

文章编号 1004-924X(2026)02-0309-13

## 基于并联双注意力的轻量级小样本矿石粒度检测

孙国栋<sup>1,2\*</sup>, 刘明轩<sup>1</sup>, 李仕成<sup>4</sup>, 吴波<sup>3\*</sup>

(1. 湖北工业大学 机械工程学院, 湖北 武汉 430068;

2. 湖北工业大学 现代制造质量工程湖北省重点实验室, 湖北 武汉 430068;

3. 中国科学院上海高等研究院, 上海 201210;

4. 湖北工业大学 底特律绿色工业学院, 湖北 武汉 430068)

**摘要:**针对传统目标检测方法在矿石粒度检测中存在的计算复杂度高、特征鲁棒性差及分类器性能受限等问题,本文提出一种小样本目标检测算法,旨在降低标注与计算成本,并提升模型在数据稀缺场景下的泛化能力。设计以 CenterNet2 为框架,采用轻量级 VoVNet 作为主干网络以保证检测速度;核心创新是设计了一个并联双注意力特征融合模块,其中通道交叉注意力模块用于重校准通道维度特征,空间分组注意力模块聚焦目标关键区域,二者协同增强判别性特征融合能力,从而精准指导查询图像检测。在矿石数据集上进行测试,所提模型平均精度(AP)达 55.2%, AP50 与 AP75 分别为 78.5% 和 66.9%,推理速度达 57 Frame/s(FPS),注意力模块参数量仅为 16.1 兆字节(MB),体现出优异的精度-效率均衡性。实验表明该方法能有效提升小样本矿石粒度检测的感知性能,且具备极高的边缘端落地潜力,为解决智能矿山中算力受限条件下的实时检测难题提供了可靠的技术方案。

**关键词:**计算机视觉;小样本目标检测;轻量化;矿石图像;实时检测

**中图分类号:** TP394.1; TH691.9 **文献标识码:** A

**doi:** 10.37188/OPE.20263402.0309

**CSTR:** 32169.14.OPE.20263402.0309

## A lightweight few-shot ore detector with parallel dual attention

SUN Guodong<sup>1,2\*</sup>, LIU Mingxuan<sup>1</sup>, LI Shicheng<sup>4</sup>, WU Bo<sup>3\*</sup>

(1. College of Mechanical Engineering, Hubei University of Technology, Wuhan 430068, China;

2. Hubei Key Laboratory of Modern Manufacturing Quality Engineering, Hubei University of Technology, Wuhan 430068, China;

3. Shanghai Advanced Research Institute, Chinese Academy of Sciences, Shanghai 201210, China;

4. Detroit Green Technology Institute, Hubei University of Technology, Wuhan 430068, China)

\* Corresponding author, E-mail: sgdeagle@163.com; wubo@sari.ac.cn

**Abstract:** To address the high computational complexity, limited feature robustness, and constrained classifier performance of conventional object detection methods in ore particle size detection, a few-shot object detection approach was proposed to reduce annotation cost and improve generalization under data-scarce conditions. The proposed method was built upon the CenterNet2 framework and employed a lightweight VoVNet as the backbone to ensure detection efficiency. A parallel dual-attention feature fusion module

收稿日期: 2025-10-11; 修订日期: 2025-12-08.

基金项目: 国家自然科学基金(No. 51775177); 湖北省揭榜制科技项目(No. 2024BEB018)

was designed as the core component. Specifically, a channel cross-attention module was introduced to recalibrate channel-wise feature responses, while a spatial group-attention module emphasized discriminative target regions. The coordinated operation of the two modules enhanced the fusion of task-relevant features and provided effective guidance for query image detection in few-shot scenarios. Experimental results on an ore dataset show that the proposed model achieved an average precision ( $AP$ ) of 55.2%, with  $AP_{50}$  and  $AP_{75}$  reaching 78.5% and 66.9%, respectively. The inference speed reached 57 frames per second (FPS), while the attention module required only 16.1 M parameters, indicating a favorable trade-off between accuracy and efficiency. Experimental results demonstrate that the proposed method effectively enhances the perception performance of few-shot ore particle size detection. Moreover, it possesses high potential for edge deployment, providing a reliable technical solution for real-time detection challenges in smart mines under computation-constrained conditions.

**Key words:** computer vision; few-shot object detection; lightweight; ore images; real-time

## 1 引 言

随着科技的进步和工业化进程的加速,对矿石资源的需求日益增长,高效、准确地识别矿石粒度大小成为提高开采效率、优化资源利用的关键环节。然而,矿石检测面临着诸多技术挑战,尤其是在复杂多变的自然环境中,矿石的形态、颜色、纹理等特征易受光照、遮挡、杂质等多种因素影响,导致检测难度增加。

近年来,深度学习技术在矿石粒度检测领域得到广泛应用。U-Net被首次引入用于破碎石块的分割<sup>[1]</sup>,验证了该网络模型的有效性;随后,有研究<sup>[2]</sup>通过改进损失函数与分水岭算法,减轻了欠分割问题;Tang等人<sup>[3]</sup>利用注意力机制和跨尺度嵌入,在提升精度与速度的同时缓解了矿石粘连问题;Sun等人<sup>[4]</sup>提出一种轻量级多层感知器框架,通过优化特征和损失函数抑制了边缘毛刺现象。有研究通过构建适用于X射线的矿石图像检测模型,提升矿石分类检测效率<sup>[5]</sup>;SaNet架构用于提升粗粒土级配识别效率<sup>[6]</sup>。然而,在矿石粒度检测任务中,与上述语义分割方法相比,目标检测法标注成本低、能有效分离粘连矿石,因此更适合用来估算矿石面积,并以此来判断矿石是否属于大块,从而解决仓内的堵料问题。Sun等人<sup>[7]</sup>提出一种轻量高效的小样本矿石粒度检测器,在减少样本量需求的同时加快检测速度,展现目标检测方法在此任务上的应用潜力,但其性能仍有大量的挖掘空间。

传统的目标检测方法可分为一阶段检测器和

两阶段检测器。一阶段检测器具有速度优势。YOLOv5<sup>[8]</sup>凭借其自适应锚框计算和更精细的数据增强策略,实现速度与精度的平衡;YOLOv7<sup>[9]</sup>通过引入可训练的策略,在不增加推理成本的情况下优化训练过程,进一步提升了模型的精度潜力。相比之下,两阶段检测器相对复杂,但DETR等系列的出现带来了新的突破。DINO<sup>[10]</sup>通过引入对比去噪训练和先进的先验知识,在精度和训练效率上实现了巨大飞跃;RT-DETR<sup>[11]</sup>更是专注于实时性,通过高效的混合编码器和不确定性最小查询,在保持高精度的同时实现了实时推理;UniAD<sup>[12]</sup>提出了一个以检测为核心的统一框架,通过任务间更优的协同设计,在多个下游任务上取得了领先性能;InternImage<sup>[13]</sup>通过引入可变形卷积与核心设计,在大规模视觉任务上超越了Swin Transformer等,提供了更强大的特征表示能力。

尽管上述目标检测方法在各自领域表现出色,但它们通常依赖于大量高质量标注样本以充分训练模型。然而,在矿石粒度检测这一特定工业场景中,获取大量精准标注的图像数据既昂贵又耗时,导致样本稀缺成为检测主要瓶颈。在此背景下,小样本目标检测技术(Few-Shot Object Detection, FSOD)提供了新的思路,其旨在通过少量的标注样本的训练来对图像中的目标进行分类和定位,为样本稀缺场景下的矿石检测提供了新的解决方案。Han等人<sup>[14]</sup>提出了一种原型匹配网络,通过联合优化小样本提案生成与细粒度分类,在多个基准上取得优异性能;另一种元学习框架通过解耦类别无关与特定组件的学习,

并采用权重预测元模型,统一处理小样本分类与定位任务<sup>[15]</sup>。Kim等人<sup>[16]</sup>结合空间推理与图卷积网络,在少量样本下实现上下文感知的新类检测;FSCE<sup>[17]</sup>通过结合监督对比学习,优化类内外差异,在多个数据集上表现领先。带注意力机制的区域提议网络(Region Proposal Network, RPN)与多关系检测器通过对比训练策略抑制背景干扰,实现了无需微调的新类别泛化<sup>[18]</sup>。ECEA<sup>[19]</sup>提出了可扩展共存注意力模块,解决因训练样本稀缺导致的目标检测不全问题。

然而,现有FSOD研究多集中于自然图像,其背景相对简单、目标特征鲜明。相比之下,矿石图像具有纹理复杂、目标密集、类间差异小等特点,且由于遮挡与粘连现象普遍,目标特征往往表达不完整,导致现有方法在矿石图像上存在显著的领域差距,难以直接迁移应用。事实上,这种高密度的非结构化群目标检测一直是小样本领域的难点。和萌等<sup>[20]</sup>在研究中发现,面对此类特征模糊且分布密集的群目标,单纯的小样本学习难以奏效,必须融合特征重组与特定的注意力机制来增强模型对微弱特征的捕获能力。矿石粒度检测面临着与之一致的挑战,亟需针对性的特征增强方案。

在技术架构方面,主流小样本检测算法主要基于Faster RCNN<sup>[21]</sup>,其两阶段检测流程计算量大,难以满足实时检测需求。此外,为弥补训练数据不足问题,现有算法往往集成复杂模块,进

一步降低了检测速度,增加内存消耗,限制其在资源受限环境下的实用性。针对这一工程瓶颈,朱威等<sup>[22]</sup>提出在轻量级网络中引入注意力机制的策略,证明了通过精简主干网络并强化关键特征,可以在大幅降低计算复杂度的同时保持检测精度。这一思路对于解决矿山边缘端设备的算力受限问题具有重要的参考价值。

基于对上述小样本群目标特性与工程算力瓶颈的分析,本文提出一种创新的小样本矿石目标检测方法,一方面摒弃冗余复杂的传统框架,构建更轻量的检测流程;另一方面,受任凤雷等<sup>[23]</sup>利用双注意力机制成功提取车道线细微特征的启发,本文针对矿石的不规则几何形态与相似纹理特征,设计了并联双注意力特征融合模块,通过创新特征表达与关联方法,使其在样本稀少时仍能克服遮挡干扰,实现可靠识别并快速响应复杂多变的检测场景。

## 2 方法

### 2.1 网络总体框架

本文提出一种基于CenterNet2<sup>[24]</sup>的小样本矿石检测框架,如图1所示。首先,为降低模型复杂度并提升检测速度,选用轻量而高效的VoVNet<sup>[25]</sup>作为特征提取骨干网络,对查询图像与支持图像分别提取特征。支持图像因目标位置已知,可直接通过区域对齐(RoI Align)获取目

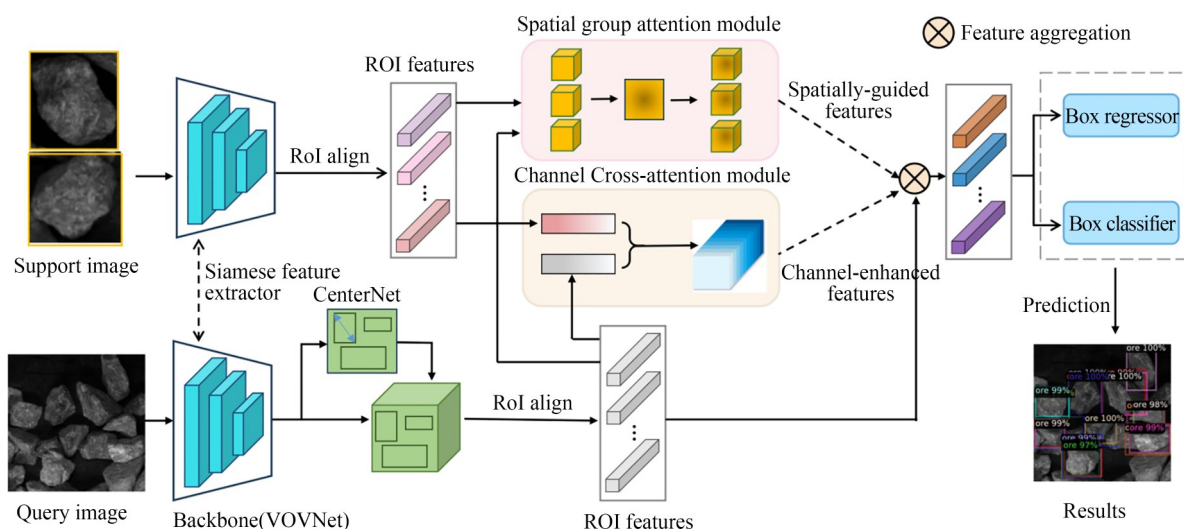


图1 小样本目标检测器整体架构图

Fig. 1 Overall framework of the few-shot object detector

标的 RoI (Region of Interest) 特征。不同于传统小样本检测中查询图像普遍依赖 RPN 生成候选框, 本方法以 CenterNet 作为候选区域生成器, 查询图像经 CenterNet 生成候选框后, 再对候选区域进行区域对齐以获取对应的 RoI 特征。接着, 为在小样本条件下实现精准匹配, 本文设计了并联双注意力特征融合模块。以支持集和查询集的 RoI 特征作为输入, 并行地计算二者在两种不同维度上的交互关系: 通道交叉注意力对特征进行通道维度的重校准, 增强判别性通道的响应, 生成通道增强的相关特征。空间分组注意力对特征进行空间分组和聚焦, 突出与支持特征关键部位相关的空间区域, 生成空间引导的相关特征。然后, 查询图像的候选框特征、通道增强特征以及空间引导特征被共同作为输入, 送入 CenterNet2 的检测头, 这三者分别提供了外观细节、通道判别性和空间先验这三个互补视角的信息, 形成一个强大且鲁棒的聚合特征表示。最后, 该聚合特征被送入最终的分类与回归预测器, 输出查询图像中每个候选框的精确类别置信度与边界框偏移量, 完成检测。

## 2.2 通道注意力模块设计

在目标检测中, 不同的特征通道往往对应不同的视觉模式或语义部分。对于小样本任务, 如何从蕴含目标类别先验信息的支持集特征中, 实现精准且高效的信息挖掘是关键。目前一些研究在 SE 模块基础上进行了改进, 在提升精度的同时却带来了较高的模型复杂度和较大的计算负担, 这不利于小样本目标检测的实际应用。基于此, 本文设计了一个轻量级的通道交叉注意力模块, 如图 2 所示, 让支持集特征来指导查询集特征的通道权重分配, 从而实现更有效的知识迁移。

该模块的核心思想是利用支持集特征的通道统计信息作为‘键’(Key), 来评估查询集特征每个通道的重要性, 并据此对查询特征进行通道维度的重校准, 从而放大与支持目标最相关的特征通道的响应。

首先, 为了获取全局表征, 分别对支持集和查询集的特征图进行全局平均池化(GAP)和全局最大池化(GMP), 并将两者的结果相加后平

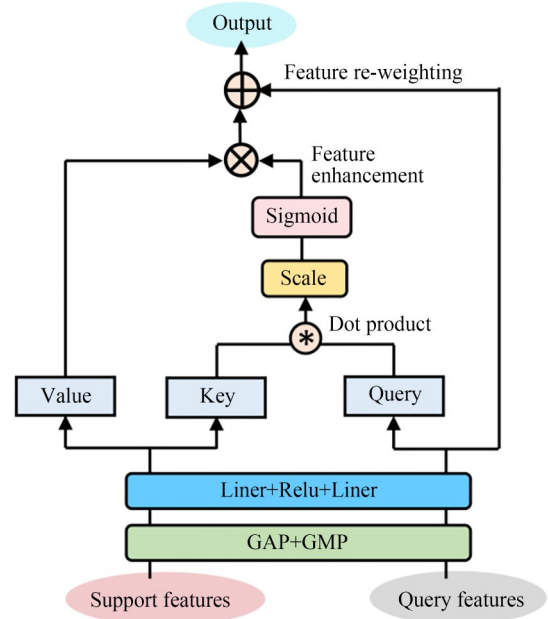


图 2 通道交叉注意力模块流程图

Fig. 2 Diagram of the channel cross-attention module

均。这种双池化策略能同时捕捉全局上下文和最显著的特征点, 形成更鲁棒的通道级描述向量。

$$z_{\text{query}} = (\text{GAP}(X_{\text{query}}) + \text{GMP}(X_{\text{query}}))/2, \quad (1)$$

$$z_{\text{support}} = (\text{GAP}(X_{\text{support}}) + \text{GMP}(X_{\text{support}}))/2, \quad (2)$$

其中:  $X_{\text{query}}$  表示查询图像的特征,  $X_{\text{support}}$  表示支持图像的特征,  $z_{\text{query}}, z_{\text{support}}$  为通道描述向量。

接着, 将这两个描述向量分别通过一个共享结构但参数独立的轻量级投影网络, 生成查询的‘提问’向量  $Q$  和支持的‘键’向量  $K$ 。

$$Q = f_{\text{proj}}^{\text{query}}(z_{\text{query}}), K = f_{\text{proj}}^{\text{support}}(z_{\text{support}}). \quad (3)$$

同时, 将支持集的描述向量直接作为‘值’向量  $V$ , 因为它本身就包含了需要传递的目标类别信息。

随后通过计算  $Q$  和  $K$  的点积, 得到一个标量相似度分数  $\alpha$ , 它代表了当前这个区域与支持目标整体的匹配置信度。然后用这个置信度对值向量  $V$  进行加权, 生成一个通道调制向量  $g$ :

$$g = \alpha \cdot V = \sigma\left(\frac{\langle Q, K \rangle}{\sqrt{C}}\right) \cdot z_{\text{support}}, \quad (4)$$

其中:  $\alpha$  为通道匹配置信度,  $\sigma$  为相似度激活函数,  $\langle Q, K \rangle$  为向量点积,  $\sqrt{C}$  为缩放因子。

最后, 将调制向量  $g$  广播为维度  $R^{B \times C \times 1 \times 1}$ ,



然后与初步的相似性特征进行逐通道乘法,完成最终的特征结合。

该设计通过一种巧妙的循环精炼结构,使相关性特征能够进行自我优化。将粗糙的初步匹配结果作为查询,主动从支持集中筛选出最相关的通道信息,从而放大判别性特征,抑制无关背景噪声,最终产生高质量、高信噪比的相关性特征图,为后续预测提供更精准的匹配信号。

### 2.3 空间注意力模块设计

由于支持特征包含了待检测目标类别的外观先验信息,直接使用该信息作为注意力引导器,可以指示查询特征应该关注哪些区域。然

而,传统的空间注意力机制通常对全局特征进行压缩,可能会丢失细粒度的分组结构信息。受SGE<sup>[26]</sup>的启发,本文将其改进为一个交叉注意力机制:将支持集特征分组后,计算其空间统计量,并生成一个分组感知的空间注意力图,作用于查询特征上。这种方法能够更精细地捕捉目标内部不同组成部分的显著性,以保障模型在有限样本下的泛化能力。

流程图如图3所示,总的来说,该模块接收从支持图像中提取出的RoI特征作为引导信号,对查询图像的RoI特征进行空间域上的重校准,从而增强与支持目标相似区域的特征响应,抑制不相关的背景噪声。

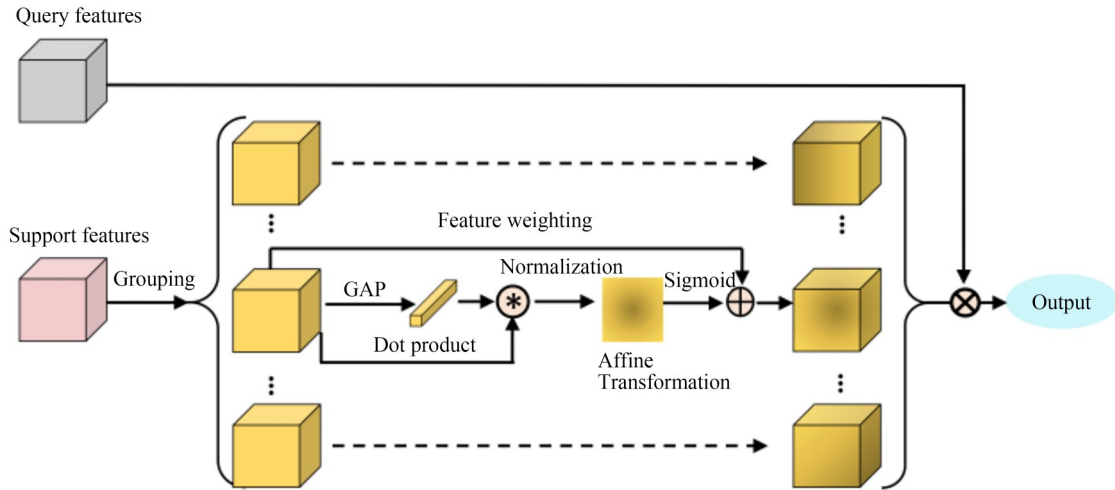


图3 空间分组注意力模块流程图

Fig. 3 Diagram of the spatial group-wise attention module

首先,将输入的支持集特征图 $X_s \in \mathbb{R}^{B \times C \times H \times W}$ 在通道维度上划分为 $G$ 个组。该设计旨在捕捉不同的特征通道组可能对应的目标不同的视觉模式或语义部分,从而增强模块的表示能力。

$$X_s' = \text{Reshape}(X_s) \rightarrow \mathbb{R}^{B \times G \times (C/G) \times H \times W}, \quad (5)$$

其中: $B$ 为批大小, $C$ 为特征通道数。 $H, W$ 为特征图的空间高度和宽度。 $G$ 为分组数量, $X_s'$ 为将支持特征在通道维度划分为 $G$ 组后的结果。

将通道维度的信息分散到分组维度上,为后续独立计算每个组的空间统计量奠定基础。

对于每个分组,计算其空间上的显著性图,以凸显每个空间位置在其所属特征组内的重要

性。由式(6)生成一个初始的组空间显著性图 $M' \in \mathbb{R}^{B \times G \times 1 \times H \times W}$ :

$$M' = \sum_{c=1}^{C/G} (X_s' \otimes \text{AdaptiveAvgPool2d}(X_s')), \quad (6)$$

其中: $\otimes$ 表示广播机制下的逐元素乘法。 $\text{AdaptiveAvgPool2d}(X_s')$ 表示对每一通道进行自适应平均池化,输出尺寸为 $\mathbb{R}^{B \times G \times (C/G) \times 1 \times 1}$ ,会沿着空间维度 $H$ 和 $W$ 广播。

求和操作沿着通道维度进行,将每个组内所有 $C/G$ 个通道的显著性响应聚合为一张代表该组整体显著性的空间图。

为了使不同样本、不同分组的注意力图具有可比性和稳定性,对其进行标准化处理,确保生

成的注意力图具有零均值和单位方差,避免梯度爆炸或消失问题,并使得注意力权重的分布更加平滑。

$$\hat{M} = \frac{M' - \mu(M')}{\sigma(M') + \epsilon}, \quad (7)$$

其中: $M'$ 为每个分组的初始空间显著性图, $\mu(M')$ , $\sigma(M')$ 是对 $M'$ 在空间维度和通道维度上按组计算的均值和标准差。 $\epsilon$ 用于避免除零。

标准化后的注意力图 $\hat{M}$ 虽然稳定,但可能丢失了其原始的分布特性。为此,对每个分组引入一组可学习的缩放参数 $\gamma \in R^{1 \times G \times 1 \times 1}$ 和偏移参数 $\beta \in R^{1 \times G \times 1 \times 1}$ 。这两个参数使得网络能够动态地恢复和调整各分组注意力图的幅度与中心,极大地增强了模块的表达力。

$$\tilde{M}^{(b,g,h,w)} = \gamma_g \cdot \hat{M}^{(b,g,h,w)} + \beta_g. \quad (8)$$

最后,该注意力掩码 $\tilde{M}^{(b,g,h,w)}$ 被直接应用于查询集特征上,通过逐元素乘法产生最终的增强特征。这一步是将支持样本所体现的空间结构模板作用于查询特征,从而在查询图像中高亮显示出所有与支持样本空间模式相似的区域,同时抑制不相关的背景区域。该模块通过构建支持样本引导的显著性空间先验,为模型在复杂背景中提供了明确的定位指引,从而显著提升目标定

位的准确性与鲁棒性。

### 3 实验与结果分析

#### 3.1 数据集与评估指标

图4是使用实验平台收集不同矿石图像,制作矿石图像数据集的过程。成像系统采用MER-U3系列CMOS工业相机垂直拍摄。相机分辨率为 $2448 \times 2048$ ,与皮带距离约300 mm,覆盖的矿石检测区域为 $365 \text{ mm} \times 365 \text{ mm}$ 。该平台搭配镜头与正面光源,确保在传送带速度为0.3 m/s的运行条件下,能稳定采集到无明显运动模糊的清晰图像,为数据采集和后续粒度分析提供了精确的物理尺寸基准。

数据集共包括了5120张不同大小和状态的矿石图像,如密集和稀疏分布,环境曝光过度和曝光不足等。研究遵循元学习范式,将检测任务构建为一系列N-way K-shot任务。由于数据集中仅包含“矿石”一个目标类别,因此任务设置为“1-way K-shot”检测。要求模型在仅有少量样本指引下,从查询图像中区分“矿石”前景与复杂背景,并完成定位。模型使用MSCOCO类别做基本训练,矿石类别用于微调训练,训练集设置为25个样本,验证集设置为1060张矿石图像。

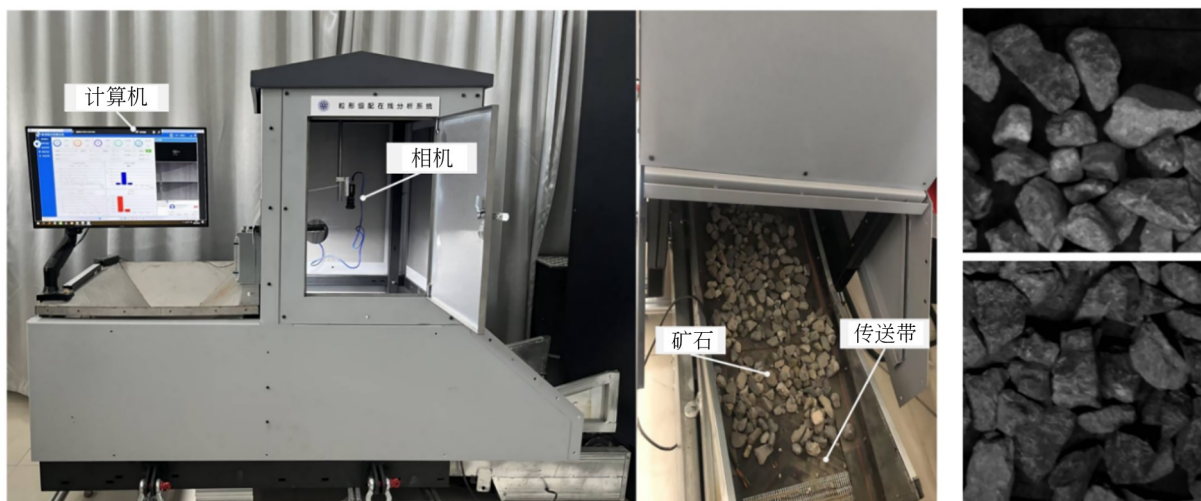


图4 矿石图像采集平台图

Fig. 4 Diagram of the ore image acquisition platform

实验采用八项指标来验证所提方法的有效性:边界框平均精度(AP)、IoU阈值为0.5时的 $AP(AP_{50})$ 、IoU阈值为0.75时的 $AP(AP_{75})$ 、中

目标 $AP(AP_m)$ 、大目标 $AP(AP_l)$ 、帧率(FPS)、推理内存消耗及模型大小。精度评估指标采用目标检测领域通用的COCO评估标准的前五项。

$AP_{\text{box}}$  表示边界框交并比(IoU)从 0.5 至 0.95(步长 0.05)计算的平均精度均值; $AP_{50}$ 与 $AP_{75}$ 则分别对应 IoU 阈值为 0.5 和 0.75 时的  $AP$  值。由于矿石目标仅包含中型和大型, $AP_m$ 和 $AP_l$ 分别表示中、大类目标——像素面积大于  $96\times 96$  的为 大目标,大于  $32\times 32$ ,小于  $96\times 96$  的为 中目标。每秒帧数(FPS)用于衡量模型计算效率。推理内存指模型在单次前向传播所占用的 GPU 显存,用于评估部署时的硬件消耗,最后两项指标对判断模型能否在低成本硬件上部署并快速响应至关重要。

3.2 实现细节

在整个实验过程中,CenterNet2 充当网络的检测框架。使用微调的方法来实现 在小样本场景下的检测,新设计的网络模块被集成到第二阶段微调流程中。在矿石数据集的微调过程中,迭代次数设置为 11 000,学习率为 0.002。使用单个 NVIDIA GTX2080Ti 图形处理器,批量大小设置为 2。为保持输入一致性,所有查询图像的短边统一缩放至 400 pixels,长边不超过 1 000 pixels;支持图像则经零填充后调整为  $320\times 320$  pixels 的统一尺寸。

3.3 消融实验

3.3.1 基于轻量级小样本检测器的消融实验

为了实现在有限的计算资源下进行高效推理,本方法选择了轻量级的 VoVNet 作为主干,如表 1 所示。

表 1 不同骨干架构对模型在矿石数据集上性能的影响  
Tab. 1 Result of different backbone

| 骨干网络     | $AP$ | $AP_{50}$ | $AP_{75}$ | FPS | 模型大小<br>/MB |
|----------|------|-----------|-----------|-----|-------------|
| Resnet50 | 56.1 | 79.8      | 68.3      | 42  | 98.8        |
| DLA      | 51.8 | 69.0      | 62.1      | 35  | 71.7        |
| VoVNet   | 55.2 | 78.5      | 66.9      | 57  | 16.1        |

与骨干网络 ResNet50 和 DLA 相比,VoVNet 在精度上相差无几,但显著提高了模型的推理速度,降低了内存和计算资源的消耗。与基线相比,所提出的模型在模型大小上减小至原来的 1/6,推理速度快了 10 帧以上,说明该模型适合资源受限的设备。本研究以 CenterNet2+VoVNet 为基准验证了设计模块的有效性。如表

2 所示,通道交叉模块与空间分组模块均能独立带来检测性能的稳定提升。单独的通道交叉模块带来了 +35.3%  $AP$  的提升,而单独的空间分组模块带来了 +46.8%  $AP$  的提升。这充分证明了在通道维度和空间维度分别引入支持集引导的注意力机制都是有效的。

表 2 设计中各模块对矿石数据集的影响  
Tab. 2 Effects of each component on the ore dataset.

| 模块     |        |        | $AP$ | $AP_{50}$ | $AP_{75}$ | FPS |
|--------|--------|--------|------|-----------|-----------|-----|
| 通道交叉模块 | 空间分组模块 | 骨干网络   |      |           |           |     |
|        |        | VoVNet | 5.3  | 11.6      | 4.2       | 55  |
| ✓      |        | VoVNet | 40.6 | 59.1      | 48.7      | 48  |
|        | ✓      | VoVNet | 52.1 | 75.3      | 63.5      | 52  |
| ✓      | ✓      | VoVNet | 55.2 | 78.5      | 66.9      | 57  |

值得注意的是,空间分组模块的提升高于通道交叉模块。分析认为,由于矿石目标的形态、尺度和空间位置变化大,因此空间位置信息相比通道信息更为关键。空间分组模块提供的精细化空间注意力能更有效地聚焦于目标区域。

当两个模块并联结合后,性能获得了最大的提升,显著超过了任一模块单独使用的效果。这表明通道注意力和空间注意力提供了互补的信息:通道交叉模块确保了正确的特征被激活,而空间分组模块确保了在正确的位置上进行聚焦。两者协同工作,共同推动了小样本矿石检测性能的进一步提升。

3.3.2 通道交叉模块和空间分组模块的有效性

为了进一步验证所提出的通道交叉注意力和空间分组注意力模块的优越性,本研究开展了一系列对比实验,将其与 SE<sup>[27]</sup>,FcaBlock<sup>[28]</sup>,GC-Block<sup>[29]</sup>,SGE<sup>[30]</sup>,PSA<sup>[31]</sup>等多种先进的注意力模块进行性能比较。所有对比模块均以相同的方式嵌入基线模型中,取代所提出的模块,并在相同的矿石小样本数据集上进行训练和测试,以确保比较的公平性。

根据表 3 的实验结果,本文方法在平均精度( $AP$ )上达到了 55.2%,高于所有单一性能模块。在  $AP_{50}$ 和 $AP_{75}$ 指标上,也分别达到了 78.5%和 66.9%,均优于对比方法。



表 3 不同注意力模块在 25-shots 设置下的性能对比

Tab. 3 25-Shots performance comparison of various attention modules

| 模块类型  | 注意力模块 | $AP$ | $AP_{50}$ | $AP_{75}$ | $AP_m$ | $AP_l$ | FPS |
|-------|-------|------|-----------|-----------|--------|--------|-----|
| 通道注意力 | SE    | 50.8 | 74.5      | 61.6      | 50.9   | 54.3   | 54  |
|       | FCA   | 54.7 | 77.7      | 66.9      | 54.7   | 59.2   | 54  |
|       | GC    | 53.2 | 76.2      | 65.1      | 53.2   | 58.0   | 54  |
| 空间注意力 | SGE   | 54.0 | 77.0      | 65.8      | 54.0   | 58.1   | 55  |
|       | PSA   | 53.8 | 76.9      | 65.7      | 53.8   | 58.0   | 51  |
| 本文方法  |       | 55.2 | 78.5      | 66.9      | 55.2   | 58.8   | 57  |

接下来,为验证所提出的并联双注意力结构的有效性。本文以其他性能优异的混合注意力模块替代所提出的并联结构,进行了性能对比。如表 4 所示,本文方法在平均精度( $AP$ ),  $AP_{50}$  和  $AP_{75}$  指标上,均优于对比方法。并且该

方法在每秒帧数(FPS)上达到了 57 帧,表明其在保持高检测精度的同时,还具有更高的计算效率。

并联双模块作用前后的效果对比可视化见图 5 所示。

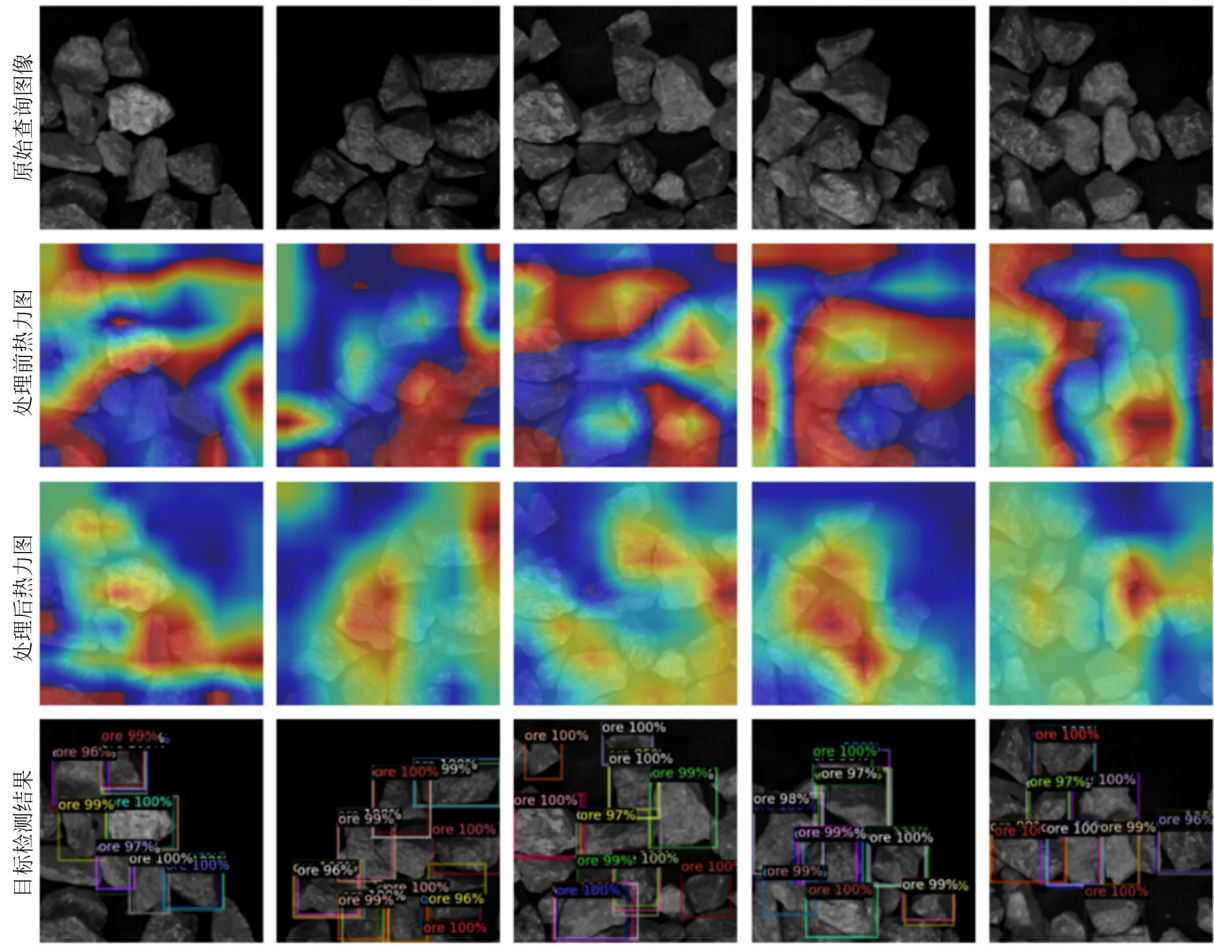


图 5 并联双模块作用前后的效果对比可视化

Fig. 5 Visualization of the effects before and after applying the parallel dual modules



表 4  并联结构在 25-shots 设置下的性能对比

Tab. 4  Performance comparison of parallel structures under the 25-Shot Setting

| 注意力<br>模块             | $AP$ | $AP_{50}$ | $AP_{75}$ | $AP_m$ | $AP_l$ | FPS |
|-----------------------|------|-----------|-----------|--------|--------|-----|
| SimAM <sup>[32]</sup> | 54.6 | 77.8      | 66.5      | 54.7   | 58.4   | 47  |
| CA <sup>[33]</sup>    | 53.1 | 75.7      | 64.8      | 53.1   | 57.1   | 52  |
| GAM <sup>[34]</sup>   | 16.6 | 27.9      | 18.2      | 16.1   | 33.6   | 39  |
| CBAM <sup>[35]</sup>  | 54.6 | 77.8      | 66.4      | 54.6   | 58.6   | 52  |
| BAM <sup>[36]</sup>   | 54.2 | 77.2      | 65.8      | 54.3   | 58.1   | 46  |
| 本文方法                  | 55.2 | 78.5      | 66.9      | 55.2   | 58.8   | 57  |

3.4  对比实验

3.4.1  与先进方法(SOTA)对比

实验通过  $AP$ ,  $AP_{50}$  和  $AP_{75}$  指标,在矿石数据集上与 TFA<sup>[37]</sup>, Meta R-CNN<sup>[38]</sup>, FsDetView<sup>[39]</sup>, MPSR<sup>[40]</sup>, FCT<sup>[41]</sup> 等多种当前最先进(State-of-the-Art, SOTA)的 FSOD 算法进行性能比较。实验训练集仅含 25 个支持样本,评估集包含 1 060 张矿石图像。

如表 5 所示,在矿石数据集上与多项 FSOD 算法进行的对比实验中,本方法取得 55.17% 的  $AP$  值,超越当前最先进的小样本目标检测算法 OreFSDet。

表 5  与 SOTA 方法在矿石数据集上的整体性能对比

Tab. 5  Overall Performance Comparison with SOTA Methods on the Ore Dataset

| 方法                        | 骨干网络        | $AP$ | $AP_{50}$ | $AP_{75}$ | FPS  | 模型大小/MB | 推理内存/MB |
|---------------------------|-------------|------|-----------|-----------|------|---------|---------|
| TFA                       | Resnet101   | 33.9 | 57.4      | 35.7      | 5.7  | 230     | 3 042   |
| FSCE                      | Resnet101   | 39.5 | 58        | 45.2      | 22.7 | 298     | 4 462   |
| FSCE                      | Resnet18    | 20.8 | 55        | 12.4      | 41.6 | 51.2    | 2 410   |
| FSCE                      | VoVNet      | 28.4 | 51.4      | 29.5      | 47.7 | 49.9    | 2 596   |
| Meta R-CNN                | Resnet50    | 22.1 | 39        | 24.1      | 26.1 | 148     | 2 422   |
| FsDetView                 | Resnet50    | 24.8 | 41.4      | 28.4      | 27.9 | 157     | 2 438   |
| MPSR                      | Resnet101   | 36.8 | 50.5      | 44.5      | 6.7  | 462     | 3 068   |
| MPSR                      | VoVNet      | 15.4 | 41.8      | 7.3       | 41.4 | 117     | 2 288   |
| Multi-RelationDet         | Resnet50    | 29.1 | 42.3      | 35.2      | 27.8 | 211     | 1 667   |
| MFR-CNN                   | Resnet101   | 27.3 | 47.3      | 29.6      | 18   | 627     | 2 303   |
| OreFSDet                  | VoVNet      | 54.1 | 78.4      | 64.7      | 50   | 19      | 1 059   |
| FCT                       | PVTv2-B2-Li | 33.6 | 55.6      | 37.9      | 10   | 93      | 1 882   |
| DCFS <sup>[42]</sup>      | Resnet101   | 39.6 | 68.9      | 41.0      | 16.1 | 334     | 2 109   |
| FS-OreDet <sup>[43]</sup> | Resnet50    | 42.5 | 71.8.     | 49.6      | 33   | 114     | 1 531   |
| BHRL <sup>[44]</sup>      | Resnet50    | 31.9 | 57.8      | 32.6      | 25   | 797     | 2 037   |
| 本文方法                      | VoVNet      | 55.2 | 78.5      | 66.9      | 57   | 16.1    | 2 463   |

在计算效率方面,该方法在 57 FPS 的实时速度下实现了最高检测精度,模型大小仅为 16.1 MB,比 OreFSDet 小 18%。推理内存消耗低于多个主流基线,说明其架构设计在内存控制上是高效的。这些结果凸显了精度与效率之间的卓越平衡。超紧凑架构与高速推理特性使该方法特别适用于计算资源受限的实时检测场景。

3.4.2  可视化

图 6 提供了该方法与 FSCE, TFA 和 OreFS-Det 在 25-shots 设置下的定性对比。如图所示, FSCE 在矿石重叠粘连严重的部分容易出现多个重叠的检测框,这对矿石粒度检测容易造成较大误差, TFA 和 OreFSDet 重复框的情况有所缓解,但在矿石密集的场景下容易出现漏检。相比之下,本文方法在漏检率和重复率上都有所改

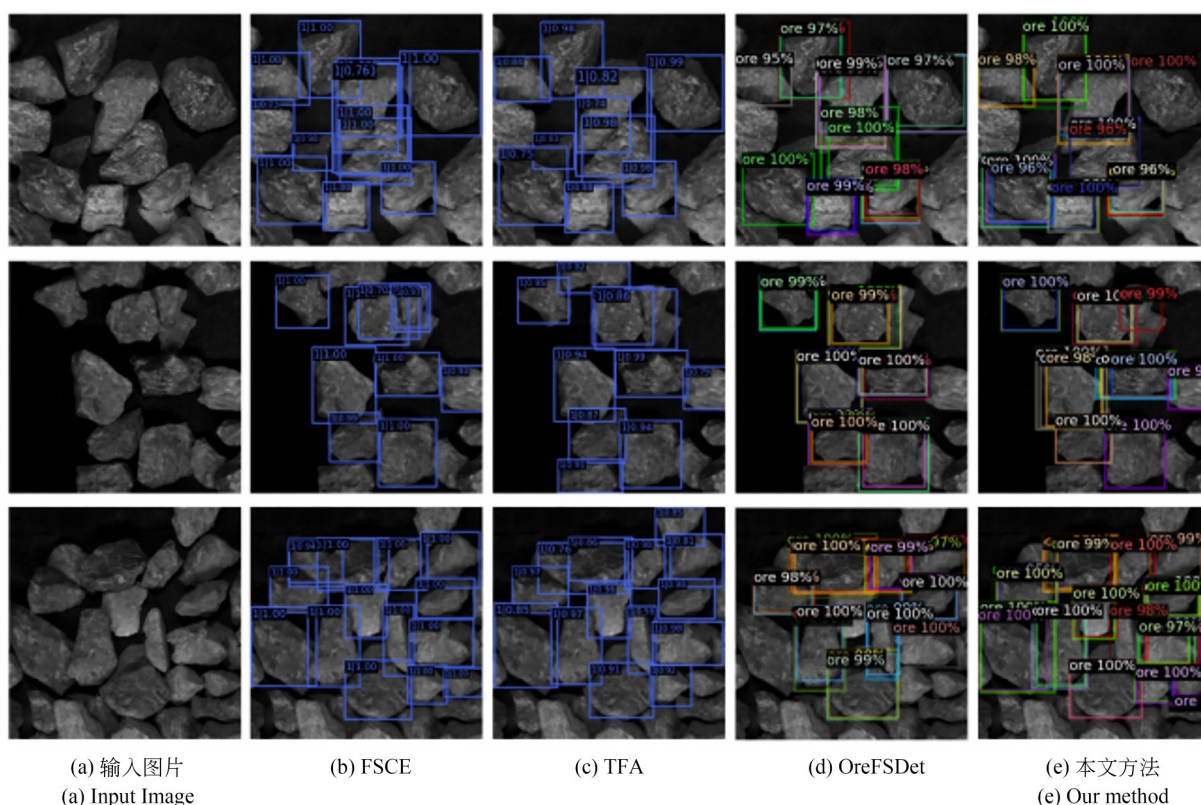


图6 25-shots设置下各方法可视化对比

Fig. 6 Visual comparison of different methods under 25-shots setting

善,得到了最干净的检测结果。

## 4 结 论

针对智能矿山矿石粒度检测中普遍存在的样本稀缺、特征粘连严重,以及传统高精度模型在端侧设备部署困难等实际工程难题,本文提出了一种基于双注意力机制的小样本矿石检测模型。核心创新在于设计了并联的通道交叉注意力模块和空间分组注意力模块,该结构能够充分利用支持样本的先验信息,在通道与空间两个互补维度上对查询特征进行精准调制与增强,实现了支持集与查询集间信息的深度交互。

实验结果表明,所提模型在矿石粒度检测数据集上达到了高检测精度,显著优于多种主流小样本检测算法。尤其是在获得高精度的同时,模型保持了优异的效率,推理速度达到 57 FPS,模

块参数量增量仅为 16.1 M,这种优异的“精度-效率”均衡性,证明了模型具备在嵌入式计算平台上实时运行的理论可行性。

尽管所提出的方法在实验室环境下展现了优异的性能,但在面对更加极端多变的工业现场时仍有优化空间。未来的工作将集中在以下两点:一是推进算法的硬件移植与现场测试,验证其在长期运行中的稳定性;二是探索结合自监督预训练技术,进一步降低模型对标注数据的依赖,提升在不同矿种场景下的泛化能力。

### 作者贡献声明:

孙国栋:方法提出、论文审核与编辑写作;  
刘明轩:实验设计、数据分析与论文撰写;  
李仕成:数据整理与分析;  
吴波:实验条件与样本采集。

## 参考文献:

- [1] 徐江川, 金国强, 朱天奕, 等. 基于深度学习U-Net模型的石块图像分割算法[J]. 工业控制计算机, 2018, 31(4): 98-99, 102.  
XU J C, JIN G Q, ZHU T Y, *et al.* Segmentation of rock images based on U-net[J]. *Industrial Control Computer*, 2018, 31(4): 98-99, 102. (in Chinese)
- [2] LI H, PAN C W, CHEN Z Y, *et al.* Ore image segmentation method based on U-Net and watershed [J]. *Computers, Materials & Continua*, 2020, 65 (1): 563-578.
- [3] TANG X X, WANG X L, YAN N, *et al.* A new ore image segmentation method based on swin-unet [C]. 2022 *China Automation Congress (CAC)*. November 25-27, 2022, Xiamen, China. IEEE, 2023: 1681-1686.
- [4] SUN G D, PENG Y T, CHENG L, *et al.* An efficient MLP-based point-guided segmentation network for ore images with ambiguous boundary [J]. *IEEE Transactions on Industrial Informatics*, 2024, 20(7): 9152-9162.
- [5] 杨文龙, 郭明钰. 轻量级注意力X射线矿石检测方法[J]. 电子测量技术, 2022, 45(18): 71-79.  
YANG W L, GUO M Y. Lightweight attention parallel X-ray ore detection algorithm[J]. *Electronic Measurement Technology*, 2022, 45 (18): 71-79. (in Chinese)
- [6] 庞元恩, 石国栋, 段煜, 等. 基于搜索分析深度学习网络(SaNet)的粗粒土级配识别[J]. 岩土工程学报, 2024, 46(9): 1984-1993.  
PANG Y E, SHI G D, DUAN Y, *et al.* Gradation recognition of coarse-grained soil based on searcher-analyzer deep learning network(SaNet)[J]. *Chinese Journal of Geotechnical Engineering*, 2024, 46(9): 1984-1993. (in Chinese)
- [7] ZHANG Y, CHENG L, PENG Y T, *et al.* Faster OreFSDet: a lightweight and effective few-shot object detector for ore images [J]. *Pattern Recognition*, 2023, 141: 109664.
- [8] KHANAM R, HUSSAIN M. What is YOLOv5: A Deep Look into The Internal Features of The Popular Object Detector [EB/OL]. 2024: *arXiv*: 2407.20892. <https://arxiv.org/abs/2407.20892>
- [9] WANG C Y, BOCHKOVSKIY A, LIAO H M. YOLOv7: trainable bag-of-freebies sets new state-of-the-art for real-time object detectors [C]. 2023 *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. June 17-24, 2023, Vancouver, BC, Canada. IEEE, 2023: 7464-7475.
- [10] ZHANG H, LI F, LIU S L, *et al.* DINO: DETR with Improved DeNoising Anchor Boxes for End-to-End Object Detection [EB/OL]. 2022: *arXiv*: 2203.03605. <https://arxiv.org/abs/2203.03605>
- [11] ZHAO Y A, LV W Y, XU S L, *et al.* DETRs beat YOLOs on Real-Time Object Detection[EB/OL]. 2023: *arXiv*: 2304.08069. <https://arxiv.org/abs/2304.08069>
- [12] HU Y H, YANG J Z, CHEN L, *et al.* Planning-oriented Autonomous Driving [EB/OL]. 2022: *arXiv*: 2212.10156. <https://arxiv.org/abs/2212.10156>
- [13] WANG W H, DAI J F, CHEN Z, *et al.* InternImage: Exploring Large-Scale Vision Foundation Models with Deformable Convolutions [EB/OL]. 2022: *arXiv*: 2211.05778. <https://arxiv.org/abs/2211.05778>
- [14] HAN G X, HUANG S Y, MA J W, *et al.* Meta faster R-CNN: towards Accurate Few-Shot Object Detection with Attentive Feature Alignment [EB/OL]. 2021: *arXiv*: 2104.07719. <https://arxiv.org/abs/2104.07719>
- [15] WANG Y X, RAMANAN D, HEBERT M. Meta-learning to detect rare objects [C]. 2019 *IEEE/CVF International Conference on Computer Vision (ICCV)*. October 27-November 2, 2019. Seoul, Korea. IEEE, 2019: 9924-9933.
- [16] KIM G, JUNG H G, LEE S W. Spatial reasoning for few-shot object detection [J]. *Pattern Recognition*, 2021, 120: 108118.
- [17] SUN B, LI B H, CAI S C, *et al.* FSCE: few-shot object detection via contrastive proposal encoding [C]. 2021 *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. June 20-25, 2021. Nashville, TN, USA. IEEE, 2021: 7348-7358.
- [18] FAN Q, ZHUO W, TANG C K, *et al.* Few-shot object detection with attention-rpn and multi-relation detector [C]. 2020 *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. June 13-19, 2020. Seattle, WA,



- USA. IEEE, 2020: 4012-4021.
- [19] XIN Z M, WU T X, CHEN S M, *et al.* ECEA: Extensible Co-Existing Attention for few-Shot Object Detection [EB/OL]. 2023: *arXiv*: 2309.08196. <https://arxiv.org/abs/2309.08196>
- [20] 和萌, 武江鹏, 梁超, 等. 融合特征重组和注意力的小样本战斗部破片群目标检测[J]. 光学精密工程, 2024, 32(12): 1929-1940.
- HE M, WU J P, LIANG C, *et al.* Few-shot warhead fragment group object detection based on feature reassembly and attention[J]. *Opt. Precision Eng.*, 2024, 32(12): 1929-1940. (in Chinese)
- [21] REN S Q, HE K M, 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*, 2017, 39(6): 1137-1149.
- [22] 朱威, 王立凯, 靳作宝, 等. 引入注意力机制的轻量级小目标检测网络[J]. 光学精密工程, 2022, 30(8): 998-1010.
- ZHU W, WANG L K, JIN Z B, *et al.* Lightweight small object detection network with attention mechanism[J]. *Opt. Precision Eng.*, 2022, 30(8): 998-1010. (in Chinese)
- [23] 任风雷, 周海波, 杨璐, 等. 基于双注意力机制的车道线检测[J]. 中国光学(中英文), 2023, 16(3): 645-653.
- REN F L, ZHOU H B, YANG L, *et al.* Lane detection based on dual attention mechanism[J]. *Chinese Optics*, 2023, 16(3): 645-653. (in Chinese)
- [24] ZHOU X Y, KOLTUN V, KRÄHENBÜHL P. Probabilistic Two-Stage Detection [EB/OL]. 2021: *arXiv*: 2103.07461. <https://arxiv.org/abs/2103.07461>
- [25] LEE Y, HWANG J W, LEE S, *et al.* An energy and GPU-computation efficient backbone network for real-time object detection [C]. 2019 *IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*. June 16-17, 2019. Long Beach, CA, USA. IEEE, 2019: 752-760.
- [26] LI X, HU X L, YANG J. Spatial Group-Wise Enhance: Improving Semantic Feature Learning in Convolutional Networks[EB/OL]. 2019: *arXiv*: 1905.09646. <https://arxiv.org/abs/1905.09646>
- [27] HU J, SHEN L, SUN G. Squeeze-and-excitation networks [C]. 2018 *IEEE/CVF Conference on Computer Vision and Pattern Recognition*. June 18-23, 2018. Salt Lake City, UT. IEEE, 2018: 7132-7141.
- [28] QIN Z Q, ZHANG P Y, WU F, *et al.* FcaNet: frequency channel attention networks [C]. 2021 *IEEE/CVF International Conference on Computer Vision (ICCV)*. October 10-17, 2021, Montreal, QC, Canada. IEEE, 2022: 763-772.
- [29] CAO Y, XU J R, LIN S, *et al.* GCNet: Non-local networks meet squeeze-excitation networks and beyond [C]. 2019 *IEEE/CVF International Conference on Computer Vision Workshop (ICCVW)*. October 27-28, 2019. Seoul, Korea. IEEE, 2019: 1971-1980.
- [30] LI Y X, LI X, YANG J. Spatial Group-Wise Enhance: Enhancing Semantic Feature Learning in CNN [M]. Computer Vision-ACCV 2022. Cham: Springer Nature Switzerland, 2023: 316-332.
- [31] LIU H J, LIU F Q, FAN X Y, *et al.* Polarized self-attention: Towards high-quality pixel-wise mapping [J]. *Neurocomputing*, 2022, 506: 158-167.
- [32] GUO M H, XU T X, LIU J J, *et al.* Attention mechanisms in computer vision: a survey[J]. *Computational Visual Media*, 2022, 8(3): 331-368.
- [33] HOU Q B, ZHOU D Q, FENG J S. Coordinate attention for efficient mobile network design [C]. 2021 *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. June 20-25, 2021, Nashville, TN, USA. IEEE, 2021: 13708-13717.
- [34] LIU Y C, SHAO Z R, HOFFMANN N. Global Attention Mechanism: Retain Information to Enhance Channel-Spatial Interactions [EB/OL]. 2021: *arXiv*: 2112.05561. <https://arxiv.org/abs/2112.05561>
- [35] WOO S, PARK J, LEE J Y, *et al.* CBAM: convolutional block attention module [C]. *Computer Vision-ECCV 2018*. Cham: Springer, 2018: 3-19.
- [36] PARK J, WOO S, LEE J Y, *et al.* BAM: Bottleneck Attention Module [EB/OL]. 2018: *arXiv*: 1807.06514. <https://arxiv.org/abs/1807.06514>
- [37] WANG X, HUANG T E, DARRELL T, *et al.* Frustratingly Simple Few-Shot Object Detection [EB/OL]. 2020: *arXiv*: 2003.06957. <https://arxiv.org/abs/2003.06957>

- [38] YAN X P, CHEN Z L, XU A N, *et al.* Meta R-CNN: towards general solver for instance-level low-shot learning [C]. 2019 *IEEE/CVF International Conference on Computer Vision (ICCV)*. October 27-November 2, 2019, Seoul, Korea. IEEE, 2020: 9576-9585.
- [39] XIAO Y, MARLET R. Few-shot object detection and viewpoint estimation for objects in the wild [C]. *Computer Vision-ECCV 2020*. Cham: Springer, 2020: 192-210.
- [40] WU J X, LIU S T, HUANG D, *et al.* Multi-scale positive sample refinement for few-shot object detection [C]. *Computer Vision-ECCV 2020*. Cham: Springer, 2020: 456-472.
- [41] HAN G X, MA J W, HUANG S Y, *et al.* Few-shot object detection with fully cross-transformer [C]. 2022 *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. June 18-24, 2022, New Orleans, LA, USA. IEEE, 2022: 5311-5320.
- [42] GAO B B, CHEN X C, HUANG Z Y, *et al.* Decoupling Classifier for Boosting Few-Shot Object Detection and Instance Segmentation [EB/OL]. 2025; *arXiv*: 2505.14239. <https://arxiv.org/abs/2505.14239>
- [43] SUN G D, CHENG L, LIU J Y, *et al.* FS-OreDet: Feature enhancement and relationship exploration for boosting few-shot object detector of ore images[J]. *Engineering Applications of Artificial Intelligence*, 2024, 133: 108437.
- [44] YANG H Q, CAI S J, SHENG H L, *et al.* Balanced and hierarchical relation learning for one-shot object detection[C]. 2022 *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. June 18-24, 2022, New Orleans, LA, USA. IEEE, 2022: 7581-7590.

## 通讯作者:



孙国栋(1981—),男,湖北天门人,博士,教授,2002年、2008年于华中科技大学分别获得学士、博士学位,主要从事机器视觉和机器学习等方面的研究。E-mail: sgdeagle@163.com

## 作者简介:



刘明轩(2000—),女,湖北武汉人,硕士研究生,2022年于湖北工业大学获得学士学位,主要从事深度学习、目标检测算法等方面的研究。E-mail: 102310158@hbut.edu.cn

## 通讯作者:



吴波(1980—),男,山东桓台人,博士,正高级工程师,2002年于华中科技大学获得学士学位、2019年于中国科学院大学获得博士学位,主要从事智能信息处理和多模态数据融合等方面的研究。Email:wubo@sari.ac.cn