

DOI:10.20033/j.1003-7241.(2026)03-0128-06

基于改进的 YOLOv8 尺度自适应大渣检测算法研究

徐福斌¹, 殷浩然¹, 任莉芸², 李哲²

(1. 国家能源集团泰州发电有限公司, 江苏 泰州 225321; 2. 北京国电富通科技发展有限责任公司, 北京 100070)

摘要:针对工业场景中渣块检测, 研究提出一种结合尺度自适应模块和特征融合自注意力机制的改进 YOLOv8 模型。该方法通过引入深度估计子网络, 利用目标的深度信息动态调整锚框的大小, 以提高模型在不同尺度下的检测能力。此外, 特征融合的自注意力机制进一步增强了对显著目标特征的关注, 减少了背景噪声的干扰。研究还使用了自制的数据集, 该数据集包含在工业排渣机场景下采集的不同深度、尺度和光照条件下的大渣块样本, 以评估所提方法的性能。实验结果表明, 改进后的 YOLOv8 模型在复杂场景中表现出显著的性能提升, 在平均精度、准确率和召回率方面均优于其他主流检测模型, 达到了 98.1% 的 mAP。这验证了所提方法在复杂工业场景中检测大渣块的有效性。与鲁棒性。

关键词:图像处理; 尺度自适应模块; 特征融合自注意力; 工业熔渣检测; 深度估计; 目标检测

中图分类号: TP391

文献标志码: A

文章编号: 1003-7241(2026)03-0128-06

Research on scale-adaptive large slag detection algorithm based on improved YOLOv8

XU Fubin¹, YIN Haoran¹, REN Liyun², LI Zhe²

(1. State Energy Group Taizhou Power Generation Co., Ltd., Taizhou 225321, Jiangsu, China;

2. Beijing Guodian Futong Technology Development Co., Ltd., Beijing 100070, China)

Abstract: In this study, an improved YOLOv8 model combining scale adaptive module and feature fusion self-attention mechanism is proposed for the detection of large slag blocks in industrial scenes. By introducing a depth estimation sub-network, this method uses the depth information of the target to dynamically adjust the size of the anchor box to improve the detection ability of the model at different scales. In addition, the self-attention mechanism of feature fusion further enhances the attention to salient target features and reduces the interference of background noise. This study also used a homemade dataset containing samples of large slag blocks at different depths, scales, and lighting conditions collected in an industrial slag discharge airport scene to evaluate the performance of the proposed method. Experimental results show that the improved YOLOv8 model shows significant performance improvement in complex scenes, and is superior to other mainstream detection models in terms of average precision, accuracy and recall, reaching 98.1% mAP. This verifies the effectiveness and robustness of the proposed method in detecting large slag blocks in complex industrial scenarios.

Keywords: picture processing; scale adaptive module; feature fusion self-attention; industrial slag detection; depth estimation; target detection

在工业应用中, 大渣块的堆积检测具有重要的意义, 尤其是在光照条件复杂、目标尺度变化大以及视角不一致的情况下, 传统的目标检测方法难以保证较高的精度^[1-3]。在排渣机的运行过程中, 渣井中常常会出现大渣块的堆积, 这些渣块如果不及时清除, 可能会对排渣机的正常工作造成严重影响, 甚至引发设备故障和安全隐患^[4,5]。排渣机通常工作在恶劣的环境中, 光照不均、粉尘遮挡等问题对大渣块的检测提出了更高的要求。因此, 如何在复杂环境中实现对渣块的实时检测, 成为工业应用中的关键挑战。

随着深度学习的发展, 卷积神经网络 (convolutional

neural network, CNN) 推动了目标检测技术的快速进步。自 R-CNN^[6] 系列问世以来, 目标检测算法逐步演化为端到端的自动化学习方式, 实现了检测精度和效率的显著提升。R-CNN 系列包括 R-CNN、Fast R-CNN^[7] 和 Faster R-CNN^[8-9], 这些方法逐步优化了区域提取和特征提取过程, 使得检测速度和精度不断提升。Faster R-CNN 通过引入区域建议网络 (region proposal network, RPN), 显著提高了检测速度, 标志着两阶段检测方法的成熟。

除了两阶段方法, 单阶段检测器如 SSD^[10] 和 YOLO^[11-12] 系列也取得了重要进展。SSD 提出了多尺度特征检测的概念, 从不同尺度的特征层中进行目标检测, 有效提高了对不同

收稿日期: 2025-01-06; 录用日期: 2025-03-18

基金项目: 河北省自然科学基金 (23280301Z)

作者简介: 徐福斌 (1970—), 男, 本科, 高级工程师, 研究方向: 火电厂热动力工程。

引用本文: 徐福斌, 殷浩然, 任莉芸, 等. 基于改进的 YOLOv8 尺度自适应大渣检测算法研究 [J]. 自动化技术与应用, 2026, 45(3): 128-132, 146. (XU Fubin, YIN Haoran, REN Liyun, et al. Research on scale-adaptive large slag detection algorithm based on improved YOLOv8 [J]. Techniques of Automation and Applications, 2026, 45(3): 128-132, 146.)

大小目标的检测效果。而 YOLO 系列则通过统一的单阶段检测框架,在检测速度和精度之间取得了良好的平衡。自 YOLOv1 到 YOLOv8,模型在网络结构、损失函数和特征提取模块上不断优化,以适应更复杂的场景需求。YOLOv8^[13] 作为该系列的最新版本,进一步优化了特征提取和网络结构,在检测精度和实时性之间取得了更优的平衡。

此外,基于 Transformer 的目标检测方法,如 DETR^[14] (DEtection TRansformer),通过引入自注意力机制实现了对目标检测任务的端到端建模。DETR 不依赖于区域提取网络,利用 Transformer 的全局建模能力,显著提升了对复杂场景中多目标的检测性能。

然而,YOLOv8 在面对目标尺度变化大、视角变化明显的工业场景中,尤其是具有明显深度差异的大渣块时,其尺度适应能力仍然存在不足。为了更好地适应这些场景的需求,提升 YOLOv8 对于渣块不同尺度的适应性,本文提出了一种结合基于深度的尺度自适应模块的改进方法。通过在网络中加入深度估计子网络,计算渣块的深度信息,利用深度信息指导特征提取和锚框的动态调整,使得 YOLOv8 能够更好地适应不同深度和尺度的目标,提高检测性能。

1 本文算法

改进后的 YOLOv8 模型整体结构如图 1 所示,主要由原 YOLOv8 主干网络、深度估计子网络和尺度自适应模块组成。首先,输入图像经过 YOLOv8 的主干网络提取特征,同时输入图像还通过深度估计子网络生成深度图。然后,利用深度图的信息来指导尺度自适应模块动态调整检测锚框和特征金字塔中的特征提取过程。深度图被输入到尺度自适应模块,动态调整检测锚框的大小以适应目标的深度差异,从而优化特征提取和检测效果。同时为了增强 YOLOv8 在复杂场景中对大渣块特征的提取能力,本文在特征融合过程中引入了自注意力机制,以提升对关键目标特征的关注度,同时抑制背景噪声。特征融合的自注意力机制在特征金字塔网络(feature pyramid network, FPN)中实现,使得模型能够在不同尺度下有效地关注重要目标区域。

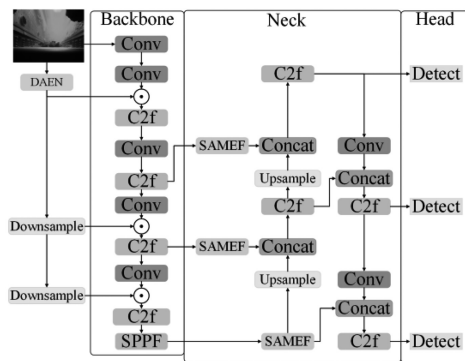


图 1 模型整体结构图

Fig. 1 Overall structure diagram of the model

1.1 模型整体结构图

深度估计子网络的目的是从输入图像中预测每个像素点的深度值。为了不增加过多的计算开销,本文选择了

一个轻量级的深度估计网络(DAEN^[15])来生成深度图。深度估计子网络的输出是一个深度图 $D \in R^{H \times W}$,其中 H 和 W 分别为输入图像的高度和宽度。

具体来说,给定输入图像 $I \in R^{H \times W \times 3}$,深度估计子网络通过一系列卷积层和反卷积层提取多尺度特征,并使用一个回归头生成深度图 D 。该深度图用于指示图像中每个目标的相对深度信息,并在后续的尺度自适应模块中使用。

1.2 尺度自适应模块

尺度自适应模块利用深度估计子网络输出的深度图来动态调整 YOLOv8 网络中的特征提取过程。具体来说,尺度自适应模块(scale adaptation module, SAM)通过深度图来确定每一层特征提取的权重,使得特征提取网络能够更加集中于目标区域,从而提高大渣块的检测效果。

1.2.1 锚框尺度调整

在 YOLOv8 中,锚框的大小会根据特征自动进行调整,但在本研究中,进一步提出了基于深度的锚框尺度自适应方法。具体而言,定义深度调节因子 $\alpha = f(D)$,其中 $f(D)$ 是一个映射函数,用于根据深度图 D 动态计算每个目标区域的尺度调节因子。

设锚框的初始尺寸为 (w_a, h_a) ,则调整后的锚框尺寸为 $(w'_a, h'_a) = (\alpha w_a, \alpha h_a)$ (1)

其中,深度调节因子 α 的计算公式为

$$\alpha = \frac{1}{1 + e^{-\beta(d-d_0)}} \quad (2)$$

式中, d 表示目标区域在深度图中的平均深度值, d_0 是深度的基准值, β 是一个控制调节因子变化速率的超参数。该公式的设计使得目标越靠近相机,锚框尺寸越大,以确保对大尺度目标的有效覆盖;而对于远处的小尺度目标,锚框尺寸会相应减小。

1.2.2 特征提取的深度加权融合

在 YOLOv8 的特征金字塔网络中,本文引入了深度加权融合策略,以增强对不同深度目标的感知能力。深度信息在特征金字塔的每一层被应用,以调整每个尺度的特征融合权重,确保远近目标的特征能够被有效区分。具体来说,在特征融合过程中,使用深度图中的深度信息对不同尺度的特征图进行加权处理。

设第 l 层的特征图为 $F^l \in R^{H_l \times W_l \times C_l}$,深度图经过下采样得到与特征图相同尺寸的深度权重图 $W^l \in R^{H_l \times W_l}$,则深度加权后的特征图 F^l_{weight} 表示为

$$F^l_{\text{weight}} = W^l \odot F^l \quad (3)$$

式中, \odot 表示逐元素相乘操作。深度权重图 W^l 通过对深度图 D 进行下采样并归一化得到,归一化过程使用公式为

$$W^l_{i,j} = \frac{D_{i,j} - D_{\min}}{D_{\max} - D_{\min}} \quad (4)$$

式中, D_{\min} 和 D_{\max} 分别表示深度图中的最小值和最大值。通过这种加权融合策略,模型在特征金字塔中能够更好地区分远近目标的特征,提高对不同深度目标的检测效果。

1.3 特征融合的自注意力机制

1.3.1 自注意力模块设计

在特征融合阶段,首先从多尺度特征图中提取关键信息。设输入特征图为 $F^l \in R^{H_l \times W_l \times C_l}$, 其中 H_l, W_l 和 C_l 分别表示特征图的高度、宽度和通道数。自注意力模块的目标是计算特征图中每个像素点的自适应权重,以增强显著特征。

自注意力模块首先将特征图 F^l 映射为查询(Query)、键(Key)和值(Value)3个矩阵,分别记为 Q^l, K^l 和 V^l , 其计算过程为

$$Q^l = W_Q F^l, K^l = W_K F^l, V^l = W_V F^l \quad (5)$$

式中, W_Q, W_K 和 W_V 分别是用于特征转换的可学习参数矩阵。接着,计算查询和键之间的相似性,通过点积操作获得自注意力得分矩阵 A^l 。

$$A_{i,j}^l = \frac{Q_i^l \cdot K_j^{lT}}{\sqrt{d_k}} \quad (6)$$

式中, Q_i^l 和 K_j^l 分别表示查询和键矩阵中的第 i 和第 j 个位置, d_k 表示键向量的维度,用于缩放点积结果以防止数值不稳定。

然后,对每个像素点的注意力得分进行 Softmax 归一化,以计算最终的权重矩阵 γ^l 。

$$\gamma_{i,j}^l = \frac{\exp(A_{i,j}^l)}{\sum_k \exp(A_{i,k}^l)} \quad (7)$$

最后,将注意力权重应用到值矩阵 V^l , 并对所有位置进行加权求和得到输出特征图 F_{out}^l 。

$$F_{out}^l = \sum_j \gamma_{i,j}^l V_j^l \quad (8)$$

1.3.2 自注意力机制与特征融合

在 YOLOv8 的特征金字塔网络中,将自注意力模块应用于每个尺度的特征图。在进行特征融合时,使用自注意力模块对不同尺度的特征图进行加权处理,使得模型能够更加专注于显著目标区域。

对于特征金字塔中每一层的特征图,首先通过自注意力模块进行处理,得到增强后的特征图 F_{fusion} , 然后将不同尺度的特征图进行融合。

$$F_{fusion} = \sum_l F_{out}^l \quad (9)$$

通过这种方式,模型在融合多尺度特征时能够更加有效地保留重要目标的特征信息,增强了模型对大渣块的检测能力。

此外,为了减少计算开销,在实现过程中使用了多头自注意力机制(multi-head attention, MHA),将特征图划分为多个子空间,并在每个子空间上独立执行注意力计算。通过这种方式,可以有效地并行化计算,同时提高模型对不同特征的关注能力。

最后,自注意力模块与 YOLOv8 主干网络进行了无缝集成,使得模型能够在复杂工业场景下具备更高的检测精度和鲁棒性。实验结果表明,引入特征融合的自注意力机制(self-attention mechanism for feature fusion, SAMFF)后,

模型对大渣块的检测性能得到了显著提升,尤其是在背景复杂和目标重叠的情况下。

2 数据集介绍

为了评估所提方法的性能,本文使用了在工业排渣机场景下采集的大量渣块堆积图像,数据集中包含不同深度、尺度和光照条件下的大渣块样本。数据集格式按照 PASCAL VOC 数据集格式建立,训练集中包含 4 000 张图像,测试集中包含 1 000 张图像,图像均使用 LabelImg 进行了手动标注然后人工复检,以保证数据集中的目标标注准确。数据集部分图像如图 2 所示。



(a) 数据集样本示例



(b) 数据集样本示例



(c) 数据集样本示例



(d) 数据集样本示例

图 2 数据集示意图

Fig. 2 Schematic diagram of the dataset

3 实验结果与分析

3.1 实验环境

实验硬件环境:i7-6700 3.4 GHz CPU、NVIDIA RTX 3080 GPU,操作系统为 Ubuntu 18.04。软件环境:编程语言为 Python3.7。

3.2 评价指标

常用的目标检测模型性能评估通常用准确率 Precision 和召回率 Recall 来评价目标检测模型的综合性能指标,mAP 值被用来描述多标签图像分类的性能,AP 是各类别平均精度(average precision, AP)的均值。这些参数用来描述检测性能。相应的计算过程如下。

1) 准确率 P (Precision)。准确率可以衡量模型的精确性,可以定义为模型预测的正类别中有多少是真正的正类别,如式(10)所示。

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \quad (10)$$

式中,TP 表示模型正确地将正例判定为正例的数量,FP 表示模型错误地将负例判定为正例的数量。

2) 召回率 R (Recall)。召回率可以衡量模型的全面性,可以定义为所有真正的正类别中有多少被模型成功检测到,如式(11)所示。

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \quad (11)$$

式中,FN 表示模型为漏报的正样本数量。

3) 平均准确率均值 mAP (mean Average Precision)。平均准确率均值是一种综合评价指标,它综合考虑了模型在不同类别上的精准率,是所有类别 AP 的均值,如式(12)所示。

$$\text{mAP} = \frac{\sum_{i=1}^d \text{AP}_i}{d} \quad (12)$$

式中, AP_i 表示单个类别的平均精度, d 表示所有类别的数量。

3.3 数据集实验结果对比分析

在本次实验中,对比本文提出的改进 YOLOv8 模型与其他几种主流目标检测模型,包括 SSD、YOLOv4、YOLOv5s、YOLOX 以及原始的 YOLOv8。实验结果如表 1 所示,本文提出的改进 YOLOv8 模型在检测精度(mAP)、准确率(Precision)和召回率(Recall)方面均取得了优异的表现,达到了 98.1% 的 mAP,相较于其他模型表现出明显的优势。

表 1 DZI 数据集对比实验

Tab. 1 Comparative experiments on DZI dataset

算法	$P/\%$	$R/\%$	mAP@0.5/%	FPS/fps
SSD	78.3	75.8	76.4	51
YOLOv4	86.2	84.5	83.5	64
YOLOv5s	93.6	92.4	92.7	72
YOLO X	89.2	88.4	87.3	49
YOLOv8	2.3	90.8	91.3	56
本文算法	98.8	97.4	98.1	48

首先,在检测精度(mAP)方面,改进后的 YOLOv8 模型显著优于其他模型。相比 YOLOv5s 的 92.7% 和原始 YOLOv8 的 91.3%,本文算法在复杂场景下能够更加准确地检测到目标,主要得益于引入了深度估计子网络和尺度自适应模块。这些改进使得模型能够更好地适应目标的深度和尺度变化,从而提高了特征提取的准确性和稳定性。其次,在准确率(Precision)和召回率(Recall)方面,改进后的 YOLOv8 模型分别达到了 98.8% 和 97.4%。这表明该模型不仅能够准确识别出正类目标,而且在处理目标遮挡和复杂背景时也有较强的鲁棒性。相比之下,YOLOv4 和 YOLOX 的准确率和召回率分别为 86.2% 和 89.2%,本文算法在应对复杂工业场景中的大渣块检测时表现得更加出色。

此外,从推理速度(FPS)来看,本文模型的检测速度为 48 fps,尽管略低于 YOLOv5s 的 72 fps,但该速度仍然满足大多数工业实时应用的需求。这种性能折衷是由于引入了深度估计子网络和自注意力机制,虽然增加了一定的计算量,但换来了更高的检测精度和鲁棒性。在工业场景中,检测的精度和稳定性往往比推理速度更为重要,因此本文方法在实际应用中具有更高的实用价值。

实验结果证明,引入深度信息和自适应模块能够有效提升目标检测模型在复杂环境中的表现。深度估计子网络使得模型能够根据目标的深度信息进行动态调整,优化了特征提取和锚框设置的过程,而自注意力机制则增强了对重要目标特征的关注度,降低了背景噪声对检测的影响。这些改进使得本文提出的模型在检测大渣块时不仅精度更高,而且在应对光照不均、目标重叠和视角变化等复杂情况时表现出更强的适应性。

综合来看,本文提出的改进 YOLOv8 模型在工业渣块检测任务中,尤其是在复杂场景下表现出了显著的优势,验证了深度估计和自适应模块对于提升目标检测性能的有效性。

图 2 展示了排渣机在复杂工业场景下的大渣块检测效果。从图中可以看出,改进后的 YOLOv8 模型能够有效地检测出不同深度和尺度的大渣块,并且在光照不均、目标遮挡和背景复杂的情况下依然表现出较高的检测精度。

具体来说,图 2 中显示的检测结果包括多个大渣块的边界框,边界框的大小和位置能够准确地反映出目标的实际尺度和位置。得益于深度估计子网络和尺度自适应模块,模型在检测过程中能够根据渣块的深度信息动态调整锚框的大小,使得靠近相机的大渣块具有较大的锚框,而远处的小渣块则具有相对较小的锚框。这种动态调整机制确保了检测结果的准确性,并有效减少了误报和漏报的情况。

此外,自注意力机制的引入使得模型在特征提取过程中能够更加专注于显著目标区域,减少了背景噪声的干扰。这一点在图 3 中也得到了充分体现,即使在背景复杂、存在大量无关物体的情况下,模型依然能够准确地定

位并检测出大渣块。这表明改进后的 YOLOv8 模型具备较强的鲁棒性和适应性,能够胜任复杂工业场景中的大渣块检测任务。

总体来看,图 2 直观地验证了本文所提方法在大渣块检测中的有效性和优越性,进一步支持了实验数据中得出的结论。改进后的 YOLOv8 模型在复杂场景下的表现,展示了其在实际工业应用中的巨大潜力。

在检测速度(FPS)方面,本文模型以较高的精度和适中的检测速度(50 fps)保持了良好的平衡。这一点对于需要在资源受限的环境中部署高精度检测模型的应用场景尤为重要。



图 3 排渣机大渣检测图

Fig. 3 Large slag detection diagram of slag discharge machine

3.4 消融实验与分析

为了验证所提出的各个模块对最终检测性能的影响,本文设计了一系列消融实验,旨在评估尺度自适应模块(SAM)和特征融合的自注意力机制(SAMFF)对模型检测性能的贡献。实验结果如表 2 所示,通过逐步增加模型模块,分析其对准确率(Precision)、召回率(Recall)和平均精度(mAP@0.5)的影响。

表 2 DZI 数据集消融实验

Tab. 2 Ablation experiments on DZI dataset

SAM	SAMFF	P/ %	R/ %	mAP@0.5/ %
—	—	93.1	92.8	93.5
✓	—	94.9	93.8	94.1
—	✓	95.6	93.4	93.2
✓	✓	98.8	97.4	98.1

首先,在未添加尺度自适应模块(SAM)和特征融合的自注意力机制(SAMFF)的基础上,YOLOv8 模型的检测性能相对较低,准确率为 93.1%,召回率为 92.8%,平均精度为 93.5%。在加入尺度自适应模块(SAM)后,模型的准确率和召回率均有显著提升,分别达到了 94.9% 和 93.8%,平均精度也提升至 94.1%。这表明,尺度自适应模块通过利用深度信息对锚框进行调整,增强了模型在不同尺度下的适应能力,提高了对不同大小目标的检测效

果。在仅引入特征融合的自注意力机制(SAMFF)的情况下,模型的准确率和召回率分别为 95.6% 和 93.4%,平均精度为 93.2%。尽管平均精度略有下降,但模型在显著目标区域的识别能力上有所增强,表明自注意力机制在特征融合过程中有助于抑制背景噪声并突出显著特征。最后,当同时引入尺度自适应模块(SAM)和特征融合的自注意力机制(SAMFF)后,模型的准确率和召回率分别提升至 98.8% 和 97.4%,平均精度达到了 98.1%。这表明两个模块的结合显著增强了模型在复杂工业场景中对大渣块的检测性能,尤其是在目标深度变化大和背景复杂的情况下,表现出更强的鲁棒性和准确性。

通过以上消融实验可以得出结论,尺度自适应模块和特征融合的自注意力机制对改进 YOLOv8 模型的检测性能均有显著作用,且二者的结合能够最大化提升检测效果。这验证了本文提出的方法在复杂工业场景中的有效性和优越性。

4 结论

本文提出了一种基于尺度自适应模块(SAM)和特征融合自注意力机制(SAMFF)的改进 YOLOv8 模型,用于解决工业场景中渣块的检测问题。通过引入深度估计子网络与自适应模块,本文模型在复杂环境下的检测性能得到了显著提升,尤其是在光照不均、目标深度变化大以及背景复杂的情况下表现出更强的适应性和鲁棒性。消融实验结果表明,尺度自适应模块和特征融合自注意力机制均在提升模型性能方面具有显著贡献。尺度自适应模块通过深度信息引导锚框调整,使得模型能够更好地适应不同深度和尺度的目标;而特征融合自注意力机制通过抑制背景噪声、增强显著目标区域的特征,进一步提高了模型的检测精度和稳定性。与现有的主流目标检测算法相比,改进后的 YOLOv8 模型在准确率、召回率和平均精度方面均取得了显著的提升,并在复杂工业场景中表现出优异的性能。这验证了所提出方法在实际应用中的可行性和有效性。未来的研究将着重于进一步提升模型的推理速度,以更好地满足实时工业检测的需求。同时,将探索其他特征融合方法与自适应策略的结合,以进一步提高模型的检测性能和鲁棒性。

参考文献

- [1]王金刚,余彪,章少山,等. 火电厂碎渣机自动清障功能的实现[J]. 安徽电气工程职业技术学院学报, 2024, 29(4): 93-98.
- [2]陈鸿鑫,马天霖,房新,等. 某 660 MW 燃煤机组干式排渣系统漏风特性及优化调整研究[J]. 洁净煤技术, 2024, 30(增刊 2): 384-392.
- [3]林悦楠. 基于双目立体视觉的干式排渣机炉渣量在线检测技术研究[D]. 南京:东南大学, 2023.
- [4]林悦楠,范永胜,李健,等. 基于双线结构光双目视觉的干排渣机排渣流量检测方法研究[J]. 节能技术, 2023, 41(2): 105-112.
- [5]冯帆,辛大欣,华瑾,等. 基于近红外视觉的排渣机械手高温炉渣定位方法[J]. 红外技术, 2019, 41(2): 194-201.

(下转第 146 页)