挡条件，并能同时检测多类防护目标。图 6(a～f)的可视化结果清晰展示了模型在不同场景下的表现：从单人到多人、从明亮到昏暗、从近景到远景、从无遮挡到被遮挡，模型均能保持高精度检测与稳定输出。由此可见，SCASFF-YOLO 能有效满足工业安全检测中对多目标、多场

景、实时性的综合要求,为施工安全监测与违规检测提供了可靠的技术支撑。

4 结论

本文提出了一种高性能目标检测模型 SCASFF-YOLO。该模型在 YOLOv10s 的基础上进行了两方面的改进:一是引入 SC-PSA 模块,利用空间与通道协同注意力机制替代传统的多头自注意力机制,从而提升特征表达能力和复杂场景下的鲁棒性;二是改进检测头结构,将自适应空间特征融合与深度可分离卷积相结合,以实现多尺度特征的高效融合,提高检测精度。实验结果表明,SCASFF-YOLO 在多类别 PPE 检测任务中取得了目前最高的精度表现,其中 M_{50} 和 M_{50-90} 均显著优于基线模型和对比方法,验证了两项改进策略在提升检测能力和适应复杂环境方面的有效性。未来研究拟从以下几个方向展开:一是探索更轻量化的结构设计,提升实时性与边缘端适用性;二是通过跨域迁移学习和大规模实地测试,进一步提升模型的通用性与可靠性。

参 考 文 献

[1] 中华人民共和国国家统计局. 中华人民共和国 2005 年国民经济和社会发展统计公报[J]. 中国统计, 2006(3): 4-8.

[2] Health and Safety Executive. Work-related fatal injuries in Great Britain, April 2023 to March 2024[EB/OL]. 2024-07-03 [2025-12-02]. https://www.gov.uk/government/statistics/work-related-fatal-injuries-in-great-britain-april-2023-to-march-2024.

[3] Nghitanwa E M, Lindiwe Z. Occupational accidents and injuries among workers in the construction industry of Windhoek, Namibia[J]. International Journal of Health, 2016, 5(1): 55-59.

[4] Yang X, Yu Y T, Ssirowzhan S, et al. Automated PPE-tool pair check system for construction safety using smart IoT[J]. Journal of Building Engineering, 2020, 32: 101721.

[5] Shin H C, Roth H R, Gao M C, et al. Deep convolutional neural networks for computer-aided detection: CNN architectures, dataset characteristics and transfer learning[J]. IEEE Transactions on Medical Imaging, 2016, 35(5): 1285-1298.

[6] 刘欣宜, 张宝峰, 符烨, 等. 基于深度学习的污染场地作业人员着装规范性检测[J]. 中国安全生产科学技术, 2020, 16(7): 169-175.

[7] Nath N D, Behzadan A H, Paal S G. Deep learning for site safety: Real-time detection of personal protective equipment[J]. Automation in Construction, 2020, 112: 103085.

[8] Delhi V S K, Sankarlal R, Thomas A. Detection of personal protective equipment (PPE) compliance on construction site using computer vision based deep learning techniques[J]. Frontiers in Built Environment, 2020, 6: 136.

[9] Wang L L, Zhang X J, Yang H L. Safety helmet wearing detection model based on improved YOLO-M[J]. IEEE Access, 2023, 11: 26247-26257.

[10] Li Z H, Huang Z W, Chen H Y, et al. Safety helmet detection in electrical power scenes based on improved lightweight YOLOv8[J]. Journal of Physics: Conference Series, 2024, 2774(1): 012043.

[11] Liu X, Li Y N. A multiscale grouped convolution and lightweight adaptive downsampling-based detection of protective equipment for power workers[J]. Electronics, 2024, 13(11): 2079.

[12] Vukicevic A M, Petrovic M, Milosevic P, et al. A systematic review of computer vision-based personal protective equipment compliance in industry practice: Advancements, challenges and future directions[J]. Artificial Intelligence Review, 2024, 57(12): 319.

[13] Liu Y, Zhang Z, Wang X, et al. A comprehensive survey on YOLO-based object detection algorithms[J]. Journal of Visual Communication and Image Representation, 2023, 92: 103742.

[14] 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 RecognitionIEEE, 2016: 779-788.

[15] Jocher G, Stoken A, Borovec J, et al. Diaconu L, et al. ultralytics/yolov5: v3.0[EB/OL]. (2020-08-13) [2026-03-17]. https://doi.org/10.5281/zenodo.3983579.

[16] Jocher G, Stoken A, Qiu J. ULTRALYTICS. YOLOv8[EB/OL]. (2023-01-10) [2025-02-28]. https://github.com/ultralytics/ultralytics.

[17] Chen H, Chen K, Ding G G, et al. YOLOv10: real-time end-to-end object detection[C]. //Proceedings of Advances in Neural Information Processing Systems 37. San Diego: Neural Information Processing Systems Foundation, Inc., 2024, 37: 107984-108011.

[18] Liu S, Qi L, Qin H F, et al. Path aggregation network for instance segmentation[C]. //Proceedings of IEEE/CVF Conference on Computer Vision and Pattern RecognitionIEEE, 2018: 8759-8768.

[19] Wang Z J, Wu Y M, Yang L C, et al. Fast personal protective equipment detection for real construction sites using deep learning approaches[J]. Sensors, 2021, 21(10): 3478.

(责任编辑:赵海军)