

DOI:10.19651/j.cnki.emt.2518796

# 基于视觉联合检测的球团矿粒径检测方法<sup>\*</sup>

龚玉洁 王旭 郭海杰 丁志星 崔雪红

(青岛科技大学信息科学技术学院 青岛 266061)

**摘要:** 针对工业球团矿图像分割任务中存在的小目标密集粘连、易受遮挡等技术难题,提出基于 YOLOv11 与 SAM2 的视觉联合检测实例分割方法(YO-SAM2)。首先,通过 CSC 模块改进 YOLOv11 的 C3k2 模块,增强网络对密集小目标特征的表达能力。其次,设计小目标混合融合特征金字塔网络(SHFPN),在 P2 层增加特征图输出以捕捉细节信息,添加跨层交互并采用基于内容引导注意力机制的混合融合策略,提升多尺度特征融合效能。此外,引入解耦空间-通道上采样模块(DSCU)替代原始上采样方法,生成更具表达力的特征表征。最后,通过嵌入可学习 Adapter 适配器对 SAM2 分割模型进行参数高效微调,有效提升模型对特定工业场景的适应性和泛化性。实验结果表明,YO-SAM2 在球团矿数据集的  $mIoU$  达 90.3%,与 Mask R-CNN、YOLOv8-seg 等主流分割算法相比,取得了最佳分割效果。该方法有效解决了工业场景下球团矿分割的精度与鲁棒性问题,为智能工业质检提供了可靠的技术方案。

**关键词:** 实例分割;多尺度特征融合;细节信息;参数微调

**中图分类号:** TP391;TN919.8 **文献标识码:** A **国家标准学科分类代码:** 520.6

## Pelletized ore particle size detection method based on vision joint detection

Gong Yujie Wang Xu Guo Haijie Ding Zhixing Cui Xuehong

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

**Abstract:** To address the technical challenges of dense target adhesion and occlusion-prone small objects in industrial pelletized ore image segmentation, this study proposes an instance segmentation method (YO-SAM2) integrating YOLOv11 and SAM2. Firstly, the CSC module is introduced to improve the C3k2 module in YOLOv11, enhancing the network's capability to represent features of densely clustered small targets. Second, a Small-Target Hybrid Fusion Feature Pyramid Network (SHFPN) is designed to augment feature map outputs at the P2 layer for fine-grained detail capture, incorporating cross-layer interactions and a content-guided attention mechanism to optimize multi-scale feature fusion. Additionally, a Decoupled Spatial-Channel Upsampling module (DSCU) is proposed to replace conventional upsampling, generating more discriminative feature representations. Finally, parameter-efficient fine-tuning of the SAM2 segmentation model is achieved via a learnable Adapter, significantly improving adaptability and generalization in industrial scenarios. Experimental results demonstrate that YO-SAM2 achieves a state-of-the-art  $mIoU$  of 90.3% on the pelletized ore dataset, outperforming mainstream segmentation algorithms such as Mask R-CNN and YOLOv8-seg. This method effectively resolves the challenges of accuracy and robustness in industrial pellet segmentation, offering a reliable technical solution for intelligent industrial quality inspection.

**Keywords:** instance segmentation; multi-scale feature fusion; detail information; parameter fine-tuning

## 0 引言

球团矿生球制造是炼钢工业中的重要环节<sup>[1]</sup>,其粒径大小对其在高炉或还原工艺中的性能有重要影响。若粒径超过 8~16 mm,则会导致塞料、返工、污染等问题,影响生产效率 and 经济效益<sup>[2-3]</sup>。传统人工取样筛分法虽能获取粒

径数据,但其作业模式存在耗时耗力、人为误差高等固有缺陷。

近年来,基于机器视觉的在线检测技术为球团矿粒径分析提供了新途径。Nanda 等<sup>[4]</sup>利用傅里叶变换分析方法对传送带上生球颗粒的边缘图像进行分析,并以此获取粒度分布测量结果。孙浠汉等<sup>[5]</sup>提出区域约束分水岭算法与

收稿日期:2025-05-13

\* 基金项目:国家自然科学基金(61472196,61672305)项目资助

LU-Net 的分割方式,提升了生球的分割效果,但分水岭算法对噪声敏感,需依赖复杂的预处理抑制过分割。Wu 等<sup>[6]</sup>提出基于 Nested U-Net 学习网络,通过卷积层简单串联的方式提高分割精度,但对微小或重叠球团的边缘分割仍存在模糊问题。Duan 等<sup>[7]</sup>提出基于 U-Net 的分割算法用于稀疏球团矿的分割,既保证了网路的轻量又提高了泛化能力,但模型高度依赖训练数据分布,若实际生产环境与训练数据差异较大,分割效果显著下降。综上,球团矿微小、边缘不规则及相互遮挡的特性导致球团矿的视觉检测仍存在着多个问题:1)依赖传统图像算法,需复杂的预处理过程并对环境要求较高;2)现有分割算法通过多层卷积与下采样生成的高层特征图因感受野过大,难以有效捕获微小目标的局部细节,尤其对不规则边缘的建模能力不足;3)深层网络的特征金字塔结构虽能融合多尺度信息,但会引入大量冗余特征,弱化针对球团矿边框的回归精度;4)算法依赖像素级标注数据,而球团矿的密集分布特性使标注成本高昂且易产生标注误差。

近两年,Kirillov 等<sup>[8]</sup>通过提示驱动机制实现了开放世界的通用分割。Chen 等<sup>[9]</sup>基于 SAM 模型并结合语义类别信息,开发了遥感图像自动实例分割模型 RSPrompter。牛津大学团队基于 SAM2 框架提出了 MedSAM-2<sup>[10]</sup>,通过置信度记忆库和加权拾取策略解锁了医学图像分割的单个提示分割。尽管 SAM 系列在常见分割任务中表现出色,其在工业领域的应用较少,在球团矿图像分割中的潜力尚未被充分挖掘。此外,以 YOLO 为代表的一阶段检测算法始终在精度与效率的协同优化方面持续推动着工业自动化发展<sup>[11]</sup>。特别是 2024 年推出的 YOLOv11<sup>[12]</sup>通过将传统 CNN 结构转向基于变换器的架构,在精度与速度间实现了最优平衡,更加契合工业生产场景中的检测需求。

基于以上情况,本文将 YOLOv11 作为 SAM2<sup>[13]</sup> 生成提示信息的合适候选者,提出 YO-SAM2 协同分割框架用于球团矿实例分割。首先通过改进的 YOLOv11 输出目标预测框,然后预测框作为提示信息输入定制的 SAM2 提示编码器,生成稀疏提示标记,最后解码器根据中间图像特征与稀疏提示标记输出分割结果。实验结果表明,该算法的  $mIoU$  达到 90.3%,为球团矿粒径检测提供了可靠解决方案。

## 1 YO-SAM2 算法

本文算法总体技术路线可以概括为两个阶段,如图 1 所示。检测阶段,改进 YOLOv11 设计了 SH\_YOLO。YOLOv11 网络结构分为 Backbone 主干网络、Neck 颈部、Head 头部 3 部分,本文主要针对 Backbone 和 Neck 做了优化改进。首先,原始 YOLOv11 使用固定卷积核进行特征提取,难以适应不同感受野的需求,为此设计 CSC 模块改进 C3k2,增强网络对多尺度特征的适应能力。其次,Neck 部分 PAFPN 结构会致使部分细节特征丢失且等权特征聚

合机制难以自适应地平衡不同尺度特征对检测任务的贡献度。为解决此问题,构建 SHFPN 替换 PAFPN,利用上下文引导下采样模块(contextguidedblock\_down,CGBD)捕捉 P2 层输出的小目标细节信息;增加网络与主干特征的跨层融合,并且引入基于内容引导注意力机制(content-guided attention,CGA)的混合融合方法聚焦关键特征信息。此外,针对网络放大特征图时易丢失小目标细节信息的问题,采用 DSCU 模块逐步上采样特征图。

分割阶段,对由图像编码器、提示编码器、掩码解码器三大部分构成的 SAM2 进行专门定制,以适应球团矿实际分割需求。SAM2 依赖于手动提示且预训练是基于大规模自然图像数据集进行的,自然图像和球团矿图像之间的域差异,尤其是球团矿的尺度、独特纹理、形状等的复杂性,很大程度上影响了模型的分割性能。故在图像编码器中嵌入 Adapter 块进行参数微调,解决域差异,增强工业场景下球团矿分割任务的适应性。

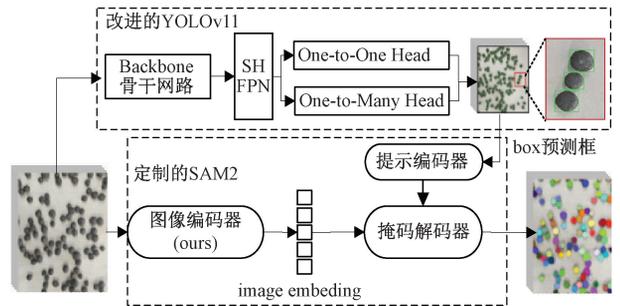


图 1 本文算法技术路线

Fig. 1 The algorithm technology roadmap of this article

## 2 算法改进

### 2.1 SH\_YOLO 算法

SH\_YOLO 算法通过 3 方面的改进显著提升了算法的检测精度,同时减少了误检和漏检。算法的架构如图 2 所示。

#### 1) C3k2 模块改进

C3k2 模块通过可变卷积核扩展感受野,尤其适用背景复杂场景和深层次特征提取任务,但其未能有效建立多尺度特征间的非线性关联,在密集小目标检测时仍面临一定挑战。为提升模型检测精度,进一步优化 C3k2 模块,设计跨尺度通道协同模块(CSC)嵌入 C3k2,得到改进后的 C3k2\_CSC,弥补模型对密集小目标检测的不足。模块优化前后结构对比如图 3 所示。

C3k2\_CSC 模块的核心设计在于 CSC,通过并行多尺度深度卷积与动态通道交互机制动态调控感受野范围。如图 4 所示,输入特征首先通过  $1 \times 1$  逐点卷积扩展通道维度,经批量归一化(BN)与 SiLU 激活函数进行非线性变换。为建立局部-全局特征关联,以 P3 为例,进入并行异构核深度卷积 DWConv 分支,核尺寸分别设置  $1 \times 1$  进行通道维

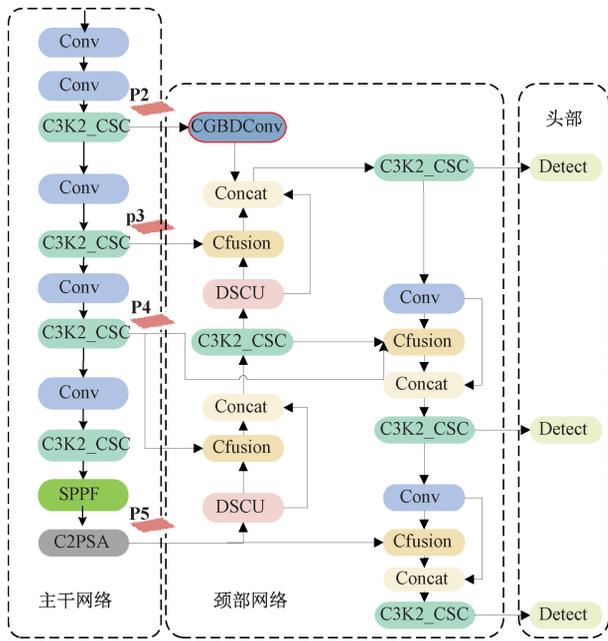


图2 SH\_YOLO 结构

Fig. 2 Architecture of SH\_YOLO

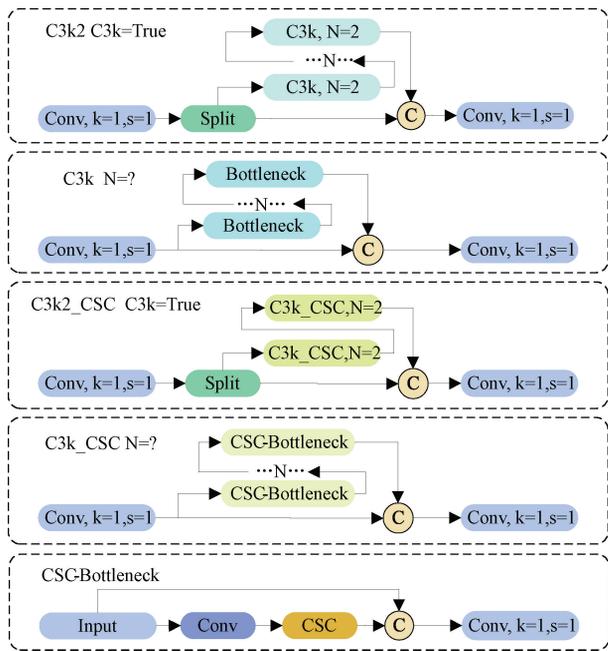


图3 C3k2 与 C3k2\_CSC 模块结构对比

Fig. 3 Comparison of C3k2 and C3k2\_CSC module structures

度的线性变换,聚焦跨通道关系,3×3 捕捉边缘纹理等高频信息,5×5 扩大感受野,获取更全局上下文。此后,为避免深度卷积对每个输入通道进行独立卷积忽略通道间的关系,输出特征经通道维度拼接后,采用通道洗牌机制(Channel\_Shuffle)跨组整合通道间关系。最终,通过1×1 卷积压缩通道至目标维度,并与输入特征通过条件残差连接融合,生成增强后的输出特征。

该模块通过多核并行设计,在单层内实现局部细节保持、边缘强化与全局上下文捕获的协同,并引入通道洗牌机制打破深度卷积的通道隔离,既能适应不同分辨率要求又能建立有效的上下文关联,很好的解决了密集小目标检测中局部细节丢失与特征混淆的问题。

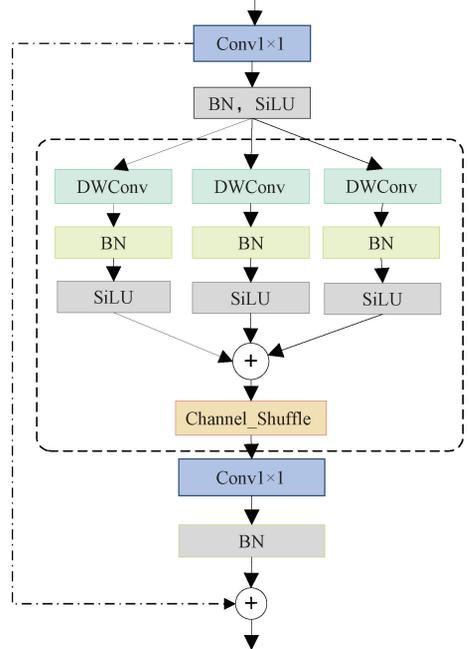


图4 CSC 结构

Fig. 4 Architecture of CSC

2) 颈部网络改进

YOLOv11 采用 PAFPN 特征金字塔结构,其通过自底向上路径和横向连接,使特征信息能够在网络中高效流动。然而 PAFPN 对密集小目标检测时仍存在不足:(1)仅依赖 P3、P4 和 P5 层,网络难以提取小目标所需的细粒度特征信息;(2)在自上而下的特征传递过程中,主干网络不同阶段的特征仅通过串行方式逐层交互,信息流动方向固定,小目标细节特征信息易发生丢失;(3)融合时所有输入特征只是简单的相加,并没有区分不同特征的重要性差异。为此,重新设计颈部网络,提出小目标混合融合特征金字塔网络(SHFPPN)。

首先,用 CGBD<sup>[14]</sup>对 P2 层特征图进行处理,增强小目标的上下文感知能力。然后将这些小目标特征与上采样层和 P3 层的特征拼接,最终输入到 C3k2\_CSC 模块。这种方法能够更好的捕捉小目标特征信息,提升密集小目标的检测精度。

其次,提出跨层特征增强机制,在基础网络结构中新增一个由主干网络至下采样层的特征直连通道,这种跨层级方法可以有效抑制小尺度目标在深层网络中的特征信息衰减问题。

最后,引入一种基于内容引导注意力机制(CGA)<sup>[15]</sup>的混合融合方法(简称 CFusion),该方法通过动态计算内容

特征的重要性权重,自适应地调整特征融合过程中的注意力分布,使网络能够聚焦于与检测任务最相关的特征信息。这样能增强网络对多源异构信息的融合能力,显著提升小尺寸目标的检测性能。

内容引导注意力机制架构如图 5 所示。其输入特征图  $X \in \mathbb{R}^{C \times H \times W}$  首先通过式(1)、(2)计算得到相应的通道特征  $W_c$  与空间特征  $W_s$ 。

$$W_c = C_{1 \times 1}(\max(0, C_{1 \times 1}(X_{GAP}^c))) \quad (1)$$

$$W_s = C_{7 \times 7}([X_{GAP}^s, X_{GMP}^s]) \quad (2)$$

其中,  $\max(0, \cdot)$  代表 ReLU 激活函数,  $C_{k \times k}$  表示内核大小为  $k \times k$  的卷积层,  $X_{GAP}^c$  表示在空间维度上执行全局平均池化操作,  $X_{GAP}^s$  和  $X_{GMP}^s$  分别表示在通道维度上执行全局平均池化和全局最大池化操作。

此后,通过相加操作将  $W_c$  与  $W_s$  进行融合,再与输入进行拼接,并通过通道洗牌操作以交替方式重新整合通道,得到最终的特征权重。

$$W = \sigma(GC_{7 \times 7}(CS([X, W_c + W_s]))) \quad (3)$$

其中,  $\sigma$  表示激活函数操作,  $GC$  表示组卷积操作,  $CS$  表示通道洗牌操作。

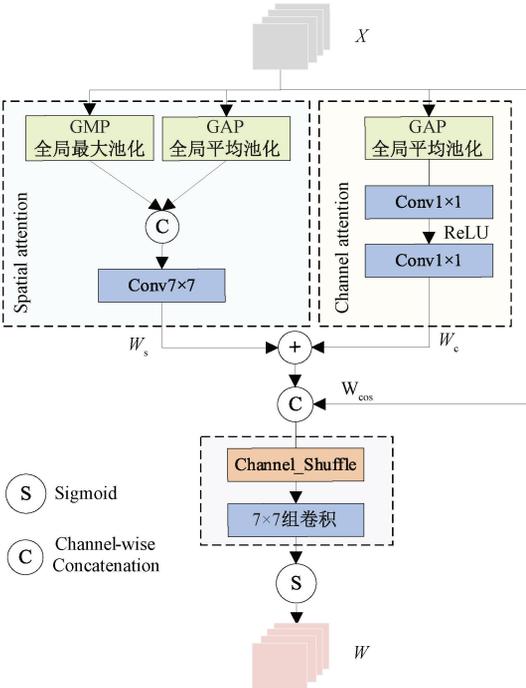


图 5 内容引导注意力(CGA)结构

Fig. 5 Architecture of content-guided attention(CGA)

而基于内容引导注意力机制的混合融合方法分为双输入与三输入。当输入为双输入时,输入特征  $F_1$  与  $F_2$  的像素矩阵逐元素相加后输入到 CGA 中得到权重  $W$ , 然后  $W$  与各特征图进行加权融合并与原始输入特征相加, 最终通过  $1 \times 1$  卷积调整通道数, 得到特征  $F_{fuse}$ , 该过程如式(4)。当输入为三输入时, 其原理与双输入相同。可如式(5),

$F_1, F_2, F_3$  分别代表同级特征图、上采样或下采样的特征图以及跨层的特征图。

$$F_{fuse} = C_{1 \times 1}(F_1 W + F_2(1 - W) + F_1 + F_2) \quad (4)$$

$$F_{fuse} = C_{1 \times 1}\left(F_1 W + \frac{F_2}{2}(1 - W) + \frac{F_3}{2}(1 - W) + F_1 + F_2 + F_3\right) \quad (5)$$

### 3) 上采样模块改进

球团矿数据通常呈现密集小目标特征, 传统的上采样方法在放大特征图时容易丢失细节信息, 难以有效捕捉边缘和纹理特征, 不利于目标的精确识别与检测。为解决此问题, 引入解耦空间-通道上采样块(DSCU)来捕捉更多关键细节特征, 生成更具表达力的特征图。该方法确保了较低的计算开销, 显著提升了模型的检测精度。DSCU 结构如图 6 所示。

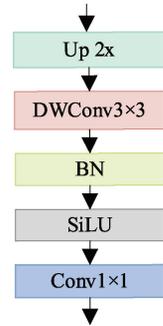


图 6 DSCU 结构

Fig. 6 Architecture of DSCU

DSCU 上采样块首先对特征图进行比例因子为 2 的上采样(UP)放大特征图, 然后通过  $3 \times 3$  深度卷积(DWConv)、批量归一化(BN)及激活函数增强放大的特征图(选择具有平滑激活特性、生成的特征图更具表达力的 SiLU 激活函数), 最后使用  $1 \times 1$  卷积(Conv)减少通道数, 确保更好的匹配下一个跳跃连接的特征图尺寸和分辨率。

其中, 采用  $3 \times 3$  深度卷积将空间特征提取与通道信息融合解耦。深度卷积仅对每个输入通道进行独立的空间滤波, 避免了普通卷积的高计算成本。后续的  $1 \times 1$  卷积则负责跨通道信息融合, 以极低的计算量调整通道维度。二者结合可显著降低计算量, 同时保留表达能力。

## 2.2 Adapter 适配器

SAM 中使用 ViT(Vision Transformer)<sup>[16]</sup> 结构作为图像编码器来提取图像特征。而 ViT 编码器基于全局注意力机制, 缺乏显式的多尺度特征提取能力, 在依赖局部细节的复杂分割任务中表现受限。与此不同, SAM2 图像编码器引入了多尺度特征层次化 Hiera<sup>[17]</sup> 骨干网络, 能够提取不同尺度的特征, 浅层包含小目标的边缘细节等局部特征, 深层含有图像的全局上下文信息。这种层次化的多尺度特征表示增强了模型在复杂场景下的分割能力。

本文在 SAM2 图像编码器中引入 Adapter 微调策略,

旨在解决自然图像与球团矿图像的域差异,提高模型对球团矿分割的适应性。Adapter 模块基于低秩自适应(LoRA)思想,其通过在原始权重矩阵的基础上添加一个低秩更新项、实现参数高效微调,即  $W' = W + AB$ ,显著减少了需要更新的参数量。如图 7(a)所示,保持图像编码器原有结构不变的前提下,在 Hiera 每个多尺度块输入路径上都添加 Adapter 块,训练时冻结图像编码器中的参数,仅更新 Adapter 模块中的参数。

如图 7(b)所示,Adapter 为 Bottleneck 结构,其包含一个用于下采样的线性层(Down)、一个 GeLU 激活函数,以及一个上采样线性层(Up)。Adapter 首先通过下采样将输入特征  $X \in R^{B \times N \times D}$  从高维空间映射到低维空间。 $W_{down}$  是下采样的权重矩阵, $b_{down}$  是对应的偏置。

$$X_{down} = X \times W_{down} + b_{down} \quad (6)$$

随后,利用激活函数 GeLU 进行非线性变换。

$$X_{act} = GeLU(X_{down}) \quad (7)$$

最后,通过上采样将处理后的特征恢复到原始维度得到最终的输出特征图。这使得 Adapter 能够在高效调整模型参数的同时,保留输入特征的关键信息。

$$X_{up} = X_{act} \times W_{up} + b_{up}$$

$$X_{out} = X + X_{up} \quad (8)$$

其中, $W_{up}$  是上采样的权重矩阵, $b_{up}$  是对应偏置。

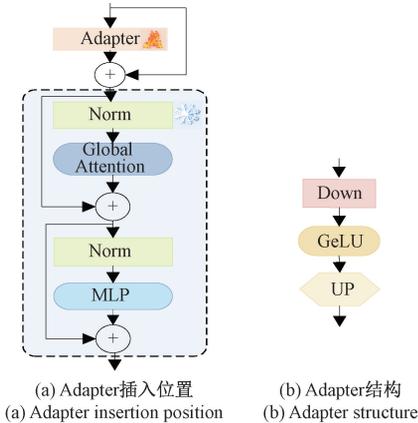


图 7 改进后的 Hiera 结构

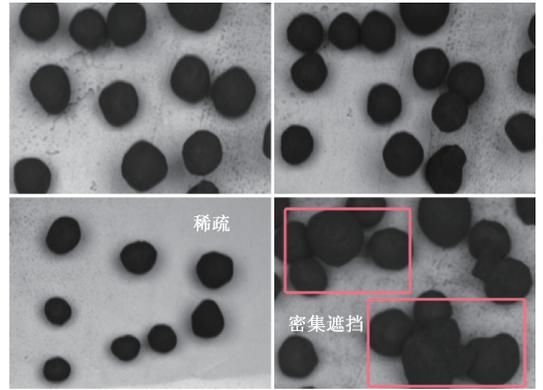
Fig. 7 Improved Hiera structure

### 3 实验结果与分析

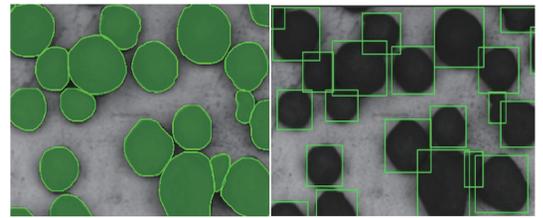
#### 3.1 实验数据集

目前关于球团矿并没有公开的数据集可供使用,将深度学习算法应用于工业球团矿似乎并不那么顺利。本文基于某钢铁公司实际生产线,收集生产车间提供的造球机出球区不同工况下的图像,构建工业球团矿数据集,该数据集共包括 468 张球团矿图像,其中包含了球团矿小目标稀疏、密集及被遮挡的图像,如图 8(a)所示。按照 8:1:1 的比例将数据集划分为训练集、验证集和测试集。球团矿多边形掩码标注示例如图 8(b)所示。此外,在检测器的训练与优

化时需将多边形掩码标注转换为最小外接矩形框,如图 8(c)所示。



(a) 球团矿图像示例  
(a) Pelletized ore image samples



(b) 多边形标注 (b) Polygonal annotation  
(c) 框型标注 (c) Bounding box annotation

图 8 数据集图像示例

Fig. 8 Sample dataset images

#### 3.2 实验设置

实验中,操作系统选择 Ubuntu20.04,GPU 选择的是 NVIDIA\_GeForce\_RTX3090,CPU 为 Intel(R) Xeon(R) Gold 6330 CPU @ 2.00 GHz。实验模型基于 PyTorch 深度学习框架,实验环境选择 3.10 版本的 python,2.5.1 的 torch 和 12.4 的 CUDA。训练过程设置 200 epoch,batch\_size 为 16,学习率为 0.017。

#### 3.3 评价指标

精确率(P)和召回率(R)常作为衡量模型检测结果可靠性和覆盖能力的关键指标。如式(9)所示,TP 代表算法正确预测的实例数,FP 代表算法误检实例数,FN 代表算法漏检的实例数。

$$\begin{cases} P = \frac{TP}{TP + FP} \times 100\% \\ R = \frac{TP}{TP + FN} \times 100\% \end{cases} \quad (9)$$

mAP(平均精度)是反映算法整体性能的综合指标。如式(10), $AP_i$  表示第 i 类的平均精度,N 为类别总数。

$$mAP = \frac{1}{N} \sum_{i=1}^N AP_i \times 100\% \quad (10)$$

本文在评价 YOLO 系列模型用于分割时,模型的 mAP 普遍呈现虚高现象,实际应用中并不能反映真实分割精度。而 IoU(交并比)常用于衡量模型生成的分割掩码与真实标注之间的重叠程度,mIoU(平均交并比)是逐类计算

$IoU$  后取平均得到的,值越大表示模型的分割结果越准确。

$$mIoU = \frac{1}{N} \sum_{i=1}^N \frac{TP}{TP + FP + FN} \quad (11)$$

所以,本文将综合  $mIoU$  与  $mAP$  两个指标来衡量算法真实精度,并利用  $P$ (精确度)、 $R$ (召回率)、 $Parameters$ (参数量)、 $GFLOPs$  衡量算法综合性能。

### 3.4 对比实验

#### 1) 金字塔网络对比实验

如表 1 所示,将当前主流金字塔网络 BiFPN 与 AFPN 与本文 SHFPN 应用在 YOLOv11 架构上进行对比实验,验证其有效性。实验结果表明,与其他金字塔网络相比,SHFPN 在保持较低参数量与计算复杂度的前提下,精确率和平均精度等性能指标均表现优异。SHFPN 通过各结构网络的设计,有效提升了模型提取关键信息及特征融合能力,显著增强了网络对小目标特征的识别能力。

#### 2) 目标检测算法对比实验

为验证本文 SH\_YOLO 算法的有效性,将其与多种先进检测算法进行对比。实验结果如表 2 所示。

表 2 不同检测方法对比实验

Table 2 Comparative experiment of different detection methods

算法	$P/\%$	$R/\%$	$mAP/\%$	$mAP50/\%$	$Param/M$	$GFLOPs$
Faster-R-CNN	76.7	77.0	75.2	76.6	63.4	369.7
YOLOX	69.8	72.9	71.3	68.5	5.1	6.5
YOLOv8n	78.3	79.1	77.9	77.4	3.0	8.2
YOLOv9n	80.7	79.8	80.3	78.8	7.3	26.8
YOLOv10n	81.6	80.5	79.8	78.9	2.7	8.4
YOLOv11	82.5	81.3	80.5	80.3	2.6	6.4
SH_YOLO	<b>85.7</b>	<b>85.9</b>	<b>84.4</b>	<b>82.5</b>	<b>2.5</b>	<b>6.4</b>

#### 3) 实例分割算法对比实验

为评估 YO-SAM2 的前沿性,更直观的评估算法的有效性,将其与多种主流实例分割算法在球团矿数据集上进行了综合对比。如表 3 所示,本文将 SH\_YOLO 联合专门定制的 SAM2 在各项指标上均表现优异:算法  $mIoU$  达到 90.3%,精确度( $P$ )和召回率( $R$ )分别为 94.6%与 87.8%,与单阶段分割模型 SOLOv2 与 YOLACT 相比,本算法的  $mIoU$  分别显著提升了 16.7%与 21.3%。与 Mask R-CNN 此类较高精度的算法相比,本文算法表现更佳, $mIoU$  领先了 14.2%。此外,相较于广受欢迎的 YOLOv8-seg,本算法  $mIoU$  也有 18.5%的提升,虽然大模型的叠加增加了算法的计算开销,但其实现了其他算法无法替代的分割效果,如图 9 所示的可视化分割效果进一步印证了本文算法的显著优势。

具体而言,在球团矿生产过程中,初始原料混合阶段常会在造球机设备添加一定量的水,以达到最佳的湿度水平帮助粘结剂充分发挥作用,因此造球现场有时会形成雾气。图 9 中第 1、2、3 行为无水雾图像,第 4、5 行为有水雾

实验数据显示,SH\_YOLO 在多个评价指标上表现出色, $mAP$  与  $mAP50$  分别达到 84.4%、82.5%,相较于原始 YOLOv11,算法的平均精度提升了 3.9%,召回率提升了 4.6%,显著降低了检测的误检率与漏检率。而相较于 YOLOv9n 与 YOLOv10n,改进后模型的平均精度分别提升了 4.1%和 4.6%,模型参数和计算量也进一步减少,模型性能显著优于其他 YOLO 系列模型。综合来看,SH\_YOLO 实现了模型精度、参数和计算量间的有效平衡且优于大多数算法,能够为分割阶段生成准确的提示信息,满足实际目标检测需求。

表 1 金字塔网络对比实验

Table 1 Pyramid network comparison experiment

网络	$P/\%$	$mAP/\%$	$Param/M$	$GFLOPs$
PAFPN(baseline)	82.5	80.5	2.6	6.4
BiFPN	78.4	80.3	2.7	6.4
AFPN	84.8	79.7	3.1	7.3
SHFPN	<b>86.6</b>	<b>82.0</b>	<b>2.5</b>	<b>6.4</b>

图像。图 9 展示了无雾条件下的分割效果图,传统的实例分割算法主要存在以下问题:Mask R-CNN 漏分割现象严重;SOLOv2 在漏分割的基础上,生球边缘分割不规范,呈现齿轮状;YOLACT、YOLOv8-seg 及 YOLOv11-seg 对生球的识别率较高,但是边缘同样存在严重的齿轮状现象,以上问题会直接影响后续粒径分布统计的准确性。相比之下,本文算法依托 SH\_YOLO 在检测阶段提供的精准位置提示,有效克服了上述主流分割算法的问题,不仅实现了对球团矿的准确识别,在复杂的图像环境中仍保持了高分割精度,展现出较强的鲁棒性。

综上所述,改进后的模型在球团矿分割任务上显著优于其他主流算法,展现出更高的分割准确率与更精细的边缘分割效果,为后续的粒径分析奠定了可靠基础。尽管模型复杂度因引入大型模型而有所增加,但实现了精度与复杂度的较优平衡,能够满足球团矿实际生产检测的要求,对生产调控影响可控。以可接受的计算代价换取显著的精度与效果提升是合理且有效的。

表3 不同分割方法对比实验

Table 3 Comparative experiment of different segmentation methods

算法	$mIoU$	$P/\%$	$R/\%$	$mAP/\%$	$mAP50/\%$	Param/M	$GFLOPs$
Mask R-CNN	76.1	79.3	76.8	82.0	84.1	6.3	23.5
SOLOv2	73.6	75.3	78.6	73.0	83.0	4.7	17.9
YOACT	69.0	99.3	99.5	99.5	99.5	5.4	20.6
YOLOv8n-seg	71.8	99.4	99.7	99.7	99.4	3.3	12.1
YOLOv11-seg	75.4	99.8	99.6	99.5	99.5	2.9	10.4
YO-SAM2 (ours)	<b>90.3</b>	<b>94.6</b>	<b>87.8</b>	<b>92.3</b>	<b>90.6</b>	40.4	105.3

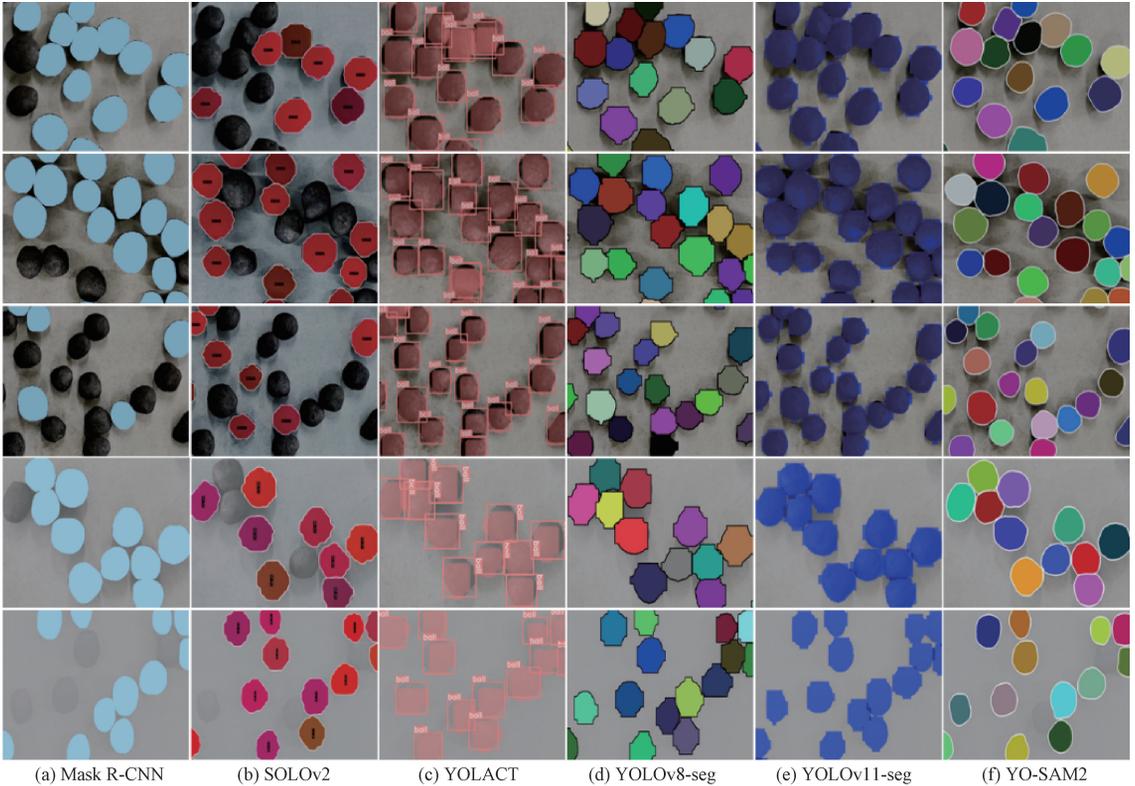


图9 不同模型分割效果对比

Fig. 9 Comparison of segmentation effects of different models

### 3.5 消融实验

为验证所提出改进方法的有效性,本文分别对基线模型YOLOv11与SAM2在球团矿数据集上进行消融实验并对其在验证集上的性能指标进行对比。具体操作为:逐步引入不同改进方法,依次评估这些改进对模型整体性能的影响。实验结果如表4、5所示。

由表4可知,各个改进模块对算法性能均有显著提升:C3k2\_CSC模块的应用使模型 $mAP$ 达到81.7%,提升了1.2%,这表明CSC模块能动态调控感受野范围,优化特征响应,增强了对局部细节的敏感度与上下文信息的有效关联。接着,SHFPN金字塔网络的进一步引入使 $mAP$ 较原来提升了3.4%,精准度与召回率分别提升2.5%和4.2%,这表明SHFPN能通过跨层特征增强机制与CFusion充分

利用细节信息,增强异构信息的融合能力,有利于密集小目标的准确识别。最终,再将DSCU上采样模块代替原始上采样,并与其他改进方法综合使用时,改进后的模型平均精度较基线模型提高了3.9%,同时模型参数量和计算复杂度也有略有下降,算法达到最佳综合性能。

在缺乏有效提示以及图像域差异带来的影响下,会导致SAM2前景与背景混淆,分割效果不佳,在球团矿实例分割任务上无法产生有意义的结果,表5的 $mIoU$ 数值也确认了这一点。而将Adapter添加到图像编码器中Hiera每个多尺度块输入路径上,仅需训练少量参数,便能较好的解决此问题,使模型的 $mIoU$ 达到70.2%,增强了模型在下游任务的适应性,分割效果如图10所示。定制后的SAM2算法能在接收到box提示后更加准确的分割出球团矿目标,为进一步粒径分析提供便捷。

表 4 SH\_YOLO 不同模块的模型性能对比

Table 4 Model performance comparison of different modules of SH\_YOLO

实验	C3k2_CSC	SHFPN	DSCU	$P/\%$	$R/\%$	$mAP/\%$	Param/M	GFLOPs
1	×	×	×	82.5	81.3	80.5	2.6	6.4
2	√	×	×	83.6	83.3	81.7	2.6	6.4
3	√	√	×	85.0	85.5	83.9	2.6	6.5
4	√	√	√	<b>85.7</b>	<b>85.9</b>	<b>84.4</b>	<b>2.5</b>	<b>6.4</b>

表 5 Adapter 的有效性

Table 5 The effectiveness of the adapter

算法	Adapter	$mIoU/\%$
SAM2	×	6.9
定制 SAM2	√	70.2

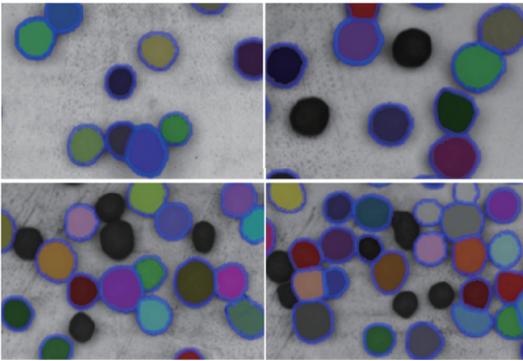
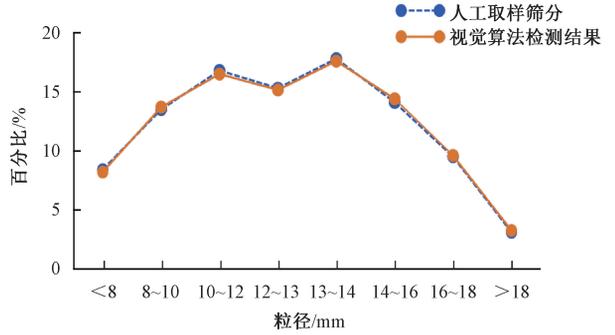
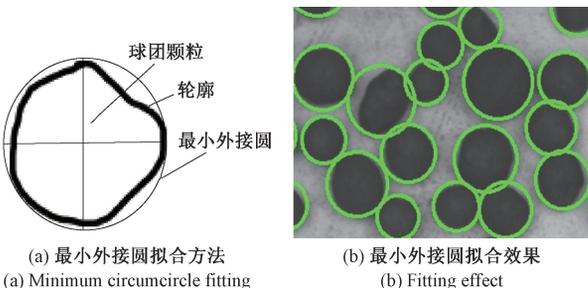


图 10 SAM2 定制后分割效果图

Fig. 10 segmentation effects of customized before and after

### 3.6 粒径分布统计

本文基于所提出算法的分割结果提取生球轮廓,对生球颗粒进行最小外接圆拟合操作,最小外接圆的直径便是生球直径。最小外接圆拟合方法如图 11(a)所示。图 11(b)展示了球团矿样本最小外接圆拟合效果,该方法能有效补偿因遮挡或重叠造成的轮廓缺失。此外,进一步将不同图像粒径分布统计与人工取样筛分结果进行比较并计算各粒径范围生球个数占总个数的百分比(图 11(c)),可以得出,本文视觉算法检测结果与人工筛分相对误差均值为 3.9%,这也进一步证明了本文提出分割算法的有效性,为工业现场在线粒度检测提供了可靠的技术方案。粒径分布统计结果如图 11(c) 所示。



(c) 粒径分布统计结果

(c) Particle size distribution statistics

图 11 圆拟合与粒径分布

Fig. 11 Circle fitting and particle size distribution

## 4 结 论

为解决球团矿在工业生产场景中存在的检测与分割难题,本文提出一种联合改进目标检测算法与定制分割模型的实例分割方法,为球团矿粒径检测及自动化控制提供前提和基础。一方面,优化 YOLOv11 网络结构,设计 C3k2\_CSC 模块以增强小尺度密集目标的特征表达能力,构建 SHFPN 特征金字塔网络以提升全局特征提取与多尺度特征的融合效能,并引入 DSCU 模块缓解上采样过程中的细节信息丢失问题,最终提出的 SH\_YOLO 模型对球团矿的检测性能显著提升。另一方面,引入适配器微调策略,有效解决了 SAM2 模型在球团矿图像域的适应性问题,实现了工业球团矿的精准分割。此外,将复杂的分割任务转化为检测任务,大大降低了数据标注与训练难度。最后,通过多个实验验证了 YO-SAM2 在多个性能指标上表现优异,后续通过最小外接圆拟合能够有效统计球团矿的粒径分布情况,有较强的实际应用价值。下一步将致力于模型轻量化部署方案,降低计算开销,以求在边缘设备上实现更高效的部署与应用。

## 参考文献

[1] 周淑伊, 刘小燕, 陈玉如, 等. 基于可调证据滤波器的球团矿生球图像裂缝检测方法[J]. 电子测量与仪器学报, 2022, 36(7): 125-135.  
ZHOU SH Y, LIU X Y, CHEN Y R, et al. Crack detection for iron ore green pellet image based on steerable evidence filter [J]. Journal of Electronic

- Measurement and Instrumentation, 2022, 36 (7): 125-135.
- [2] 李振铠, 高冰, 贾彬, 等. 粒度对熔剂性球团冶金性能及矿相结构的影响[J]. 烧结球团, 2024, 49(2):59-64. LI ZH K, GAO B, JIA B, et al. Effect of particle size on metallurgical properties and mineral phase structure of fluxed pellets[J]. Sintering and Pelletizing, 2024, 49(2):59-64.
- [3] 陈许玲, 吕浩, 甘敏, 等. 焦粉粒度对烧结混合料制粒的影响[J]. 钢铁研究学报, 2022, 34(4):309-316. CHEN X L, LYU H, GAN M, et al. Effect of coke granularity on granulation of sintered mixture [J]. Journal of Iron and Steel Research, 2022, 34 (4): 309-316.
- [4] NANDA A, PATRA A K, PANIGRAHI S. Pellet sphericity and size analysis by image processing technique [C]. International Conference on Applied Machine Learning. Bhubaneswar, 2019: 222-227.
- [5] 孙浠汉. 球团矿生球粒径智能检测系统研究[D]. 长沙:湖南大学, 2022. SUN X H. Intelligent particle size inspection system for iron ore green pellets [D]. Changsha: Hunan University, 2022.
- [6] WU X, LIU X, DUAN J. Online size distribution measurement of dense iron green pellets using an efficient and multiscale nested U-net method [J]. Powder Technology, 2021, 387:584-600.
- [7] DUAN J, LIU X, WU X, et al. Detection and segmentation of iron ore green pellets in images using lightweight U-net deep learning network [J]. Neural Computing and Applications, 2019, 32(10):5775-5790.
- [8] KIRILLOV A, MINTUN E, RAVI N, et al. Segment anything [C]. IEEE/CVF International Conference on Computer Vision, 2023: 4015-4026.
- [9] CHEN K Y, LIU CH Y, CHEN H, et al. RSPrompter: Learning to prompt for remote sensing instance segmentation based on visual foundation model [J]. IEEE Transactions on Geoscience and Remote Sensing, 2024:62: 1-17.
- [10] ZHU J, QI Y, WU J. Medical sam 2: Segment medical images as video via segment anything model 2[J]. ArXiv preprint arXiv:2408.00874, 2024.
- [11] 吴葛, 朱宇凡, 叶天成. 改进 YOLOv10 的钢材表面缺陷检测方法 [J]. 电子测量技术, 2025, 48 (4): 158-168. WU G, ZHU Y F, YE T CH. Improving the steel surface defect detection method of YOLOv10 [J]. Electronic Measurement Technology, 2025, 48 (4): 158-168.
- [12] KHANAM R, HUSSAIN M. YOLOv11: An overview of the key architectural enhancements [J]. ArXiv preprint arXiv:2410.17725, 2024.
- [13] RAVI N, GANEUR V, HU Y T, et al. Sam 2: Segment anything in images and videos [J]. ArXiv preprint arXiv:2408.00714, 2024.
- [14] WU T, TANG S, ZHANG R, et al. Cgnet: A lightweight context guided network for semantic segmentation [J]. IEEE Transactions on Image Processing, 2020, 30: 1169-1179.
- [15] CHEN Z X, HE Z W, LU ZH M. DEA-Net: Single image dehazing based on detail-enhanced convolution and content-guided attention [J]. IEEE Transactions on Image Processing, 2024, 33: 1002-1015.
- [16] DOSOVITSKIY A, BEYER L, KOLESNIKOV A, et al. An image is worth 16 × 16 words: Transformers for image recognition at scale [J]. ArXiv preprint arXiv:2010.11929, 2020.
- [17] RYALI C, HU Y T, BOLYA D, et al. Hiera: A hierarchical vision transformer without the bells-and-whistles [C]. International Conference on Machine Learning, PMLR, 2023: 29441-29454.

## 作者简介

**龚玉洁**, 硕士研究生, 主要研究方向为目标检测、图像分割。

E-mail: gongyujie0303@163.com

**王旭**, 硕士研究生, 主要研究方向为目标检测。

E-mail: 17860273050@163.com

**郭海杰**, 硕士研究生, 主要研究方向为图像分割。

E-mail: 15890354319@163.com

**丁志星**, 硕士研究生, 主要研究方向为目标检测。

E-mail: 1410452670@qq.com

**崔雪红**(通信作者), 博士, 副教授, 主要研究方向为人工智能、计算机视觉。

E-mail: 19853226869@163.com