

基于改进 YOLOv10 的安全帽检测算法研究

黄 昆, 李天明, 尹建华, 曹 奔, 曹 刨, 胡文杰

(湖北三江航天江北机械工程有限公司, 湖北 孝感 432000)

摘要: 安全帽佩戴检测是工厂安全重要组成部分, 其采用模式识别的方法对监控中工人的安全帽佩戴情况进行检测, 进而实现智能监控; 针对工地工厂环境下由于工人在监控中的尺度不同, 场景复杂特征提取较难等问题, 对 YOLOv10n 进行研究提出了 Helmet-YOLO 算法; 在主干网络设计了 SimC2f 模块, 加强了算法在复杂场景下对安全帽特征的提取和表达; 在颈部采用动态选择性注意力机制, 使特征融合过程中充分利用长距离语义信息; 在上采样部分引入轻量化动态上采样算子, 提高了上采样的质量; 实验结果表明该算法在复杂场景 SHWD 数据集 mAP_{50} 和 mAP_{50-95} 分别取得了 91.5% 和 58.2% 的检测效果, 在仅升高 0.3 GFLOPS 的情况下, 与 YOLOv10n 相比分别提高了 2.2% 和 1.3%, 检测效果取得了提升。

关键词: 安全帽佩戴检测; YOLOv10n; Helmet-YOLO; SimC2f 模块; 动态选择性注意力; 轻量化动态上采样算子

Research on Safety Helmet Detection Algorithm Based on Improved YOLOv10

HUANG Kun, LI Tianming, YIN Jianhua, CAO Ben, CAO Zhao, HU Wenjie

(Hubei Sanjiang Aerospace Jiangbei Machine Engineering Co., Ltd., Xiaogan 432000, China)

Abstract: Safety helmet wearing detection is a critical part of factory safety, which uses pattern recognition methods to monitor the wearing of safety helmets by workers during monitoring, thereby achieving intelligent detection; In construction and factory environments, it is difficult to extract the features with different scales and complex scenes in monitoring workers; Thereby, research on the YOLOv10n is conducted to propose a Helmet-YOLO algorithm; The SimC2f module is designed in the backbone network to enhance the algorithm's ability to extract and represent the features of safety helmets in complex scenes; A dynamic selective attention mechanism is adopted in the neck network to ensure that the long-range semantic information is fully utilized during feature fusion. A lightweight dynamic upsampling operator is introduced in the up-sampling part to improve the quality of upsampling. Experimental results show that the algorithm achieves the mAP_{50} by 91.5% and the mAP_{50-95} by 58.2% on the SHWD dataset, respectively. With an increase of only 0.3 GFLOPS, the mAP_{50} and mAP_{50-95} of the algorithm increase by 2.2% and 1.3% than those of the YOLOv10n, respectively with an improvement in detection performance.

Keywords: safety helmet wearing detection; YOLOv10n; Helmet-YOLO; SimC2f; dynamic selective attention; light-weight dynamic upsampling

0 引言

中国作为工业大国, 工业生产是经济发展必不可少的推动力。然而在工厂的场景下通常存在很多安全隐患, 为了尽可能保护工人的生命财产安全, 确保生产安全有序的进行, 正确佩戴安全帽尤为重要^[1]。目前监督佩戴安全帽的方式主要是人工监督, 但人工监督存在雇佣人员成本高, 精力有限无法实行全面监控的问题, 检测的效果不太理想。随着深度神经网络技术和高性能计

算设备的持续进步, 目标检测算法以速度快、效率高的优势为安全帽检测提供了新的解决方案, 研究如何实时高精度的实现安全帽检测也具有重要的研究意义。

在安全帽检测领域, 不少学者曾尝试运用传统的图像分析技术以及手动提取图像特征的方式来构建检测模型。例如, 刘晓慧团队^[2]便采用 Hu 矩来提取图像特征, 并通过分析安全帽与头发、皮肤之间的颜色差异, 进而判断工人是否正确佩戴了安全帽; Chiverton 等^[3]则采用先定位人体目标, 再提取头部区域的 HOG、

收稿日期:2024-10-22; 修回日期:2024-12-09。

作者简介:黄 昆(1997-), 男, 硕士。

引用格式:黄 昆, 李天明, 尹建华, 等. 基于改进 YOLOv10 的安全帽检测算法研究[J]. 计算机测量与控制, 2025, 33(11): 42

SIFT 特征来达成安全帽佩戴状态的识别; Mneymneh 等^[4]则是通过匹配工人与安全帽的空间位置关系, 借助特征匹配的对应点来识别工人是否规范佩戴了安全帽。尽管这类方案在一定程度上实现了检测并降低了计算资源消耗, 但在检测精度和环境适应性方面仍存在明显不足, 仅能在预设的特定环境中完成对安全帽的识别, 难以应对工厂复杂多变的作业场景下的实际检测要求。

在深度学习技术取得重大进展的背景下, 基于深度学习的目标检测方法正逐步占据安全帽检测领域的主导地位。这类算法借助卷积神经网络来提取具有更高层次和更强表达能力的安全帽特征, 从而切实提升了检测工作的精度。基于深度学习的目标检测算法按检测流程分为二阶 (two-stage) 目标和一阶 (one stage) 目标检测算法两种类型。

二阶方法的检测过程通常包含两个环节, 首先对图像进行候选框区域提取, 然后再对这些区域进行分类和回归, 其中 R-CNN 系列^[5-7]为最具代表性的二阶算法。在这一研究方向下, 邓开发等人^[8]将 Faster R-CNN 与深度特征流技术相融合, 用于施工现场监控场景中的安全帽识别任务; 徐守坤等人^[9]则在 Faster RCNN 的基础上, 通过增加锚点数量并引入在线困难样本挖掘技术对施工现场安全帽佩戴情况进行有效检测。二阶算法虽精度可观, 但是检测过程复杂导致检测速度较慢, 并不适合实时性需求高的应用场景。

一阶目标检测算法定位和分类一起进行, 不仅在检测精度上表现出色, 其处理速度也相当可观。YOLO 算法^[10]作为最经典的一阶算法近年来吸引很多学者进行研究。杨永波等人^[11]基于 YOLOv5s 提出了 YOLO-M3 算法, 采用 Mobilenetv3、CBAM 注意力机制提升了安全帽的检测效果; 韩锟栋等人^[12]对 YOLOv5 进行改进, 采用微尺度检测层进行特征提取, 在特征融合网络中引入坐标注意力机制实现安全帽检测; 冯勇等人^[13]在 YOLOv8 主干引入 GhostNet 网络并对下采样进行轻量化改进实现实时的安全帽检测; 韩博等人^[14]采用 EMA 注意力机制和 VoVGSCSP 模块对 YOLOv8 的进行改进, 提升了对安全帽遮挡目标的检测效果。虽然上述方法使检测性能取得了一定的提升, 但在目标太小, 周边环境复杂的情况下仍存在一定的漏检和误检的问题, 仍存在较大挑战。

为了解决上述问题, 以 YOLOv10n 为基础, 面向复杂场景下的安全帽佩戴检测提出了改进的 YOLOv10n 算法, 并通过实验验证了该算法既可以在提高检测精度, 也能保证检测的实时性。

1 YOLOv10 算法介绍

YOLOv10^[15]是 2024 年提出的 YOLO 算法。其总

体结构与 YOLOv8 类似。与 YOLOv8 相比, 其改进主要体现在以下几个方面: 首先在主干和颈部神经网络的模块中引入了 SCDown 下采样结构, 该结构通过先采用点卷积操作调整特征图的通道维度, 随后利用深度卷积实现空间维度的压缩, 在有效维持模型原有检测精度的前提下, 减少了计算复杂度和模型参数规模; 在主干网络的尾部, YOLOv10 引入了局部自注意力机制 PSA^[16], 提高了主干网络的特征提取能力; 在检测头方面, YOLOv10 将检测头分散成了一对多和一对一检测头, 这两检测头同时训练, 使算法可以利用一对多分配的全面监督, 从而提高算法模型的学习能力和准确性。

2 Helmet-YOLO 算法

针对工业生产环境中目标尺寸过小、特征信息提取困难导致的检测精度不足问题, 在 YOLOv10 算法基础上进行优化设计了 Helmet-YOLO, 其网络架构如图 1 所示。首先, 在主干部分设计了 SimC2f 模块提高网络的特征提取能力, 加强网络对目标关键区域的聚焦程度, 剔除不相干因素的干扰; 其次, 在颈部引入了 LSK 动态选择性注意力机制, 使网络能够更为全面的使用图像蕴含的上下文信息和背景方面的先验知识; 最后对颈部上采样部分进行优化, 使用 DySample 上采样算子对进行改进, 提高了上采样质量且没有造成额外的计算开销, 接下来将对这些改进点进行详细介绍。

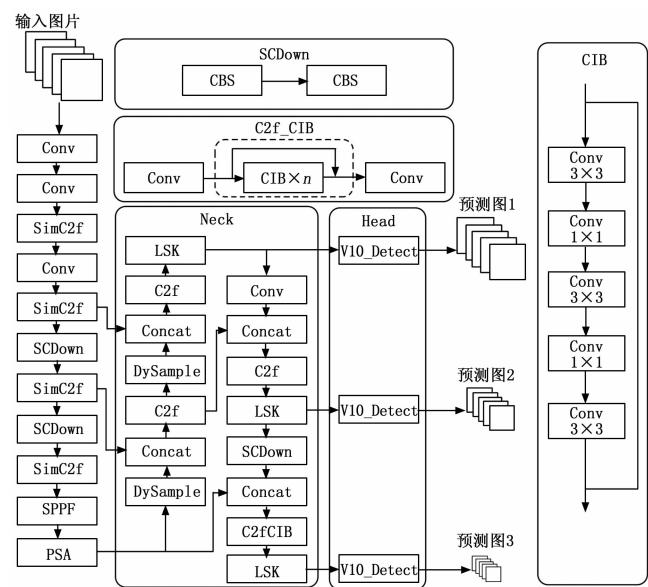


图 1 Helmet-YOLOv10 网络结构

2.1 SimC2f 模块

在工厂工地等复杂作业环境中, 安全帽的外观特征常与周围诸多物体存在较高相似度, 例如一些呈半球形的工具、机械部件及零散零件等, 这些无关物品很容易

对目标检测的精确性造成干扰。为解决此问题, 在主干网络架构中以 SimAM 注意力机制^[17]为基础进行改进, 设计了 SimC2f 模块。该模块能够增强对图像中关键区域及目标的关注度, 有效排除图像中不相关的干扰内容, 提高对检测目标特征提取能力。

如图 2 所示为 SimAM 注意力的结构图。SimAM 作为一种运用 3D 权重的非参数化注意力机制, 与 CBAM 注意力^[18]中简单地将空间和通道注意力相组合的方式存在差异。SimAM 能够衡量各维度神经元间的线性可分离性, 推断出特征映射的三维注意力权重, 通过对空间位置特征和通道信息的全面评估来实现对关键神经元的重点关注, 以此增强模型对图像整体及上下文特征的捕捉能力。各维度注意力权重的计算公式如 (1) 和 (2) 所示:

$$\begin{cases} e_t^* = \frac{4(\hat{\sigma}^2 + \lambda)}{(t - \hat{\mu})^2 + 2\hat{\sigma}^2 + 2\lambda} \\ w_t = \frac{1}{e_t^*} \end{cases} \quad (1)$$

$$X' = \text{sigmoid}(\frac{1}{E}) \odot X \quad (2)$$

公式 (1) 为神经元能量函数, e_t^* 代表神经元的能量值, t 为目标神经元, λ 为正则化系数, $\hat{\mu}$ 为输入特征中每个通道上的均值, $\hat{\sigma}^2$ 为输入特征图每个通道上的方差。神经元的能量值越低, 则该神经元与相邻神经元的差异越大, 则表示该神经元的重要程度也就越高, 该神经元也将被赋予更高的权值。然后通过公式 (2) 采用 sigmoid 函数限制能量函数的范围得到神经元注意力权重, 这里 E 为所有神经元的最小能量函数构成的集合。最终, 将这些注意力权重与初始输入的特征图 X 进行融合处理, 从而生成新的输出特征图 X' 。

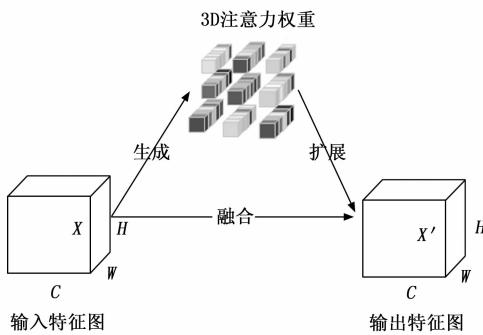


图 2 SimAM 注意力机制

为了尽可能在不增加运算量的情况下加强算法对安全帽特征的提取和表达能力。基于 SimAM 设计了 SimC2f 模块, 代替主干网络的 C2f 模块用于特征提取工作。如图 3 所示, 对输入特征使用 1×1 卷积调整通道数, 再用 Split 操作对特征图进行切分处理, 通过注意力瓶颈模块 SimBottleneck 选择性的提取图像特征,

加强对关键区域及目标特征的聚焦程度, 抑制与目标无关的背景区域特征。与原本的 C2f 相比, SimC2f 能更精准的提取安全帽目标特征, 而且没有增加模型参数和浮点运算量。

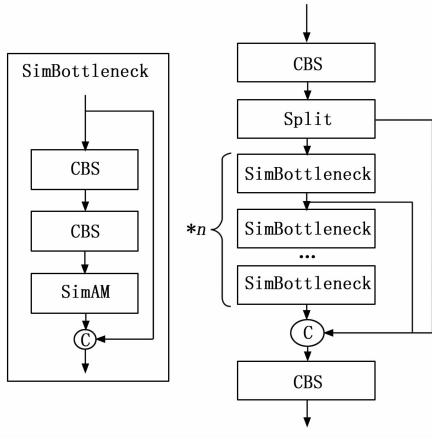


图 3 SimC2f 结构图

2.2 颈部引入 LSK 注意力机制

在工厂工地的复杂场景下, 由于目标距离、目标遮挡、拍摄视角和图像分辨率等变化都会对安全帽的外观特征产生很大影响, 因此充分利用图像的上下文信息和背景先验知识对精准检测非常重要。然而普通的卷积神经网络感受野不足, 无法有效利用图像上下文信息, 在目标密集重叠、目标太小时容易出现漏检的问题。针对这个问题, 采用 LSK 动态选择性注意力^[19]对 YOLOv10 的颈部进行改进。LSK 可以通过空间选择机制动态的调整感受野, 获取动态且丰富的上下文信息, LSK 模块的结构如图 4 所示。

首先将一个大核卷积拆分为一系列深度可分离卷积, 其卷积核可以逐渐扩张, 目的是为了在不增加模型参数量的同时获取更多上下文信息。其中第 i 个深度可分离卷积的卷积核大小 k 、扩张率 d 和感受野 RF 的参数可定义如公式 (3) 和 (4) 所示。通过增加卷积核的尺寸和扩张速率, 能够有效实现感受野的快速扩张。同时设定了扩张率的上限, 以此来避免在特征图之间产生多余的空隙:

$$k_{i-1} \leq k_i, d_1 = 1, d_{i-1} < d_i \leq RF_{i-1} \quad (3)$$

$$RF_i = k_i, RF_i = d_i(k_i - 1) + RF_{i-1} \quad (4)$$

将特征图 X 输入到具有不同感受野的深度可分离卷积 F_{i+1}^{dw} , 从不同范围的输入中获得丰富的上下文信息特征, 并在每个深度可分离卷积后面加上一个 1×1 的卷积 $F_i^{1 \times 1}$ 进行通道融合:

$$U_0 = X, U_{i+1} = F_{i+1}^{dw}(U_i) \quad (5)$$

$$\tilde{U}_i = F_i^{1 \times 1}(U_i), i \in [1, N] \quad (6)$$

将由不同的深度可分离卷积得到的不同感受野的特征图进行拼接操作:

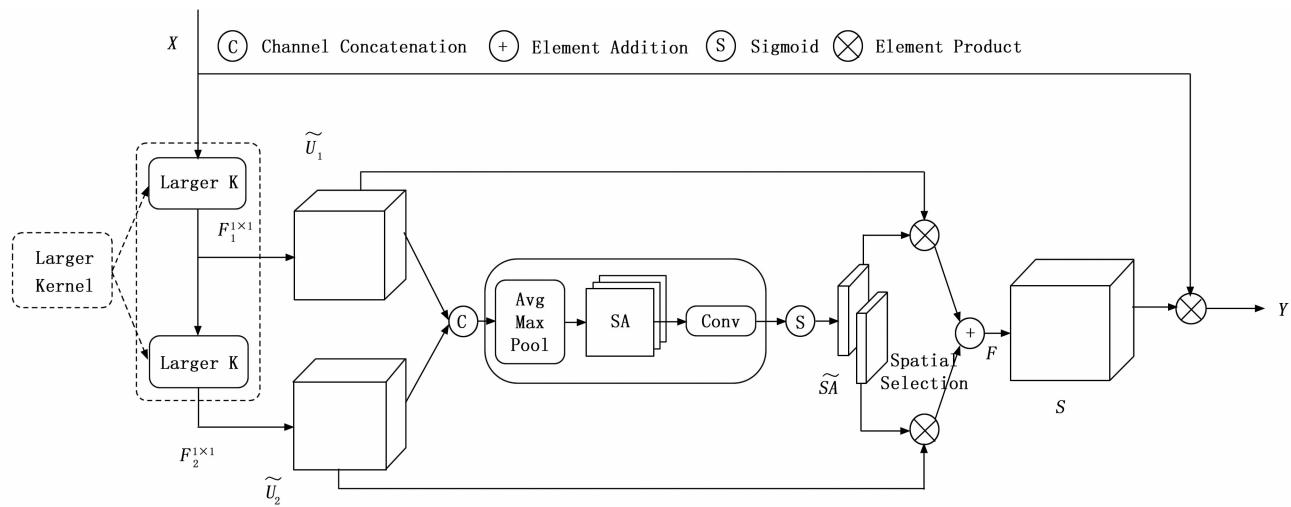


图 4 LSK 注意力机制结构图

$$\mathcal{U} = [\mathcal{U}_1, \dots, \mathcal{U}_i] \quad (7)$$

然后对 \mathcal{U} 采用基于通道的平均和最大池化得到特征图 SA_{avg} 和 SA_{max} 。为了使不同空间特征进行信息交互用, 接下来使用卷积操作 $F^{2 \rightarrow N}$ 将 SA_{avg} 和 SA_{max} 进行拼接, 经过激活函数后, 生成空间注意力特征图 SA 。在此基础上, 对空间注意力特征图与深度可分离卷积序列所生成的特征图实施加权操作, 随后借助卷积层 $F(\cdot)$ 进行融合处理, 进而获取 LSK 最终的注意力特征 S 。以上过程如下所示:

$$SA = F^{2 \rightarrow N}([SA_{avg}; SA_{max}]) \quad (8)$$

$$SA = \sigma(SA) \quad (9)$$

$$S = F\left(\sum_{i=1}^N (SA_i \cdot \mathcal{U}_i)\right) \quad (10)$$

最后将输入特征图 X 与最终注意力特征 S 进行逐元素的相乘, 得到 LSK 注意力机制的最终输出:

$$Y = X \cdot S \quad (11)$$

将 LSK 模块使用在了颈部网络, 使颈部网络在特征融合的过程中通过动态选择性注意力机制对图像上下文信息进行理解, 有效的提取不同目标上下文信息的细

微差别, 很好的提升了安全帽佩戴检测的效果。

2.3 DySample 轻量化动态上采样算子

上采样模块是 YOLO 颈部网络的重要的组成部分, 主要用于恢复高层级特征图的图像分辨率。在工厂和工地环境下对尺寸较小的目标有着极高的检测精度要求, 然而, 原始的 YOLO 算法所采用的上采样技术, 像最近邻插值和双线性插值等, 其上采样核的生成完全依赖于像素点的空间坐标位置, 未能充分运用特征图所蕴含的语义信息, 同时也忽略了图像中周围特征点可能产生的相互影响, 且感知范围较为有限, 导致上采样处理后的图像质量欠佳, 进而造成了小尺寸目标特征的丢失与变形。近几年出现了很多改进方法, 如 CARAFE^[20]、FADE^[21] 和 SAPA^[22] 等以内核为基础的动态上采样技术, 虽然提高了上采样质量, 但也明显增加了计算负担, 这是由于动态卷积本身的运算复杂度以及生成动态卷积内核所需的附加子网络造成的。因此引入了一种动态点采样的上采样算子 DySample^[23], 不同于以内核为基础的动态上采样方法, DySample 能够在输入特征图上动态地选择采样点, 既提高了上采样质量又不会造成额外的计算开销。DySample 的上采样过程如图 5 所示。

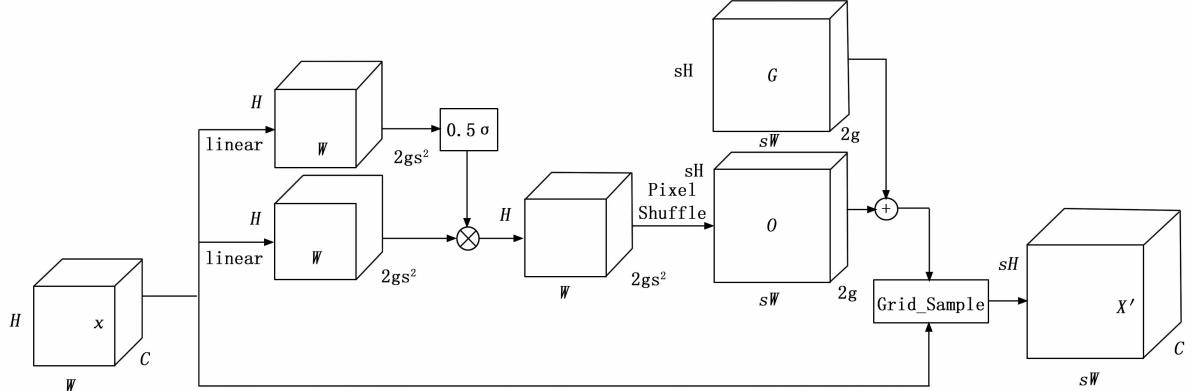


图 5 DySample 上采样算子结构

DySample 的上采样过程主要分为生成点采样集和根据点采样集进行上采样两个步骤。对于给定一个 $C \times H \times W$ 的低分辨率输入特征图, 首先通过一个线性层来计算每个上采样点的偏移量, 这些偏移量将用于确定上采样操作后该点的新坐标位置。同时有另一个线性层负责生成偏移量的动态范围因子, 经过 sigmoid 处理后与偏移量相乘得到具有动态特性偏移量。这种动态调整机制使偏移量能根据输入特征图的具体内容进行相应变化, 进而提升了整体操作的灵活性, 这个过程如公式 (12) 所示:

$$O = 0.5 \text{sigmoid}[\text{linear}_1(\chi)] \cdot \text{linear}_2(\chi) \quad (12)$$

将动态偏移量 O 通过像素重组运算 (Pixel Shuffle) 重塑为 $2g \times sH \times sW$ 的形状, 其中 s 为上采样倍率, 第一维度 $2g$ 为要上采样点的横纵坐标信息, 这样就得到了各点的偏移量及上采样前后点的对应关系信息。接着, 需将偏移量 O 整合至原始采样网格 G 中, 从而构建出最终的点采样集 δ , 如公式 (13) 所示。最后由网格采样函数 Grid_Sample 基于 δ 位置信息将输入特征图选择性上采样得到大小为 $C \times sH \times sW$ 的上采样图, 如公式 (14) 所示:

$$\delta = O + G \quad (13)$$

$$\chi' = \text{Grid_Sample}(\chi, \delta) \quad (14)$$

通过上述步骤, DySample 实现了从低分辨率特征图到高分辨率特征图的动态上采样, 其中采样点的生成和位置的调整都是根据输入特征图的内容动态确定的, 有效保证了上采样过程的效率和效果。为了解决传统上采样造成的特征图失真使小目标特征丢失的问题, 引入 DySample 作为 Helmet-YOLO 的上采样模块, 在没有增加算力负担的情况下有效提高了上采样图的质量, 更好的保留了小目标的特征信息, 进而提升安全帽佩戴检测的性能。

3 实验分析与比较

3.1 SHWD 数据集介绍

SHWD (Safety Helmet Wearing Dataset) 数据集是一个用于安全头盔佩戴的公共数据集。该数据集包含 7 581 张图像, 其中涵盖 9 044 个正确佩戴安全帽目标对象和 111 514 个未佩戴安全帽的头部目标对象。该数据集涵盖了丰富的场景环境与多变的人体姿态表现, 主要用于工地安全和工人行为规范检测等领域, 对于安全帽佩戴检测算法的效果验证具有较高的参考价值。如图 6 所示为该数据集部分图片示例。

3.2 实验设置与评价指标

3.2.1 实验设置

实验设备配置如表 1 所示。

实验参数配置如表 2 所示。



图 6 SHWD 数据集

表 1 实验设备配置

名称	配置
操作系统	Ubuntu 20.04
CPU	AMD EPYC 9654
内存	60 G
GPU	RTX3090 × 1
显存	24 G
CUDA	12.2
开发工具	PyCharm
Python	3.8
PyTorch	1.11.0

表 2 实验参数配置

名称	配置
数据集	SHWD
划分比例	7:2:1
图片尺寸	640 × 640
Epoch	100
Batch Size	16
优化器	SGD
学习率	0.01

3.2.2 评价指标

在对安全帽检测算法进行性能评估时, 使用平均精度均值 (mAP) 用作衡量模型表现的关键指标。其在对目标检测算法进行评估时, 能够提供一种全面且综合的衡量标准。当置信度阈值为 50% 时所计算出的平均精度均值称为 mAP_{50} ; 当置信度阈值范围在 50% 至 95% 区间后计算出的平均精度均值称为 mAP_{50-95} :

$$P = \frac{TP}{TP + FP} \times 100\% \quad (15)$$

$$R = \frac{TP}{TP + FN} \times 100\% \quad (16)$$

$$AP = \int P(R) dR \quad (17)$$

其中: TP 为检测结果中正确检测目标的数量; FP 为检测结果中错误检测的目标数量; FN 为模型误

检及漏检的数量; P 为精确率, R 为召回率, P 和 R 的关系曲线 $P(R)$ 所围成的的面积就是某一类别的平均精度 AP , 所有类别的 AP 计算得到的平均结果就是平均精度均值 mAP 。

使用模型参数量 (Parameters) 和浮点运算量 (GFLOPS) 衡量模型大小和复杂度。模型参数量指的是神经网络模型中的可学习参数数量; 浮点运算量是神经网络的乘法一加法运算的总次数 (MACs, multiply-accumulate operations), 全连接层和卷积层的 MACs 计算方式分别如下:

$$FLOP_{s_{rc}} = N_{in} \cdot N_{out} \quad (18)$$

$$FLOP_{s_{conv}} = K \cdot H \cdot W \cdot C_{in} \cdot C_{out} \quad (19)$$

其中: B_{in} 是线性层输入神经元数量; N_{out} 是输出神经元数量; K 是卷积核大小; 卷积层输出特征图的高度与宽度可分别用 H 和 W 来表示; 卷积层输入与输出的通道数量分别由 C_{in} 和 C_{out} 来表示。

3.3 结果分析

3.3.1 对照试验

为了验证算法 Helmet-YOLO 在安全帽检测任务的有效性, 在 SHWD 公共数据集上进行了对照试验。

表 3 和图 7 为对照实验结果。可以看出 Helmet-YOLO 算法在 mAP_{50} 和 mAP_{50-95} 上分别取得了 91.5% 和 58.2% 的结果, 与基线算法相比在 mAP_{50} 和 mAP_{50-95} 分别提升了 2.2% 和 1.3%; 检测性能高于 YOLOv3、YOLOv3-tiny、YOLOv5n、YOLOv8n、SSD、Retinanet 等经典算法。与近几年的其他安全帽检测算法, 如 FEV-YOLOv8n、YOLOv8-LE、YOLO-M3 和 Improve-YOLOv5 等相比 Helmet-YOLO 也具有更好的检测效果。虽然检测性能与 RT-DETR-l、RTDETR 和 Gold-YOLO 相比还有些差距; 但 Helmet-YOLO 在浮点运算量、模型参数量和帧率等指标上更加优越, 在检测的实时性上具有明显的优势。

以上结果表明 Helmet-YOLO 算法有效提高了 YOLOv10n 的安全帽检测效果, 并有效地控制了计算开销, 为注重性能与资源消耗平衡的应用场景提供了更灵活和可行的解决方案。

3.3.2 消融实验

为了验证各改进组件对检测性能的提升效果, 以 YOLOv10n 为基线, 在 SHWD 数据集上, 对 SimC2f、LSK 注意力机制、DySample 上采样算子进

表 3 各算法在 SHWD 数据集上检测效果

models	mAP_{50} /%	mAP_{50-95} /%	GFL- OPs	Param- eters /M	FPS /fps
YOLOv10n(baseline)	89.3	56.9	8.6	2.6	353
YOLOv3-tiny	79.3	50.2	19.1	11.6	345
YOLOv5n	89.0	56.7	7.8	2.5	385
YOLOv8n	89.1	56.2	8.9	3.0	363
YOLOv8n-p2	90.1	58.1	17.4	3.2	256
SSD	84.5	57.8	41.0	34.3	—
RT-DETR-l	92.6	59.0	108.0	31.3	—
RetinaNet	80.7	49.8	50.0	45.7	—
EFV-YOLOv8n ^[14]	90.4	57.0	6.6	4.8	337
YOLOv8n-LE ^[13]	88.4	54.8	7.7	2.4	458
YOLO-M3 ^[11]	89.5	—	—	8.4	—
Improve-YOLOv5 ^[12]	90.9	—	—	—	—
Poly-YOLO ^[12]	87.1	—	—	—	—
TPH-YOLOv5 ^[12]	89.6	—	—	—	—
DEYOv2 ^[24]	91.8	58.8	32.9	11.9	208
Gold-YOLO ^[25]	92.1	58.4	23.9	9.8	231
Contextual Trans-YOLOv10n ^[26]	90.9	57.5	9.3	2.8	256
GlobalContext-YOLOv10n ^[27]	90.9	57.4	8.6	2.6	244
NAM-YOLOv10n ^[28]	90.3	56.9	8.6	2.6	300
SpatialGroupEnhance-YOLOv10n ^[29]	90.8	57.2	8.6	2.6	141
TripletAttention-YOLOv10n ^[30]	90.5	57.4	8.7	2.6	206
SKA-YOLOv10n ^[31]	91.1	57.8	15.3	4.8	182
EffectiveSE-YOLOv10n ^[32]	90.8	57.2	8.6	2.7	267
Fasternet-YOLOv10n ^[33]	87.7	53.5	6.2	1.9	769
Helmet-YOLO(ours)	(↑ 2.2)	(↑ 1.3)			

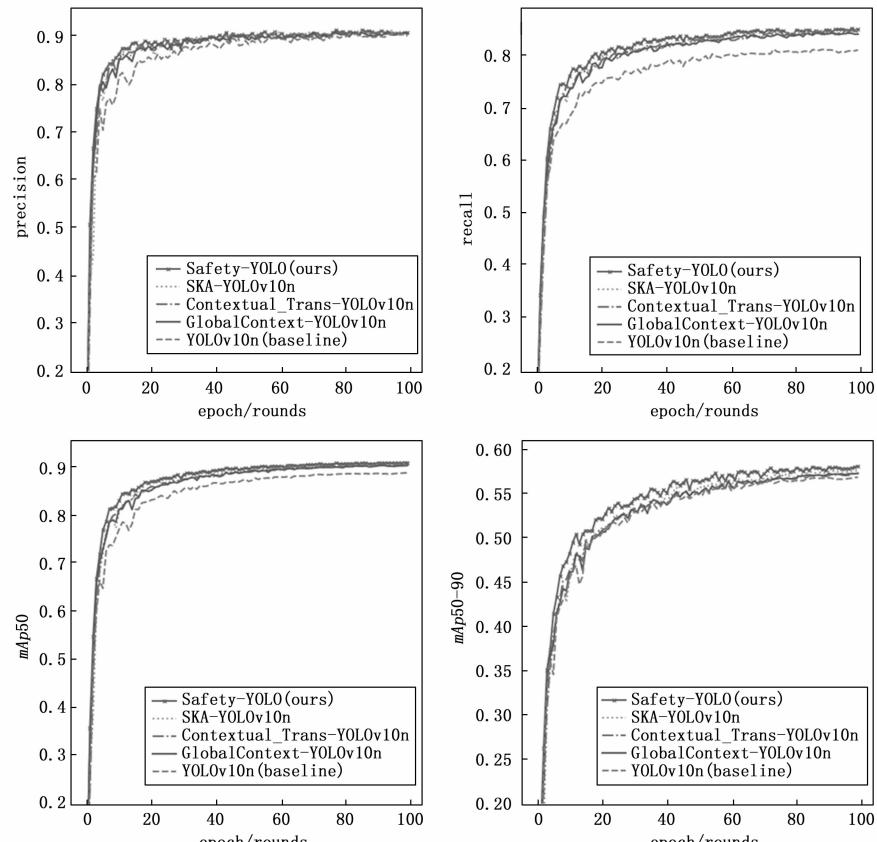


图 7 Helmet-YOLO 与其他改进方法检测效果对比

行消融实验。

如表 4 为消融实验结果, 可以看出 SimC2f 效果明显, 使检测性能 mAP_{50} 和 mAP_{50-95} 提升了 1.9% 和 0.9%, 浮点运算量和模型参数量没有增加, 证明了 SimC2f 模块在没有增加计算成本的情况下有增强了主干网络的提取图像特征的能力; 在颈部引入 LSK 注意力机制后使检测性能上提升了 1.4% 和 0.3%, 证明了 LSK 可以使网络充分利用图像上下文信息, 充分使用图像背景先验知识; 使用 DySample 上采样算子后在 SHWD 上提升了 0.2% 和 0.2%, 浮点运算量和模型参数量并没有明显提升, 证明了 DySample 上采样算子可以通过动态的选择上采样点, 有效提高上采样图质量, 更好的保留了上采样图的细节特征。

表 4 模块消融实验

models	$mAP_{50} / \%$	$mAP_{50-95} / \%$	GFL-OPs	Parameters/M
YOLOv10n(baseline)	89.3	56.9	8.6	2.62
YOLOv10n+SimC2f	91.2(↑ 1.9)	57.8(↑ 0.9)	8.6	2.62
YOLOv10n+LSKneck	90.7(↑ 1.4)	57.2(↑ 0.3)	8.9	2.77
YOLOv10n+DySample	89.5(↑ 0.2)	57.1(↑ 0.2)	8.6	2.63
YOLOv10n+SimC2f +LSKneck	91.3(↑ 2.0)	58.2(↑ 1.3)	8.9	2.77
YOLOv10n+SimC2f +DySample	91.3(↑ 2.0)	57.8(↑ 0.9)	8.6	2.63
YOLOv10n+LSKneck +DySample	90.9(↑ 1.6)	57.5(↑ 0.6)	8.9	2.78
Helmet-YOLO(ours)	91.5(↑ 2.2)	58.2(↑ 1.3)	8.9	2.78

然后将各模块组合使用。Helmet-YOLO 使用 SimC2f+LSK+DySample 使检测性能在 SHWD 数据集上提升了 2.2% 和 1.3%; 当 SimC2f 与 LSK 注意力机制协同使用时, 其检测性能分别提高了 2.0% 和 1.3%, 这一提升幅度略低于 Helmet-YOLO 所取得的成果, 从而有力地证明了 DySample 上采样算子有效性; 当 SimC2f 与 DySample 协同使用时, 其检测性能分别提高了 2.0% 和 0.9%, 这一提升幅度略低于 Helmet-YOLO 所取得的成果, 从而有力地证明了 LSK 注意力机制的有效性; 当 LSK 注意力机制与 DySample 协同使用时, 其检测性能分别提高了 1.6% 和 0.6%, 这一提升幅度略低于 Helmet-YOLO 所取得的成果, 从而有力地证明了 SimC2f 模块的有效性。

3.4 检测效果展示

如图 8 所示展示了 YOLOv10n 和 Helmet-YOLO 算法的检测效果对比。其中第一行为原始图片, 第二行为经典 YOLOv10n 的检测效果, 第三行为 Helmet-YOLO 的检测效果。由第一列、第三列和第五列可以看出 Helmet-YOLO 对工人画面不全、目标较小的情况下个体漏检的情况改善了很多, 检测性能明显优于原始

YOLOv10n。由第二列和第四列可以看到原始 YOLOv10n 错误的将图像背景部分的其他物品检测成了未佩戴头盔的人, 出现了误检的情况, 而 Helmet-YOLO 并没有出现这种状况, 验证了 Helmet-YOLO 算法在特征提取能力和检测性能上的有效性。



(a) 原始图片 (b) YOLOv10n检测效果 (c) Safety-YOLO检测效果

图 8 Helmet-YOLO 检测效果展示

4 结束语

Helmet-YOLO 通过对 YOLOv10n 的主干网络, 颈部网络和颈部上采样算子进行改进, 提高了网络在复杂场景下的特征提取能力, 解决了安全帽佩戴检测的误检漏检问题。实验结果表明, 该算法不仅提升了检测效果, 同时也维持了较高的处理速度, 这对于推动智能安防系统的创新应用以及终端设备的智能化升级具有积极推进作用。

然而, 该算法在一些极端光照情况下, 比如背光、摄像头前方光源太强、光照条件太暗情况下的检测效果并不理想。因此接下来将对数据集进行扩充, 构建极端光照条件下的安全帽数据集以提高算法的泛化性; 并考虑配合红外摄像头和红外数据集一起进行检测, 进一步提高安全帽佩戴检测的效果。

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 (6): 65–69.
- [2] 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.
- [3] CHIVERTON J. Helmet presence classification with motorcycle detection and tracking [J]. IET Intelligent Transport Systems, 2012, 6 (3): 259–269.
- [4] MNEYMNEH B E, ABBAS M, KHOURY H. Automated hardhat detection for construction safety applications [J]. Procedia Engineering, 2017, 196: 895–902.
- [5] SHAOQING R, KAIMING H, ROSS G, 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).
- [6] REN S, HE K, GIRSHICK R, et al. Faster R-CNN: to-

- wards real-time object detection with region proposal networks [J]. Advances in Neural Information Processing Systems, 2015, 39 (6): 1137–1149.
- [7] GIRSHICK R. Fast R-CNN [J]. Computer Science, 2015: 1440–1448.
- [8] 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.
- [9] 徐守坤,王雅如,顾玉宛,等. 基于改进Faster RCNN的安全帽佩戴检测研究 [J]. 计算机应用研究, 2020, 37 (3): 901–905.
- [10] 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 Recognition, 2016: 779–788.
- [11] 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.
- [12] 韩锟栋,张涛,彭玻,等. 基于改进YOLOv5的安全帽检测算法 [J]. 现代电子技术, 2024, 47 (5): 85–92.
- [13] 冯勇,杨思卓,徐红艳. 基于YOLO v8的轻量化安全帽佩戴检测算法 [J/OL]. 计算机应用, 1–9. <http://kns.cnki.net/kcms/detail/51.1307.TP.20240418.1338.002.html>.
- [14] 韩博,张婧婧,鲁子翱. FEV-YOLOv8n: 轻量化安全帽佩戴检测方法 [J/OL]. 计算机测量与控制, 1–11. <http://kns.cnki.net/kcms/detail/11.4762.TP.20240506.1527.010.html>.
- [15] WANG A, CHEN H, LIU L, et al. Yolov10: real-time end-to-end object detection [J]. ArXiv Preprint ArXiv: 2405.14458, 2024.
- [16] ZHANG H, ZU K, LU J, et al. EPSANet: an efficient pyramid squeeze attention block on convolutional neural network [C] // Proceedings of the Asian Conference on Computer Vision, 2022: 1161–1177.
- [17] YANG L, ZHANG R Y, LI L, et al. Simam: a simple, parameter-free attention module for convolutional neural networks [C] // International Conference on Machine Learning, PMLR, 2021: 11863–11874.
- [18] WOO S, PARK J, LEE J Y, et al. Cbam: convolutional block attention module [C] // Proceedings of the European Conference on Computer Vision (ECCV), 2018: 3–19.
- [19] LI Y, HOU Q, ZHENG Z, et al. Large selective kernel network for remote sensing object detection [C] // Proceedings of the IEEE/CVF International Conference on Computer Vision, 2023: 16794–16805.
- [20] WANG J, CHEN K, XU R, et al. Carafe: content-aware reassembly of features [C] // Proceedings of the IEEE/CVF International Conference on Computer Vision, 2019: 3007–3016.
- [21] LU H, LIU W, FU H, et al. FADE: fusing the assets of decoder and encoder for task-agnostic upsampling [C] // European Conference on Computer Vision, Cham: Springer Nature Switzerland, 2022: 231–247.
- [22] LU H, LIU W, YE Z, et al. SAPA: similarity-aware point affiliation for feature upsampling [J]. Advances in Neural Information Processing Systems, 2022, 35: 20889–20901.
- [23] LIU W Z, HAO L, FU H T, et al. Learning to upsample by learning to sample [J]. 2023 IEEE/CVF International Conference on Computer Vision (ICCV), 2023: 6004–6014.
- [24] OUYANG H. DEYOV2: rank feature with greedy matching for end-to-end object detection [J]. ArXiv Preprint ArXiv: 2306.09165, 2023.
- [25] WANG C, HE W, NIE Y, et al. Gold-YOLO: efficient object detector via gather-and-distribute mechanism [J]. Advances in Neural Information Processing Systems, 2024, 36.
- [26] LI Y, YAO T, PAN Y, et al. Contextual transformer networks for visual recognition [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2022, 45 (2): 1489–1500.
- [27] 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: 0–0.
- [28] LIU Y, SHAO Z, TENG Y, et al. NAM: normalization-based attention module [J]. ArXiv Preprint ArXiv: 2111.12419, 2021.
- [29] LI X, HU X, YANG J. Spatial group-wise enhance: Improving semantic feature learning in convolutional networks [J]. ArXiv Preprint ArXiv: 1905.09646, 2019.
- [30] MISRA D, NALAMADA T, ARASANIPALAI A U, et al. Rotate to attend: convolutional triplet attention module [C] // Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision, 2021: 3139–3148.
- [31] LI X, WANG W, HU X, et al. Selective kernel networks [C] // Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2019: 510–519.
- [32] LI X, HU X, YANG J. Spatial group-wise enhance: Improving semantic feature learning in convolutional networks [J]. ArXiv Preprint ArXiv: 1905.09646, 2019.
- [33] CHEN J, KAO S, HE H, et al. Run, don't walk: chasing higher FLOPS for faster neural networks [C] // Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2023: 12021–12031.