

文章编号: 1000-4734(2024)01-0011-13

花岗岩矿物正交偏光镜下图像人工智能识别研究

王名越, 狄永军*, 张春禹

(中国地质大学(北京)地球科学与资源学院, 北京 100083)

摘要: 人工智能在地球科学领域中的应用是近年来研究的热点, 对地球科学的发展有着重要意义, 其应用之一就是使用计算机视觉技术实现岩石或矿物的自动化识别分类。然而, 目前大多研究直接对岩石薄片图像进行分类, 不能够精细定位并识别薄片多且复杂的矿物目标。虽然现已有许多学者将目标检测技术应用于岩石矿物的图像识别分类中, 但这些方法识别的对象大多是岩石手标本图像, 只能对图像中的单一对象检测。在识别分类研究领域中, 缺少对岩石薄片镜下图像识别的算法及质量较好的相关数据集。为了解决这些问题, 本文首先采集了 3000 张正交偏光镜下花岗岩薄片图像, 标注矿物样本 10000 余个, 并通过数据增广方式对数据集进行增强, 建立了一个质量较好、具有多样性的数据集。其次本文提出基于 YOLOv5x 的改进算法 RDB-YOLOv5x。这种方法在特征提取过程中添加了密集连接方式, 使用密集连接残差模块(RDB)替代传统的残差结构, 有效地保留了图像的语义和位置信息细节。实验结果表明该方法泛化能力较好, 在对图像中小尺寸、特征模糊的矿物颗粒的识别中表现出优秀的性能, 可以准确有效地对花岗岩中的五类目标矿物(石英、黑云母、白云母、斜长石、钾长石)进行识别, 平均精度均值 mAP 达到 94.1%。较对比方法, 在 IoU 阈值为 50% 时提高了 0.5%, 阈值为 95% 时提高了 1%。

关键词: 人工智能识别; 岩石矿物; RDB-YOLOv5x; 镜下图像

中图分类号: P575; TP391.4

文献标识码: A

doi: 10.16461/j.cnki.1000-4734.2023.43.063

第一作者: 王名越, 女, 1997 年生, 硕士研究生, 地质工程专业. E-mail: 2101200023@cugb.edu.cn

A study on identification of minerals in crossed polarizer images of granites using the artificial intelligence

WANG Mingyue, DI Yongjun*, ZHANG Chunyu

(School of Earth Sciences and Resources, China University of Geosciences Beijing, Beijing 100083, China)

Abstract: The application of artificial intelligence in field of geosciences becomes a research hot spot in recent years. It is of great significance to the development of geosciences. One of applications of artificial intelligence is to achieve automated identification and classification of rocks or minerals using the computer vision technique. However, most researches are generally focused on the classification of rock type rather than the directly precise identification of multiple and complex minerals in the rock based on its thin section images. Although the object detection technique has been applied by many scholars to identify and classify types of rocks and minerals based on images, its application objects are mostly rock hand specimen images, and it can only be used to detect a single object in the image. In the research field of identification and classification, there is a lack of good quality algorithms and datasets for identifying and classifying minerals in the rock based on the rock thin section images. In order to solve these problems, firstly, we have collected more than 3000 images of thin sections of granite under crossed polarizer microscope, have labeled more than 10000 mineral samples on those images, have enhanced the dataset by means of data augmentation and have established a dataset with good quality and diversity. Secondly, we have proposed an improved algorithm of the RDB-YOLOv5x based on the YOLOv5x. In this method, the dense connection method was added in the feature extraction process, and the residual dense block (RDB) was used to have

收稿日期: 2022-06-09

基金项目: 中国地球物理场与成矿关系(编号: KD-[2019]-XZ-064; KD-[2020]-XZ-044; KD-[2021]-XZ-049)

* 通信作者, E-mail: diyongjun@cugb.edu.cn

replaced the traditional residual structure. Thus, the semantic and location information details of images can be effectively preserved by using this method. Experimental results showed that this method had good generalization capability, and excellent performance for identifying small sized and fuzzy characterized mineral grains in images. By using this method, we have accurately and effectively identified five kinds of targeting minerals (quartz, biotite, muscovite, plagioclase, potassium feldspar) in the granite, with the mean average precision (mAP) up to 94.1%. Compared with the optimal comparison method, the mAP values are increased 0.5% at the IoU threshold of 50% and 1% at the threshold of 95%, respectively

Keywords: artificial intelligence identification; rocks and minerals; the RDB-Yolov5x; images taken under crossed polarizer microscope

造岩矿物的识别与分类一直是地球科学领域研究中的重要内容之一,矿物的准确识别是岩性识别和许多研究工作的基础,更有一些指示性矿物的识别在地质科学研究、地质资源勘探、岩土工程等領域都十分重要^[1-2]。通过矿物在偏光显微镜下表现出的光学特征差异识别矿物组合类型是一种传统可靠的方法。在单偏光镜下和正交偏光镜下观察岩石薄片,可以获得矿物的晶形、解理、颜色(多色性)和突起,最高干涉色级序,消光类型和双晶类型等多种特征,这些特征是识别矿物的重要依据。但镜下观察识别矿物组合需要耗费大量的人工成本和时间成本,且其正确性也往往会受到观察者的经验和专业限制所影响^[3]。近年来,随着人工智能的迅速发展,尤其在计算机视觉领域有了重大的突破,基于深度学习的图像识别技术已经发展成熟,相关理论和成果已经广泛应用于无人驾驶,人脸识别,医学影像等领域^[4]。在地球科学研究工作中,人工智能的应用也得到了快速的发展。将人工智能方法应用于地球科学研究,使用计算机视觉技术实现岩石或矿物的自动化识别分类可以有效地解决岩矿鉴定费时费力、准确度受限等问题。

目前国内外已经有许多学者对岩石矿物图像的自动分类进行了尝试和研究。一些学者采用了传统的图像分割方法,先提取图像特征参数,再进行分类。Thompson 等^[5]首次尝试利用人工神经网络模型,使用了包括矿物颜色,矿物结构等 27 个不同的特征输入,对 10 种矿物图像进行分类。模型采用了标准的前馈网络及反向传播校正误差,并使用 GA 遗传算法进行优化,对单一矿物的识别精度达到 93%。在网络输入的参数选择中,一些学者选择对矿物的结构特征进行分类。Marmo 等^[6]建立了一个多层感知神经网络模型,提出一种处理图像的数值计算方法,该方法以灰度图像作为输入,数值特征值作为输出,对 500 多张碳酸盐岩薄片图像的岩石结构类型进行分类,测试集上识别精度达到 93.5%。Singh 等^[7]利用神经网络,对玄武岩的薄片图像进行识别,提取图像中的矿物结构特征,识别精度为 92.22%。一些学者在研究过程中则更关注矿物的颜色特征。Baykan 等^[8]使用关于颜色表现的参数作为输入,并使用人工神经网络对矿物进行分类,对石英,白云母,黑云母、绿泥石和不透明矿物进行分类训练,在测试集上取得了 81%~98%的识别精度。Aligholi 等^[9]采用了 CIELab 颜色空间方法对岩石薄片图像进行处理以引入与颜色相关的特征参数。并将局部二元模式算子(LBP)方法和 CIELab 颜色空间方法进行结合,对黑云母、角闪石、石英和方解石矿物图像分类精度分别为 99%、98%、96%和 95%。Patel 等^[10]采用了概率神经网络(PNN)模型,以彩色图像直方图作为输入,对石灰岩图像进行分类,总分类误差在 6%以下。研究中,将分类器与图像分割方法结合也是许多学者的研究方向。刘延保等^[11]提出了一种基于 LS-SVM 的岩石细观图像分析方法,对花岗岩岩石细观力学试验图像的处理结果精度达 96.82%。刘珏先等^[12]采用基于协同表示的分类器(CRC)对岩石薄片中的矿物成分进行分类识别,该方法的识别速度较基于稀疏的分类器识别方法提高 300%,识别率提高 2%。除此之外 K-means 聚类算法也常被用在岩石矿物图像的分类研究中,杨艳梅等^[13]在 Spark 平台借助优化后的 K-means 算法处理岩石图像,通过特征提取大幅降低图像识别难度。

一些研究者则使用卷积神经网络 (CNN), 将岩石矿物图像直接作为网络的输入, 实现对岩石矿物图像的自动分类。这种方法避免了传统识别算法中复杂的特征提取和数据重建, 近年来在地球科学的图像处理领域被广泛应用。白林等^[14]收集了 15 种常见岩石的图像数据, 利用卷积神经网络对岩石的识别精度达到 63%。张野^[15]基于 Inception-v3 深度卷积神经网络模型, 建立了岩石图像集分析的深度学习迁移模型, 对 173 张花岗岩图像、152 张千枚岩图像和 246 张角砾岩图像进行了分类研究, 分类精度达到 90% 以上。徐述腾等^[16]设计采用了 U-NET 卷积神经网络模型, 对黄铁矿、黄铜矿、方铅矿、闪锌矿等几种矿物图像进行训练识别, 在测试集上精度高于 90%。郭艳军^[3]等采取 Resnet-18 卷积神经网络进行实验, 使用单偏光镜下图像对石英、角闪石、黑云母、石榴子石、橄榄石等常见矿物进行分类, 模型测试中取得了 89% 的准确率。Xu 等^[17]提出了一种基于 Faster R-CNN 结构的岩石矿物图像智能识别模型, 可以识别岩石手标本图像中的岩性信息和位置信息, 模型识别精度达到 99.19%, F1 值为 96.6%。

尽管前人在岩石矿物人工智能识别领域的应用中取得了许多成果, 但目前仍存在一些问题: 1) 目前缺少使用目标检测技术对岩石薄片镜下图像的识别研究。目前对岩石薄片镜下图像的自动化识别分类大多是针对剪裁后, 只含有单类别矿物的图像, 且不能给出目标矿物的位置信息, 无法适应在矿物类型复杂, 矿物目标多的现实场景中应用。2) 在使用目标检测方法对岩石进行类别和位置预测的研究中, 对象均为岩石手标本图像, 检测图像的内容比较单一。目前应用目标检测方法实现岩石的自动识别分类研究中缺少质量较好的岩石薄片镜下图像数据集和对应解决方案。3) 花岗岩薄片镜下图像信息比较丰富, 图像中矿物目标多, 矿物尺寸差异大, 小尺寸矿物目标多。当前在岩石矿物识别领域应用的目标检测算法对于花岗岩岩石薄片中的矿物不能很好地检测, 尤其对花岗岩岩石薄片中小尺寸矿物颗粒的检测能力较弱。

花岗岩属于酸性岩浆岩中的深成侵入岩, 是岩浆岩中最具代表性的一类岩石。其主要矿物为石英、钾长石和酸性斜长石, 次要矿物有云母、角闪石及少量辉石^[18]。花岗岩中的矿物识别为花岗岩的分类定名, 岩石成因及成岩成矿作用演化研究等提供了重要的信息。偏光显微镜下的花岗岩薄片图像具有矿物种类多样、信息复杂、特征相似度高、图像噪声大等特点。因此, 基于深度学习方法, 对花岗岩正交偏光镜下图像中矿物目标的识别检测仍是一项具有挑战性的自动化图像检测任务。基于上述问题, 本文提出了一种针对花岗岩正交偏光图像中主要矿物的目标检测方法, 该方法可以对一张偏光显微镜下的花岗岩正交偏光图像中, 包括石英、白云母、黑云母、斜长石、钾长石的多种矿物有效识别。在信息复杂的高分辨率镜下图像识别任务中, 表现出了优秀的保留图像细节和定位的性能, 提高了对于石英、云母、斜长石等矿物颗粒的检测精度和定位效果。改进模型具有较好的鲁棒性和泛化能力。可以直接对一般花岗岩岩石薄片正交偏光镜下图像检测, 无需对图像做预处理, 有较高的实际应用意义。

1 研究方法

本次研究方法技术路线如图 1 所示。在首先进行的数据集的准备工作中, 本文采集了花岗岩薄片的正交偏光镜下图像, 对每张图像中特征明显的多个目标矿物进行标注, 并使用数据增广技术形成一个可观的数据集。接下来, 本次研究对 YOLOv5x 算法做出改进, 在网络骨干中添加了 RDB 残差密集链接模块。并使用该算法对数据集进行训练和验证, 得到各项精度指标。最后, 本文将训练得到的检测模型与 YOLOv5x 作对比, 并对两种模型的精度及在小尺寸矿物颗粒上的检测能力进行比较分析。

2 数据集的准备

2.1 数据集的采集

在偏光显微镜下观察, 花岗岩中的主要矿物在单偏光镜下可以对矿物的突起程度和解理进行区分, 正交偏光镜下图像则可以提供矿物的干涉色级序, 双晶律等更多矿物信息。

为达到对一般花岗岩中主要矿物的有效识别,本研究采集并筛选了花岗岩岩石薄片的正交偏光镜下图片共 3003 张,其中包含 30%正长花岗岩,40%二长花岗岩和 30%花岗闪长岩。采集过程中使用物镜放大倍数为 4 倍和 10 倍,采集图像分辨率均为 3600×3600 。使用软件 Labellmg 对 3003 张图片中的矿物进行信息标注。因同一种矿物在正交偏光镜下的不同角度下会表现出不同的光学特征,而不同矿物在同一角度下也可能有相似的表现,为最大化类间差异,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,最终对石英、白云母、黑云母在正交偏光镜下的干涉色级序特征(图 2a,b,c),碱性长石中的条纹长石、微斜长石以及酸性斜长石的双晶律特征(图 2d; 图 2e)进行标注。每个标注包含该矿物的边界框坐标信息和类别信息。以图 2 为例,1 个白色框即为 1 个标注边界框。

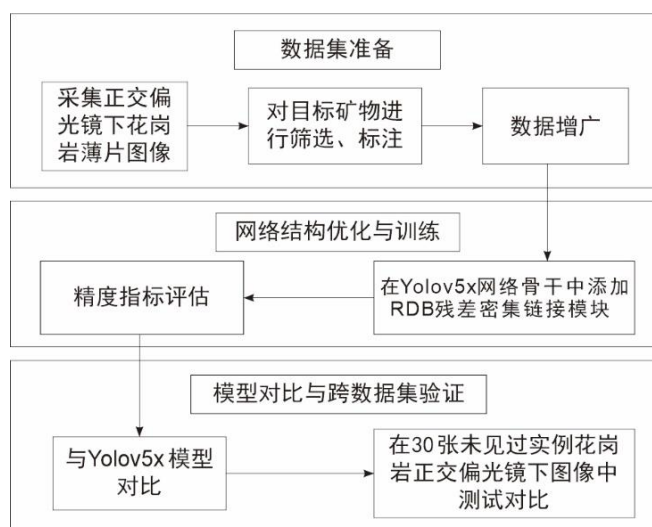
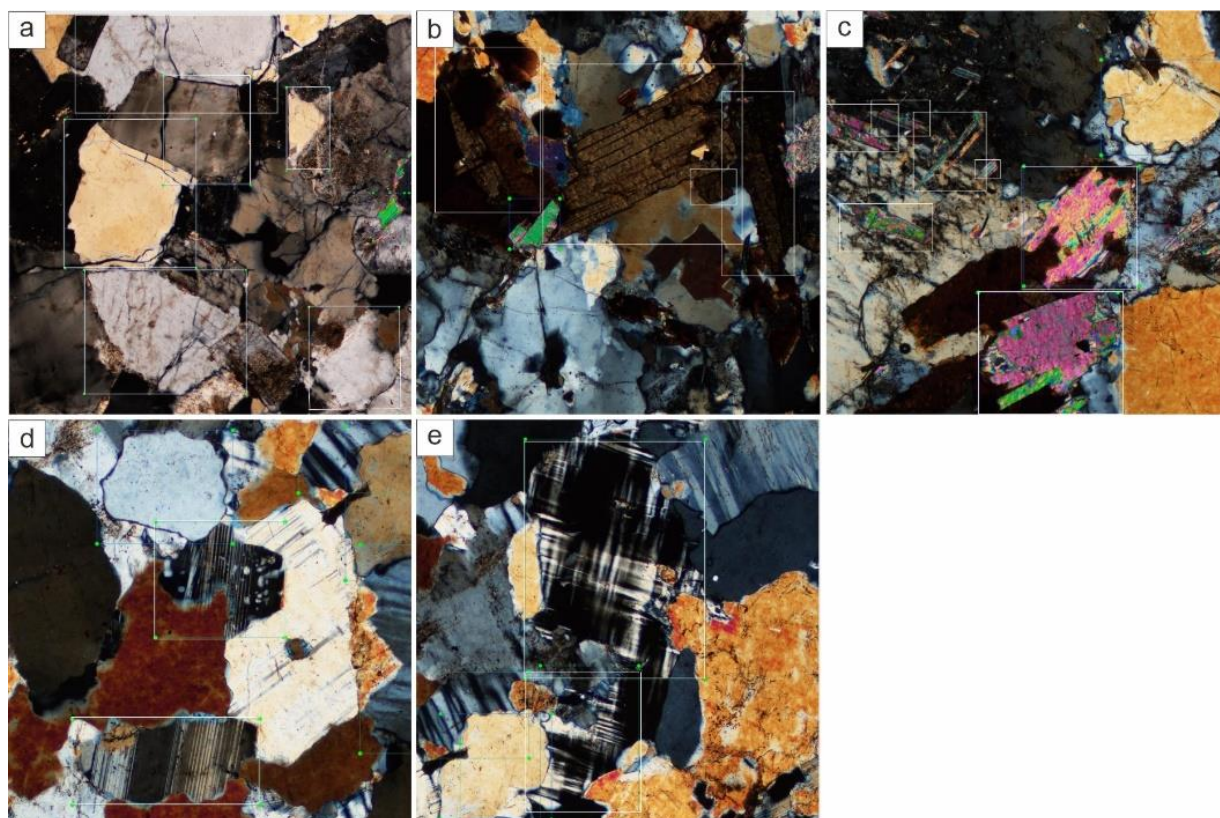


图 1 技术路线图

Fig. 1. Technological roadmap.



a. 石英颗粒的标注示例; b. 黑云母颗粒的标注示例; c. 白云母颗粒的标注示例; d. 具有明显聚片双晶特征的斜长石标注示例; e. 具有明显格子双晶特征的钾长石标注示例

图 2 正交偏光镜下标注示例图

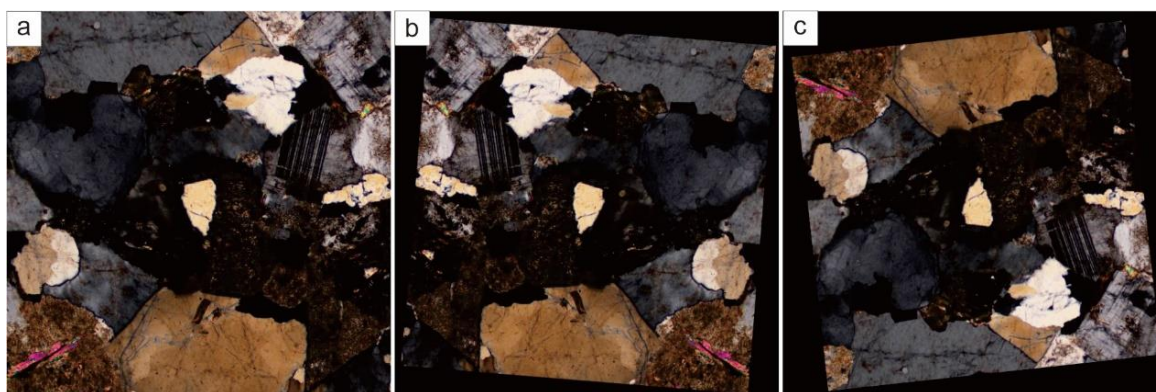
Fig. 2. Examples for labeling quartz, biotite, muscovite, plagioclase and K-feldspar under crossed polarizer microscope.

图 2 为每种矿物颗粒的单一标注示例。实际标注时, 会对图像中出现的具有显著特征的所有种类矿物进行标注。标记过程原则尽可能标注出颗粒新鲜完好, 边界完整清晰, 特征明显的矿物。共标注样本 10578 个。其中钾长石 1660 个, 斜长石 1390 个, 石英 4802 个, 白云母 1332 个, 黑云母 1394 个。

2.2 数据集增广

花岗岩中矿物普遍类内差异大, 而类间区分度小。如石英颗粒常边界不规则, 不同石英颗粒表现的干涉色差异也较大。为此, 本次研究采用数据增广方法对图像进行数据增强, 解决样本不足, 样本分布不均衡的问题, 提高了网络对矿物颗粒特征提取能力, 避免对不相关信息的学习。

数据增广方法以图 3 为例, 图 3a 为原始图像, 第 1 步: 对图像进行左右镜像处理, 再以图像中心为原点随机角度旋转并适当放缩, 得到图 3b; 第 2 步: 对图像 3b 进行上下翻转, 再以图像中心为原点随机角度旋转并适当放缩, 得到图 3c。在数据增广过程中, 随机旋转角度小于 30° 。最终将原始 3003 张图像扩充至 9009 张。



a. 原始训练图像; b. 对图 a 左右镜像、旋转、适当放缩后图像; c. 对图 b 上下翻转、旋转、适当放缩后图像

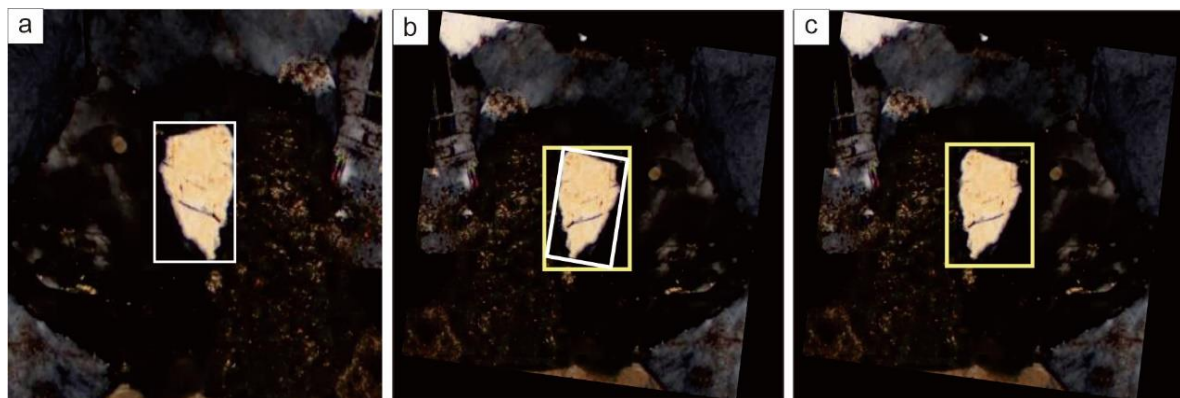
图 3 数据增广示例图

Fig. 3. Data augmentation examples.

在对图像进行增广的同时, 对每张图像中的标签信息进行读取, 提取的每个标签包含的 5 个参数: 边界框的左上角横、纵坐标, 边界框的右下角横、纵坐标以及该标签的类别。通过左上、右下角坐标计算出边界框中心点坐标, 并使用序列容器打包图像的变换操作顺序, 按照该顺序对目标边界框坐标进行相同的变换, 并通过计算生成适应新图像的标签边界框。以图 4 中的石英颗粒为例, 图 4a 中边界框随图像变换后得到图 4b 中内侧的白色边界框。再对图 4b 中的白色框使用三角函数计算, 求得白色边界框的最小外接矩形 (黄色框)。对最小外接矩形的左上、右下角坐标及类别信息进行保存, 则得到新的标签。图 4c 即为图 4a 数据增广后新得到的图像和标注。

旋转变换后, 使用最小外接矩形的标注有一定的误差。但对于数据集中大多是尺寸较小、长宽相近的标注误差影响较小。此外, 为进一步减小误差, 本文将增广中的随机旋转角度控制在 30° 内。

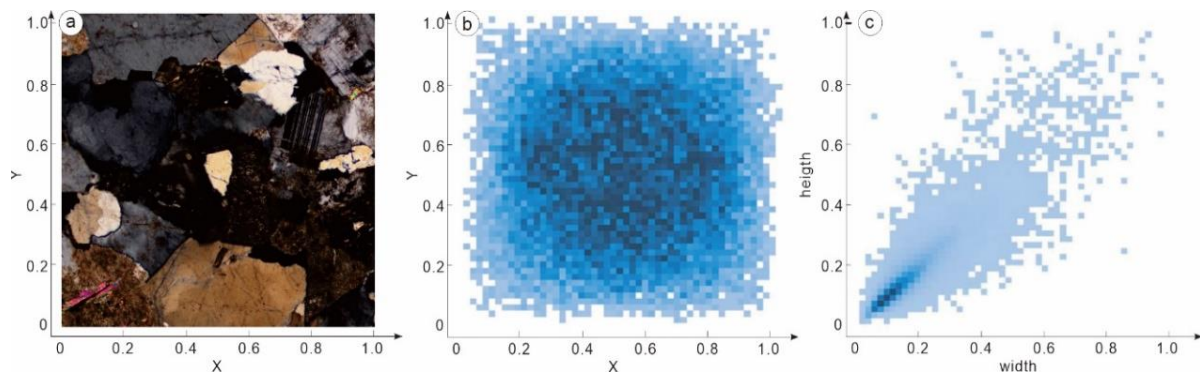
增广后数据集目标边界框统计基本情况如图 5 所示。以训练图像的左下角为原点, 图像的宽和高分别为横、纵坐标建立坐标系, 并对坐标轴作归一化处理, 如图 5a 所示。图 5b 中蓝色的点代表着目标边界框的中心点, 点落在图中的位置代表了实际数据集中目标边界框中心在图像中的分布, 区域颜色越深代表该区域的标注分布越集中。坐标轴归一化处理后, 图 5c 横坐标反映了目标边界框的宽与训练图像宽之比, 纵坐标反映了目标边界框的高与训练图像的高之比。图 5c 中的点多分布在第一象限对角线附近, 说明标注边界框形状大多近似正方形, 图中的点多集中在原点附近, 说明数据集中标注的小尺寸边界框较多。因此, 该数据集比较适合本文提出的增广方法。



a. 白色框为人工标注石英颗粒边界框；b. 白色框为与图像同步变换后边界框，黄色框为白色框的最小外接矩形；c. 增广后只保留最小外接矩形（黄色框）作为新的边界框

图 4 数据增广中边界框的处理示例图

Fig. 4. Augmentation examples for bounding boxes.



a. 图像坐标轴的建立及归一化处理示例；b. 目标边界框中心在训练图像中的分布情况；c. 目标边界框的尺寸和形状情况

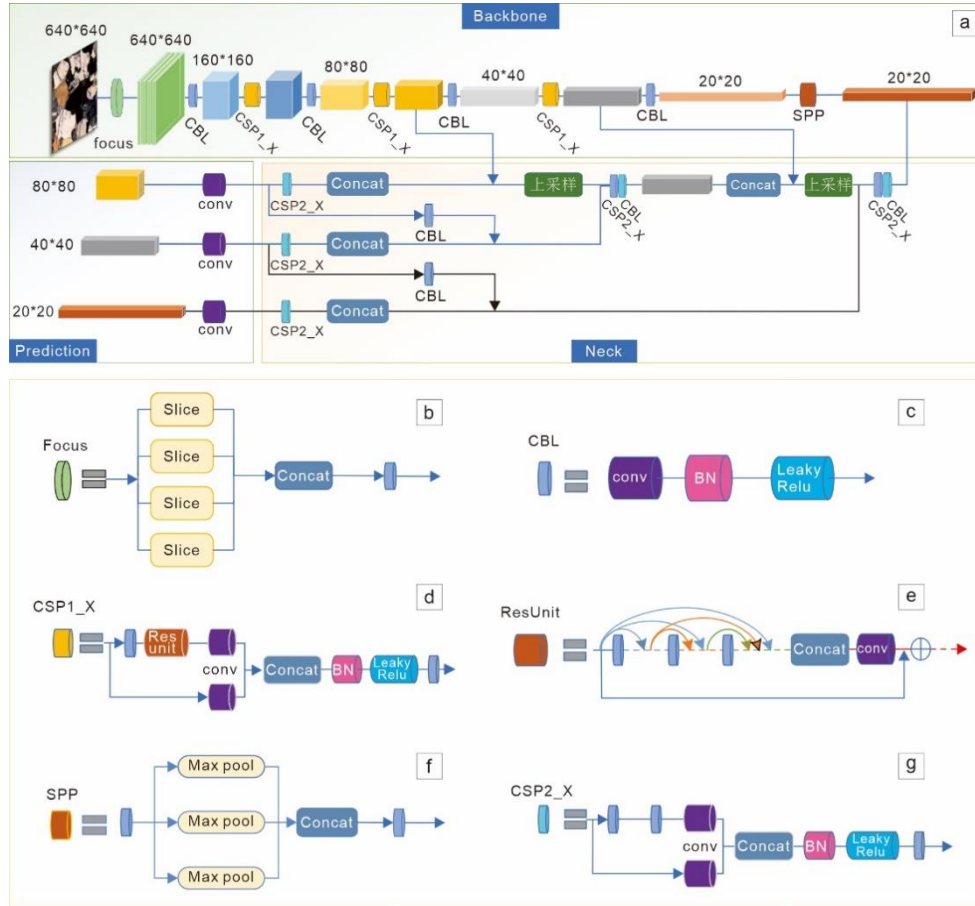
图 5 数据集标注基本情况图

Fig. 5. Basic information of dataset annotation.

3 智能识别算法：改进的 RDB-Yolov5x 模型

目标检测算法是一种利用深度学习方法自动提取输入特征实现对图像目标高精度快速分类和预测的自动化方法^[19-20]。目前基于深度学习的目标检测算法主要分 2 类：One-Stage 与 Two-Stage。One-Stage 算法直接从图像中进行预测回归，获取边界框坐标和种类概率^[4, 21]。Redmon 等^[22]提出的 Yolo 系列算法是 One-Stage 优秀算法代表之一，通过定义类别对图像中多个对象进行分类和定位，Yolo 将特征提取部分与特征金字塔网络结构相结合实现了多尺度检测^[23-24]。Yolov5 是 Yolo 系列中最新提出的网络结构，其模型结构有 4 种配置方式，Yolov5x 是其中精度最高的模型配置方式。Huang 等^[25-26]提出的 DenseNet 算法采用了密集连接的方式将前后层全部连接起来，实现了特征重用，有效地减少了因网络深度加深造成的信息丢失。为了更好地适应岩石薄片正交偏光镜下图像的检测需求，本文参考 DenseNet 结构，在 Yolov5x 的基础上融入残差密集连接，提出了 RDB-Yolov5x 模型。

RDB-Yolov5x 模型保留了 Yolov5x 网络结构的主要框架^[29]，由主干网络（Backbone），特征融合结构（Neck），预测输出（Prediction）3 个部分组成（图 6a）。3 个部分分别负责特征提取，特征融合，对预测结果的边界框进行得分排序分类。



a. Yolov5x 基本网络结构; b. Focus 层内部结构; c. CBL 层内部结构; d. CSP1_X 层内部结构; e. ResUnit 层内部结构; f. SPP 层内部结构; g. CSP2_X 层内部结构

图 6 RDB-Yolov5x 网络结构

Fig. 6. The RDB-Yolov5x network structure.

模型输入端的 Mosaic 数据增强方法, 将训练中的 4 张原始图像进行拼接, 这种方式可容纳更大的批样本数, 同时也使每张图像有更大概率包含小目标。进入主干网络的图像统一为 640×640 的三通数据。设计特征提取过程中采用了 CSP-darknet 作为主干网络结构。CSP-darknet 包含 5 个基本组件: Focus 层、CBL 层、CSP1_X 层、RDB 模块和 SPP 池化层。

其中, Focus 层 (图 6b) 对输入图像的相邻 4 个位置的像素进行切片 (slice) 后堆叠 (concat), 将输入的宽和高信息堆叠到通道数中, 形成 $320 \times 320 \times 12$ 大小的特征映射。这使高分辨率图像中的信息可以重构到低分辨率图像中, 减少了原图像的信息损失^[27]。

CBL 层 (图 6c) 是 Yolov5x 中的基本卷积单元, 由卷积层、BN 批量归一化层和 LeakyRelu 激活函数组成, 主要作用是提取特征并调整输出维度。

CSP1_X 层 (图 6d) 是主干网络中基于 CSPnet 构建的特征提取组件, 包括 1 个 CBL 层, x 个 RDB 残差模块和 1 个卷积层。CSP1_X 层将输入的特征图分割为两部分, 通过跨阶段分层结构对两部分进行合并实现更丰富的梯度组合。

ResUnit 是 CSP1_X 层中的子模块, 中心思路是采用残差结构来构建深层网络。残差结构采用了跳跃连接的方法 (Skip-connection), 从而使底层信息可以传递到更深层的网络中, 避免了随网络加深产生的梯度爆炸和梯度弥散问题。本文在结构设计中将传统的 ResUnit 模块优化为 RDB (Residual-dense-block) 模块 (图 6e)。RDB 将残差模块中前层的每一层输出传递到当前层的输出上, 使每一层之间存

在密集连接，密集连接的每一层都可以直接从损失函数和原始输入信号中获取梯度，经过 **add** 方式叠加后能充分利用前层特征信息，实现隐含的深度监控。本文在 3 个 CSP1_X 层中添加了 RDB 密集链接模块，设置密度层数为 4，以期减少小目标特征信息的丢失。在特征提取过程中，改进后的 RDB 模块明确区分了添加到网络中的信息（改变的信息）和被保护的信息（重用的信息），对重要的信息进行了保护，加强特征的传递。

SPP 层，空间金字塔池化组件（图 6f）采用了 5/9/13 三种尺度的最大池化，并对输出的 3 种尺度特征图进行堆叠，有效地提高了每个点的感受野。通过特征提取后进入特征融合部分的 3 个特征输出的尺寸分别为 20×20 ， 40×40 和 80×80 。

模型的特征融合部分采用了特征金字塔的思路对深浅层特征进行融合，主要由 2 个结构组成：FPN（Feature Pyramid networks）结构和 PAN（Path Aggregation Network）结构（图 7）。

FPN 是自上而下的特征融合，主干网络输出的 20×20 的特征经上采样后与 40×40 尺寸的输出融合，融合后再次进行上采样与 80×80 尺寸的输出进行融合，得到新的 3 个尺寸的输出，大小分别为 20×20 ， 40×40 和 80×80 。FPN 的跨尺度融合方式综合考虑了深浅层的特征，很好地保留了高层的语义信息。

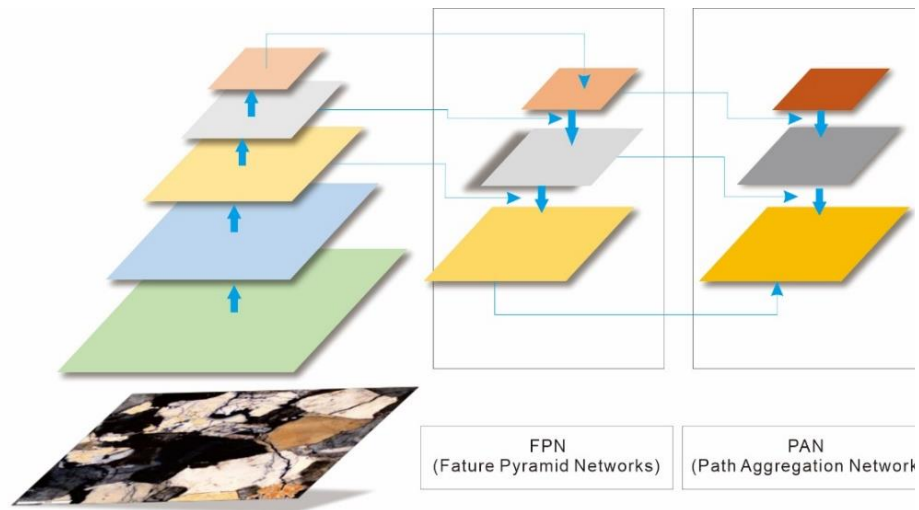


图 7 RDB-Yolov5x 特征融合结构

Fig. 7. The RDB- Yolov5x featured fusion structure.

PAN 是自下而上的特征融合。PAN 结构的输入是 FPN 结构的输出。其中 80×80 的特征下采样后与 FPN 中相同尺寸的输出进行融合，得到尺寸为 40×40 的新特征。对新特征再次下采样并与 FPN 小尺寸的输出融合，得到尺寸为 20×20 的新特征。经过 PAN 后得到 20×20 ， 40×40 和 80×80 等 3 个尺度的输出进入到预测部分，对输入图像进行预测类别和边界框回归。特征融合部分中用于特征提取的组件为 CSP2_X 层（图 6g），其结构与 CSP1_X 类似。

在预测输出部分中，使用 $\text{CIoU}_{\text{loss}}$ 作为边界框回归预测的损失函数。目标检测的后处理过程采用了 NMS 非极大值抑制方法（Non-Maximum Suppression）解决针对多目标边界框的筛选。 $\text{CIoU}_{\text{Loss}}$ 的计算基于目标边界框与预测框的交并比（IoU），具体计算方法如下^[28]：

$$\text{IoU} = \frac{(\mathbf{A} \cap \mathbf{B})}{(\mathbf{A} \cup \mathbf{B})} \quad (1)$$

式(1)中， $\mathbf{A}$ 为目标边界框覆盖区域； $\mathbf{B}$ 为预测框覆盖区域。

$$\text{CIoU}_{\text{Loss}} = 1 - \text{IoU} + \frac{\rho^2(b, b^{\text{gt}})}{c^2} + av \quad (2)$$

其中， ρ 为预测框与目标边界框中心点之间的式距离； C 为预测框与目标边界框最小外接矩的对角线的

长度。式(2)中, v 是目标边界框与预测框对角线的倾斜角度, 即目标边界框, a 是权重系数, 分别表示为:

$$v = \frac{4}{\pi^2} \left(\arctan \frac{w^{gt}}{h^{gt}} - \arctan \frac{w}{h} \right)^2 \quad (3)$$

$$a = \frac{v}{(1-\text{IoU})+v} \quad (4)$$

其中, w^{gt} 、 h^{gt} 、 w 、 h 分别代表目标边界框和预测框的宽和高。

4 模型训练及精度评价

4.1 训练环境及超参数选择

数据集以 9 : 1 比例随机划分为训练集和测试集。

优化后的 RDB-Yolov5x 网络训练设定初始学习率为 0.01, 采用 warm up 的学习策略, batchsize 为 8, epoch 训练总轮次为 100 次。

实验平台软硬件为:

CPU: E5-2680 V4 , GPU: GeForce RTX 3080。

RDB-Yolov5x 网络的深度系数为 1.33, 宽度系数为 1.25。结构中各部分超参数设置如图 8。

4.2 精度评价指标

为测试新模型是否可以较好地检测识别矿物目标, 采用了包括精确率 (Precision)、召回率 (Recall)、平均精度均值 mAP (mean Average Precision) 3 个指标^[29-30]。

在训练中, 根据真实类别与预测类别将样例分为真正例 TP (True

Positive); 假正例 FP (False Positive); 真反例 TN (True Negative); 假反例 FN (False Negative)。通过样例类型建立的分类结果混淆矩阵可以比较直观地反映模型的训练分类情况。混淆矩阵如表 1^[30]:

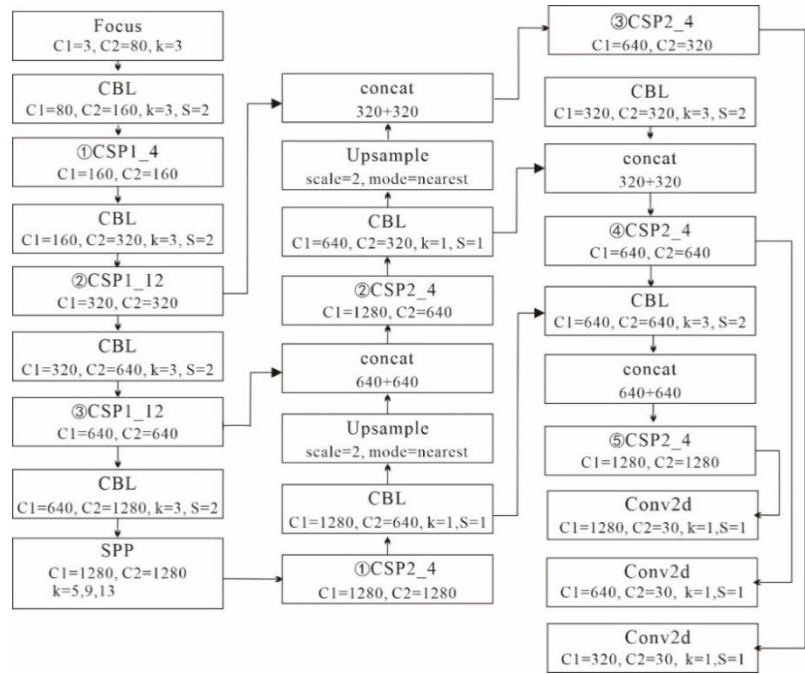
精度 P 与召回率 R 的计算方法为^[30]:

$$P = \frac{TP}{(TP+FP)} \quad (5)$$

$$R = \frac{TP}{TP+FN} \quad (6)$$

以精度 P 为纵轴, 召回率 R 为横轴作图得到 P - R 曲线, 则对于离散的 P - R 曲线, 平均精度 AP (Average Precision) 计算方法^[30]:

$$AP = \sum_{k=1}^n P(k) \Delta r(k) \quad (7)$$



C1-输入通道数; C2-输出通道数; k-卷积核尺寸; s-卷积操作中的步长; 上采样 (Upsample) 模块中的 scale=2 表示上采样倍数为 2 倍, mode=nearest 表示上采样采用最邻近插值方法。叠加 (concat) 模块中的加法运算表示叠加的两个通道数

图 8 RDB-Yolov5x 网络中超参数设置

Fig. 8. Hyperparameter setting of the RDB-Yolov5x network.

平均精度均值 mAP 计算方法为^[30]:

$$\text{mAP} = \frac{\sum_{q=1}^Q \text{AP}(q)}{Q} \quad (8)$$

4.3 实验结果分析

RDB-Yolov5x 训练后各项指标如下:

图 9 为 RDB-Yolov5x 训练结果的混淆矩阵。矩阵显示了各类矿物被正确识别的概率。混淆矩阵反映出, 识别过程中, 误识别概率较高的情况发生在石英和背景区分之时, 概率是 0.57。由于石英在花岗岩中含量较高, 在人工标注时无法将所有矿物标注。验证时, 模型仍能将未标注出的石英颗粒, 即默认为背景的石英颗粒正确识别, 此种情形仍可以被认为识别较准确。

经过 100 轮训练后, 验证集上各类目标精度指标平均精度 AP 如下: 石英 (quartz): 0.938; 白云母 (muscovite): 0.942; 黑云母 (biotite): 0.965; 斜长石 (plagioclase): 0.924; 钾长石 (potassium feldspar): 0.937, IoU

阈值在 0.5 时的平均精度均值 mAP 为 0.941。图 10 为 RDB-Yolov5x 模型与 Yolov5x 模型在验证集上 P-R 曲线的对比, RDB-Yolov5x 模型下对如小尺寸的黑云母、白云母、斜长石等矿物颗粒的识别精度有明显提高 (表 2)。且在 IoU 阈值在 0.5 时, mAP 提高 0.5%, IoU 阈值在 0.95 时, mAP 提高 1%, 说明 RDB-Yolov5x 模型对矿物颗粒的位置回归性能更加准确。

表 2 不同方法下各类矿物平均精度对比

Table 2. Average accuracy values of various minerals identified by using different methods

方法	AP					mAP	
	quartz	muscovite	biotite	plagioclase	potassium feldspar	@0.5	@0.95
RDB-Yolov5x	0.938	0.942	0.965	0.924	0.937	0.941	0.775
Yolov5x	0.940	0.940	0.961	0.915	0.926	0.936	0.765

图 11 为 RDB-Yolov5x 和 Yolov5x 的均值平均精度对比。横坐标为训练轮次, 纵坐标为均值平均精度 mAP, 图中显示出, 由于添加了密集连接结构使网络结构复杂, RDB-Yolov5x 在训练初期的收敛速度比 Yolov5x 慢, 但在 25 轮次左右时开始在精度上超越 Yolov5x, 并在 45 轮次左右时 mAP 稳定 90% 以上。

表 1 分类结果混淆矩阵

Table 1. Confusion matrix of classification results

真实情况	预测结果	
	正例	反例
正例	TP (真正例)	FN (假反例)
反例	FP (假正例)	TN (真反例)

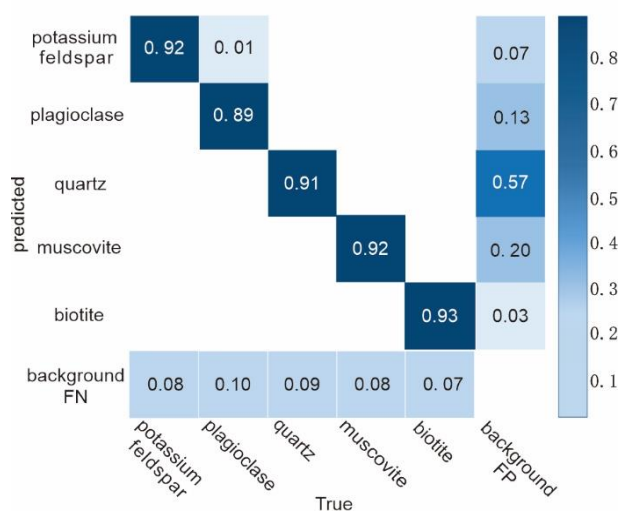


图 9 RDB-Yolov5x 模型下混淆矩阵

Fig. 9. Confusion matrix under the RDB-Yolov5x model.

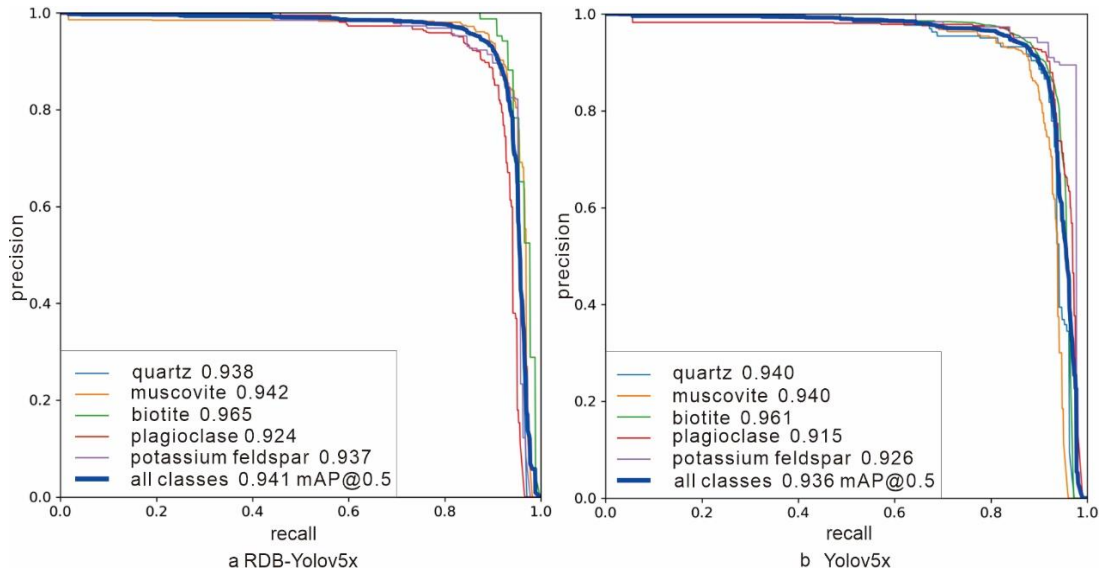


图 10 模型训练在验证集上的 P-R 曲线

Fig. 10. The P-R curves for model trainings.

4.4 模型泛化能力跨数据集验证实验

为验证模型的泛化能力, 又选取了 30 张未参与实验的花岗岩的正交偏光图像进行验证实验, 验证实验统计结果如表 3。同 Yolov5x 模型相比, RDB-Yolov5x 在基数较高的精度下, 验证集上的 mAP 仍提高 0.5%。在实际检测中, RDB-Yolov5x 也展现出了明显更好的效果。RDB-Yolov5x 模型对于边界清晰, 表面较新鲜的矿物颗粒漏检率为 11.5%, 较 Yolov5x 降低 3.7%。

从该 30 张图像中选取了 4 组具有代表性的对照图像 (图 12)。

通过上述实验和验证结果可知:

1) 较 Yolov5x 方法相比, RDB-Yolov5x 模型漏检数少, 可以对图像中的更多目标成功检测, 且改进方法的

检测框定位更准确。如图所示, 改进模型 (图 12b) 检测到了原模型 (图 12a) 未检测到的小尺寸石英颗粒及大尺寸钾长石; 改进模型 (图 12d) 检测到了原模型 (图 12c) 未检测到的斜长石。这表明 RDB-Yolov5x 模型在同一图像上存在多目标、且目标尺寸差异较大时有较好的检测性能。

2) 较 Yolov5x 方法相比, RDB-Yolov5x 模型对边界模糊, 特征不明显的矿物检测能力提高。图 12 显示, 改进模型 (图 12f) 在一张矿物信息丰富的高分辨率图像中, 能够检测出原模型 (图 12e) 未检测出的边图像边缘的石英颗粒。

3) 较 Yolov5x 方法相比, RDB-Yolov5x 模型针对图像中小尺寸矿物颗粒的检测能力更好。图 12 显示, 改进模型 (图 12h) 检测到了原始方法 (图 12g) 未检测到的图像边缘处较小的白云母颗粒。

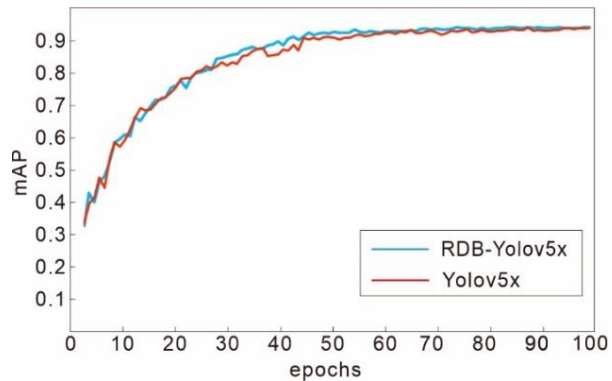


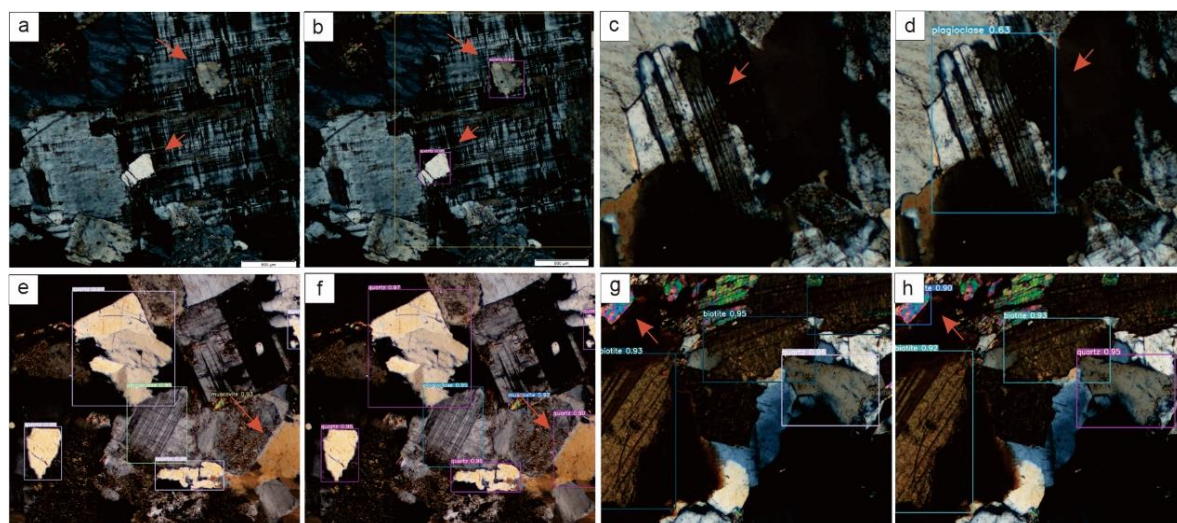
图 11 不同方法下训练 mAP 比较

Fig. 11. Comparison of mAP values in trainings by using different methods.

表 3 验证实验统计

Table 3. Validation experiment statistics

方法	检测样本数量	实际样本数量	漏检样本数量	漏检率/%
RDB-Yolov5x	382	432	50	11.5
Yolov5x	366	432	66	15.2



a. Yolov5x 未检测到钾长石、石英; b. RDB-Yolov5x 检测到钾长石、石英; c. Yolov5x 未检测到斜长石; d. RDB-Yolov5x 检测到斜长石;
e. Yolov5x 未检测到图像边缘石英; f. RDB-Yolov5x 检测到图像边缘石英; g. Yolov5x 未检测到图像边缘白云母;
h. RDB-Yolov5x 检测到图像边缘白云母

图 12 验证实验结果对照

Fig. 12. Comparison of verification experimental results by using different methods.

5 讨论与结论

花岗岩岩石薄片镜下图像矿物的智能识别是一件具有挑战性的任务。不同于常见的目标检测图像，岩石薄片图像常富含非常复杂的信息，矿物颗粒图像具有很大的识别困难。而目标检测对图像进行特征提取时，随着网络加深，高层次低分辨率的图像会具有高语义特征，具有较大感受野，但位置信息模糊，小目标丢失严重。低层次高分辨率的图像语义信息少，但目标位置准确，细节信息保存较好。这使得在薄片识别时对特征不明显的矿物颗粒，小尺寸矿物颗粒丢失严重；对边缘不清晰，表面被蚀变的矿物检测以及对这些矿物的精准位置回归效果差。

本文在 Yolov5x 算法的基础上提出的 RDB-Yolov5x 算法解决了这一问题。实验研究表明，该方法具有以下几点优势：

1) 可以准确有效地对一张花岗岩正交偏光镜下图像中的石英、钾长石、斜长石、黑云母、白云母矿物颗粒的类别和位置进行识别。

2) RDB 模块的改进有效削弱了注意力模块带来的信息频繁变动。该方法的 IoU 阈值为 0.5 时，均值平均精度 mAP 可达 94.1%，较 Yolov5x 提高了 0.5%。IoU 阈值 0.95 时，均值平均精度 mAP 可达 77.5%，较 Yolov5x 提高了 1%。

3) 方法在跨数据集上验证有效，针对小尺寸的云母，长石及石英颗粒也能够精准地检测，且在定位上可以直观地显现出优势。

4) 检测方法减少了对图像的预处理工作，模型可以将任意一张常见的花岗岩岩石薄片正交偏光镜下图像作为输入，无需经过挑选和切割裁剪，这也提高了人工智能在岩石矿物鉴定中实际应用意义。

RDB-Yolov5x 方法的建立基于花岗岩薄片的正交偏光镜下图像，但其展现出的小尺寸矿物识别优势未来也可以应用于其他矿物类型中。该方法使自动化对隐藏在大批量图像中的小尺寸、肉眼难鉴定的稀有矿物的搜寻和筛选工作成为可能，可以极大地提高地质工作者的效率和精准率。

但优化后的 RDB-Yolov5x 仍然存在一些不足之处：同许多人工智能矿物识别方法一样，模型的特征训练对图像的数据标签仍有极强依赖性，而一些人工标定数据的过程会受到个人主观因素的影响，

具有不确定性, 为检测能力的提高带来困难。

基于人工智能的岩石矿物自动识别系统可以有力地促进地球科学领域相关工作的开展, 根据矿物形态特征分析矿物结构与成分, 迅速给出辅助判断, 可以大大提高从业者的工作效率, 期待更多优秀的算法和数据集可以应用在地球科学中。

参考文献:

- [1] López A J A, Ramil A A, Pozo-Antonio J S B, et al. Automatic Identification of Rock-Forming Minerals in Granite Using Laboratory Scale Hyperspectral Reflectance Imaging and Artificial Neural Networks (Article)[J]. Journal of Nondestructive Evaluation, 2017, 36(3): 52.
- [2] Rajendran S, Hersi O S, Al-Harthy A, et al. Capability of advanced spaceborne thermal emission and reflection radiometer (ASTER) on discrimination of carbonates and associated rocks and mineral identification of eastern mountain region (Saih Hatat window) of Sultanate of Oman[J]. Carbonates and Evaporites, 2011, 26(4): 351-364.
- [3] 郭艳军, 周哲, 林贺洵, 等. 基于深度学习的智能矿物识别方法研究[J]. 地学前缘, 2020, 27(5): 39-47.
- [4] Tiwari V, Rawat A, Kumar D. An overview on object detection using Open CV - Python[J]. Journal of Interdisciplinary Cycle Research, 2020, 12(6).
- [5] Thompson S, Fueten F, Bockus D. Mineral identification using artificial neural networks and the rotating polarizer stage[J]. Computers and Geosciences, 2001, 27(9): 1081-1089.
- [6] Marmo R, Amodio S, Tagliaferri R, et al. Textural identification of carbonate rocks by image processing and neural network: Methodology proposal and examples[J]. Computers and Geosciences, 2005, 31(5): 649-659.
- [7] Singh, Naresh, Singh T.N, Tiwary, et al. Textural identification of basaltic rock mass using image processing and neural network[J]. Computational Geosciences, 2010, 14(2): 301-310.
- [8] Baykan,Nurcan Akhan,Yilmaz, Nihat. Mineral identification using color spaces and artificial neural networks[J]. Computers and Geosciences, 2010, 36(1): 91-97.
- [9] Aligholi S, Khajavi R, Razmara M. Automated mineral identification algorithm using optical properties of crystals[J]. Computers and Geosciences, 2015, 85 (Part A): 175-183.
- [10] Ashok Kumar Patel, Snehamoy Chatterjee. Computer vision-based limestone rock-type classification using probabilistic neural network[J]. Geoscience Frontiers, 2016,7(1): 53-60.
- [11] 刘延保, 曹树刚, 刘玉成. 基于 LS-SVM 的岩石细观图像分析方法探讨[J]. 岩石力学与工程学报, 2008, 27(5): 1059-1065.
- [12] 刘珏先, 滕奇志, 王正勇, 等. 基于协同表示的多特征融合岩石分类[J]. 计算机应用, 2016, 36(3): 854-858.
- [13] 杨艳梅, 柳娜, 程国建, 等. 基于 Spark 平台的岩石图像聚类分析[J]. 西安石油大学学报(自然科学版), 2016, 31(6): 114-118.
- [14] 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.
- [15] 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.
- [16] 徐述腾, 周永章. 基于深度学习的镜下矿石矿物的智能识别实验研究[J]. 岩石学报, 2018, 34(11): 3244-3252.
- [17] Xu Z H, Ma W, Lin P, et al. Deep learning of rock images for intelligent lithology identification[J]. Computers and Geosciences, 2021, 154.
- [18] 张喜松. 华南早古生代火山岩与紫苏花岗岩研究: 对陆内造山岩浆起源和结晶分异的启示[D]. 南京: 南京大学, 2020.
- [19] Zhang X, Zhao Ch, Luo HZ, et al. Automatic learning for object detection[J]. Neurocomputing, 2022, 484: 260-272.
- [20] Xie J, Pang Y W, Nie J, et al. Latent Feature Pyramid Network for Object Detection[J]. IEEE Transactions on Multimedia, 2022. 1520-9210.
- [21] Ray A, Bhardwaj A, Malik Y K, et al. Artificial intelligence and Psychiatry: An overview[J]. Asian J Psychiatr, 2022, 70: 103021.
- [22] Redmon J, Divvala S, Girshick R, et al. You only look once: Unified, real-time object detection [C]//Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2016: 779-788.
- [23] Pestana, Danie, Miranda, Pedro R, et al. A Full Featured Configurable Accelerator for Object Detection with YOLO[J]. IEEE Access, 2021, 9(1): 75864-75877.
- [24] Redmon J, Farhadi A. YOLO9000: Better, Faster, Stronger[C] // 30th Ieee Conference on Computer Vision and Pattern Recognition (CVPR 2017), 2017: 6517-6525.
- [25] Huang G, Chen D L, Li T H, et al. Multi-scale dense convolutional networks for efficient prediction [Z]. arXiv, 2017: 11.
- [26] HuangG,Liu Zh, Van DerMaaten L, et al. Densely Connected Convolutional Networks[C] // 30th Ieee Conference on Computer Vision and Pattern Recognition (CVPR 2017), 2017: 2261-2269.
- [27] Jiang L M, Fan H,Li J. A multi-focus image fusion method based on attention mechanism and supervised learning[J]. Applied Intelligence, 2022, 52(1): 339-357.
- [28] Xue J, Cheng F, Li Y, et al. Detection of Farmland Obstacles Based on an Improved YOLOv5s Algorithm by Using CIoU and Anchor Box Scale Clustering[J]. Sensors (Basel, Switzerland), 2022, 22(5): 1790.
- [29] Jiang D B. Complete explanation of core knowledge of Yolov5[EB/OL]. CSDN, 2020. <https://blog.csdn.net/nan355655600/article/details/107852353>.
- [30] Zhou Z H. Machine learning[M]. Beijing: Tsinghua University Press, 2016.