

# 基于深度学习的岩石薄片矿物智能识别研究\*

孙林<sup>1,2,3</sup>, 李岩<sup>1,3</sup>, 姚旭龙<sup>2,3</sup>, 陶志刚<sup>4</sup>, 来有邦<sup>5</sup>, 曹冲<sup>6</sup>

(1.华北理工大学 人工智能学院, 河北 唐山市 063210;

2.华北理工大学 矿业工程学院, 河北 唐山市 063210;

3.河北省矿山绿色智能开采技术创新中心, 河北 唐山市 063210;

4.深部岩土力学与地下工程国家重点实验室, 北京 100083;

5.河北钢铁集团司家营研山铁矿有限公司, 河北 唐山市 063210;

6.华北理工大学 应急管理与安全工程学院, 河北 唐山市 063210)

**摘要:**随着矿物资源勘探技术的不断进步,岩石矿物的智能识别在矿物成分分析领域变得日益重要。为分析岩石薄片图像复杂纹理结构和多变矿物形态对智能识别技术的影响,提出了一种基于改进YOLOv8算法的岩石矿物智能识别模型(Mineral-YOLO模型)。Mineral-YOLO模型创新性地集成了LSK模块,以增强模型对不同目标与背景信息差异的识别能力;引入ODConv技术,以减轻背景干扰的影响,从而提升卷积网络的性能;优化了损失函数,以提升边界框定位的准确性。在模型训练时,利用组合增强技术对自建数据集进行扩展,使数据集的样本更加丰富。利用验证集对训练好的模型进行验证,结果表明,所提出的矿物智能识别模型在识别6种矿物时,平均精度均值为83.3%, $F_1$ 为78%,相较于YOLOv8模型分别提升了3个百分点和1个百分点,证明了Mineral-YOLO模型在岩石矿物智能识别中的高效性和准确性。

**关键词:**岩石薄片图像;智能识别;深度学习;Mineral-YOLO模型;目标检测

中图分类号:TD98;TP391.41

文献标识码:A

文章编号:1005-2763(2025)10-0224-12

## Research on Intelligent Identification of Rock Thin Section Minerals Based on Deep Learning

SUN Lin<sup>1,2,3</sup>, LI Yan<sup>1,3</sup>, YAO Xulong<sup>2,3</sup>, TAO Zhigang<sup>4</sup>, LAI Youbang<sup>5</sup>, CAO Chong<sup>6</sup>

(1.College of Artificial Intelligence, North China University of Science and Technology, Tangshan, Hebei 063210, China;

2.College of Mining Engineering, North China University of Science and Technology, Tangshan, Hebei 063210, China;

3.Mine Green Intelligent Mining Technology Innovation Center of Hebei Province, Tangshan, Hebei 063210, China;

4.State Key Laboratory of Deep Geotechnical Mechanics and Underground Engineering, Beijing 100083, China;

5.Sijiaying Yanshan Iron Mine Co., Ltd., Hebei Iron and Steel Group, Tangshan, Hebei 063210, China;

6.School of Emergency Management and Safety Engineering,

North China University of Science and Technology, Tangshan, Hebei 063210, China)

**Abstract:** With the continuous progress of mineral resources exploration technology, the intelligent identification of rock minerals has become increasingly important in the field of mineral composition analysis. In order to analyze the influence of complex texture structure and variable mineral morphology of rock thin section images on intelligent identification technology, an intelligent identification model of rock minerals based on improved YOLOv8 algorithm (Mineral-YOLO model) was proposed. The Mineral-YOLO model innovatively integrates the LSK module to enhance the identification capacity of the model for different target and background information differences. The ODConv technology is introduced to reduce the influence of background interference, thereby improving the performance of the convolutional network. The loss function is optimized to improve the accuracy mAP of bounding box positioning. In the model training, the self-built data set was extended using the combination enhancement technology, so that the samples of the data set were more abundant. The validation set was used to verify the trained model. The results show that the mean average accuracy of the proposed mineral intelligent identification

\* 收稿日期:2024-10-31

基金项目:国家自然科学基金项目(52074123)

作者简介:孙林(1981—),男,河北唐山人,博士,副教授,主要从事数字信号采集、处理相关研究工作。E-mail:sunlin@ncst.edu.cn

通信作者:姚旭龙(1983—),男,河北张家口人,博士,副教授,主要从事智能矿山、灾害智能预警研究。E-mail:yaoxulong@ncst.edu.cn

model is 83.3% and  $F_1$  is 78% when identifying 6 kinds of minerals. Compared with the YOLOv8 model, it is increased by 3 percentage points and 1 percentage points respectively, which proves the high efficiency and accuracy of the Mineral-YOLO model in the intelligent identification of rock minerals.

**Key words:** Rock thin section image, Intelligent identification, Deep learning, Mineral-YOLO model, Object detection

## 0 引言

在地质勘探、矿产资源开发和环境监测等领域,矿物成分的识别工作至关重要,它不仅为科学研究提供了基础数据和信息,而且对经济的可持续性发展具有深远的影响。然而,传统的矿物成分识别方法依赖于专业人员的鉴定,对鉴定者的专业知识要求较高,随着矿物种类的增加和复杂性的提高,这些方法在效率和准确性方面的局限性日益凸显<sup>[1-3]</sup>。

深度学习技术在计算机视觉领域的应用取得了显著进展,尤其是在岩矿分类方面,王李管等<sup>[4]</sup>采用预训练后的神经网络进行迁移学习,实现对围岩、钨矿石和石英三类矿物的识别任务。张利军等<sup>[5]</sup>提出了一种分别基于不同尺度显微图像的深度学习模型,用于岩石-矿物的鉴定。彭伟航等<sup>[6]</sup>以 16 种常见的矿物作为研究对象,用改进后的 Inception-V3 网络模型进行识别,识别准确率高达 86%。李雷等<sup>[7]</sup>提出了一种利用迁移学习策略并引入通道注意力机制的改进 ConvNeXt 模型,准确率和精确率分别达到了 98.58% 和 98.62%。陈博勋等<sup>[8]</sup>基于卷积神经网络、优化算法等,对采集的辉石、石英、角闪石、橄榄石、斜长石 5 种矿物图像进行研究,精确率达到了 98.48%。宛鹤等<sup>[9]</sup>在 MobileNet 模型的基础上引入协调注意力机制,采用迁移学习的方法对其进行预训练,获得了较高的准确率和  $F_1$ 。深度学习方法在提高矿物识别的效率和准确性方面具有巨大潜力,基于深度学习的目标检测算法以其快速和高精度的特点,有望弥补传统方法的不足,并在减少人力投入的同时实现高效准确的智能识别<sup>[10-12]</sup>。

在岩矿识别领域,雷明峰等<sup>[13]</sup>对 Mask R-CNN 模型进行针对性改造,基于像素占比分析实现各矿物成分含量的量化统计,在 95% 置信度下含量识别误差率仅为 6.9%。王名越等<sup>[14]</sup>提出了改进的 RDB-Yolov5 算法,使用密集连接残差模块替代传统残差结构,有效识别了花岗岩中的 5 类目标矿物。桂方俊等<sup>[15]</sup>通过卷积注意力模块、双向特征金字塔网络结构等方法对 YOLOv5 模型进行改进。

PAVITHRA 等<sup>[16]</sup>使用 YOLOv7 对煤矸石进行识别,试验结果表明,YOLOv7 可以在不牺牲识别速度的前提下,具有较高的精度。ZHANG 等<sup>[17]</sup>对 YOLOv7 进行改进,提高了煤矸石目标检测的速度和精度。王建才等<sup>[18]</sup>将 CBAM 注意力机制和 Transformer 添加到 YOLOv5 网络模型中,对煤岩图像进行检测。

尽管 YOLO 模型已被应用于岩矿识别领域开展了很多探索性的工作,但岩石薄片图像在进行矿物识别时依然存在问题:岩石薄片图像通常具有复杂的纹理和结构,要求目标检测模型具有较强的特征提取能力和高效的数据处理能力;岩石薄片图像具有同类矿物镜下数据变化大的特点,会影响模型对目标矿物的准确定位;矿物的大小和形状千变万化,且相同矿物在单偏光和正交偏光下会呈现出不同的外表特征,均会导致模型的识别准确率低。同时,这些问题致使基于岩石薄片图像数据集的矿物识别研究相对匮乏。针对上述问题,本文提出一种基于 YOLOv8 的岩石矿物智能识别模型,通过集成 LSK(Large Selective Kernel Network)模块动态调整感受野范围,提高模型识别精度;引入全维动态卷积(Omni-Dimensional Convolution, ODConv)技术,使卷积核能够更好地适应输入数据的特征,可以提高模型的特征提取能力和数据处理能力;优化损失函数为 Focal-EIoU(Focal Enhanced Intersection over Union),提高模型的收敛速度、定位精度,克服了传统方法在效率和准确率上的局限性。

## 1 关键技术与模型构建

### 1.1 YOLOv8 模型

YOLOv8 模型作为一种高效的实时目标检测和图像分割方法,在多类别目标检测中具有快速率和精确性,能够直接输出目标类别概率及其位置信息,被广泛应用于各种计算机视觉应用中<sup>[19]</sup>。YOLOv8 模型由输入端、主干网络、颈部网络以及输出端组成。

输入端能有效处理输入数据,为模型提供适当

的信息,从而实现有效的目标检测任务。主干网络用于提取图像特征,引入 C2f(Concatenate with 2 Fusion)模块<sup>[20]</sup>,用梯度分流的思想,使网络能更有效地表达更深层次的特征,并且保留了快速空间金字塔池化(Spatial Pyramid Pooling-Fast, SPPF)模块<sup>[21]</sup>,在解决输入图像尺寸不同问题的同时保证模型轻量化。在颈部网络中使用特征金字塔网络(Feature Pyramid Network, FPN)结构<sup>[22]</sup>和路径聚合网络(Path Aggregation Network, PAN)结构<sup>[23]</sup>。FPN结构通过自顶向下进行上采样,使底层特征图包含更强的语义特征,而 PAN 结构通过自底向上进行下采样,使顶层特征包含更强的定位特征。将两种特征进行融合,使不同尺寸的特征图均包含强语义特征和强定位特征,保证了对不同尺寸图像的准确预测。输出端使用了目前主流的解耦头结构,解耦头结构将分类和检测头分离,独立优化每个任务,使网络的训练和推理更加高效。

### 1.2 矿物智能识别模型

由于岩石薄片图像具有复杂的纹理和结构、同类矿物镜下数据变化大以及矿物的大小和形状千变万化的特点,为了更好地适应岩石薄片这种特殊领域的目标检测任务,矿物智能识别模型在 YOLOv8 模型的基础上进行改进,将改进模型命名为 Mineral-YOLO 模型,主要改进工作如下:

- (1) 在主干网络和颈部的 3 个输出端上添加 LSK 模块,提高处理不同目标所需的背景信息差异的能力;
- (2) 将颈部的降采样卷积替换为 ODConv,减轻背景干扰的影响,增强卷积网络的性能;
- (3) 基于 Focal-EIoU 损失函数的关注困难样本机制和超强的边界框匹配能力来优化 YOLOv8,提高边界框定位的准确性。

本文技术路线和 Mineral-YOLO 模型结构如图 1、图 2 所示。

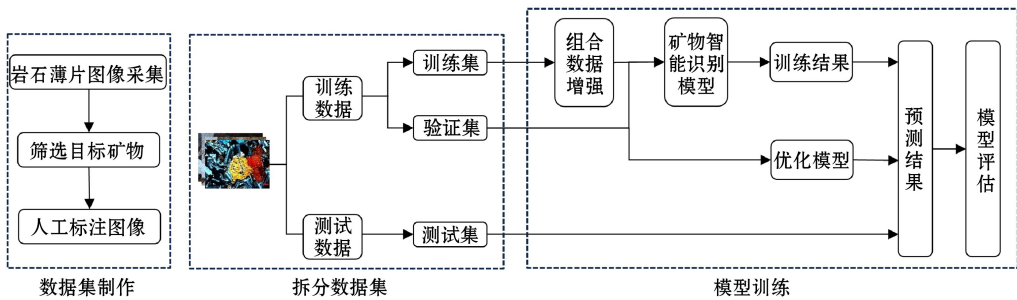


图 1 技术路线

Fig.1 Technical route

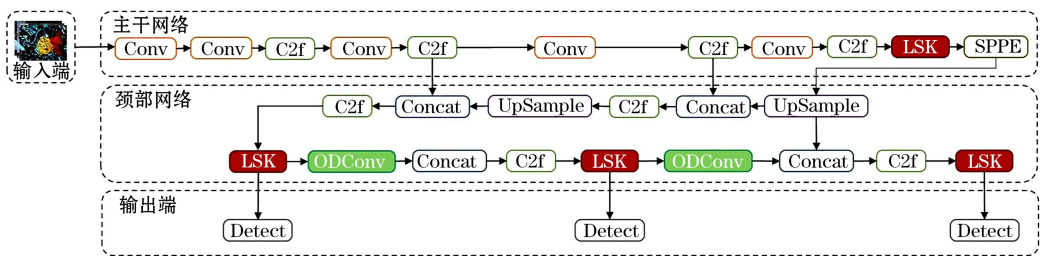


图 2 Mineral-YOLO 模型结构

Fig.2 Structure of Mineral-YOLO model

#### 1.2.1 基于 LSK 的特征提取网络

岩石薄片图像通常具有复杂的纹理和结构,这要求目标检测模型具有较强的特征提取能力和高效的数据处理能力,而目前的主干特征提取网络忽视了图像中丰富的背景先验知识。为解决该问题,在 YOLOv8 的主干网络以及颈部的 3 个输出端上添加

LSK 模块。LSK 模块由大核卷积序列和空间核选择机制两部分组成,如图 3 所示。由图 3 可知,LSK 模块首先通过大核卷积序列生成多种不同大小感受野的特征向量;然后利用空间核选择机制从不同尺度的大卷积核中对特征向量进行空间选择,增强网络关注最相关空间上下文区域的目标检测能力<sup>[24]</sup>。

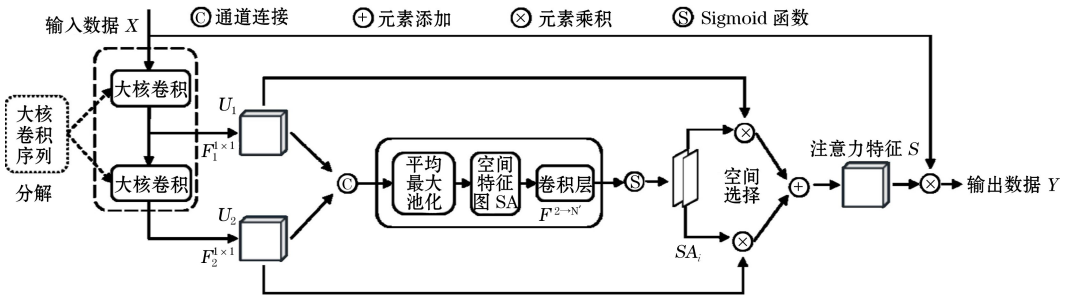


图3 LSK 模块结构

Fig.3 Structure of LSK module

(1) 大核卷积序列。

大核卷积序列将更大的卷积核分解成一系列深度可分离卷积,对输入数据  $X$ , 赋值为  $U_0 = X$ , 使用一系列深度卷积以获取不同区域丰富的上下文信息特征。将  $N$  个分解核通过 1 个  $1 \times 1$  卷积  $F^{1 \times 1}(\cdot)$  进一步处理得到空间特征向量  $U_i$ , 最后将全部空间特征向量进行通道融合。

$$U_i = F^{1 \times 1}(U_i), i \in [1, N] \quad (1)$$

(2) 动态选择机制。

动态选择机制从多尺度的大卷积核中对特征向量  $U_i$  进行空间选择,使网络关注最相关的空间上下文区域,以增强目标的检测能力,计算过程如下。

首先,对从不同核获得的特征与不同感受野范围拼接得到特征向量  $U_i$ ,对  $U_i$  使用通道平均池化和最大池化以提取空间关系,得到  $SA_{avg}$  和  $SA_{max}$ 。再使用卷积层将两个通道的池化特征转换为  $N'$  个空间注意力图:

$$SA = F^{2 \rightarrow N'}([SA_{avg}; SA_{max}]) \quad (2)$$

其次,对每个空间特征图 SA 使用 Sigmoid 激活函数,得到每个分解大核的单独空间选择掩码  $SA_i$ 。将分解的大核序列中的特征向量  $U_i$  进行对应的空间选择掩码加权,再通过卷积层  $F(\cdot)$  进行融合,得到注意力特征  $S$ 。

$$S = F\left(\sum_{i=1}^{N'} (SA_i \cdot U_i)\right) \quad (3)$$

最后,将输入数据  $X$  和注意力特征  $S$  逐元素相乘,得到最终输出  $Y$ 。

LSK 模块在特征提取过程中,能根据输入图像的动态特性自适应调整感受野范围,从而更有效地处理不同目标所需的背景信息差异,为特征分析提供所需的上下文信息。本文将其运用在岩石矿物成分智能识别模型中,能更有效地提取图像中的特征,显著提高岩石薄片图像的认识效率。

1.2.2 基于 ODConv 的数据适应性改进策略

岩石薄片图像具有同类矿物镜下数据变化大的特点,影响模型对目标矿物的准确定位,因此,将模型颈部中的降采样卷积替换为全维动态卷积 (ODConv),使卷积核能够更好地适应输入数据的特征,提高模型的特征提取能力和数据处理能力。

全维动态卷积采用多维注意力机制和并行策略,针对核空间的 4 个维度(空间核尺度、输入通道数、输出通道数、卷积核数量)学习卷积核的互补注意力<sup>[25]</sup>,大幅提升卷积的特征提取能力。与传统的卷积操作相比,全维动态卷积的卷积核的大小和形状可以根据不同大小的输入进行调整,在处理时更加灵活,能提供更好的性能以捕获丰富的上下文信息<sup>[26]</sup>。ODConv 的计算公式如下:

$$y = \sum_{i=1}^n (W_i \odot \alpha_{si} \odot \alpha_{ci} \odot \alpha_{fi} \odot \alpha_{wi}) * x \quad (4)$$

式中: $x$  表示输入特征; $y$  表示输出特征; $W_i$  表示卷积核; $\alpha_{si}, \alpha_{ci}, \alpha_{fi}$  分别表示沿着卷积核  $W_i$  的空间维度、输入通道维度、输出通道维度引入的注意力标量,  $\alpha_{si} \in R^{k \times k}, \alpha_{ci} \in R^{c_{in}}, \alpha_{fi} \in R^{c_{out}}$ ;  $\alpha_{wi}$  表示整个卷积核  $W_i$  的注意力标量,  $\alpha_{wi} \in R$ ;  $\odot$  表示沿着核空间的不同维度的乘法操作;  $*$  表示卷积运算。全维动态卷积运算流程如图 4 所示。

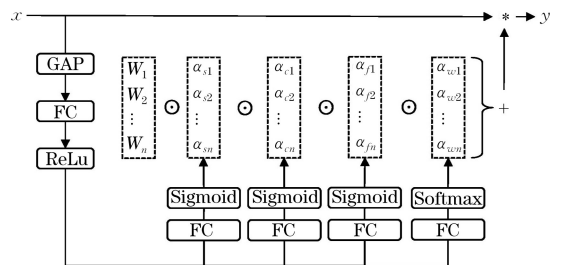


图4 ODConv 运算流程

Fig.4 ODConv operation flowchart

ODConv 通过并行策略采用多维注意力机制沿

核空间的 4 个维度学习互补性注意力。4 个维度中,卷积核维度可以在卷积时细化矿物的局部特征,空间位置维度能精确确定矿物所在位置区域,输入通道维度可以提取矿物的结构性特征,输出通道维度能避免矿物特征的丢失。ODConv 能够帮助模型更好地关注图像中的目标区域,减轻对背景干扰的影响,提高卷积网络的性能,加强特征的适应性。

### 1.2.3 优化损失函数

损失函数是在模型训练过程中衡量模型预测结果与真实结果之间差异的函数,在训练过程中,对模型参数进行优化的目标是使该损失达到最小。YOLOv8 所采用的 CIoU 损失函数通过引入中心点距离和宽高比来衡量边界框之间的差异<sup>[27]</sup>,其计算公式如下:

$$L_{\text{CIoU}} = \text{IoU} + \frac{\rho^2(b, b^{\text{gt}})}{(\omega^c)^2 + (h^c)^2} + \alpha v \quad (5)$$

式中:IoU 表示边界框交并比; $b$  为预测框; $b^{\text{gt}}$  为真实框; $\rho(b, b^{\text{gt}})$  表示两中心点的欧氏距离; $\omega^c$ 、 $h^c$  表示最小包围框的宽和高; $v$  为评估预测框和真实框宽高比相似性的值; $\alpha$  为  $v$  的权重系数。 $\alpha$  和  $v$  的定义如下:

$$\alpha = \frac{v}{(1 - \text{IoU}) + v} \quad (6)$$

$$v = \frac{4}{\pi^2} \left( \arctan \frac{\omega^{\text{gt}}}{h^{\text{gt}}} - \arctan \frac{\omega}{h} \right)^2 \quad (7)$$

式中: $\omega$  和  $\omega^{\text{gt}}$  分别表示预测框和目标框的宽; $h$  和  $h^{\text{gt}}$  分别表示预测框和目标框的高。由于 CIoU 损失函数中的  $v$  所反映的不是宽、高分别与其置信度的真实差异,会阻碍模型有效的优化相似性。

为解决该问题,引入 Focal-EIoU 损失函数<sup>[28]</sup>,将宽高比拆开,从梯度的角度出发,引入了动态平衡参数,提高 EIoU 损耗的性能,Focal-EIoU 的表达式如下:

$$L_{\text{Focal-EIoU}} = \text{IoU}^\gamma L_{\text{EIoU}} \quad (8)$$

式中, $\gamma$  为控制异常值抑制程度的参数。

EIoU 的表达式如下:

$$L_{\text{EIoU}} = L_{\text{IoU}} + L_{\text{dis}} + L_{\text{asp}} = 1 - \text{IoU} + \frac{\rho^2(b, b^{\text{gt}})}{(\omega^c)^2 + (h^c)^2} + \frac{\rho^2(\omega, \omega^{\text{gt}})}{(\omega^c)^2} + \frac{\rho^2(h, h^{\text{gt}})}{(h^c)^2} \quad (9)$$

式中: $\rho(\omega, \omega^{\text{gt}})$  表示预测框和目标框宽度的差值; $\rho(h, h^{\text{gt}})$  表示预测框和目标框高度的差值。

同一种矿物的岩石薄片图像在单偏光和正交偏光下会呈现出不同的外表特征,且光照变化和噪声等干扰都会对图像造成影响,Focal-EIoU 的关注困难样本机制和超强的边界框匹配能力,能更好地解决因岩石薄片图像中矿物大小和形状差异所导致的识别效率低的问题,以提高模型的定位精度和识别效率。

## 2 数据集的构建与预处理

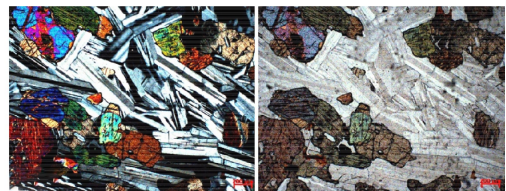
### 2.1 数据集采集

偏光显微镜根据不同的偏振光可以分为单偏光和正交偏光,在单偏光下,根据矿物的形态、解理、突起程度和糙面程度等特征对矿物进行区分,在正交偏光下,根据矿物的干涉色级序、消光类型、双晶等特征对矿物进行区分,通过对单偏光和正交偏光结合观察才能准确地确定矿物种类。

由于缺乏高质量且开源的数据集,本文采用自主收集的方式,利用偏光显微镜拍摄华北理工大学岩石薄片教学样品来获取研究所用图像,共包含来自山东、浙江等地区的 8 种岩浆岩单偏光图像和正交偏光图像,分别是二长花岗岩、橄榄苏长辉岩、黑云母花岗闪长岩、辉橄榄岩、辉石角闪岩、辉石岩、辉长岩、石英闪长岩(见图 5)。



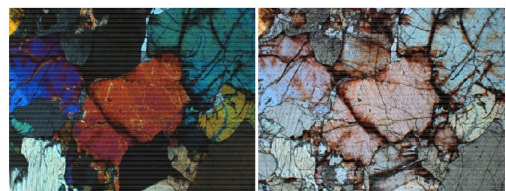
(a) 二长花岗岩



(b) 橄榄苏长辉石



(c) 黑云母花岗闪长岩



(d) 辉橄榄岩

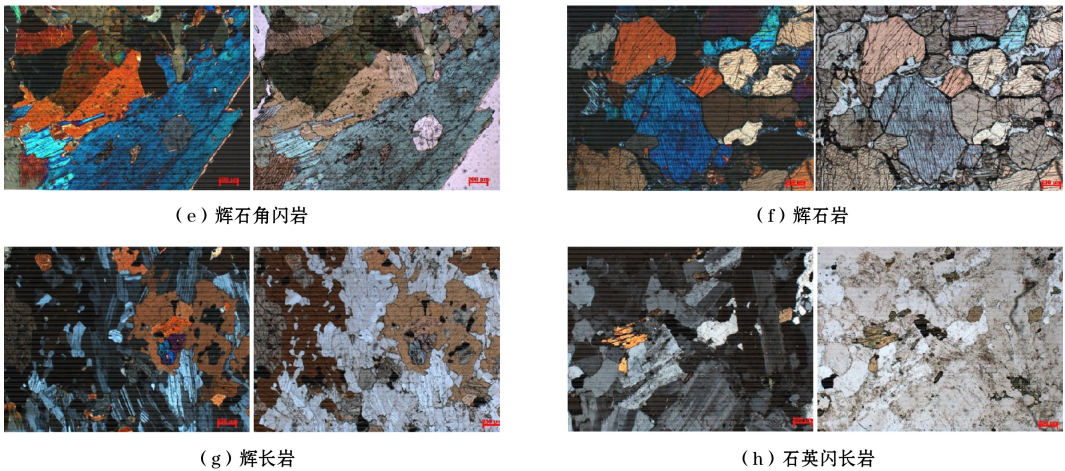


图5 8种岩浆岩正交偏光图像和单偏光图像

Fig.5 Orthogonal polarized images and single polarized images of eight types of magmatic rocks

采集过程中通过移动薄片和旋转载物台,拍摄单偏光图像和正交偏光图像共2 080张,每幅图像的尺寸为1 292像素×968像素。将所拍摄到的岩石薄片图像逐一采用图像标注工具进行矿物标注,标注时依据最大化特征向量空间距离原则,选择关注信息载量最多、差异最明显的特征,最终共标注了长石(包括钾长石和斜长石)、石英、云母(包括黑云母和白云母)、角闪石、辉石和橄榄石6类矿物,将其作为深度学习目标检测模型的数据集。6种矿物的偏光显微图像如图6所示。

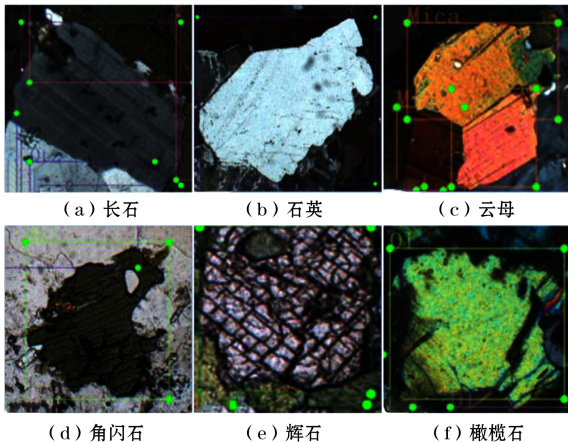


图6 6种矿物的偏光显微图像

Fig.6 Polarized microscopic images of six minerals

每个标注包含该矿物的边界框坐标信息和类别信息,实际标注时,一张图像中会包含多类矿物,对图像中出现的显著特征的所有种类矿物进行标注,标注时尽可能多得标注出颗粒完整、边界清晰、特征明显的矿物。共标注标签22 994个,其中长石7 441个,石英6 212个,云母1 161个,角闪石2 483

个,辉石3 823个,橄榄石1 874个。

## 2.2 数据集划分

在矿物成分智能识别领域,数据集的划分是至关重要的步骤。首先,按照9:1的比例将原始数据集随机划分为训练数据和测试数据,以确保充分利用数据进行模型构建和评估。然后,将训练数据按照9:1的比例随机划分为训练集和验证集。其中,训练集用于完成矿物成分智能识别模型的训练,验证集则用于优化模型参数并验证训练结果的有效性。测试数据独立作为测试集用于模型识别准确率的计算,全面评估模型在未知岩石薄片矿物上的识别能力。选择9:1的比例不仅能够确保大部分数据用于训练,帮助模型学习丰富的特征,还能保证验证集的规模足以提供可靠的性能评估。

通过这种数据划分策略,可以充分利用岩石薄片图像数据,避免模型对特定矿物样本过度拟合,增强模型的稳定性和可靠性,从而能更加科学地构建和优化矿物成分智能识别模型,为矿物成分智能识别领域的研究和应用提供有力支持。

## 2.3 岩石薄片图像数据增强

深度学习可以通过对大量数据的学习采集到更多的特征,是否具备充分的数据样本是保证训练模型效果好坏与泛化能力强弱的基础,但对于收集到的岩石薄片图像,依然存在数据集样本不足和数据类别不平衡的问题。对此,本研究利用数据增强相关理论,针对训练集中的图像进行数据增强,研究岩石薄片图像数据的增强优化方法,以实现岩石薄片矿物智能识别样本数量的增加和多样性的提升。

常见的数据增强技术分为单样本数据增强和多

样本数据增强,其中,单样本数据增强指的是在对某样本进行增强时,全部围绕该样本本身进行操作,包括翻转、旋转、缩放、裁剪、移位、变形和缩放等。而多样本数据增强是利用多个样本来产生新的样本,包括合成少数类过采样技术、混合样本技术(Mixup)和马赛克数据增强技术(Mosaic)等。这些操作可以降低模型对目标位置、色彩、尺度和视角差异等信息的敏感度,从而解决网络过拟合和样本不

平衡问题,提高模型的泛化能力。

本试验采取随机选择增强方式的策略,综合单样本和多样本数据增强方法,随机从添加噪声、调整亮度、旋转、裁剪、平移、翻转 6 种数据增强技术中选择一种或几种进行单样本数据增强,每张图片增强 5 次,在数据增强时,针对选定的数据增强技术,随机确定增强的程度或参数。图 7 为对二长花岗岩正交偏光图像进行随机单样本数据增强的示例。

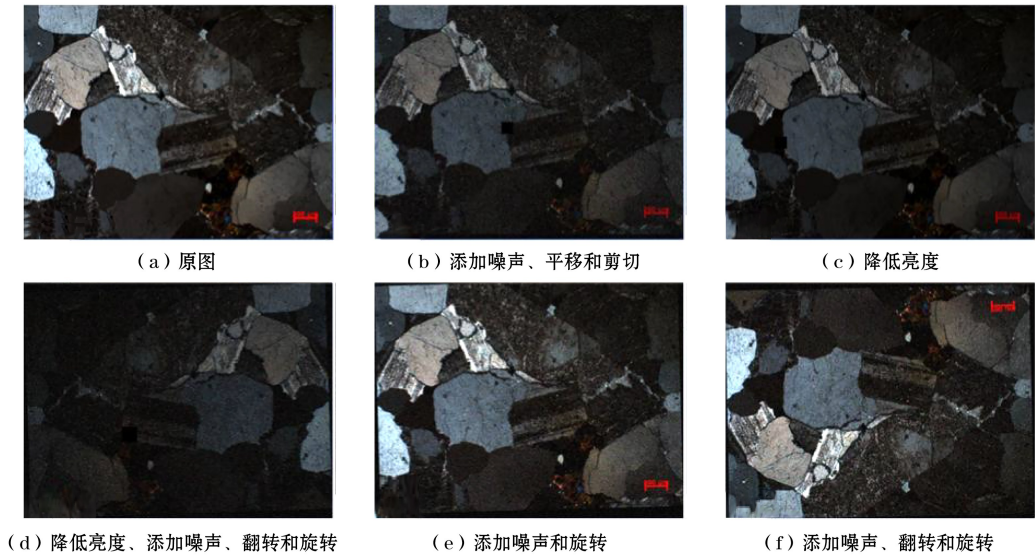


图 7 随机单样本数据增强示例

Fig.7 Example of random single sample data augmentation

多样本数据增强选用 Mixup 方法,从岩石薄片图像数据集中随机选择两张图像,分别对这两张图像进行翻转、缩放、平移、色域变化等操作,最终将两者的图像和真实框按照一定的比例融合。Mixup

数据增强技术可以将数据集中不同类别的图像进行混合,使得对数据的理解更具有线性化,消除了因数据不均匀带来的误差。图 8 为运用 Mixup 方法进行多样本数据增强的示例。

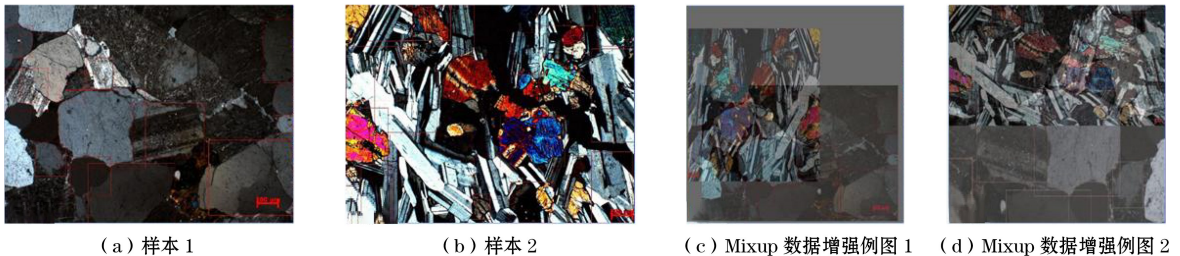


图 8 Mixup 多样本数据增强示例

Fig.8 Example of Mixup multi sample data augmentation

这种融合单样本和多样本数据增强的方式,不仅能够借助多种单样本变换来增加样本的多样性,从而提高模型对不同种类岩石薄片图像的适应性,还可以通过多样本增强方法混合样本,进一步扩充数据集,丰富矿物类别的分布,强化模型对类别边界的学习能力。在图像增广过程中,还需对每张图像

的标签信息进行精确处理,并使用序列容器记录图像变换操作的顺序,按照相同的顺序对目标边界框坐标进行一致变换,最终生成与新图像相匹配的标签边界框。通过综合应用这些数据增强操作,将训练集中原始的 1 684 张图像扩充到 13 504 张,将原始图像中的 18 643 个标签扩充到 177 646 个。因

此,该数据增强策略不仅显著增加了样本的多样性,还有效提升了模型对复杂样本的判别能力,使其在实际应用中表现出更高的稳定性和可靠性。

### 3 训练与结果分析

#### 3.1 环境配置

本次试验的操作系统为 Windows11,框架选用 Pytorch,采用 YOLOv8 模型进行训练。在模型训练时,选择随机梯度下降(SGD)优化器,初始学习率为 0.01,批次大小设为 16,训练迭代次数为 150 次,在最后 10 次关闭 Mosaic 数据增强,训练和测试时输入图像尺寸均设为 640 像素×640 像素。试验主要超参数见表 1。

表 1 试验超参数

Table 1 Experimental hyperparameters

类别	条目	版本
硬件环境	处理器	Intel(R) Xeon(R) Platinum 8358P
	显卡	RTX 3090
软件环境	操作系统	Windows11
	深度学习框架	Pytorch 1.11.0
	CUDA	11.3
	Python	3.8
	编辑软件	VSCode

#### 3.2 精度评价指标

为了综合评估模型的效果,本文采用精确率  $P$ 、召回率  $R$ 、 $F_1$ 、 $P$ - $R$  曲线和平均精度均值  $mAP$  等评价指标。这些指标不仅能够反映模型的分类准确程度,还能提供详细的分类结果分析,可用于深入理解模型在不同类别上的表现。

在训练中,预测结果与真实类别将样本分为真正例  $TP$ 、假正例  $FP$ 、真反例  $TN$ 、假反例  $FN$ 。精确率  $P$ 、召回率  $R$  和  $F_1$  的计算公式如下:

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

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

$$F_1 = \frac{2 \times P \times R}{P + R} = \frac{2TP}{2TP + FP + FN} \quad (12)$$

以精确率  $P$  为纵轴,召回率  $R$  为横轴作图得  $P$ - $R$  曲线,曲线下的面积为平均精度  $AP$  的值,用来衡量对一个类检测的好坏。对于离散的  $P$ - $R$  曲线,平均精度  $AP$  的计算公式为:

$$AP = \sum_{k=1}^n P(k) \Delta r(k) \quad (13)$$

式中: $P(k)$  表示第  $k$  个样本的精确度; $\Delta r(k) = R(k) - R(k-1)$ , $R(k)$  表示第  $k$  个样本的召回率; $n$  为样本总数。

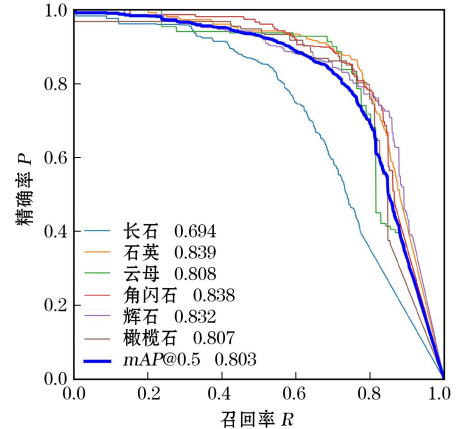
平均精度均值  $mAP$  指所有类别  $AP$  的平均值,用来衡量多类别目标检测的好坏, $mAP$  的计算公式为:

$$mAP = \frac{1}{Q} \sum_{q=1}^Q AP(q) \quad (14)$$

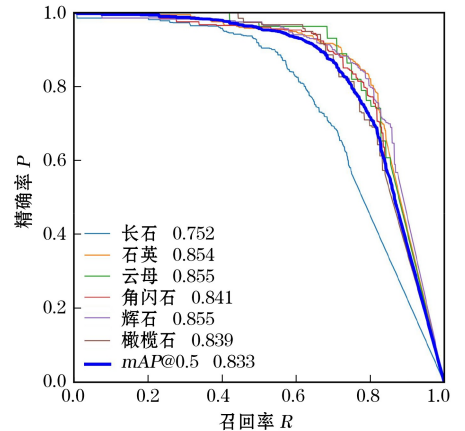
#### 3.3 结果分析

##### 3.3.1 模型改进前后性能对比

为验证 Mineral-YOLO 模型的优越性,本文将其与原始 YOLOv8 模型进行了详细的对比试验,图 9、图 10 为两个模型在验证集上的  $F_1$  的差异以及  $P$ - $R$  曲线的变化情况,其中不同颜色的曲线代表不同的矿物类别(颜色标识见电子版)。



(a) YOLOv8 模型



(b) Mineral-YOLO 模型

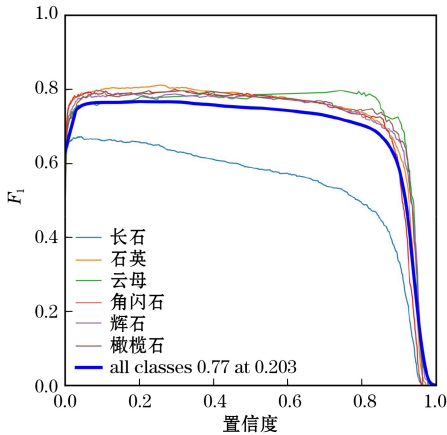
图 9 两个模型在验证集上的  $P$ - $R$  曲线对比

Fig.9 Comparison of  $P$ - $R$  curves of two models on validation set

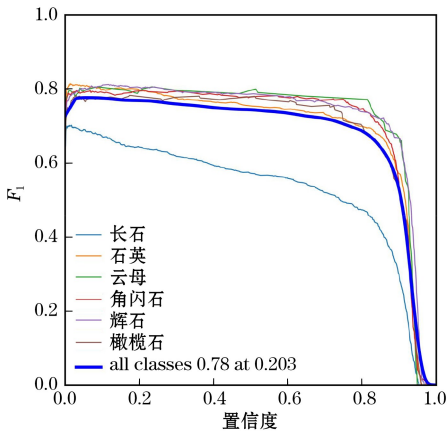
由图 9 可知,Mineral-YOLO 模型经过 150 轮次训练后,在验证集上各类目标的平均精度为:长石

75.2%，石英 85.4%，云母 85.5%，角闪石 84.1%，辉石 85.5%，橄榄石 83.9%。在  $IoU$  阈值设定为 0.5 时，模型的  $mAP$  达到 83.3%，相比改进前提升了 3 个百分点。表明 Mineral-YOLOv8 模型能够对 6 类矿物精准识别。

Mineral-YOLO 模型对长石、云母、橄榄石等矿物的识别精度有明显提高，尤其是在云母识别上，模型表现出最佳的识别效果，而长石的识别精度明显低于其他几种矿物。通过对偏光显微镜下云母和长石图像特征的深入分析发现，云母大多呈片状和板状，节理缝密集且干涉色较高，使得模型能够更高效、准确地提取云母的显微特征。相比之下，尽管长石多为柱状，形状特征较为明显，但其轮廓模糊、糙面微弱且干涉色单一（主要为白色），特征差异细微，增加了模型提取图像特征的难度，导致长石的分类精度较低。



(a) YOLOv8 模型



(b) Mineral-YOLO 模型

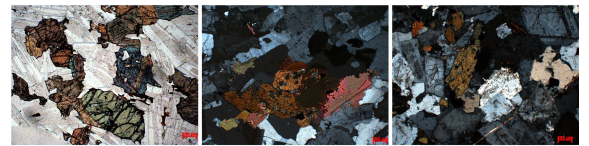
图 10 两个模型在验证集上的  $F_1$  对比

Fig.10  $F_1$  comparison of two models on validation set

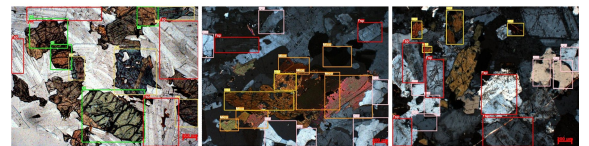
由图 10 可知，Mineral-YOLO 模型在验证集上的  $F_1$  为 78%，比改进之前(77%)相比， $F_1$  提高了 1 个百分点，增强了对少数类的识别能力，能更加有效

地衡量模型的整体性能。与 YOLOv8 模型相比，Mineral-YOLO 模型在识别角闪石、辉石和橄榄石等特征相似且识别难度较大的矿物上时表现出显著优势，其  $F_1$  明显高于原始模型。这表明 Mineral-YOLO 模型在少数类和复杂矿物的识别上有了显著的增强，其性能提升的关键在于引入了 ODConv 技术，使卷积核能够更好地适应输入数据特征，从而更准确地捕捉偏光显微镜下矿物的特征细节。尽管只提升了 1 个百分点，但这也能表明 Mineral-YOLO 模型在保持较高精度的基础上，进一步优化了模型的整体性能。

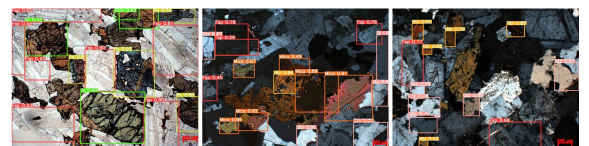
图 11 为原始图像、人工标注图、YOLOv8 模型预测和 Mineral-YOLO 模型预测的效果对比。由图 11 可知，Mineral-YOLO 模型在岩石薄片的目标检测任务中表现优异，其预测结果与人工标注图高度一致，漏检现象显著减少，能够精确识别并分类不同矿物。预测的边界框位置与标签框位置高度吻合，定位精度无明显偏差，表现出出色的目标空间定位能力。此外，Mineral-YOLO 模型预测的边界框形状与人工标注的边界框基本一致，进一步增强了其在岩石薄片图像这种复杂场景中的适用性和识别的准确性。



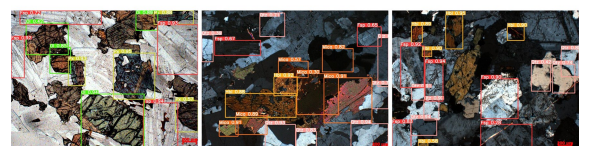
(a) 原始图像



(b) 人工标注图



(c) YOLOv8 模型预测



(d) Mineral-YOLO 模型预测

图 11 模型预测效果对比

Fig.11 Comparison of prediction performance of model

### 3.3.2 消融试验分析

通过消融试验验证各改进部分对整体模型的影响,采用相同的数据集和软硬件设备,以 YOLOv8 为基准算法,以  $F_1$ 、 $mAP$  和计算量(GFLOPs)作为评价指标,结果见表 2。

表 2 消融试验结果

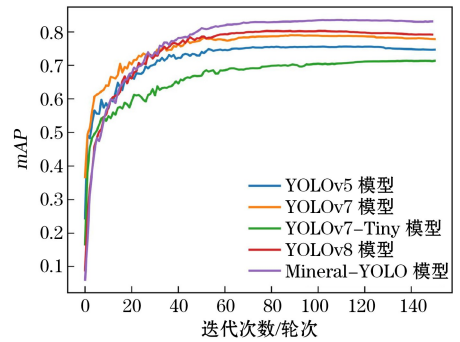
Table 2 Results of ablation experiments

模型	$F_1/\%$	$mAP/\%$	GFLOPs
YOLOv8	77	80.3	257.4
YOLOv8+LSK	77	81.7	262.9
YOLOv8+ODConv	77	81.3	251.6
YOLOv8+Focal-EIoU	75	79.6	257.4
YOLOv8+LSK+ODConv	79	82.9	257.1
YOLOv8+ODConv+Focal-EIoU	76	81.0	251.6
YOLOv8+LSK+Focal-EIoU	77	82.2	262.9
Mineral-YOLO	78	83.3	257.1

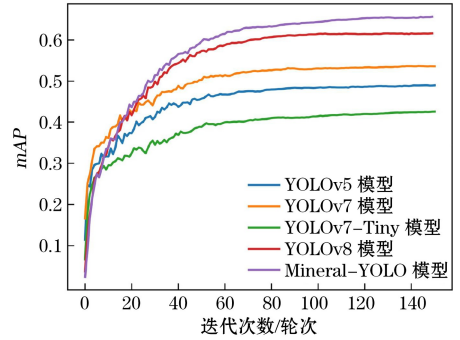
由表 2 可知,在 YOLOv8 模型中添加 LSK 模块后, $mAP$  提升了 1.4 个百分点(由 80.3% 提升至 81.7%),这表明该模块增强了模型对不同目标与背景信息的差异,提升了识别的精度。引入 ODConv 技术后,使模型更有效地聚焦于图像中的目标区域,减少了背景干扰的影响,相比于 YOLOv8 模型, $mAP$  提高了 1 个百分点。将损失函数替换为 Focal-EIoU 后,虽然  $mAP$  略有下降(79.6%),但通过结合 LSK、ODConv 和 Focal-EIoU,Mineral-YOLO 模型最终实现了  $mAP$  为 83.3% 的最佳性能。通过对比发现 Focal-EIoU 只有与 LSK 模块结合使用时才能有效发挥作用,与只结合 LSK 和 ODConv 的模型相比,Mineral-YOLO 模型  $mAP$  提高了 0.4 个百分点,表明 LSK 模块能帮助 Focal-EIoU 损失函数更好地提高边界框定位的准确性。因此,Mineral-YOLO 模型不仅显著提升了识别精度,还表现出了更好的效率和适应性,能够更加可靠、精确地完成岩石薄片复杂矿物的智能识别任务,验证了改进策略在实际任务中的有效性与可行性。

### 3.3.3 不同模型精度对比

为验证 Mineral-YOLO 模型的有效性,在相同试验条件下分别与 YOLOv5、YOLOv7、YOLOv7-Tiny、YOLOv8 模型进行对比训练。 $IoU$  阈值分别为 0.5 和 0.95 时  $mAP$  的变化情况如图 12 所示,其中不同颜色的曲线代表不同模型(颜色标识见电子版)。所有模型均在训练集上训练 150 轮次,将训练后的模型应用于测试集,得到的试验结果见表 3。



(a)  $IoU$  阈值为 0.5



(b)  $IoU$  阈值为 0.95

图 12 不同  $IoU$  阈值下的  $mAP$  对比

Fig.12 Comparison of  $mAP$  under different  $IoU$  thresholds

由图 12 可知,当  $IoU$  阈值为 0.5 时,Mineral-YOLO 模型在训练到 60 轮次后, $mAP$  收敛至 83% 左右。在同等条件下,YOLOv8、YOLOv7 和 YOLOv5 模型的  $mAP$  分别在 60 轮次后收敛于 80%、76% 和 74% 左右,YOLOv7-Tiny 模型在迭代 80 轮次后, $mAP$  收敛于 70% 左右。当  $IoU$  阈值为 0.95 时,Mineral-YOLO 模型在迭代了 80 轮次后收敛至 64% 左右,比 YOLOv7-Tiny 模型高出约 22 个百分点,表现出明显的优势。在训练过程中, $IoU$  阈值为 0.5 时,Mineral-YOLO 模型  $mAP$  约从 40 轮次后开始高于其他对比模型; $IoU$  阈值为 0.95 时,仅经过 20 轮次, $mAP$  就已显著领先于其他模型。表明 Mineral-YOLO 模型在不同  $IoU$  阈值下均具备更高的精度与较强的收敛性,验证了其在复杂矿物识别任务中的准确性。

由表 3 可知,不同模型在测试集上的评估指标表现出了显著差异,Mineral-YOLO 模型的精确率为 82.9%,明显高于其他模型,表明 Mineral-YOLO 模型在准确区分不同矿物类别方面具有更优的分类能力。在不同阈值的  $mAP$  指标上,Mineral-YOLO 模型均取得了最高值,表明该模型能稳定且准确地识别目标。在  $F_1$  指标上,Mineral-YOLO 模型达到

了 77%，与 YOLOv7 模型较为接近，但综合来看，Mineral-YOLO 模型在平衡准确率和召回率方面更加稳定。因此，Mineral-YOLO 模型更适用于岩石薄片矿物智能识别，能够有效满足复杂矿物分类与检测的需求。

表 3 测试集下不同模型效果对比

Table 3 Comparison of performance of different models in the test set

模型	P/%	R/%	mAP@50/ %	mAP@50:95/ %	F <sub>1</sub> / %
YOLOv5	75.8	71.4	74.9	49.0	73
YOLOv7	77.5	73.9	76.5	52.9	76
YOLOv7-Tiny	69.9	67.0	69.9	42.1	68
YOLOv8	80.1	70.6	78.2	60.0	75
Mineral-YOLO	82.9	71.6	81.3	64.0	77

## 4 结论

(1) 通过自主收集方式创建岩石薄片图像数据集，利用综合数据增强技术扩增数据集，增加样本的多样性，丰富矿物类别的分布。

(2) 提出了引入 LSK 模块、ODConv 技术和优化损失函数 3 种改进方法的 Mineral-YOLO 岩石薄片矿物智能识别模型，提升了矿物识别的准确性。

(3) 与 YOLOv8 模型相比，Mineral-YOLO 模型对 6 种矿物的识别平均精度均值  $mAP$  为 83.3%，提高了 3.0 个百分点， $F_1$  提高了 1.0 个百分点，表明 Mineral-YOLO 模型在岩石薄片矿物领域具有出色的识别能力。

(4) 通过与其他模型在测试集上的对比验证表明，Mineral-YOLO 模型的精确率  $P$  为 82.9%， $mAP@50$  为 81.3%， $F_1$  为 77%，均高于其他模型，更适用于岩石薄片矿物智能识别。

### 参考文献(References):

[1] 张野,李明超,韩帅.基于岩石图像深度学习的岩性自动识别与分类方法[J].岩石学报,2018,34(2):333-342.  
ZHANG Ye, LI Mingchao, HAN Shuai. Automatic identification and classification in lithology based on deep learning in rock images[J]. Acta Petrologica Sinica, 2018, 34(2): 333-342.

[2] 白林,姚钰,李双涛,等.基于深度学习特征提取的岩石图像矿物成分分析[J].中国矿业,2018,27(7):178-182.  
BAI Lin, YAO Yu, LI Shuangtao, et al. Mineral composition analysis of rock image based on deep learning feature extraction[J]. China Mining Magazine, 2017, 27(7): 178-182.

[3] 郝慧珍,顾庆,胡修棉.基于机器学习的矿物智能识别方法研究进展与展望[J].地球科学,2021,46(9):3091-3106.  
HAO Huizhen, GU Qing, HU Xiumian. Research advances and prospective in mineral intelligent identification based on machine learning[J]. Earth Science, 2021, 46(9): 3091-3106.

[4] 王李管,陈斯佳,贾明滔,等.基于深度学习的黑钨矿图像识别选矿方法[J].中国有色金属学报,2020,30(5):1192-1201.  
WANG Ligan, CHEN Sijia, JIA Mingtao, et al. Beneficiation method of wolframite image recognition based on deep learning[J]. The Chinese Journal of Nonferrous Metals, 2020, 30(5): 1192-1201.

[5] 张利军,鲁文豪,张建东,等.基于深度学习的镜下岩石、矿物薄片识别[J].地学前缘,2024,31(3):498-510.  
ZHANG Lijun, LU Wenhao, ZHANG Jiandong, et al. Rock and mineral thin section identification based on deep learning[J]. Earth Science Frontiers, 2024, 31(3): 498-510.

[6] 彭伟航,白林,商世为,等.基于改进 InceptionV3 模型的常见矿物智能识别[J].地质通报,2019,38(12):2059-2066.  
PENG Weihang, BAI Lin, SHANG Shiwei, et al. Common mineral intelligent recognition based on improved InceptionV3[J]. Geological Bulletin of China, 2019, 38(12): 2059-2066.

[7] 李雷,卢才武,江松,等.基于改进 ConvNeXt 网络的矿物图像智能识别[J].地质通报,2024,43(7):1266-1275.  
LI Lei, LU Caiwu, JIANG Song, et al. Intelligent mineral image recognition based on improved ConvNeXt network[J]. Geological Bulletin of China, 2024, 43(7): 1266-1275.

[8] 陈博勋,王宏铭,王玲,等.基于深度学习的镜下矿物图像识别方法[J].矿业研究与开发,2022,42(11):163-170.  
CHEN Boxun, WANG Hongming, WANG Ling, et al. Identification method of mineral image under microscope based on deep learning [J]. Mining Research and Development, 2022, 42(11): 163-170.

[9] 宛鹤,张金艳,屈娟萍,等.基于改进 MobileNet V3 的矿物智能识别模型[J].金属矿山,2024(1):174-181.  
WAN He, ZHANG Jinyan, QU Juanping, et al. Intelligent recognition model of mineral based on improved MobileNet V3[J]. Metal Mine, 2024(1): 174-181.

[10] 李德崎.深度学习在岩石薄片序列图像分割和目标识别中的应用研究[D].北京:中国地质大学(北京),2022.  
LI Deqi. Deep Learning in rock slice sequence image segmentation and target recognition applied research [D]. Beijing: China University of Geosciences (Beijing), 2022.

[11] 许振浩,马文,李术才,等.岩性识别:方法、现状及智能化发展趋势[J].地质论评,2022,68(6):2290-2304.  
XU Zhenhao, MA Wen, LI Shucan, et al. Lithology identification: method, research status and intelligent development trend [J]. Geological Review, 2022, 68(6): 2290-2304.

[12] 李雷.基于机器视觉的矿物识别与移动端应用研究[D].西安:西安建筑科技大学,2024.

- LI Lei. Research on mineral recognition and mobile application based on machine vision [D]. Xi'an: Xi'an University of Architecture and Technology, 2024.
- [13] 雷明锋,张运波,王卫东,等.岩石岩性 Mask R-CNN 智能识别方法与应用研究[J].铁道科学与工程学报,2022,19(11):3372-3382.
- LEI Mingfeng, ZHANG Yunbo, WANG Weidong, et al. Investigation and application on lithology intelligent recognition method based on Mask R-CNN[J]. Journal of Railway Science and Engineering, 2022, 19(11): 3372-3382.
- [14] 王名越,狄永军,张春禹.花岗岩矿物正交偏光镜下图像人工智能识别研究[J].矿物学报,2024,44(1):11-23.
- WANG Mingyue, DI Yongjun, ZHANG Chunyu. A study on identification of minerals in crossed polarizer images of granites using the artificial intelligence [J]. Acta Mineralogica Sinica, 2024, 44(1): 11-23.
- [15] 桂方俊,李尧.基于 CBA-YOLO 模型的煤矸石检测[J].工矿自动化,2022,48(6):128-133.
- GUI Fangjun, LI Yao. Coal gangue detection based on CBA-YOLO model[J]. Journal of Mine Automation, 2022, 48(6): 128-133.
- [16] PAVITHRA D, POOVIZHI P, VIGNESH N, et al. Identification of coal Gangue in Coal using YOLO V7[C]//International Conference on Trends in Electronics and Informatics. New York:IEEE,2023:1677-1682.
- [17] ZHANG Yuanhao, JI Liang, REN Shuwen, et al. YOLOv7-based Object detection method for coal gangue [C]//International Conference on Real-time Computing and Robotics. New York:IEEE,2023:786-792.
- [18] 王建才,李瑾,李志军,等.基于改进 YOLOv5 的煤岩图像检测识别研究[J].煤矿机械,2022,43(9):204-208.
- WANG Jiancai, LI Jin, LI Zhijun, et al. Research on coal rock image detection and recognition based on improved YOLOv5[J]. Coal Mine Machinery, 2022, 43(9): 204-208.
- [19] HUSSAIN M. YOLO-v1 to YOLO-v8, the rise of YOLO and its complementary nature toward digital manufacturing and industrial defect detection [J]. Machines, 2023, 11(7):11070677.
- [20] LI Yiting, FAN Qingsong, HUANG Haisong, et al. A modified YOLOv8 detection network for UAV aerial image recognition[J]. Drones, 2023, 7(5): 7050304.
- [21] 张磊,孙志鹏,陶虹京,等.改进 YOLOv8 的矿井人员防护装备实时监测方法研究[J/OL].煤炭科学技术:1-12[2024-10-25]. <https://link.cnki.net/urlid/11.2402.td.20240527.1700.004>.
- ZHANG Lei, SUN Zhipeng, TAO Hongjing, et al. Research on real-time monitoring method of mine personnel protective equipment with improved YOLOv8[J/OL]. Coal Science and Technology: 1-12 [2024-10-25]. <https://link.cnki.net/urlid/11.2402.td.20240527.1700.004>.
- [22] LIN Tsungyi, DOLLAR P, GIRSHICK R, et al. Feature pyramid networks for object detection[C]//Conference on Computer Vision and Pattern Recognition. New York:IEEE,2017:2117-2125.
- [23] LIU Shu, QI Lu, QIN Haifang, et al. Path aggregation network for instance segmentation[C]//Conference on Computer Vision and Pattern Recognition. New York:IEEE,2018:8759-8768.
- [24] LI Y X, HOU Q B, ZHENG Z H, et al. Large selective kernel network for remote sensing object detection[C]//International Conference on Computer Vision. New York:IEEE,2023:16794-16805.
- [25] MA Wenbi, GUAN Zheng, WANG Xue, et al. Research on reflective clothing recognition algorithm based on combining omni-dimensional dynamic convolution and partial convolution [J]. Engineering Applications of Artificial Intelligence, 2024, 137(Part B): 109180.
- [26] WEI Qiang, HU Xiaoyang, HOU Qinxian. Dynamic-YOLOv5: an improved aerial small object detector based on YOLOv5 [C]//International Conference on Neural Networks Information and Communication Engineering. New York:IEEE,2023:679-683.
- [27] 孙林,陈圣,姚旭龙,等.煤矿井下残缺信息的多目标检测方法研究[J/OL].煤炭科学技术:1-13[2024-10-25]. <https://link.cnki.net/urlid/11.2402.TD.20231204.2016.014>.
- SUN Lin, CHEN Sheng, YAO Xulong, et al. Research on multi-objective detection method for incomplete information in coal mine underground[J]. Coal Science and Technology: 1-13[2024-10-25]. <https://link.cnki.net/urlid/11.2402.TD.20231204.2016.014>.
- [28] ZHANG Yifan, REN Weiqiang, ZHANG Zhang, et al. Focal and efficient IOU loss for accurate bounding box regression[J]. Neurocomputing, 2022, 506: 146-157.