

刘毅, 高海海, 韩英杰, 等. 基于结构化剪枝的矿区地质灾害检测算法[J]. 智能计算机与应用, 2025, 15(1): 158-164. DOI: 10.20169/j.issn.2095-2163.250124

基于结构化剪枝的矿区地质灾害检测算法

刘毅¹, 高海海¹, 韩英杰¹, 张文杰², 李鹏越²

(1 华晋焦煤有限责任公司, 山西 吕梁 033000; 2 太原理工大学 电气与动力工程学院, 太原 030024)

摘要: 本文提出基于 YOLOv5s 模型的结构化剪枝目标检测算法, 解决矿区无人机巡检中常规算法过大、参数多、难以部署的问题。通过遍历网络中的 BN 层, 对 γ 进行排序, 并设定全局阈值评估通道重要性, 剔除低于阈值的通道。实验结果显示, 相较于 YOLOv5s, 该算法模型减小 52.9%, 检测时间降低 18.1%, 平均精度仅下降 1.5%。

关键词: 矿区地质灾害; YOLOv5s; 目标检测; 结构化剪枝

中图分类号: TP183; P694

文献标志码: A

文章编号: 2095-2163(2025)01-0158-07

Geological hazard detection algorithm in mining area based on structured pruning

LIU Yi¹, GAO Haihai¹, HAN Yingjie¹, ZHANG Wenjie², LI Pengyue²

(1 Huajin Coking Coal Co., Ltd, Lvliang 033000, Shanxi, China;

2 College of Electrical and Power Engineering, Taiyuan University of Technology, Taiyuan 030024, China)

Abstract: In this paper, a structured pruning target detection algorithm based on YOLOv5s model is proposed to solve the problems of excessive size, many parameters and difficult deployment of conventional algorithms in UAV inspection in mining areas. By traversing the BN layer in the network, the γ are sorted and the global threshold is set to evaluate the importance of the channel and exclude channels below the threshold. The experimental results show that compared with YOLOv5s, the algorithm model is reduced by 52.9%, the detection time is reduced by 18.1%, and the average accuracy is only reduced by 1.5%.

Key words: Geological hazards in mining areas; YOLOv5s; object detection; structured pruning

0 引言

近年来, 矿区开采规模不断扩大, 导致矿区地质灾害发生的频率逐渐增高。矿区地质灾害是由于矿床开采过程中庞大的采掘井巷导致地表破坏和岩土体变形, 以及矿区地质、水文地质环境条件和自然环境的改变而产生的灾害^[1]。矿区地质灾害不仅破坏采矿工程设备, 还对矿区的资源环境产生负面影响, 对当地居民的生命财产安全构成威胁^[2]。因此, 有效的地质灾害检测是降低损失的重要措施。

传统的地质灾害识别主要依赖于人眼的观察和判断, 这种方式存在着主观性强、效率低下等问题。随着基于卷积神经网络 (CNN) 的目标检测算法迅

猛发展, 为地质灾害识别带来了新的解决方案。深度学习算法可以通过大规模数据的训练和学习, 自动学习地质灾害目标的特征信息, 提高了识别的准确性和稳定性。其次, 深度学习目标检测技术能够实现大规模图像数据的高速处理, 大大提高了地质灾害识别的效率。Amit 等^[3]设计了一个 CNN 网络, 其中包括 3 个卷积层和两个全连接层, 该网络通过对样本进行训练, 可以提取遥感图像中的滑坡或泥石流区域的特征信息。Austin 等^[4]对多种方法进行了评估, 其中包括多层前馈神经网络、径向基神经网络和随机森林方法, 以确定其在 2010 年海地地震破坏情况检测方面的有效性。Vetrivel 团队^[5]提出了 3D 点云特征与 CNN 网络融合的分类框架来检测

基金项目: 山西省自然科学基金(202103021224056)。

作者简介: 刘毅(1968—), 男, 学士, 高级工程师, 主要研究方向: 地质灾害检测与防治; 高海海(1974—), 男, 学士, 工程师, 主要研究方向: 矿区地质灾害检测; 张文杰(1995—), 男, 博士, 讲师, 主要研究方向: 深度学习。

通信作者: 李鹏越(1990—), 男, 博士, 讲师, 硕士生导师, 主要研究方向: 计算机视觉。Email: lipengyue@tyut.edu.cn。

收稿日期: 2023-08-21

受灾建筑物的损害情况。近些年, CSPDarkNet53^[6]、ResNet101^[7] 等大型骨干网络的提出, 增加了网络的深度, 提高了特征提取能力, 但也带来了一些问题, 即深度学习模型的参数量和体积不断增加, 需要更多的计算资源。使得在边缘设备上部署深度学习模型变得困难, 尤其是在无人机等移动设备上, 由于计算资源、存储能力和功耗的限制, 大型目标检测算法的运行变得更加困难^[8]。因此, 需要寻找一种能够在计算资源受限的边缘设备上运行的轻量级目标检测算法。这样的算法能够在保持较高准确性的同时, 减少模型的参数量和体积, 提高运行效率。

综上所述, 为提高矿区地质灾害目标检测算法的准确性与实时性, 本文以 YOLOv5s 网络为基础, 提出一种基于 BN 层缩放因子的结构化剪枝方法。该方法对 BN 层的尺度缩放因子 γ 进行 L1 正则化约束, 将缩放因子推向 0, 进而实现通道稀疏化^[9]; 对尺度缩放因子进行排序, 以此来衡量通道的重要性, 通过与预先设置的全局阈值比较, 剔除尺度缩放因子较小的通道, 实现了模型剪枝。由于模型剪枝会减少大量的参数, 导致模型的精度下降, 故本文对剪枝后的模型进行微调训练来恢复精度。实验结果表

明, 剪枝后的模型相较于基础模型 YOLOv5s 体积更小, 检测单幅图像所需时间更少。

1 模型剪枝

1.1 YOLOv5 简介

YOLOv5 是 Ultralytics 公司提出的一种基于单阶段目标检测器的深度学习算法, 在目标检测领域取得了显著成就。本文采用的 YOLOv5s 网络是 YOLOv5 系列中的一种轻量级网络, 具有更小的网络规模和参数量, 模型网络结构如图 1 所示。YOLOv5 的网络结构采用了一种 Backbone+Neck+Head 的模块化设计, 其中 Backbone 负责提取图像特征, 通常基于骨干网络。如 CSPDarknet53 或 EfficientNet^[10] 等, Neck 用于融合不同尺度的特征, 以便更好地适应不同大小的目标, 而 Head 则负责生成目标检测的结果, 包括类别预测和边界框回归。YOLOv5 包括两种不同的 CSP 网络结构, 即 CSP1_X 结构和 CSP2_X 结构^[11]。同时, 与 YOLOv4^[6] 模型类似, YOLOv5 模型利用特征图金字塔网络^[12] (FPN) 和路径聚合网络来融合特征图信息, 结构如图 2 所示。

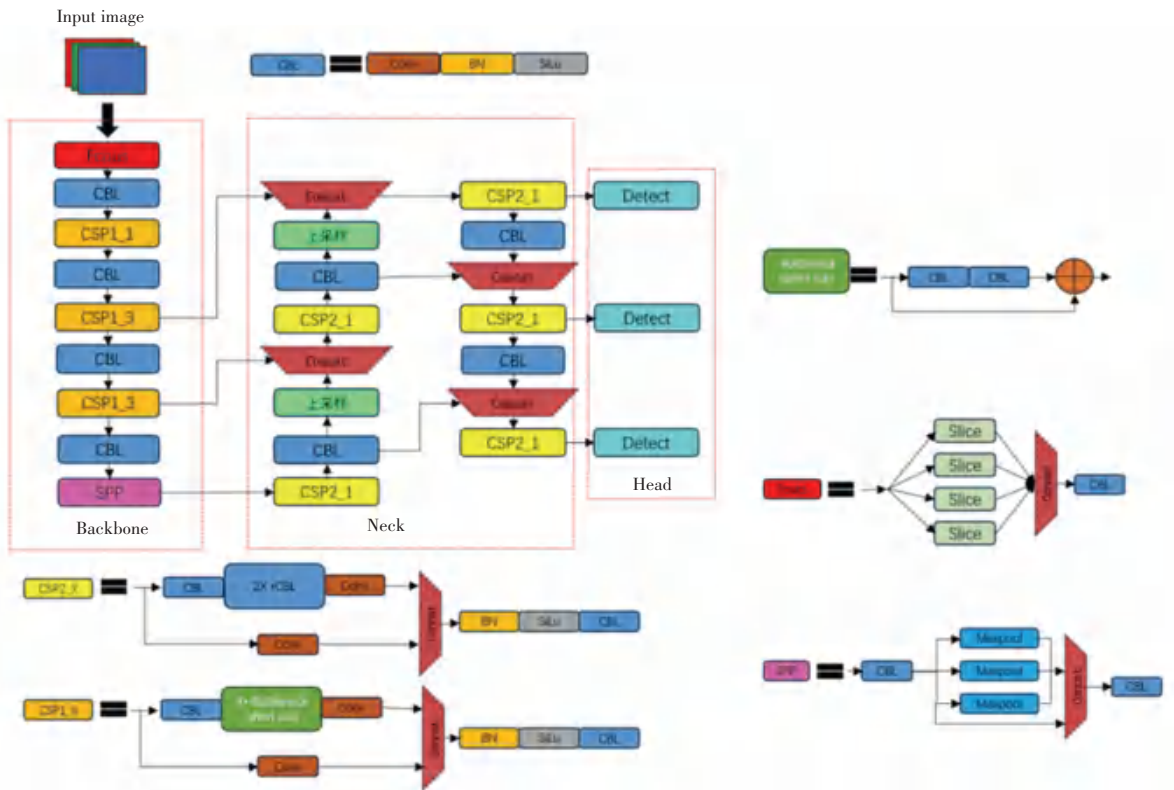


图 1 YOLOv5s 网络结构

Fig. 1 YOLOv5s network structure

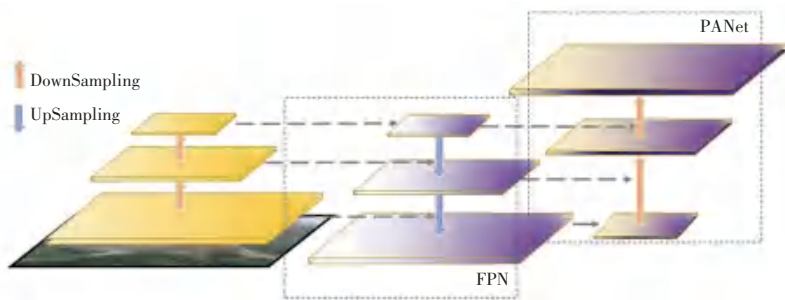


图 2 FPN 及 PANet 结构

Fig. 2 FPN and PANet structures

1.2 模型剪枝

卷积神经网络(CNN)^[13]通常是由卷积层、池化层、全连接层组成,卷积层的运算量在整个神经网络的计算过程中占很大比例,随着网络的加深,卷积层数量的增加,网络参数的冗余程度变高。而模型剪枝是一种用于减小深度学习模型规模的技术^[14],通过去除冗余参数和连接来实现模型的压缩和加速,使得模型在算力有限的边缘设备上完成部署。目前,模型的剪枝方法主要分为结构化剪枝和非结构化剪枝两种^[15]。非结构化剪枝是通过对权重矩阵中的权重值进行修剪,将要修剪的值设置为 0,使得

权重矩阵变为稀疏矩阵。结构化剪枝是一种粗粒度的剪枝方式^[16],主要是对滤波器和通道进行修剪。通过设置规则来评价通道的重要性,将一些冗余的通道移除,从而实现剪枝。由于非结构化剪枝对硬件平台和计算库的性能有一定的要求,因而本文采用结构化剪枝。

剪枝的具体流程如图 3 所示。首先对 YOLOv5s 模型进行基础训练;然后对模型的权重文件进行稀疏化训练;之后将稀疏化的模型进行结构化剪枝操作;接着将剪枝后的模型进行微调训练;最后输出剪枝完成的模型。



图 3 剪枝流程

Fig. 3 Pruning process

1.3 基于 BN 层缩放因子的结构化剪枝

在对模型剪枝之前,首先要在训练过程中加入稀疏化。稀疏化可以在神经网络中的不同地方实现,例如权重、通道、层级等。然而,权重稀疏虽然可以使网络的灵活性和通用性提高,但是需要一些硬件加速器来实现,对于层级的稀疏,由于修剪的部分较多,灵活性不高。因此,本文采用具有一定灵活性且方便实现的通道稀疏,来对模型进行修剪。要实现通道稀疏,需要修剪与通道相连接的一些通路,例如输入、输出连接,由于输入输出连接端上的权重值不一定都接近于 0,所以本文在稀疏化的同时加入 L1 正则化,以便于将通道的连接全部推向于 0,从而实现通道稀疏化。本文做法是为每个通道引入一个缩放因子 γ ^[17] 乘以通道的输出,联合训练权重和缩放因子,然后对缩放因子较小的通道进行修剪,具体的损失函数如下式所示。式中,左侧一项是 CNN 中的正常的训练损失值,右侧一项是对缩放因子的稀

疏惩罚项, λ 是为了平衡左右两项。

$$L = \sum_{(x,y)} l(f(x,W),y) + \lambda \sum_{\gamma \in \Gamma} g(\gamma) \quad (1)$$

批量归一化(Batch Normalization, BN)已经被广泛用于现代 CNN 中,可以加快网络的训练速度,提高模型的稳定性^[18]。BN 层可以描述如下:

$$\hat{z} = \frac{z_i - \mu}{\sqrt{\sigma^2 + \epsilon}} \quad Z_o = \gamma \hat{Z} + \beta \quad (2)$$

其中, z_i 是 BN 层的输入; μ 和 σ 是输入的平均值和标准差; ϵ 是防止标准差为 0 而加入的正则化参数; γ 和 β 是可训练的仿射变换参数; z_o 是经过 BN 层后的输出。

在此,本文直接将 BN 层中的缩放因子 γ 作为网络剪枝所需要的缩放因子,因为其不需要引入额外的参数,减少资源的浪费。具体来讲,就是每个通道都引入一个 BN 层的尺度缩放因子 γ , 与通道的输出相乘,并对其进行 L1 正则化。通过遍历网络中

所有的 BN 层,对所有的尺度缩放因子 γ 进行排序。为了避免过度剪枝,本文设置了一个剪枝的安全阈值,即 $percent = 0.4$ 。这意味着本文方法至少保留 60% 的通道,以确保模型的稳定性和性能。通过 $percent$ 参数值与 γ 的总数相乘,得到一个 γ 的位置索引,将该位置获取的值作为全局阈值 δ 。通过 γ 与 δ 的比较,对尺度缩放因子 γ 小于全局阈值 δ 的通道进行剔除。通道剪枝示意图如图 4 所示,蓝色和橙色分别为第 i, j 卷积层的特征图,经过剪枝模块,对网络进行剪枝处理。其中,橙色的箭头是经剪枝处理后仍连接的通路,绿色的箭头是经过剪枝后剔除的连接。

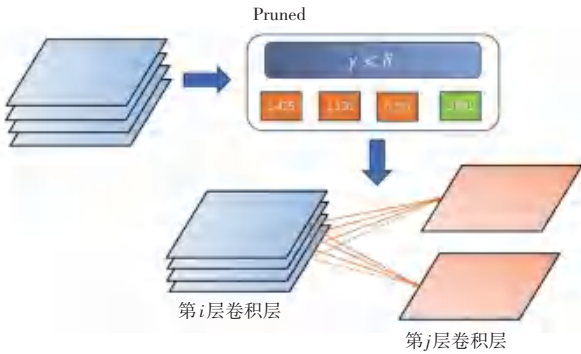


图 4 通道剪枝示意图

Fig. 4 Schematic diagram of channel pruning

在 YOLOv5s 模型的剪枝处理过程中,本文注意到骨干网络 (Backbone) 中 CSP 模块的 BottleNeck 存在残差连接,而预测部分 (Head) 中则不存在。由于通道数相同才可以进行相加操作,为了确保 CSP 结构中的 Otleneck 的张量维度可加 (如图 5),本文不对残差结构进行剪枝处理。

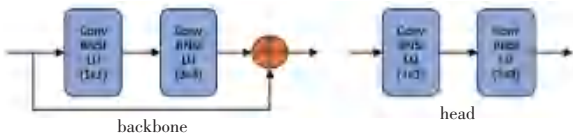


图 5 CSP 结构中的 BottleNeck 结构图

Fig. 5 BottleNeck structure diagram in the CSP structure

通过这样的剪枝策略,可以减少模型的参数量和计算开销,同时保持足够的模型表达能力,能够实现高效的模型压缩和加速,同时在保持较高准确性的前提下减少模型的复杂度,提高推理速度。

2 实验结果分析

2.1 数据集与实验参数设置

当前阶段,人们面临着矿区地质灾害数据集缺乏的挑战。缺乏数据集的原因主要有两方面,一是

获取这些数据往往需要投入大量的人力和时间进行地质调查和采样,其次矿区地理环境的复杂性也增加了数据获取的难度。同时,一些矿区企业对其地质灾害数据持保留态度,考虑到隐私和安全问题,限制了数据的公开和共享。

鉴于上述情况,为克服这些挑战,本文自建了一个专门用于地质灾害目标检测的数据集。其中,一部分图片来自于矿区巡查时,使用大疆 M350RTK 无人机并搭载赛尔 PSDK 102S 五镜头相机采集所得,另一部分是从公开的数据集中获取的与地质灾害相关的高质量图像。这些图像包括卫星图像、航拍图像、现场摄影图像,其目的是可以提供具有挑战性的数据样本,以帮助本文更好地评估和改进地质灾害的目标检测算法的鲁棒性和泛化能力。为保证数据集的准确性和可靠性,本文进行了严格的数据筛选。首先排除了低质量的图像样本,然后使用精灵标注助手,对每个图像的地质灾害进行标注,用于识别和定位地质灾害的类别和位置。为了适应 YOLO 模型的需求,将 XML 文件格式的标注结果转换为符合 YOLO 模型要求的 txt 文本格式。

本文将自建的数据集按照 7 : 3 的比例划分为训练集和验证集,以便于目标检测算法的训练和评估及测试。自建数据集中图像的分辨率为 $1\ 200 \times 900$,包含矿区地质灾害常见的 3 个类别:泥石流、塌陷和滑坡。其中,泥石流有 305 个样本,塌陷有 618 个样本,滑坡有 1 486 个样本。图 6 展示了训练集中的部分样本。从这些样本中可以看出,本文的数据集 (见表 1) 涵盖了不同的地质灾害类别,并且包括了不同季节的地质灾害图像数据。

实验使用了 Intel (R) Xeon (R) CPU E5-2620 v4 CPU 和 TITAN XP 显卡作为实验平台。设置训练轮数为 100, Batch-size 为 32, 动量衰减系数为 0.937, 权重衰减系数为 0.000 5, 初始学习率为 0.001, 优化器选择 Adam, 稀疏训练的全局阈值和稀疏系数分别为 0.4 和 0.001。本文首先对数据进行预处理,将图像缩放至 512×512 以适应 YOLOv5 模型的输入要求;其次,由于采集到的地质灾害图像数量有限,所以对样本进行了数据增强操作,包括随机裁剪、水平变换、颜色转换以增强数据的多样性和模型的鲁棒性。在数据预处理完成后,本文使用 YOLOv5s 的预训练权重作为基础参数来对地质灾害数据集进行了训练,并在训练过程中使用 Adam^[19] 优化器,该策略能够根据训练过程自适应调整学习率,有助于模型更好地收敛。



图6 训练集地质灾害样本

Fig. 6 Training set geological hazard samples

表1 数据集样本类型

Table 1 Dataset sample type

地质灾害类别	训练集	验证集	总计	标签
泥石流	1 185	301	1 486	Debrisflow
滑坡	243	62	305	Landslide
塌陷	507	111	618	Sinking

2.2 模型评价指标

模型评价指标是用来衡量机器学习模型在解决特定任务上的性能和表现的度量标准,本文采用平均精确率均值、参数量与浮点运算量,来衡量地质灾害目标检测模型的准确性和效率。

平均精度均值 (mAP) 是一种综合评估模型性能的指标,用于评估目标检测器在不同目标类别上的性能。其能够综合考虑模型在所有类别上的准确性和召回率, mAP 值越高表示模型在多个类别上表现越好。平均精度均值的计算公式如下:

$$mAP = \frac{1}{n} \sum_{i=1}^n AP_i \quad (3)$$

其中, n 为检测的种类, AP_i 是单一类别的精度。

AP 的计算公式如下:

$$AP = \int_0^1 P(r) dr \quad (4)$$

P 表示精确率,计算公式如下:

$$P = \frac{TP}{TP + FP} \times 100\% \quad (5)$$

其中, TP 表示模型正确地检测到了实际存在的目标个数, FP 表示模型将不存在的目标误认为存在的个数。

参数量是指模型可训练参数的数量,模型的权重文件越小,所包含的参数量越少。浮点运算量 (GFLOPs) 是浮点运算次数,反应模型的复杂程度,模型越复杂,浮点运算量越大。

2.3 实验结果定量分析

在完成剪枝处理后,相较于 YOLOv5s 模型,本文的模型 Trim-Net 的参数量减少了 53.31%;模型的权重大小减少了 52%;浮点运算次数减少 26.6%。在检测精度方面,为了改善由于通道减少带来的精度下降问题,本文使用剪枝后的模型权重和模型配置文件对其进行了 100 轮次的微调训练,使 mAP 达到了 78.6%,并且 Trim-Net 模型在单张图片的检测时间上减少了 0.4 ms,大幅降低了计算资源的消耗。不同模型的各项指标见表 2。

表2 不同模型性能对比

Table 2 Performance comparison of different models

模型名称	mAP / %	推理耗时/ ms	GFLOPs	权重大小/ MB	参数量
YOLOv3-Tiny	60.1	3.4	13.0	17.4	8 674 496
YOLOv5s	79.8	2.2	16.5	14.4	7 068 936
Trim-Net	78.6	1.8	12.1	6.9	3 300 313

实验结果表明,剪枝后的模型 Trim-Net 在矿区地质灾害测试数据集上表现与 YOLOv5s 模型效果相当。本文的剪枝模型 Trim-Net 在减少模型参数量和检测速度方面取得了显著的成果,相较于 YOLOv5s 模型,本文的模型权重文件更小,参数更少,能够更好地部署在算力较小的边缘设备上,提高运算效率与实时性。这些结果表明,剪枝技术在优化深度学习模型方面具有重要的应用潜力。

2.4 实验结果定性分析

在模型的训练过程中,损失函数的变化如图 7 所示,展示了在基础训练过程中损失函数的变化曲线。这些曲线反映了模型在训练过程中的优化情况,本文可以通过观察损失函数的变化趋势来评估模型的训练效果。根据图 7 的结果,可以观察到在经过 100 轮的基础训练后,损失值基本收敛。

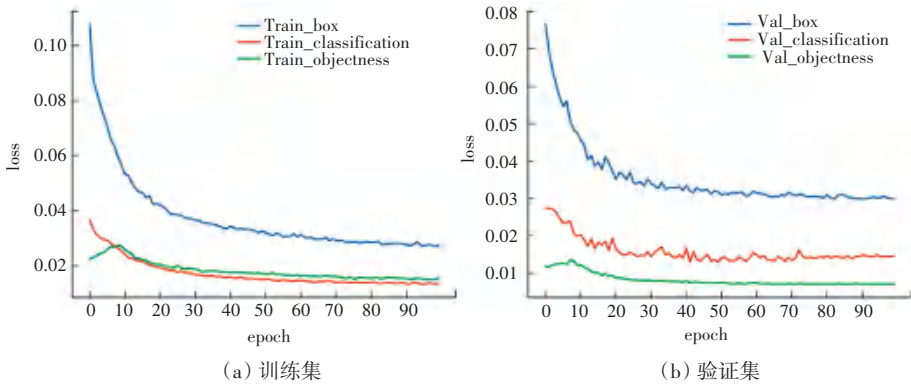


图 7 训练过程中损失函数变化情况

Fig. 7 Loss function changes during training

为了评估 Trim-Net 模型的性能, 本文使用 YOLOv5s 模型和 YOLOv3-tiny 模型, 对矿区不同的地质灾害进行检测, 检测效果如图 8 所示。第一行的图像为地质灾害原始图像, 从左到右分别为泥石流、滑坡、塌陷, 第二行的图像为 YOLOv3-tiny 模型的检测结果, 第三行的图像为 YOLOv5s 模型的检测结果, 最后一行为 Trim-Net 模型的检测结果。

两种地质灾害进行检测时更有优势。对比 YOLOv5s 模型和 Trim-Net 模型的检测效果, 本文提出的 Trim-Net 模型在置信度上虽不如 YOLOv5s, 但在检测速度方面明显高于 YOLOv3-tiny 与 YOLOv5s 模型, 并且该模型参数量更少, 模型权重文件更小, 可以减少计算资源的消耗。

3 结束语

在使用目标检测算法对矿区地质灾害进行检测的过程中, 由于常规的目标检测算法存在参数量过多, 模型复杂等问题, 本文提出了一种基于 YOLOv5s 网络的模型剪枝算法。该算法能够在保持较高准确率的同时, 显著降低硬件设备的计算成本, 并加快检测速度。经过实验测试, 与原始 YOLOv5s 模型相比, 该剪枝模型的参数量减少了 53.31%, 而平均精度仅下降了 1.5%。

这种基于剪枝模型的检测算法为利用无人机搭载的边缘设备实时监测矿区地质灾害提供了理论和技术上的支持。通过减少模型参数量和计算量, 本文可以在有限的硬件资源下实现高效的目标检测。同时, 该算法的实时性能使其具备在实际应用中进行实时监测的能力。

总之, 基于 YOLOv5s 网络剪枝模型的检测算法为矿区地质灾害监测^[20]提供了一种高效、准确且实时的解决方案, 并具有应用于无人机等边缘设备的潜力。这将极大地降低人工成本, 提高监测效率, 为矿区安全管理提供有力支持。

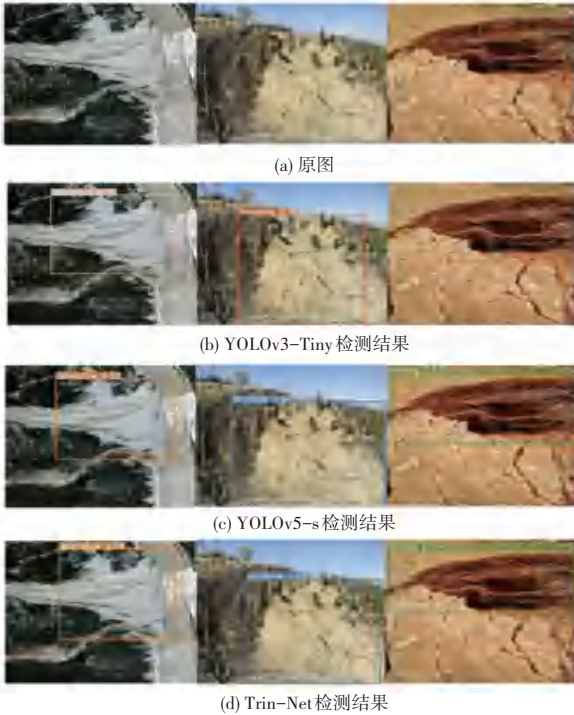


图 8 不同模型对地质灾害的检测效果

Fig. 8 Detection effect of different models on geological hazards

通过对不同模型的检测结果分析, 所有的模型均能够识别出泥石流、滑坡、塌陷这 3 种矿区常见的地质灾害。在对泥石流的检测中, YOLOv3-tiny 模型相较于 YOLOv5s 和 Trim-Net 模型更加准确, 而 Trim-Net 模型和 YOLOv5s 模型在对滑坡和塌陷这

参考文献

[1] 刘晓玲, 魏奥林, 王毅, 等. 浅析陕北煤矿矿区地质灾害发育特征及其成灾过程[J]. 中国地质灾害与防治学报, 2016, 27(4): 70-73.
 [2] 张永庭. 矿山地质灾害智能监测与评价[D]. 北京: 中国地质

- 大学(北京),2022.
- [3] AMIT S. SHIRAIISHI T. INOSHITA T, et al. Analysis of satellite images for disaster detection[C]//Proceedings of the 2016 IEEE International Geoscience and Remote Sensing Symposium. Piscataway,NJ;IEEE,2016; 5189-5192.
- [4] AUSTIN J C, YANG S, JAMES C. Detection of urban damage using remote sensing and machine learning algorithms: Revisiting the 2010 Haiti earthquake[J]. Remote Sensing,2016, 8(10): 868.
- [5] VETRIVEL A, GERKE M, KERLE N, et al. Disaster damage detection through synergistic use of deep learning and 3D point cloud features derived from very high resolution oblique aerial images, and multiple-kernel-learning[J]. ISPRS Journal of Photogrammetry and Remote Sensing, 2018, 140: 45-59.
- [6] BOCHKOVSKIY A, WANG C Y, LIAO H Y M. Yolov4: Optimal speed and accuracy of object detection[J]. arxiv preprint arxiv,2004.10934, 2020.
- [7] HE K, ZHANG X, REN S, et al. Deep residual learning for image recognition[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Piscataway,NJ;IEEE, 2016: 770-778.
- [8] 魏昊坤,刘敬一,陈金勇,等. 基于L-FPN的无人机上小目标识别模型轻量化方法[J]. 航空兵器,2024,31(1):97-102.
- [9] LIU Z, LI J, SHEN Z, et al. Learning efficient convolutional networks through network slimming[C]//Proceedings of the IEEE International Conference on Computer Vision. Piscataway, NJ: IEEE, 2017: 2736-2744.
- [10] TAN M, LE Q V. Efficientnet: Rethinking model scaling for convolutional neural networks[C]// Proceedings of International Conference on Machine Learning. PMLR, 2019: 6105-6114.
- [11] 王慧赢. 基于深度学习的车辆未礼让行人的违规行为判定方法研究[D]. 石家庄: 河北科技大学,2022.
- [12] LIN T Y, DOLLÁR P, GIRSHICK R, et al. Feature pyramid networks for object detection[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Piscataway,NJ;IEEE,2017; 2117-2125.
- [13] LECUN Y, BOTTOU L, BENGIO Y, et al. Gradient-based learning applied to document recognition[J]. Proceedings of the IEEE, 1998, 86(11): 2278-2324.
- [14] 毛远宏,曹健,贺鹏超,等. 深度神经网络剪枝方法综述[J]. 微电子学与计算机,2023,40(10):1-8.
- [15] 姜晓勇,李忠义,黄朗月,等. 神经网络剪枝技术研究综述[J]. 应用科学学报,2022,40(5):838-849.
- [16] 李屹,魏建国,刘贯伟. 模型剪枝算法综述[J]. 计算机与现代化,2022,325(9):51-59.
- [17] HE Y, DONG X, KANG G, et al. Asymptotic soft filter pruning for deep convolutional neural networks[J]. IEEE Transactions on Cybernetics, 2019, 50(8): 3594-3604.
- [18] IOFFE S. Batch normalization: Accelerating deep network training by reducing internal covariate shift[J]. arXiv preprint arXiv,1502.03167, 2015.
- [19] KINGMA D P. Adam: A method for stochastic optimization[J]. arXiv preprint arXiv,1412.6980, 2014.
- [20] 龙林丽,刘英,张旭阳,等. 无人机在矿区表土特征及地质灾害监测中的应用[J]. 煤田地质与勘探,2021,49(6):200-211.