

基于改进 YOLOv8 的化工厂安全装置检测算法

曹鑫泉, 李海涛, 张俊虎

(青岛科技大学 信息科学技术学院, 山东 青岛 266061)

摘要: 针对工厂工业安全检测场景安全装置目标检测存在的漏检与误检问题, 提出了 SAG-YOLOv8 改进架构; 通过移位卷积替换原 C2f 模块中的常规卷积构建 C2f-SWC 新模块以增强多尺度特征表达能力; 采用 AIFI 模块替代传统空间池化金字塔结构来强化图像语义理解能力; 引入 GFPN 网络架构通过增强跨层多尺度交互与同层横向连接促进小目标特征传播; 实验数据显示 SAG-YOLOv8 算法 $mAP@0.5$ 指标较原始 YOLOv8 提升了 3.4%, 精确度和召回率也有一定提升; 该方法显著提高了化工厂安全装置中目标检测的精准度和稳定性, 为其安全运行提供了有力的技术保障。

关键词: 工业安全检测; YOLOv8; Shift-Wise Conv; AIFI; GFPN

Chemical Plant Safety Equipment Detection Algorithm Based on Improved YOLOv8

CAO Xinquan, LI Haitao, ZHANG Junhu

(College of Information Science and Technology, Qingdao University of Science and Technology,
Qingdao 266061, China)

Abstract: To address the issues of missed detection and false detection on industrial safety device detection scenarios in a factory, an improved architecture, SAG-YOLOv8, is proposed. The conventional convolution in the original C2f module is replaced with shifted convolution, a new C2f-SWC module is constructed to enhance the ability of multi-scale feature representation. Additionally, the AIFI module replaces the traditional spatial pyramid pooling structure to strengthen the ability of image semantic understanding. The GFPN network architecture is introduced to improve small target feature propagation by enhancing cross-layer multi-scale interactions and same-layer lateral connections. Experimental results show that compared to the original YOLOv8, the SAG-YOLOv8 algorithm improves the $mAP@0.5$ by 3.4%, with noticeable enhancements in precision and recall. This method significantly enhances the accuracy and stability of safety device target detection in chemical plants, providing a strong technical support for safe operations.

Keywords: industrial safety inspection; YOLOv8; shift-wise Conv; AIFI; GFPN

0 引言

化工厂的装卸区存在易燃易爆的气体和液体泄漏风险, 为了确保设备和作业人员的安全, 必须采取一系列的安全措施。要求操作人员在设备运行过程中将控制装卸车钥匙挂置在固定区域, 还需使用安全绳、佩戴安全帽, 确保鹤管和四步梯等安全设施的正确使用。在实际作业中, 可能由于操作人员疏忽、管理不到位或设备维

护不及时, 导致安全装置未按规定使用或放置。这些问题可能带来设备操作不当或作业人员安全隐患, 进而影响整体安全性。

由于人工监管难以做到全程监控, 存在管理漏洞。对此, 结合深度学习目标检测算法和现场防爆摄像头, 能够实现对化工厂作业过程的实时监控, 可以自动检测设备操作状态及作业人员的安全防护情况, 包括安全帽是否佩戴、鹤管是否正确连接、车辆钥匙是否正确挂置

收稿日期: 2025-02-14; 修回日期: 2025-03-21。

基金项目: 山东省重点研发计划(科技示范工程)课题(2021SAGC0701); 青岛市海洋科技创新专项(22-3-3-hygg-3-hy)。

作者简介: 曹鑫泉(1999-), 男, 硕士。

张俊虎(1974-), 男, 博士, 副教授, 硕士生导师。

通讯作者: 李海涛(1978-), 男, 博士, 副教授, 硕士生导师。

引用格式: 曹鑫泉, 李海涛, 张俊虎. 基于改进 YOLOv8 的化工厂安全装置检测算法[J]. 计算机测量与控制, 2026, 34(2): 57-64.

等, 最大程度减少人为疏忽, 提高工厂作业的安全性和效率。

随着深度学习的不断发展, 深度卷积神经网络在目标检测领域中的应用愈加广泛^[1], 现已被应用于安全检测领域^[2]。深度学习中的目标检测模型主要分为两种: 一种是双阶段算法, 如 Regin-CNN (R-CNN)^[3]和 Fast-erR-CNN^[4]。双阶段算法通过先生成候选框再分类回归的方式, 具有检测精度高、错误率低、速度较快且适应性强等方面优势, 是目前目标检测领域的主流算法之一。另一种是单阶段算法, 如 YOLO 系列^[5-9]和 SSD^[10]等。单阶段目标检测方法直接预测目标框和类别, 通过回归分析将区域提取与分类回归融合, 达到端到端的检测效果。该方法只需一次前向传播即可完成目标检测, 且在小物体检测方面表现尤为突出。

目前, 很多研究者将基于深度学习的目标检测算法应用到各种实时检测领域, 并取得了显著的效果。文献 [11] 通过在 YOLOv8 主干网络中引入 DSCConv 和 EMA 注意力机制, 并优化损失函数, 使模型更加聚焦于与不规范行为相关的身体部位或动作, 提高了工人不规范行为检测的精度和实时性。文献 [12] 针对安全帽佩戴检测在复杂场景下的漏检、误检问题, 基于 YOLOv5s 提出一种改进算法。通过采用串联形式的混合池化优化 SPP 模块、增加浅层输出以及在切片模块中嵌入坐标注意力机制 (CA) 等方式, 增强了特征提取与表达, 提高了小目标安全帽佩戴检测性能和准确率, 降低了误检率。文献 [13] 通过将 C3 模块融合 ECA 注意力来改进 YOLOv5 算法, 有效提升了对安全帽特征信息的提取能力, 使用 EIOU 损失函数来提升训练效果, 最后在降低模型大小的同时提高了准确率。文献 [14] 通过采用多空间金字塔池化和压缩激励通道注意力机制, 提升了特征融合网络的能力, 并引入深度可分离卷积减少了网络参数量, 从而提高了目标检测的准确性与实时性。此外, 文献 [15] 针对施工现场安全帽检测中遮挡目标检测难度大、误检漏检率高的问题, 使用 K-means++ 聚类算法重新设计匹配安全帽数据集的先验锚框尺寸并采用 Swin Transformer 作为骨干网络提取特征, 以更好地适应不同尺度的目标检测任务。文献 [16] 提出了基于 YOLOv8 的安全帽佩戴检测模型 CCG-YOLOv8。采用轻量级上采样 CARAFE 替换最近邻插值上采样, 在保持轻量化的同时提高细节信息利用率并使用 GIOU 优化边界框回归损失函数, 准确衡量预测框和真实框之间的重叠情况并加快模型收敛速度。

当前, 以 YOLOv8 为代表的目标检测算法在化工厂安全装置检测中暴露出显著缺陷。在光线昏暗的装卸区, 易漏检未佩戴安全帽的作业人员, 事故发生时头部重伤风险极大; 在管道设备交错区域, 常将形似鹤管的

管道误检为已正确连接, 一旦鹤管实际未连接到位, 易燃易爆物料泄漏风险剧增, 遇明火或致爆炸, 严重威胁化工厂安全。先前研究通过注意力机制嵌入、网络轻量化设计等手段提升目标检测性能, 但特征融合机制仍受限于固定卷积核的静态感受野、池化操作的细节丢失及传统特征传递架构, 难以充分发挥多尺度特征信息的协同效应。为解决复杂背景下安全装置的漏检、误检问题, 本文提出引入 C2f-SWC 模块、AIFI 模块和 GFPN 网络架构, 构建 SAG-YOLOV8 (Shift-Wise Adaptive Generalized-YOLOV8) 架构, 以此增强多尺度特征的表达和语义理解能力, 有效提升特征融合与理解水平。

1 YOLOv8 算法原理

YOLOv8 作为 YOLO 系列算法的杰出代表, 在以往版本的基础上进行了创新和改良, 从而显著提升了模型的性能。该模型的结构原理由特征提取模块和目标检测模块组成。整体网络结构是由输入层、主干网络、颈部和头部四部分构成。主干网络结合了 Conv、SPPF 和 C2f 模块, 其中 C2f 模块借鉴了 C3 和 ELAN 模块的设计思路, 增加了更多残差连接, 优化了梯度流。头部部分采用了解耦头结构, 将分类任务与检测任务分开, 并使用无锚框架, 对目标的中心点及其宽高比例进行直接预测。

YOLOv8 在目标检测架构层面实现了多维度创新: 主干网络设计融合 YOLOv7 的 ELAN 分层特征融合思想, 通过 C2f 模块替代传统 C3 结构以增强梯度传播效率, 并采用差异化通道配置策略提升特征表达能力; 损失函数体系创新性整合动态正样本分配机制 (Task Aligned Assigner) 与分布聚焦损失 (Distribution Focal Loss)^[17], 通过自适应样本匹配策略和置信度校准方法提升分类性能; 训练策略方面实施两阶段数据增强方案, 前期采用 Mosaic 多图拼接增强提升数据多样性, 后期冻结增强操作以稳定特征学习。与之前的 YOLO 版本相比, YOLOv8 在检测精度和速度上都有显著提升, 是一种全新的 SOTA 模型。

2 YOLOv8 改进算法

2.1 C2f-SWC 模块提高多尺度特征融合

基于大卷积核的持续演变, 通过小卷积核与位移操作的结合来模拟大卷积核的效果, 从而在保持计算效率的同时捕获更广阔的感受野。该方法的核心在于将传统的大卷积核分解为多个小卷积核, 并通过位移操作来实现这些小卷积核输出的特征图之间的错位对齐, 以此达到大卷积核所具有的覆盖范围。改进的模块结构如图 1 所示。

大卷积核可以通过以下变换分解 (分解结构大的卷积核), 表示为公式 (1)。为了通用性, 假设 $kw > kh$ 。

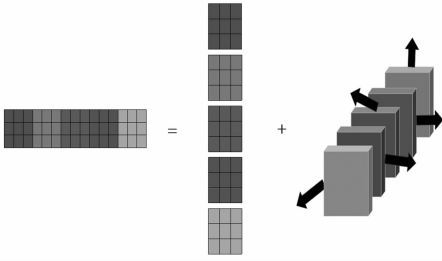


图 1 小卷积核位移操作模块结构

其中 (i, j) 表示图像或特征图上滑动窗口的位置。 k_w 和 k_h 分别表示卷积核的宽度和高度。 p 表示位置信息。 w 和 x 表示对应位置的权重和特征:

$$y(p_{(i,j)}) = \sum_{k=0}^{\lfloor \frac{k_w}{k_h} \rfloor} \sum_{m=0}^{k_w} \sum_{n=0}^{k_h} w[p_{(\Delta m, \Delta n)} + \Delta p] \cdot x[p_{(i,j)} + p_{(\Delta m, \Delta n)} + \Delta p]$$

$$\Delta m = m - \frac{k_w}{2}; \Delta n = n - \frac{k_h}{2}$$

$$\Delta p = kh \times k, \left\{ k \in \left[0, \left\lfloor \frac{k_w}{k_h} \right\rfloor \right] \right\} \quad (1)$$

式中, $\lfloor \cdot \rfloor$ 是向上取整。从公式很容易看出分解后的卷积核, 即 $k_w > k_h$ 的情况。可以用多个正规小卷积核来替换。这需要将对超参数进行对齐, 如偏移控制和填充设置。比如说一个 15×3 的卷积可以等价于 $5 \times 3 \times 3$ 的卷积。对于这种情况, 需要使用移位操作。与之前的情况不同, 移位需要在—维上移动得更远, 并使它与大小的网格对齐。

Shift-Wise 方法通过剪枝和组移位操作优化网络结构^[18]。在训练中, 剪枝去除不重要连接, 获得稀疏组卷积, 随后应用组移位模拟多卷积核大小, 创建多个输出分支。从每组中选择一个通道形成 Identity 分支, 最后通过 add 操作将所有分支合并, 保持总通道数不变。该方法从数据流形角度优化依赖关系, 提高计算效率, 同时维持网络结构稳定。

移位运算符是一种更通用的形式, 可以进一步放宽限制。这包括卷积核的大小、形状和信息融合的方法。由于移位和稀疏性, 引入了关注长度和关注宽度, 来对应于 SLaK 中 Kernel 大小的 m 和 n , 也就是卷积核的长度和宽度。在这里, 定义移位运算符为一种方法, 首先指定一个具有关注长度和关注宽度的矩形区域, 然后选择大小小于或等于关注宽度的卷积核, 并调整这些卷积核的分布进行信息融合。关注长度不需要小于当前特征图大小。对于这种情况, 稍微调整公式 (1) 得到公式 (2):

$$y(p_{(i,j)}) = \prod_{k=0}^{f(k_w, k_h, A)} \sum_{m=0}^A \sum_{n=0}^A w[p_{(\Delta m, \Delta n)} + \Delta p] \cdot x[p_{(i,j)} + p_{(\Delta m, \Delta n)} + \Delta p]$$

$$\Delta m = m - \frac{A}{2}; \Delta n = n - \frac{A}{2}$$

$$\Delta p = g(kh, k), \{k \in [0, f(k_w, kh, A)]\} \quad (2)$$

式中, A 为小卷积核的大小。 k_w 和 k_h 分别为关注长度和关注宽度。 $f(k_w, kh, A)$ 代表与 (k_w, kh, A) 相关的函数。相应地, 权重和特征偏移量 Δp 是与 (kh, k) 相关的函数。将该函数记为 g 。

Shift-Wise Conv 在性能上的提升主要得益于其位移操作和稀疏依赖性的引入。具体来说, 我们将大卷积核分解为多个小卷积核, 并通过固定移位量 5 将小卷积核的输出对齐, 覆盖大卷积核的原始感受野。在浅层网络中, 较大的卷积核能够覆盖更广的感受野, 捕捉全局上下文信息 (如大目标的轮廓), 而在深层网络中, 较小的卷积核则更关注局部细节 (如小目标的边界)。这种层级化的设计使模型能够在浅层网络中捕捉全局上下文, 在深层网络中聚焦局部细节, 从而提升全尺度目标检测的能力。

本研究通过引入 Shift-Wise Conv 对 YOLOv8 中的 C2f 结构进行改进。具体做法是用 Shift-Wise Conv 替换 C2f 中 Bottleneck 模块中的 Conv 层, 从而构建出新的 C2f-SWC 结构。并根据实验来对比不同层次网络中大卷积核大小作对比。改进后的结构不仅减少了卷积的计算量, 而且通过位移操作实现了大卷积核的覆盖效果, 提升了模型的学习能力和表达能力。同时, 引入的稀疏性机制增强了特征的表达能力, 使得模型能够更准确地感知和融合多尺度的重要特征, 从而进一步提升了整体性能。

2.2 AIFI 提升图像内容的理解与识别能力

在工厂的工业安全检测中, 准确识别安全装置中的各类目标物体是核心任务, 其中小目标检测和多目标重叠问题尤为常见。然而, 现有的目标检测算法如 YOLOv8 在处理小目标物体时, 常常面临漏检和错检的问题。这主要是由于传统方法在多尺度特征融合和特征表达能力方面存在不足, 无法有效捕捉和处理小目标物体的细节特征。为了进一步提高多尺度特征理解和特征表达能力, 将 AIFI^[19] 引入替换原有 YOLOv8 的 SPPF 模块。

AIFI 模块是通过引入基于多头自注意 (MHSA) 机制的高级特征处理模块, 当 MHSA 处理图像的特定区域时, 动态地关注相关区域的信息。这种方法能有效捕获丰富的语义信息, 提升高级特征的表达能力, 并增强模型对复杂图像内容的理解和识别能力。该模块的核心计算单元采用多头自注意力机制, 计算多个注意力头之间的查询 (Q)、键 (K) 和值 (V) 之间的相似性, 然后对这些值进行加权求和生成最终的特征图 F 。其计算过程可表述为如公式 (3) 和 (4) 所示:

$$Att(Q, K, V) = Softmax\left(\frac{QK^T}{\sqrt{d}}\right)V \quad (3)$$

$$F = Reshape[Att(Q, K, V)] \quad (4)$$

式中，对于给定输入特征序列 $X \in R^{n \times d}$ ，其中 n 表示 token 数量， d 表示每个 token 的嵌入维数。输入 X 通过 3 个可学习的权重矩阵 W^Q 、 W^K 和 W^V 映射到查询 (Q)、键 (K) 和值 (V) 空间，每个权重矩阵的维度为 $R^{d \times d}$ 。然后，通过点积计算 Q 和 K 中每对 token 之间的注意力得分，接着通过 Softmax 函数进行归一化和加权，以量化 token 之间的交互作用。加权后的交互结果再与 V 相乘，生成最终的加权输出。每个自注意力层包含 h 个注意力头，使模型能够关注输入的不同表示方面。该机制允许模型自适应地构建全局像素级关联，突破传统卷积局限于局部感受野的限制。

注意力头的数量 h 是一个关键的超参数，它决定了模型能够同时关注输入特征的不同表示方面的能力。理论上，增加注意力头的数量可以增强模型捕捉多样化特征的能力，但过多的头数量可能导致计算复杂度增加和过拟合风险。因此，选择合适的注意力头数量需要在模型表达能力和计算效率之间进行权衡。根据文献 [20] 的研究，多头注意力机制中头数量通常设置为 8 或 16，这在不同任务中表现出较好的性能。此外，文献 [21] 的研究表明，头数量的选择应与输入特征的维度 (d) 相匹配，通常满足 d 除以 h 的商等于 0，以确保每个头能够均匀地分配特征维度。基于这些理论依据，我们在 AIFI 模块中选择了 $h=8$ ，以平衡模型的表达能力和计算效率。

为增强特征非线性表达能力，AIFI 模块在 MHSA 层后加入前馈神经网络 FFN (Feed-Forward Network)，整个处理流程遵循“二维特征展平、MHSA、FFN、维度还原”的架构，特别针对经过骨干网络输出的高层特征进行优化。AIFI 的结构如图 2 所示。

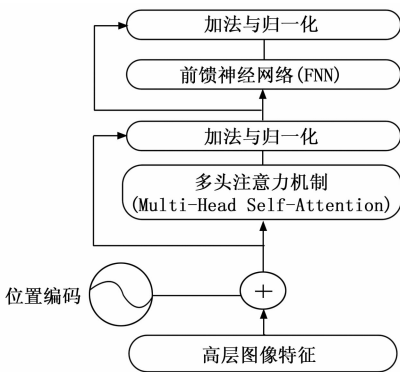


图 2 AIFI 结构图

相较于被替换的 SPPF 模块，AIFI 在特征交互机制上实现了本质改进。SPPF 模块依赖多尺度池化操作获取固定模式的上下文信息，而 AIFI 通过自注意力动态建模全局依赖关系，显著增强了对长距离语义关联的捕获能力。并且在特征表达维度上，MHSA 机制可同

步学习位置敏感特征和语义关联特征，有效缓解了 SP-PPF 在多目标重叠场景下的特征混淆问题，有效减少工业安全检测中的多目标重叠和小目标检测困难问题。

2.3 GFPN 提高特征信息的捕捉和融合能力

在特征金字塔网络 (FPN) 中，多尺度特征融合通过整合不同分辨率的特征来提高检测精度。YOLOv8 的 Neck 部分在 FPN 基础上增加了自底向上的路径，从而进一步增强了特征融合的效果，如图 3 (a) 所示。尽管 FPN 和 PAN 在多尺度目标检测中表现优异，但在工业安全检测中，由于复杂背景、目标密集性高以及小物体检测的不足，导致它们的性能在这些特殊环境下受到了一定的限制，无法有效应对复杂的场景，特别是在低分辨率或目标被遮挡的情况下。

为了克服这些问题，引入了广义特征金字塔网络 (GFPN, generalized feature pyramid network)，并增加了跳跃连接和跨尺度连接机制^[22]。GFPN 通过 $\log_2 n$ 跳跃连接有效地连接同层特征图，而跨尺度连接则优化了不同层次之间特征图的联系，从而提高了多尺度特征融合的效果。GFPN 的结构如图 3 (b) 所示，能够在复杂场景中更好地捕捉细节信息，提升了目标检测的准确性和稳定性。

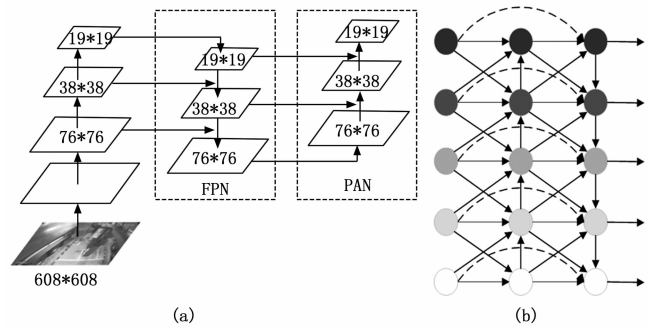


图 3 特征融合图

在 $\text{Log}_2 n$ -link 中，最多会进行 $\log_2 l + 1$ 层的级联。它的机制就是通过层级连接，将前几层的特征图逐级传递和融合。其中在级别 k 中，第 l^k 层会接收最多来自 $\log_2 l + 1$ 个先前层的特征图。这意味着连接的先前层数是与当前层编号 l 的对数成正比的。这些输入层的结构可以通过底数为 2、深度为 i 的指数展开来表示。输出特征图的计算如公式 (5) 所示：

$$P_k^l = Conv[Cat(P_k^{l-2^i}, \dots, P_k^{l-2^1}, P_k^{l-2^0})] \quad (5)$$

为了更有效地捕捉不同尺度下的工业安全装置和车辆特征，GFPN 通过为每个特征图分配不同的通道数，优化了多尺度特征的处理方式，提升了对不同尺寸目标的感知能力，尤其是小目标的识别。然而，直接替换 YOLOv8 的 Neck 部分后，基于卷积的跨尺度特征融合方法仍然存在性能瓶颈，无法充分提取各尺度间的细节信息。为了优化这一问题，本研究提出对 GFPN 的连接机

制进行改进, 引入 CSPStage 模块替代传统的卷积融合方法。这一改进不仅能够更加精确地融合不同尺度的特征, 提升小目标检测的能力, 还能有效应对复杂场景中的各种变化, 从而大幅提高网络在多样化环境中的准确性和稳定性。CSPStage 模块通过替换 Neck 中的 C3 模块, 结合 RepConv 的 BasicBlock_3×3_Reverse 网络, 能够更高效地融合来自不同层级和尺度的特征信息, 进一步提升检测精度、加快推理速度, 并增强网络整体性能。其中的 BasicBlock_3×3_Reverse 网络由 RepConv 类^[23]构成, RepConv 是 RepVGG 网络中的关键组成部分, 包含标准卷积层 (Conv)、批量归一化 (BN) 和修正线性单元 (ReLU)。CSPStage 模块如图 4 所示。

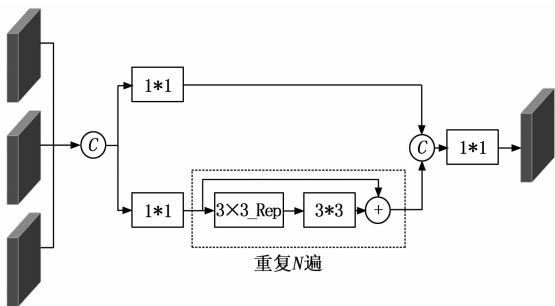


图 4 CSPStage 模块

拼接后的特征图分别进入两路分支处理。一路分支中的特征图先经过 1×1 卷积, 调整通道数以适配后续操作。接着进入包含重复 Rep 模块和 3×3 卷积的流程, 对特征进行深度提取与变换, 随后将该路径的处理结果相加, 进一步融合特征。另一路分支同样先让特征图经过 1×1 卷积。

最后, 这两路处理后的特征图进行拼接, 并再次通过 1×1 卷积, 输出最终的特征图。这种结构设计能够充分融合多源特征, 有效增强特征表达, 进而提升模型在相关任务中的性能表现。

通过 3 个部分的模块改进, 得到最终的 SAG-YOLOv8 模型架构。SAG-YOLOv8 网络结构如图 5 所示。

3 实验结果与分析

3.1 实验数据集

本文实验使用的数据集来源于化工厂装卸车作业区的防爆摄像头的实时监控录像, 包含不同车间下, 白天自然光和夜间照明灯光的实际作业场景, 将拍摄下的数据进行手动筛选后标注, 最终得到的图片数据共计 6 000 张, 分别为钥匙 (Key)、安全帽 (Helmet)、工人 (Worker)、鹤管 (Cranetube)、车牌 (Plate)、制动块

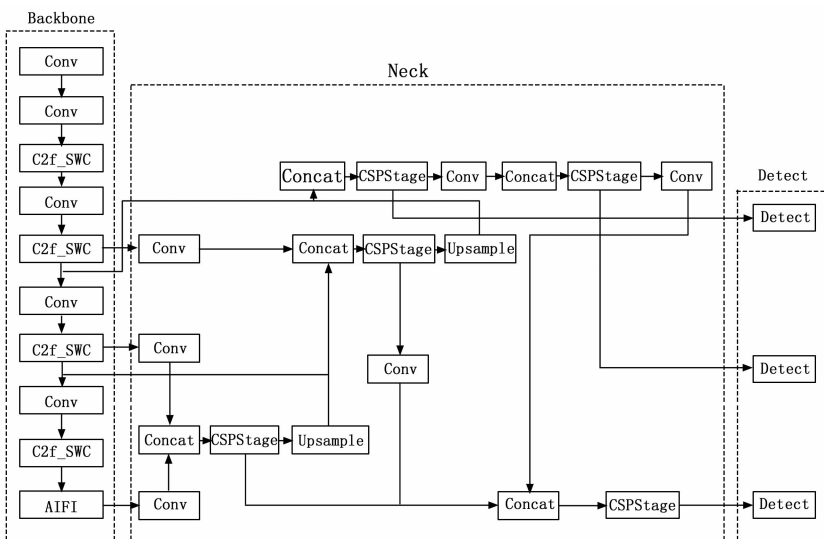


图 5 SAG-YOLOv8 网络结构

(Stopblock)、轮胎 (Tire)、车辆 (Car) 和梯子 (Ladder) 等 16 种类别。并按照 8 : 1 : 1 的比例分成训练集、验证集、测试集 3 部分。实验数据集如图 6 所示。



图 6 实验数据集

3.2 实验配置

本文实验是基于 PyTorch 和 CUDA 框架在 Linux 操作系统下进行的, 具体的软硬件配置和环境配置如表 1 所示。

表 1 实验配置表

参数	版本
操作系统	Ubuntu 18.04
CPU	Intel(R) Core(TM) i7-12700KF CPU @3.60 GHz
GPU	NVIDIA GeForce RTX 3080 Ti
Python	3.8.1
PyTorch	PyTorch 1.10.1+cu113
CUDA	11.3
CuDNN	8.2

3.3 评估指标

本文实验采用准确率 (P, precision)、召回率 (R, recall)、计算量 (FLOPs) 和平均精度均值 (mAP, mean average precision) 作为评估指标。其中准确率和

召回率的公式分别为下, P 表示每类识别准确率, TP 表示将每类识别正确的样本数量, FP 表示将不是某类的目标识别成另外类别的数目, FN 表示将某类识别成另外类别的样本数目, R 表示每类目标检测的召回率:

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

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

平均准确度以及 AP 表示精度的平均值 map 如下:

$$X_{AP} = \int_0^1 P(R) \quad (8)$$

$$m_{AP} = \frac{\sum_{i=1}^K X_{AP_i}}{K} \quad (9)$$

X_{AP_i} 表示对第 i 类物体的检测平均准确率, K 为类别, 在本文实验中以 IoU 阈值为 0.5 的情况作为评价指标 $X_{AP@0.5i}$ 代表第 i 类物体在 IoU 阈值为 0.5 时的平均准确度:

$$m_{AP@0.5} = \frac{\sum_{i=1}^K X_{AP@0.5i}}{K} \quad (10)$$

3.4 模型训练

在模型训练中, 设置的迭代次数为 300 轮, 实验结果表明模型在 300 轮时已收敛。每次训练使用 32 个样本, 权重衰减系数为 0.000 5, 动量因子为 0.937, 初始学习率为 0.01。置信度阈值设置为 0.5, 只有当置信度 ≥ 0.5 时, 预测结果才被视为有效。使用 ADAM 优化器, 它具有自适应学习率、快速收敛、对超参数相对不敏感且泛化性好的特点, 在很多情况下能取得较好性能, 泛化性较好。

3.5 实验结果

本文改进后的算法与原始 YOLOv8 算法在相同训练策略下进行对比, 试验结果如图 7 所示, 横坐标为训练轮数, 纵坐标为百分比指标。从对比曲线可以看出, 改进后的模型在各项性能上均优于 YOLOv8 算法。与原 YOLOv8 相比, 本模型的精度和平均精度均值分别提高了 5.1% 和 3.4%。

3.6 改进点相关参数的实验分析

为了探寻提升模型检测能力的有效方法, 实验旨在剖析不同大卷积核在网络不同位置的表现差异, 通过固定移位量 5 对齐小卷积核输出, 使其覆盖大卷积核原始感受野, 期望借助这种层级化设计, 实现模型在浅层网络捕捉全局上下文, 深层网络聚焦局部细节的效果。采用 $X_1-X_2-X_3-X_4$ 的命名方式, 用以表示 C2f-SWC 在 4 个改进位置上顺序对应的不同大卷积核大小。其中, X_1 对应第 1 个改进位置所采用的大卷积核大小, X_2 对应第 2 个改进位置, 以此类推。实验结果汇总如表 2 所示。

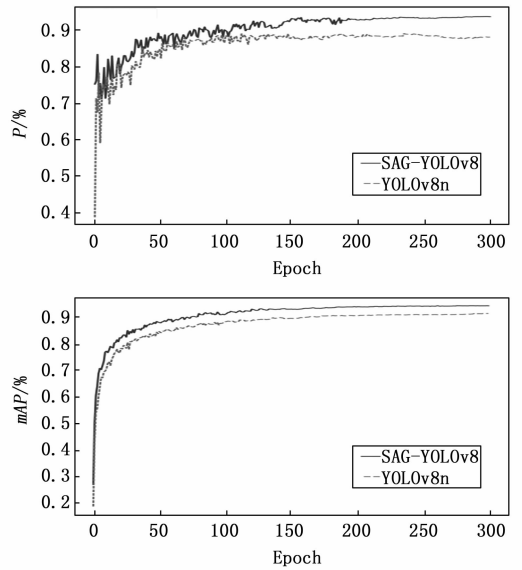


图 7 实验结果对比曲线图

表 2 C2f-SWC 对比实验结果

模型参数	P/%	R/%	mAP@0.5/%	FLOPs
15-13-11-9	91.3	84.2	90.3	11.4
13-11-9-7	92.1	85.4	91.7	11.2
11-9-7-5	93.0	86.9	93.5	10.8
9-7-5-5	91.7	84.7	91.2	10.4

在对比实验中, “11-9-7-5” 这一实验设置展现出显著优势。从实验结果来看, 其精确率、召回率和平均精度均值均高于其他几种实验设置, 表明该设置在目标检测的精准度和召回效果上表现更优, 能更有效地检测出目标。同时, 其每秒浮点运算次数为 10.8, 相对其他设置更低, 意味着在计算资源消耗上更具优势, 计算效率更高, 兼顾了检测性能与计算成本, 因此选择 “11-9-7-5” 这一参数设置, 在提升模型检测能力方面具有较好的综合表现。

同时, 又对 GFPN 中不同通道数设置在网络不同位置的效果差异进行实验对比, 试图找寻能优化模型性能的通道数分配方案。同样采用 $X_1-X_2-X_3-X_4-X_5$ 的命名方式, 对顺序位置的 CSPStage 模块使用的通道数设定。实验过程中, 设置了多组不同的通道数组合方案。在特征融合环节, 针对不同通道数的特征图, 通过恰当的卷积、上采样或下采样操作, 实现不同尺度特征图在空间维度上的精准对齐, 以保障特征融合的有效性。实验结果如表 3 所示。

在实验中, 512-512-256-512-1024 的通道数组合通过层级化设计显著提升了模型的性能与效率。浅层 (X_1-X_2) 采用 512 通道, 保留高分辨率特征图的细节信息, 增强小目标检测能力。中层 (X_3) 缩减至 256 通道, 基于特征冗余理论减少计算量, 同时通过跨层连

表 3 CSPStage 对比实验结果

模型参数	P/%	R/%	mAP@0.5/%	FLOPs
256-512-512-512-1024	93.3	84.2	91.7	13.7
256-256-512-512-1024	92.5	85.1	90.1	13.2
512-1024-256-512-1024	91.9	83.5	90.5	9.7
512-512-256-1024-1024	93.4	84.2	91.2	10.4
512-512-256-256-1024	92.6	84.0	90.3	7.9
512-512-256-512-1024	93.8	83.2	91.8	8.3

接机制补偿信息损失, 确保特征融合完整性。深层 (X_4-X_5) 分别设置为 512 和 1 024 通道, 增强高层语义特征的表达能力。这种设计在计算效率与性能间取得平衡, X_3 层的 256 通道构成计算瓶颈, 降低计算成本, 同时通过跨层融合确保信息完整性, 适合计算资源有限的工业场景部署。

3.7 对比实验分析

对比实验中, 本研究提出的 SAG-YOLOv8 算法与其他几种主流目标检测算法进行了性能比较。实验结果如表 4 所示, 本模型在召回率、准确率和平均精度均值 3 个关键指标上均优于其他算法。具体来看, 本模型的大类别平均准确率为 94.2%, 召回率为 86.3%, $mAP@0.5$ 为 93.9%, 相较其他算法有显著提升。Faster R-CNN 模型虽然在召回率上表现较好, 但其准确率较低, 且计算量较高。RT-DETR 模型虽然其准确率和召回率较高, 但是在计算量上相对较高, 这表明其在实时检测精度上存在局限。与 YOLOv11 相比, 本模型的准确率提高了 1.0%, $mAP@0.5$ 提高了 1.1%。与文献 [24] 提出的算法进行对比, 虽然本模型参数量要比其高, 但准确率、召回率和平均精度都优于其算法, 且仍然适用于实时检测任务。

表 4 对比实验结果

模型	P/%	R/%	mAP@0.5/%	FLOPs
SSD	89.5	74.3	80.2	61.9
Faster R-CNN	67.3	72.4	70.9	91.7
Yolov3	84.6	80.7	82.3	156.7
Yolov4	86.5	82.5	84.4	110.5
Yolov5n	90.7	84.1	90.1	15.8
Yolovx	87.4	88.3	89.7	147.9
Yolov7	91.2	85.6	90.2	105.1
RT-DETR	93.8	89.3	93.2	103.5
Yolov10	91.5	85.9	92.1	8.3
Yolov11	93.2	86.1	92.8	7.9
文献 24	91.1	84.0	90.5	7.6
本模型	94.2	85.3	93.9	11.3

3.8 消融实验分析

在消融实验部分, 我们系统地评估了 SAG-YOLOv8 算法中每个关键组件的贡献。实验结果汇总如

表 5 所示, 研究了各改进模块对工业安全装备装置检测的有效性。

表 5 消融实验

模型	C2f-SWC	AIFI	GFPN	P/%	R/%	mAP@0.5/%	FLOPs
YOLOv8n	×	×	×	89.1	80.7	90.5	8.1
	√	×	×	93.0	86.9	93.5	10.8
	×	√	×	92.7	85.0	92.1	8.0
	×	×	√	93.8	83.2	91.8	8.3
	√	√	×	92.3	86.1	92.7	10.7
	√	×	√	93.7	85.6	92.4	11.0
	×	√	√	93.4	84.7	91.6	8.2
SAG-YOLOv8	√	√	√	94.2	85.3	93.9	11.3

从表 5 可以明显看出, 基础的 YOLOv8n 模型已经具备了较高的准确率和 $mAP@0.5$, 分别为 89.1% 和 90.5%, 但是召回率为 80.7%, 表明在检测目标物体时可能会遗失一部分有效检测。随着 C2f-SWC、AIFI 和 GFPN 组件的逐步加入, 模型的准确率、召回率和 $mAP@0.5$ 均呈现上升趋势。具体来说, C2f-SWC 组件的加入显著提升了模型的召回率至 86.9%, 表明其在增强模型对各种类别目标检测的敏感性方面起到了关键作用。AIFI 组件虽然计算量减少, 但准确率的提升表明它在改善模型对图像内容的理解方面有积极作用。GFPN 组件的引入将模型的准确率提升至 93.8%, 这证明了其在融合不同尺度特征方面的优势。当所有改进组件集成到 SAG-YOLOv8 模型中时, 我们观察到准确率、召回率和 $mAP@0.5$ 分别达到了 94.2%、85.3% 和 93.9%, 这证明了多组件协同作用的显著效果。尽管 FLOPs 从 YOLOv8n 的 8.1 增加到 SAG-YOLOv8 的 11.3, 但考虑到性能的显著提升, 这一增长是合理的, 并且模型仍然适用于实时检测任务。

为了更直观地看到模型改进前后的性能差异, 观察改进前后对不同车间、环境和光照下的工厂实时监控图片, 现选取未参与训练过程的数据进行测试, 验证结果如图 8 所示。对比图 8 可以看出, 在不同工厂环境、自然光和黑夜灯光以及多目标重叠的情况下, 与 YOLOv8n 算法相比, SAG-YOLOv8 模型算法明显检测性能更强, 识别准确率获得较大提高, 一定程度上减少了漏检的情况发生。

4 结束语

本文在 YOLOv8 算法的基础上进行了改进, 在主干网络中, 使用 C2f-SWC 模块替换了原有的 C2f 模块, 能有效增强特征表达能力和多尺度特征融合能力, 然后使用 AIFI 替换空间池化金字塔, 能更好地提高模型对图像语义理解能力, 使网络更轻量化。最后在 Neck 部分加入 GFPN 替换头部的 C3 网络, 通过自底向上和自

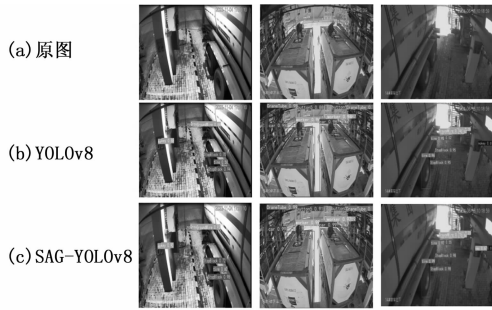


图 8 对比结果图

顶向下两种方式联合进行多尺度特征融合。实验结果表明,改进后的算法在准确率上有所提升,能够满足工厂安全装置装备的识别精度和实时检测需求,尤其在夜间照明不足和复杂背景的环境下,识别效果得到了显著改善,并减少了漏检情况。未来将优化模型轻量化,提升检测性能,推动其在实际业务中应用。

参考文献:

- [1] 巴建磊. 信息化在安全生产和应急管理的应用探讨 [J]. 化工设计通讯, 2024, 50 (11): 164-167.
- [2] 徐彦威, 李 军, 董元方, 等. YOLO 系列目标检测算法综述 [J]. 计算机科学与探索, 2024, 18 (9): 2221-2238.
- [3] CAI Z, VASCONCELOS N. Cascade R-CNN: delving into high quality object detection [C] //Computer Vision and Pattern Recognition, 2018: 6154-6162.
- [4] REN S, HE K, 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, 2016, 39 (6): 1137-1149.
- [5] REDMON J. You only look once: Unified, real-time object detection [C] //Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016: 779-788.
- [6] FARHADI A, REDMON J. Yolov3: An incremental improvement [C] //Computer Vision and Pattern Recognition, Berlin/Heidelberg, Germany: Springer, 2018, 1804: 1-6.
- [7] BOCHKOVSKIY A, WANG C Y, LIAO H Y M. Yolov4: optimal speed and accuracy of object detection [J]. Arxiv Preprint Arxiv: 2004.10934, 2020.
- [8] 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.
- [9] WANG A, CHEN H, LIU L, et al. Yolov10: Real-time end-to-end object detection [J]. Advances in Neural Information Processing Systems, 2024, 37: 107984-108011.
- [10] LIU W, ANGUELOV D, ERHAN D, et al. Ssd: single shot multibox detector [C] //Computer Vision-ECCV

2016: 14th European Conference, Amsterdam, The Netherlands, Proceedings, Part I 14. Springer International Publishing, 2016: 21-37.

- [11] 贺纪桦, 张月峰, 刘悦云. 基于改进 YOLOv8 的智慧工厂工人不行为规范检测 [J]. 制造技术与机床, 2025 (2): 185-193.
- [12] 祁泽政, 徐银霞. 改进 YOLOv5s 算法的安全帽佩戴检测研究 [J]. 计算机工程与应用, 2023, 59 (14): 176-183.
- [13] 李永辉, 袁 亮, 何 丽, 等. 化工园区场景下安全帽佩戴状态检测算法研究 [J]. 计算机工程与应用, 2024, 60 (12): 111-117.
- [14] 王国刚, 李泽欣, 董志豪. 基于注意力机制和多空间金字塔池化的实时目标检测算法 [J]. 计算机测量与控制, 2024, 32 (2): 56-64.
- [15] 郑楚伟, 林 辉. 基于 Swin Transformer 的 YOLOv5 安全帽佩戴检测方法 [J]. 计算机测量与控制, 2023, 31 (3): 15-21.
- [16] 王 贞, 邱 杭, 吴 斌, 等. 基于 CCG-YOLOv8 的施工场景下安全帽佩戴检测 [J]. 武汉理工大学学报, 2024, 46 (6): 73-80.
- [17] LI X, WANG W, WU L, et al. Generalized focal loss: learning qualified and distributed bounding boxes for dense object detection [J]. Advances in Neural Information Processing Systems, 2020, 33: 21002-21012.
- [18] LI D, LI L, CHEN Z, et al. Shift-ConvNets: small convolutional kernel with large kernel effects [J]. Arxiv Preprint Arxiv: 2401.12736, 2024.
- [19] ZHAO Y, LV W, XU S, et al. Detrs beat Yolos on real-time object detection [C] //Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2024: 16965-16974.
- [20] VASWANI A, SHAZEER N, PARMAR N, et al. Attention is all you need [J]. Advances in Neural Information Processing Systems, 2017: 30.
- [21] ZHAO H, JIA J, KOLTUN V. Exploring self-attention for image recognition [C] //Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2020: 10076-10085.
- [22] JIANG Y, TAN Z, WANG J, et al. GiraffeDet: a heavy-neck paradigm for object detection [J]. Computer Science, 2022, 2 (6): 1-6.
- [23] SOUDY M, AFIFY Y, BADR N. RepConv: a novel architecture for image scene classification on Intel scenes dataset [J]. International Journal of Intelligent Computing and Information Sciences, 2022: 1-11.
- [24] 王东升, 聂建军. 一种基于改进 YOLOv8 网络模型的安全帽佩戴检测算法 [J]. 中原工学院学报, 2024, 35 (5): 1-8.