

基于改进 YOLOv8 的梯子检测与定位

王晓君¹, 王梓栋¹, 彭亮亮², 薛伟²

(1. 河北科技大学 信息科学与工程学院, 河北 石家庄 050000;
2. 国药集团威奇达药业有限公司, 山西 大同 037000)

摘要: 在医药与化工行业生产中, 工人往往通过反应罐中的梯子来标识液位高度。反应罐中环境复杂多变, 传统的图像处理算法对于梯子的检测与定位存在准确率不高的缺陷。基于深度学习的 YOLOv8 算法检测精度较高, 但庞大的计算量限制了其在工业现场的应用。为此, 设计一种基于 YOLOv8 的轻量级梯子自动检测与定位算法 HFDL-YOLO。具体来说, 首先, 使用 HGNetv2 替换 YOLOv8n 的主干网络, 用于减少参数量和计算量, 在主干网络中嵌入无参注意力机制 SimAM, 用于提高检测精度, 在此基础上, 结合部分卷积思想, 设计 C2f-FCP 模块代替 Neck 部分的 C2f 模块, 实现多尺度特征融合; 其次, 选用一种轻量级且高效的上采样器 Dysample 来改进 Neck 部分特征融合网络的上采样算子, 以进一步提高算法性能; 最后, 引入一种轻量级共享卷积检测头 LSCD-head, 用于减少模型尺寸, 提高检测效率。实验结果表明, HFDL-YOLO 算法的平均检测精度相较传统 YOLOv8n 提高了 1.1%, 而模型大小、GFLOPs、参数量分别降低了 54.0%、51.9%、57.0%。在亮度不均、角度各异的情况下, HFDL-YOLO 算法以更低的运算成本更精准地实现了反应罐中梯子的自动检测与定位。

关键词: 目标检测; YOLOv8; 轻量化; HGNetV2; 共享卷积

DOI: 10.11907/rjtk.241810

扫描二维码阅读全文: 

中图分类号: TP391.4

文献标识码: A

文章编号: 1672-7800(2025)012-0213-08

Ladder Detection and Positioning Based on Improved YOLOv8

WANG Xiaojun¹, WANG Zidong¹, PENG Liangliang², XUE Wei²

(1. School of Information Science and Engineering, Hebei University of Science and Technology, Shijiazhuang 050000, China;
2. National Pharmaceutical Group Weiqida Pharmaceutical Co., Ltd., Datong 037000, China)

Abstract: In the pharmaceutical and chemical industries, workers often use ladders in reaction tanks to indicate the liquid level height. The environment in the reaction tank is complex and ever-changing, and traditional image processing algorithms have the drawback of low accuracy in detecting and locating ladders. The YOLOv8 algorithm based on deep learning has high detection accuracy, but its large computational complexity limits its application in industrial fields. To this end, design a lightweight ladder automatic detection and localization algorithm HFDL-YOLO based on YOLOv8. Firstly, HGNetv2 is used to replace the backbone network of YOLOv8n, in order to reduce the number of parameters and computation. A parameter free attention mechanism SimAM is embedded in the backbone network to improve detection accuracy. On this basis, combined with the idea of partial convolution, a C2f FCP module is designed to replace the C2f module of Neck, achieving multi-scale feature fusion; Secondly, a lightweight and efficient upsampler, Dysample, is selected to improve the upsampling operator of the Neck partial feature fusion network, in order to further enhance the algorithm performance; Finally, a lightweight shared convolutional detection head LSCD head is introduced to reduce model size and improve detection efficiency. The experimental results show that the average detection accuracy of the HFDL-YOLO algorithm has improved by 1.1% compared to traditional YOLOv8n, while the model size GFLOPs, The parameter count decreased by 54.0%, 51.9%, and 57.0%, respectively. In the case of uneven brightness and different angles, the HFDL-YOLO algorithm achieves more accurate automatic detection and positioning of ladders in reaction tanks with lower computational costs.

Key Words: object detection; YOLOv8; lightweight; HGNetV2; shared convolution

收稿日期: 2024-09-18

作者简介: 王晓君(1973-), 男, 博士, 河北科技大学信息科学与工程学院教授, 研究方向为导航基带信号处理技术、数字信号处理算法及工程实现; 王梓栋(1998-), 男, 河北科技大学信息科学与工程学院硕士研究生, 研究方向为图像处理、目标检测; 彭亮亮(1984-), 男, 国药集团威奇达药业有限公司工程师, 研究方向为微生物发酵; 薛伟(1984-), 男, 国药集团威奇达药业有限公司工程师, 研究方向为生物发酵。本文通讯作者: 王梓栋。

0 引言

医药产业承担着保障人民健康、推动经济发展等多重任务。在许多药品生产过程中,发酵是一个重要环节,由于不断搅拌会使药品溶液产生浮沫导致液位升高,需要实时监测并控制液位高度。工人通常通过反应罐中梯子的高度来标识罐内液位高度,依赖工人主观判断,不同工人视力、经验、观察角度不同,对液位与梯子对应高度判断有差异。工人无法时刻守在罐旁,两次巡检间液位变化难察觉,可能会错过控制时机。此外,靠近反应罐观察液位,周围高温高压、有毒气体等危险易引发事故。因此,如何准确且自动地检测和定位梯子以确保后续工作有序进行是一个关键问题^[1-2]。

随着机器视觉技术的快速发展,使用相机代替人眼定位梯子的目标检测方法成为一个不错的选择。例如,张金水等^[3]采用基于 Adaboost 算法与 Haar 特征的方法对梯子进行检测,但其受光照强度和角度的影响较大;郭鑫等^[4]采用基于方向梯度直方图与支持向量机的方法对目标进行检测,但其鲁棒性较差;Liu 等^[5]采用形态学滤波器,通过计算原始图像与预测背景之间的差值来增强对目标的检测性能,但是需要较高的图像信噪比,且仅适用于简单且均匀的背景;Sun 等^[6]采用小波变换的方法进行目标检测,但是容易受到复杂动态背景噪声的干扰。由以上研究可知,传统的目标检测算法受到环境变化和角度限制的影响,检测精度往往不尽如人意。

YOLO (You Only Look Once) 是一种成功的实时物体检测算法,检测精度大大提高,但具有参数量较大的问题^[7-11]。为此,姜香菊等^[12]采用 FasterNet 网络替代 YOLOv8s 的主干网络 (Backbone), 设计 FasterBlock 模块替代颈部网络 (Neck) 的 C2f 模块,重新设计 BiFPN-A 特征融合结构并在每一层 FasterBlock 模块前引入无参注意力机制,改进算法的参数和计算量有所减小,但精度也随之降低。马超凡等^[13]使用改进的 HGNetV2 替换 YOLOv8 的主干网络,采用 GSConv 卷积模块减少模型计算量,同时采用 Wise-IoU 损失函数提高检测准确率,改进算法的平均检测精度提高 0.3%,但参数量与 FLOPs 仍然很大。由以上研究可知,参数量小的模型精度会降低,精度高的模型参数量较大,较难平衡精度与检测速度之间的关系。

为此,本文基于 YOLOv8n 提出一种轻量化目标检测算法 HFDL-YOLO (HGNetv2-FCP-DySample-LSCD-YOLO)。首先,引入 HGNetv2 网络作为主干网络,并引入无参注意力机制 SimAM,在不增加参数量的前提下提高检测精度;其次,在颈部网络引入动态上采样器 DySample 来替换上采样模块,融合部分卷积思想设计一个名为 C2f-FCP (C2f-FasterConvPconv) 的模块对 C2f 模块进行改进,以降低模型计算复杂度;最后,引入一种轻量级卷积检测头 LSCD-

head (Lightweight Shared Convolutional Detection Head-head), 以减小模型尺寸,提高检测效率^[14]。所提算法在保证检测精度的同时提高了检测效率。

1 HFDL-YOLO 算法

YOLOv8n 模型由以下几个关键部分组成:①输入端。负责接收原始图像数据;②主干特征提取网络。用于提取图像的关键特征;③特征强化提取网络。用于进一步增强和细化特征信息;④检测头。负责根据提取的特征进行目标识别与定位。具体结构如图 1 所示。

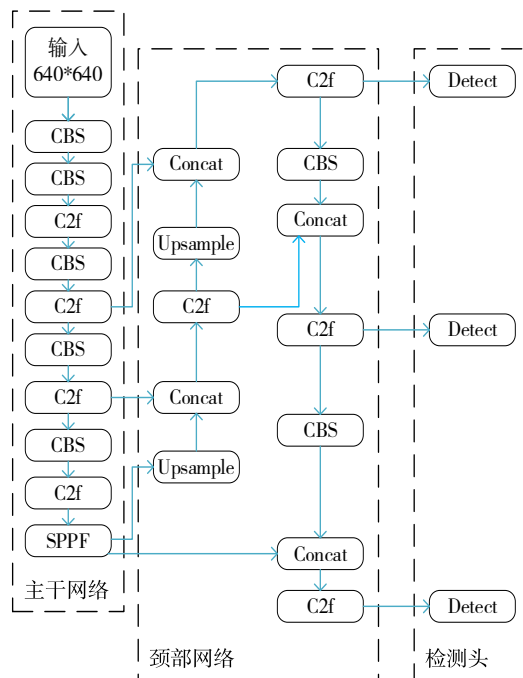


Fig. 1 YOLOv8n structure

图 1 YOLOv8n 结构

YOLOv8n 的主干网络由 Darknet 改进而来,由 CBS (Convolutional Block with Shortcut) 模块、C2f 模块和 SPPF (Spatial Pyramid Pooling Fast) 模块组成,主要作用为按照不同类型聚合图像信息,以提取图像特征,最终输出包含目标位置、类别等信息的特征图,但是其庞大的参数规模和复杂的模型结构可能导致目标推理进程延迟^[15]。因此,本文采用 HGNetv2 替换 YOLOv8n 模型的主干网络,以降低参数量。由于复杂的生产环境会影响检测准确率,本文还将一种无参注意力机制 SimAM 引入主干网络。

YOLOv8n 中的 C2f 模块是对 YOLOv5 中 C3 模块的升级,其汲取了 YOLOv7 中 ELAN (Efficient Layer Aggregation Network) 的设计理念,强化了特征融合能力,并显著提升了推理速度。然而,由于工业环境中液体飞溅会导致目标被遮挡,给模型带来了显著的噪声干扰,导致梯子检测精度偏低。同时,C2f 模块的引入亦增加了模型的参数量和复杂度。因此,本文引用部分卷积 (Partial Convolution,

PConv)思想设计 EMSCPC_Bottleneck 来替换 C2f 模块中的 Bottleneck 结构,在明显提高检测精度的同时大大减少了模型计算量。

在颈部网络部分, YOLOv8n 对 PAN(Path Aggregation Network)结构进行了优化,移除了上采样后的卷积步骤,但其上采样模块使用最近邻插值上采样,其只是简单复制最近邻像素的值,无法根据特征图的内容自适应地选择合适的采样点,可能会造成语义信息丢失的问题。因此,本文在颈部网络部分引入轻量级动态上采样算子 Dysample,该算子能根据输入特征图动态调整偏移量,更好地保留图像的细节信息。

在检测头部分,本文采用最新的解耦头结构,将分类头与检测头分离,有效缓解了复杂场景中可能出现的定位不准确和分类错误问题^[16]。然而,这种设计具有一定局限性:首先检测头的参数量较大;其次各个检测头之间缺少信息交互,可能会导致部分信息丢失。针对这两个问题,本文引入一种轻量级卷积检测头 LSCD-head,用于减小模型尺寸,提高检测效率。综合以上改进措施,本文提出的 HF DL-YOLO 模型结构见图 2。

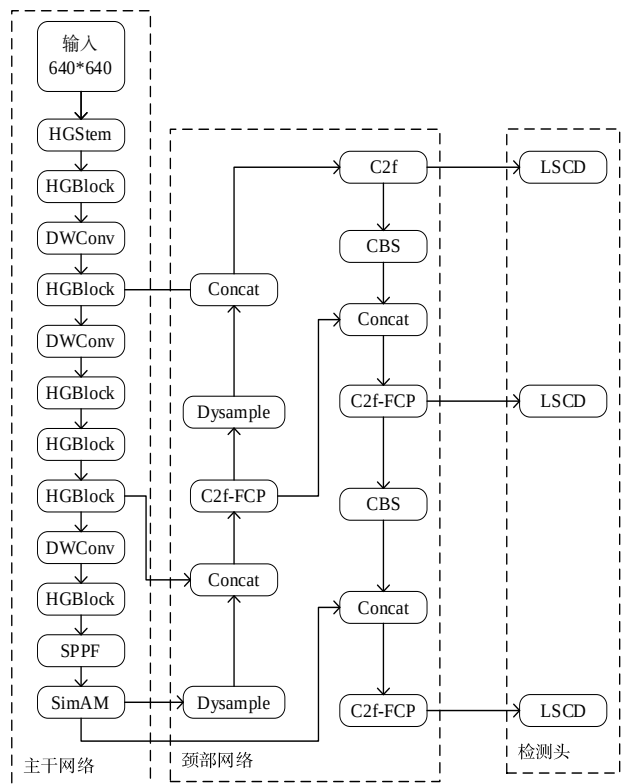


Fig. 2 HF DL-YOLO model structure

图 2 HF DL-YOLO 模型结构

1.1 HGNetv2 模块

为降低参数量, HF DL-YOLO 采用 HGNetv2 替换 YOLOv8n 模型的主干网络。HGNetv2 主要由 HGStem、HGBlock 和 DWConv 3 个模块构成^[17]。HGStem 作为网络的入口,通过卷积层对输入图像进行初步特征提取,并利用最大池化来降低图像分辨率,从而在多尺度上捕获图像特

征,结构如图 3(a)所示。HGBlock 由一系列不同大小的卷积层构成,旨在捕获图像中丰富多样的特征信息,结构如图 3(b)所示。DWConv 作为 HGNetv2 中的一个创新点,通过深度可分离卷积技术替代传统卷积,不仅降低了模型的计算复杂度,提升了推理速度,而且使每个通道能够独立进行特征学习。

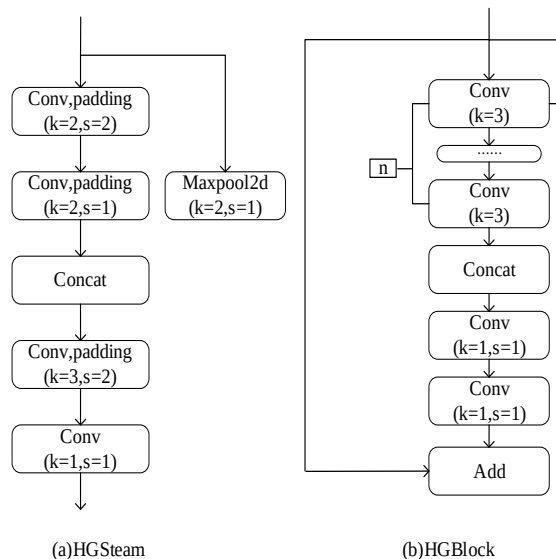


Fig. 3 HGStem and HGBlock structure

图 3 HGStem 和 HGBlock 结构

1.2 SimAM 注意力机制

本文将无参数注意力机制 SimAM 引入主干网络来提高复杂生产环境下的目标检测准确率。相较于现有通道和空间注意力模块,该模块创新性地避免了向网络添加额外参数,通过在单一网络层内推断特征图,直接计算出三维注意力权重,在简化模型的同时提高了特征提取效率和准确性^[18]。其原理如图 4 所示。

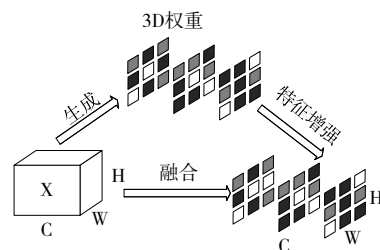


Fig. 4 Principle of SimAM

图 4 SimAM 原理

SimAM 增加了模型的感知范围,且不需要额外的参数,因而减少了模型参数量,提高了计算效率,尤其在资源受限的设备中有助于加速推理。在 YOLOv8n 的主干网络部分引入 SimAM 注意力模块可提升特征图的表征能力,同时减少由于通道数增多而造成的特征信息冗余问题,提高了模型检测精度。

SimAM 注意力机制通过能量函数来实现,其对每个神经元进行评估,判断每个神经元的重要程度。能量函数表示为:

$$e_i(w_i, b_i, y, x_i) = (y_i - \hat{t})^2 + \frac{1}{M-1} \sum_{i=1}^{M-1} (y_0 - \hat{x}_i)^2 \quad (1)$$

式中: $\hat{t} = w_i t + b_i$ 和 $\hat{x}_i = w_i x_i + b_i$ 分别表示输入特征中同一个通道中神经元 t 和其他神经元 x_i 的线性变换; M 表示神经元个数; w_i 和 b_i 分别表示某个神经元变换时的权重和偏差。

通过引入二元标签并结合正则化项来对能量函数进行最小化处理, 目的为强化同一通道中神经元 t 与其他神经元之间的区分能力。简化并重新定义的能量函数为:

$$e_i(w_i, b_i, y, x_i) = \frac{1}{M-1} \sum_{i=1}^{M-1} (-1 - (w_i x_i + b_i))^2 + (1 - (w_i t + b_i))^2 + \lambda w_i^2 \quad (2)$$

通过对上式进行解析可得权重 w_i 和偏差 b_i 为:

$$w_i = -\frac{2(t - \mu_i)}{(t - \mu_i)^2 + 2\delta_i^2 + 2\lambda} \quad (3)$$

$$b_i = -\frac{1}{2}(t + \mu_i)w_i$$

式中: μ_i 表示均值。计算公式为:

$$\mu_i = \frac{1}{M-1} \sum_{i=1}^{M-1} x_i \quad (4)$$

δ_i^2 表示方差。计算公式为:

$$\delta_i^2 = \frac{1}{M-1} \sum_{i=1}^{M-1} (x_i - \mu_i)^2 \quad (5)$$

因此, 推导出最终的能量最小化表达式为:

$$e_i^* = \frac{4(\hat{\sigma}^2 + \lambda)}{(t - \hat{u})^2 + 2\hat{\sigma}^2 + 2\lambda} \quad (6)$$

由式(6)可知, 能量越低, e^* 值越小, 表示神经元 t 与周围神经元的区别越大, 重要程度也越高, 因此神经元的重要性可以通过 $1/e^*$ 得到。使用 Sigmoid 函数对重要特征进行增强处理, 得到新的输出特征 X' 。表示为:

$$X' = \text{sigmoid}\left(\frac{1}{E}\right) \cdot X \quad (7)$$

式中: X 为输入特征; E 为输入特征所有神经元能量值之和。

1.3 C2f-FCP

C2f 模块主要包括两个卷积层和若干个 Bottleneck 模块, 其中 Bottleneck 模块是一种残差块, 用于增加网络的深度和复杂度, 作用原理是通过卷积操作和 Bottleneck 模块的迭代, 对输入特征图中的信息进行提取与融合, 进而得到一个更加丰富的特征图。为构建更轻量级的检测模型, 同时提高检测精度, 本文引入 PConv 思想, 设计 EMSCPC_Bottleneck 来替换 C2f 模块中的 Bottleneck 结构, 在明显提高模型检测精度的同时大大减少计算量^[19]。

Pconv 卷积通过在输入数据的一部分通道上执行标准卷积操作来对空间特征进行提取, 同时保持其他通道的维度不变。假设 EMSCPCConv 模块接收具有 c 个通道的输入特征图, 并首先通过 Split 层将其分为两个具有 $c/2$ 通道的特征图, 前半部分特征图的通道保持不变, 不经过卷积操

作, 而后半部分特征图的通道均分为两组。第一组特征图应用 3×3 的卷积核进行卷积操作以提取特征, 第二组特征图则使用 5×5 的卷积核进行卷积。后半部分获得的结果与前半部分获得的特征连接起来, 且连接的特征在每个通道中彼此独立。最终, 这两个特征图的组合将通过一次 1×1 卷积, 以确保输出特征图的通道数与原始输入特征图的通道数一致。这一过程不仅增强了特征的表达能力, 而且通过逐点卷积保持了通道维度的一致性, 为后续的网络层提供了丰富的特征信息。EMSCPCConv 模块结构见图 5。

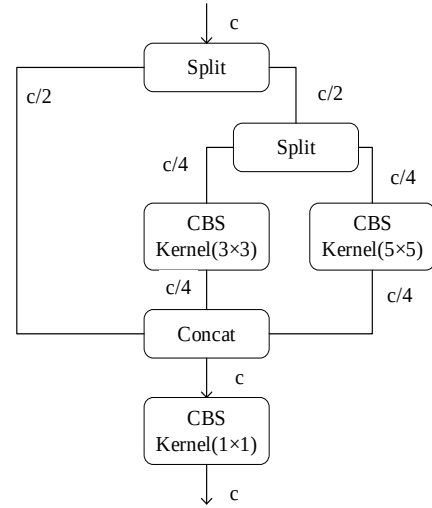


Fig. 5 EMSCPCConv architecture

图 5 EMSCPCConv 结构

一个常规卷积模块的计算量为:

$$h \times w \times k^2 \times c^2 \quad (8)$$

EMSCPCConv 模块的计算量为:

$$h \times w \times k_1^2 \times (c/4)^2 + h \times w \times k_2^2 \times (c/4)^2 \quad (9)$$

式中: h 、 w 、 c 、 k 分别表示输入特征图的高、宽、通道数和卷积核大小。

由式(8)和式(9)可以求出, 当 EMSCPCConv 模块卷积核 k_1 和 k_2 大小分别为 3×3 和 5×5 时的计算量为一个同等通道数 3×3 卷积核计算量的 $1/3$ 。将 Bottleneck 中第二个 CBS 模块替换为 EMSCPCConv, 利用 EMSCPC_Bottleneck 替换 C2f 中的 Bottleneck 部分, 记作 C2f-FCP。C2f-FCP 模块不仅减少了网络模型参数量, 而且能充分提取不同尺度的特征信息, 有效提高了模型在梯子检测与定位中的准确性。Bottleneck 部分结构和替换后的 C2f-FCP 结构分别如图 6 和图 7 所示。

1.4 Dysample

在目标检测任务中, 为使模型能够准确识别不同尺寸和位置的物体, 需要通过上采样对输入的特征图尺寸进行调整, 使其与原始图像的尺寸相匹配。传统的上采样方法通常通过双线性插值技术来实现^[20]。然而, 该方法存在一些固有缺陷, 可能会造成图像的重要特征丢失。此外, 基于核函数的上采样技术在计算和参数需求上较为繁重, 这在构建轻量级网络架构时可能会成为瓶颈^[21]。

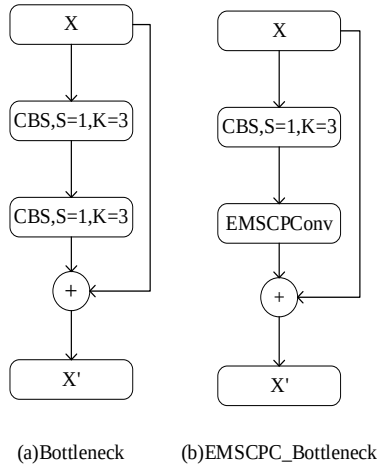


Fig. 6 Structure of Bottleneck part

图 6 Bottleneck 部分结构

在现实的工业场景中,由于药液或蒸汽等因素造成检测镜头不清晰,可能会出现像素失真等问题,导致识别任务中细粒度细节丢失和特征学习困难。为解决这一问题,本文采用 Dysample 替换上采样模块。其是一种高度轻量级和高效的动态上采样器,可增强模型对低分辨率

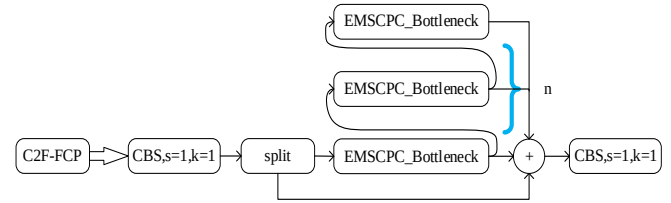


Fig. 7 C2f-FCP structure

图 7 C2f-FCP 结构

图像的检测能力,同时减少假阳性和假阴性现象的发生^[22-23]。

Dysample 采用基于点的采样技术和学习采样视角,避免了使用动态卷积运算和额外的子网络,能够在不增加额外负担的情况下提升图像分辨率,故能在降低计算成本的同时有效提升模型的检测效率。具体结构如图 8 所示。其中,图 8(a)为基于采样的动态上采样结构;图 8(b)为 Dysample 中的采样点生成器结构。图中的 X 、 X' 、 G 、 O 分别表示输入特征、上采样特征、生成的偏移量和原始网格;符号 σ 表示 Sigmoid 函数; sh 表示采样高度; sw 表示采样宽度; gs^2 表示特征图通过线性层后的通道数。

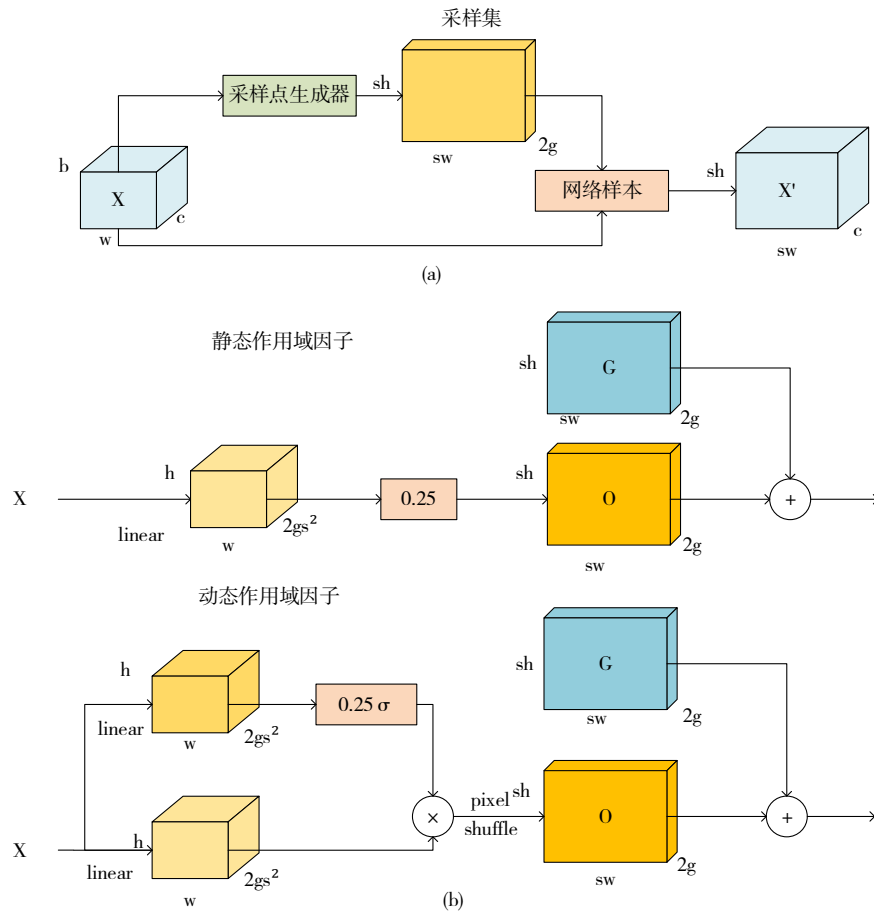


Fig. 8 Dysample structure

图 8 Dysample 结构

采样集 S 由原始采样网格 (O) 和生成的偏移量 (G) 组成,其中偏移量使用“linear + pixel shuffle”方法生成,其范

围可以通过静态和动态因素确定。以静态因子抽样方法为例,对于给定大小的特征图 $c \times h \times w$ 和上采样因子 s ,特

征图首先经过一个输入和输出通道分别为 c 和 $2s^2$ 的线性层;其次使用 pixel shuffle 方法将其重塑为 $2 \times sh \times sw$, 其中 2 表示坐标维度,用于存储每个采样点的二维空间坐标 (x, y) ;最后生成大小为 $c \times sh \times sw$ 的上采样特征图。

1.5 LSCD-head

YOLOv8 的原始检测头有一些局限性:首先,检测头的参数量较大,占整个算法计算量的 1/5,且 3 个检测头均需要通过 2 个 3×3 卷积和 1 个 1×1 卷积提取图像信息,这种结构导致算法参数量显著增加;其次,各个检测头之间缺少信息交互,可能会导致部分信息丢失,检测效果下降。为

解决上述问题,本文引入 LSCD-head 结构,采用 Group Normalization 技术来增强检测头在目标定位和类别识别方面的性能。该结构的一个显著特点是共享卷积层的设计,即使用共享卷积 GroupNorm 替换 3 个检测头使用的两种常见卷积。该设计允许 3 个不同分辨率的特征图在网络中进行有效融合,这种多尺度特征融合不仅促进了特征间的信息交流,而且通过参数共享降低了模型复杂性。共享卷积层的引入显著降低了模型参数量,减少了对内存的需求,并减轻了计算负担,这对于工业现场的应用尤为重要。LSCD-head 结构如图 9 所示。

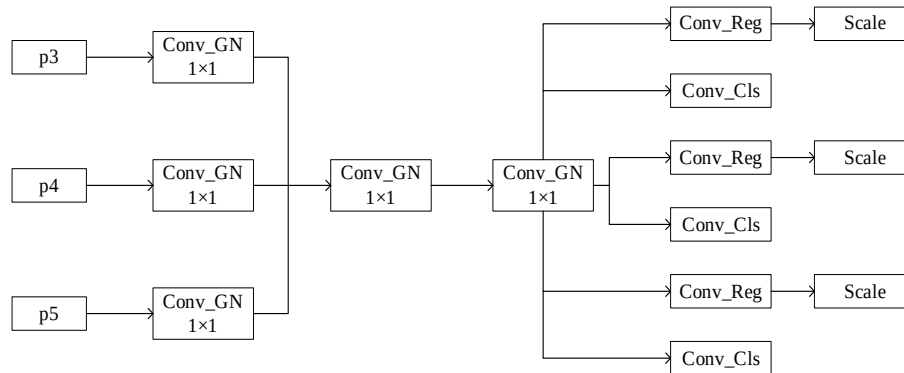


Fig. 9 LSCD-head structure

图 9 LSCD-head 结构

2 实验方法与结果分析

2.1 实验环境与参数配置

使用 Ubuntu 20.24 作为操作系统,PyTorch 作为深度学习框架,选择 YOLOv8n 作为基准网络模型。实验平台的 GPU 配置为 RTX 3090(24 GB),编程语言为 Python 3.8。具体配置见表 1。

Table 1 Experimental environment setting

表 1 实验环境配置

环境项	环境规格
操作系统	Ubuntu 20.24
编程语言	Python 3.8
深度学习框架	PyTorch 1.11.0
CUDA	11.3
CPU	Intel(R) Xeon(R) Platinum 8255C
GPU	RTX 3090(24 GB)
内存	43 GB

在进行训练与验证时,初始学习率设置为 0.01,模型的输入样本尺寸均转换为 640×640 ,动量设置为 0.937,采用 SGD 优化器,训练时 batch size 大小为 16,Epoch 设置为 150,Weight Decay 设置为 0.000 5。

2.2 数据集

本文采用自制梯子数据集。对自行采集的不同照度、不同大小、不同摆放角度的梯子图像进行组合,获取原始

图像 500 张。为模拟反应罐中的复杂环境,提高模型鲁棒性和泛化能力,对梯子图像进行添加噪声、随机裁剪、镜像翻转、低照度等方式的增强处理,最终得到 1 500 张梯子图像,按照 9:1 的比例划分为训练集和验证集。自制梯子数据集示例见图 10。

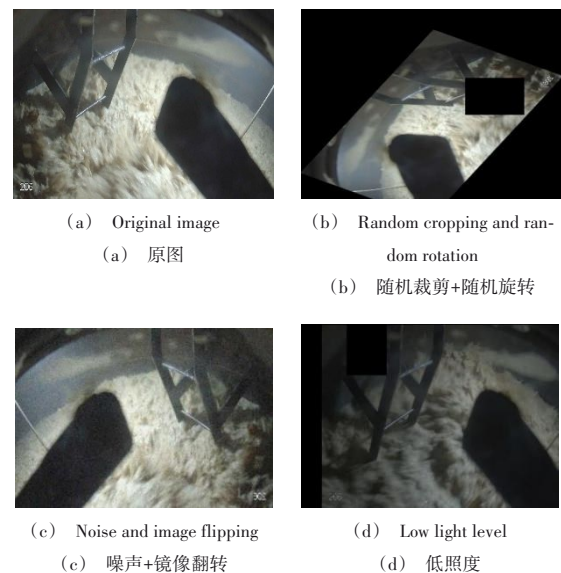


Fig. 10 Dataset example

图 10 数据集示例

2.3 评估指标

为考察模型性能,本文采用平均精度均值(Mean Average Precision, mAP)、模型参数量(Params)、浮点运算量

(Giga Floating-point Operations Per Second, GFLOPs)、模型尺寸大小(MB)作为评价指标。选用 mAP@0.5 评价模型的识别精度,其为阈值等于 0.5 时的 mAP。参数量、浮点运算量和模型尺寸大小用于评价模型的存储需求和计算资源消耗。

2.4 比较实验

为验证本文模型的优势,在自制梯子数据集上将其与主流算法 YOLOv3-Tiny、YOLOv5n、YOLOv6n、YOLOv8s、YOLOv8n 进行比较。实验结果见表 2。可以看出,HFDL-YOLO 算法在大幅度降低模型参数量和计算量的前提下精度也有小幅度提升。

Table 2 Comparative experimental results on the self-made ladder dataset
表 2 在自制梯子数据集上的比较实验结果

模型	mAP@0.5/%	模型大小/MB	GFLOPs/G	Params/M
YOLOv3-Tiny	90.8	24.4	18.9	12.12
YOLOv5n	90.0	5.3	7.1	2.50
YOLOv6n	90.6	8.7	11.8	4.23
YOLOv8s	91.2	22.5	28.4	11.13
YOLOv8n	90.2	6.3	8.1	3.00
HFDL-YOLO	91.3	2.9	4.9	1.29

为验证本文模型的泛化性能,在 VOC2007 数据集上与同样的对照模型进行比较。实验结果见表 3。可以看出,HFDL-YOLO 模型在更大规模和更丰富多样的数据集上进行训练时,能够更好地学习并捕捉目标特征,从而提高了在验证集上的检测准确性和鲁棒性。

Table 3 Comparative experimental results on the VOC2007 dataset
表 3 在 VOC2007 数据集上的比较实验结果

模型	mAP@0.5/%	模型大小/MB	GFLOPs/G	Params/M
YOLOv3-Tiny	46.3	24.4	18.9	12.12
YOLOv5n	44.3	5.3	7.1	2.50
YOLOv6n	42.2	8.7	11.8	4.23
YOLOv8s	47.9	22.5	28.4	11.13
YOLOv8n	46.9	6.3	8.1	3.00
HFDL-YOLO	47.1	2.9	4.9	1.29

3 结语

为对药品生产过程中反应罐的梯子进行准确检测,本文提出一种基于 YOLOv8n 的轻量化目标检测算法模型 HFDL-YOLO。首先,为占用更少资源,引入 HGNetv2 作为 HFDL-YOLO 的主干网络,同时引入无参注意力机制 SimAM 来解决轻量化带来的检测精度下降问题;其次,在颈部网络部分引入动态上采样器 DySample 来替换上采样模块,同时融合部分卷积思想设计一个名为 C2f-FCP 的模块,以降低模型计算复杂度;最后,引入一种轻量级卷积检测头 LSCD-head,以减小模型尺寸,提高检测效率。为提

Table 3 Comparative experimental results on the VOC2007 dataset
表 3 在 VOC2007 数据集上的比较实验结果

模型	mAP@0.5/%	模型大小/MB	GFLOPs/G	Params/M
YOLOv3-Tiny	46.3	24.4	18.9	12.12
YOLOv5n	44.3	5.3	7.1	2.50
YOLOv6n	42.2	8.7	11.8	4.23
YOLOv8s	47.9	22.5	28.4	11.13
YOLOv8n	46.9	6.3	8.1	3.00
HFDL-YOLO	47.1	2.9	4.9	1.29

2.5 消融实验

为详细验证改进模型的可行性和各个模块的作用,在共同应用 YOLOv8n 框架的前提下对每个模块进行消融实验,结果如表 4 所示。其中,模型 1 为 YOLOv8n 模型;模型 2 为利用 HGNetv2 替换 YOLOv8n 主干网络的模型,实验结果表明 HGNetv2 在不影响检测精度的前提下可有效减少模型参数量和计算量;模型 3 在模型 2 的基础上加入 C2f-FCP 模块,实验结果表明该模块的引入使模型在提高检测精度的同时可有效减少模型参数量和计算量;模型 4 在模型 3 的基础上加入无参注意力机制 SimAM,实验结果表明该机制的引入可以在不增加模型参数量和计算量的前提下提高检测精度;模型 5 在模型 4 的基础上利用 Dysample 模块替换 YOLOv8 模型的上采样模块,实验结果表明 Dysample 模块的引入使模型在轻微增加参数量和计算量的情况下提高了检测精度;模型 6 在模型 5 的基础上引入 LSCD-head,即 HFDL-YOLO 模型,实验结果表明该检测头的引入使模型在损失很小检测精度的情况下大幅度减少了参数量和计算量。

Table 4 Results of ablation experiment
表 4 消融实验结果

模型	HGNETv2	C2f- FCP	SimAM	Dysample	LSCD	mAP@0.5 /%	模型大小/MB	GFLOPs /G	Params /M
1	×	×	×	×	×	90.2	6.3	8.1	3.00
2	√	×	×	×	×	90.3	5.0	6.9	2.35
3	√	√	×	×	×	91.2	4.7	6.7	2.20
4	√	√	√	×	×	91.4	4.7	6.7	2.20
5	√	√	√	√	×	91.7	4.8	6.7	2.21
6	√	√	√	√	√	91.3	2.9	4.9	1.29

高模型的泛化能力和鲁棒性,本文还创建标注了复杂工业环境下的梯子数据集。实验结果表明,与 YOLOv8n 模型相比,HFDL-YOLO 模型在大程度降低参数量和计算量的前提下,精度也有小幅度提升,实现了内存占用、计算量与准确性的良好平衡。

然而,本文仅针对梯子检测与定位算法进行改进,尚未在移动设备上实现硬件部署。后续可尝试将模型移植到移动平台,同时结合实际生产环境的特定需求对模型进行优化与调整,以提高其实用性。

参考文献:

[1] AN G C, LIU J J, WANG Y, et al. Liquid level detection algorithm based

- on template matching and roughness texture index[J]. *Electronic Measurement Technology*, 2023, 46(23): 168-174.
- 安国臣, 刘娟娟, 王雁, 等. 基于模板匹配及粗糙度纹理指标的液位检测算法[J]. *电子测量技术*, 2023, 46(23): 168-174.
- [2] LIU J X. Liquid level measurement and control system based on machine vision [D]. Shijiazhuang: Hebei University of Science and Technology, 2021.
- 刘嘉鑫. 基于机器视觉的液位测量与控制系统[D]. 石家庄: 河北科技大学, 2021.
- [3] ZHANG J S, WANG Q P, FAN T Y. Safety testing method for ship pilot ladder[J]. *Marine Engineering*, 2019, 48(3): 179-182.
- 张金水, 王群朋, 范天佑. 船舶引航梯安全检测方法[J]. *航海工程*, 2019, 48(3): 179-182.
- [4] GUO X, HAO Q, YANG F. Apple multi object detection method based on improved HOG and SVM [J]. *Foreign Electronic Measurement Technology*, 2022, 41(11): 154-159.
- 郭鑫, 郝骞, 杨凤. 基于改进 HOG 和 SVM 的苹果多目标检测方法[J]. *国外电子测量技术*, 2022, 41(11): 154-159.
- [5] LIU R, WANG D, ZHOU D. Point target detection based on multiscale morphological filtering and an energy concentration criterion [J]. *Applied Optics*, 2017, 56(24): 6796-6805.
- [6] SUN Y Q, TIAN J W, LIU J. Background suppression based on wavelet transformation to detect infrared target [C]//*International Conference on Machine Learning and Cybernetics*, 2005: 4611-4615.
- [7] 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.
- [8] REDMON J, FARHADI A. YOLO9000: better, faster, stronger [C]//*IEEE Conference on Computer Vision and Pattern Recognition*, 2017: 6517-6525.
- [9] FARHADI A, REDMON J. Yolov3: an incremental improvement [DB/OL]. <http://arxiv.org/pdf/1804.02767>.
- [10] 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>.
- [11] 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.
- [12] JIANG X J, FENG H Z, LI T. Improved railway foreign object intrusion rapid detection method based on FasterNet and YOLOv8s [J]. *Journal of Beijing Jiaotong University*, 2024, 48(5): 39-48.
- 姜香菊, 冯海照, 李涛. 基于 FasterNet 和 YOLOv8s 改进的铁路异物入侵快速检测方法[J]. *北京交通大学学报*, 2024, 48(5): 39-48.
- [13] MA C F, LI X, WANG X X. Lightweight non motorized vehicle object detection based on heavy parameterization [J]. *Computer Engineering and Applications*, 2024, 60(19): 190-198.
- 马超凡, 李翔, 王晓霞. 基于重参数化的轻量化非机动车目标检测[J]. *计算机工程与应用*, 2024, 60(19): 190-198.
- [14] WANG X Q, GAO H B, JIA Z M. Improving the road defect detection algorithm of YOLOv8 [J]. *Computer Engineering and Applications*, 2024, 60(17): 179-190.
- 王雪秋, 高焕兵, 郝泽萌. 改进 YOLOv8 的道路缺陷检测算法[J]. *计算机工程与应用*, 2024, 60(17): 179-190.
- [15] ZHU Y, DONG E, TONG J, et al. Deep neural network based object detection algorithm with optimized detection head for small targets [C]//*IEEE International Conference on Mechatronics and Automation*, 2023: 2378-2382.
- [16] YANG J, FU X, HU Y, et al. PanNet: a deep network architecture for pan-sharpening [C]//*Venice: IEEE International Conference on Computer Vision*, 2017.
- [17] WANG C C, LI M. Research on fire smoke detection algorithm based on YOLOv8 [J]. *Journal of Beijing Union University*, 2023, 37(5): 69-77.
- 王晨灿, 李明. 基于 YOLOv8 的火灾烟雾检测算法研究[J]. *北京联合大学学报*, 2023, 37(5): 69-77.
- [18] ZHANG L, ZHENG J, LI C, et al. CCDNETR: a detection transformer based on constrained contrast denoising for multi-class synthetic aperture radar object detection [J]. *Sensors*, 2024, 24(6): 1793.
- [19] YANG L, ZHANG R, LI L D, et al. SimAM: a simple, parameter-free attention module for convolutional neural networks [C]//*Proceedings of the 38th International Conference on Machine Learning*, 2021: 11863-11874.
- [20] MA X, GUO F, NIU W, et al. PCONV: the missing but desirable sparsity in DNN weight pruning for real-time execution on mobile devices [DB/OL]. <https://arxiv.org/pdf/1909.05073>.
- [21] LI Y, MA C, LI L, et al. Lightweight tunnel obstacle detection based on improved YOLOv5 [J]. *Sensors*, 2024, 24(2): 395.
- [22] WANG J, CHEN K, XU R, et al. CARAFE: content-aware reassembly of features [C]//*International Conference on Computer Vision*, 2019: 3007-3016.
- [23] LIU W, LU H, FU H, et al. Learning to upsample by learning to sample [C]//*International Conference on Computer Vision*, 2023: 6027-6037.

(责任编辑: 尹晨茹)