

# 基于多特征融合的露天矿区道路负障碍检测

阮顺领<sup>1,2,3</sup>, 鄢盛钰<sup>1,2</sup>, 顾清华<sup>1,2,3</sup>, 卢才武<sup>1,3</sup>

(1. 西安建筑科技大学 资源工程学院, 陕西 西安 710055; 2. 西安建筑科技大学 矿山系统工程研究所, 陕西 西安 710055; 3. 西安市智慧工业感知计算与决策重点实验室, 陕西 西安 710055)

**摘要:**随着智慧矿山概念的逐步落实, 智能化、无人化逐渐在矿区落实, 露天矿卡车无人驾驶日益成为矿山智能化建设的主要内容, 为解决露天矿区复杂多变的道路因坑洼、塌陷等路面小部分下陷出现的非规则负障碍而导致矿区无人车、重载卡车侧翻等安全难题, 提升矿区安全驾驶系数, 提出一种多特征融合的露天矿区道路负障碍检测方法。该方法使用 BiFPN 特征融合模块, 提高小型负障碍检测权重占比; 引入空间和通道双注意力机制提高对负障碍边缘的特征提取和特征融合能力, 从而提高对道路小尺度负障碍的检测精度; 采用 SIOU Loss 作为模型边界框损失函数并使用 K-means++ 方法优化 Anchor 以提高负障碍检测模型的收敛速度和边界框定位效果, 并基于遗传算法优化超参数让模型更贴合矿区场景, 最终实现对矿区道路负障碍的快速精准识别。实验表明该检测模型能快速准确识别复杂背景下露天矿区道路负障碍目标, 对道路负障碍目标的检测精度、召回率、平均精确度均值分别达到了 96.9%、89.9%、95.3%, 且该模型大小仅有 12.7 MB。对比其他主流检测网络, 该网络模型更适合复杂环境下露天矿区非结构化道路行驶安全需求, 且该检测模型的鲁棒性好, 可适配于多种情况的露天矿区, 为实际环境复杂多变的露天矿区非结构化道路负障碍检测提供了可行的方法, 为露天矿无人卡车安全生产运输提供安全预警。

**关键词:** 露天矿区; 无人卡车; 负障碍检测; 特征融合; 安全预警

中图分类号: TD804 文献标志码: A 文章编号: 0253-9993(2024)05-2561-12

## Negative obstacle detection on open pit roads based on multi-feature fusion

RUAN Shunling<sup>1,2,3</sup>, YAN Shengyu<sup>1,2</sup>, GU Qinghua<sup>1,2,3</sup>, LU Caiwu<sup>1,3</sup>

(1. School of Resource Engineering, Xi'an University of Architecture and Technology, Xi'an 710055, China; 2. Institute of Mine System Engineering, Xi'an University of Architecture and Technology, Xi'an 710055, China; 3. Xi'an Key Laboratory of Perceptual Computing and Decision Making for Smart Industry, Xi'an 710055, China)

**Abstract:** With the gradual implementation of intelligent mine concept, intelligence and unmanned operation are gradually implemented in mining area. Unmanned driving of open pit mine trucks is increasingly becoming the main focus of intelligent mine construction. In order to solve the safety problems of the overturn of unmanned vehicles and heavy-duty trucks due to irregular negative obstacles appearing in some parts of road surface such as potholes and collapses in open pit mines, and to improve the safe driving coefficient in mines, a multi-feature fusion method of detecting negative obstacles in open pit mine roads is proposed. The method uses the BiFPN feature fusion module to improve the weight proportion of small-scale negative obstacle detection, introduces the spatial and channel dual attention mechanism to improve the feature extraction and feature fusion ability of negative obstacle edges, so as to improve the detection accuracy

收稿日期: 2023-04-25 修回日期: 2023-06-26 责任编辑: 王凡 DOI: 10.13225/j.cnki.jccs.2023.0539

基金项目: 国家自然科学基金资助项目 (52074205); 陕西省自然科学基金资助项目 (2022JM-201)

作者简介: 阮顺领 (1981—), 男, 河南周口人, 教授, 博士。E-mail: ruanshunling@xauat.edu.cn

通讯作者: 鄢盛钰 (1999—), 男, 陕西安康人, 硕士研究生。E-mail: yanshengyu@xauat.edu.cn

引用格式: 阮顺领, 鄢盛钰, 顾清华, 等. 基于多特征融合的露天矿区道路负障碍检测[J]. 煤炭学报, 2024, 49(5): 2561-2572.

RUAN Shunling, YAN Shengyu, GU Qinghua, et al. Negative obstacle detection on open pit roads based on multi-feature fusion[J]. Journal of China Coal Society, 2024, 49(5): 2561-2572.



移动阅读

of small-scale negative obstacles on the road. Also, the SIOU Loss is adopted as the loss function of the model bounding box, the Anchor by using the K-means++ method is used to improve the convergence speed and boundary frame localization effect of the obstacle detection model, the hyperparameters are optimized based on genetic algorithm to make the model more suitable for the mine scene, and finally the fast and accurate recognition of negative obstacles on the mine road is realized. The experiments show that the detection model can quickly and accurately identify the negative road obstacle targets in the complex background of the open pit mine, and the detection accuracy, recall rate, and mAP of the negative road obstacle targets reach 96.9%, 89.9%, and 95.3%, respectively, and the size of the model is only 12.7 MB. Compared with other mainstream detection networks, the network model is more suitable for the safety needs of unstructured road driving in open pit mining areas under complex environment, and the robustness of the detection model is good, which can be adapted to a variety of situations in open pit mining areas, providing a feasible method for the detection of negative obstacles on unstructured roads in open pit mining areas where the actual environment is complex and variable, and providing some safety warnings for the safety of unmanned trucks in open pit mines.

**Key words:** open pit mines; unmanned trucks; negative obstacle detection; feature fusion; safety warning

随着矿山智能化建设的快速发展,无人驾驶技术在矿区的应用也日益成熟,但由于露天矿区非结构化道路时常出现路面坑洼、塌陷、凹陷积水等负障碍,这些障碍特征与道路特征相似度较高,在无人矿卡行进的过程中很难发现,很容易导致重载矿卡的车身倾斜和颠簸,并给无人矿卡的生产运输带来极大安全隐患,因此亟待研究一种能对矿区负障碍进行准确、快速自动检测的新方法。

国内外部分学者已对结构化道路和非结构化道路上的负障碍进行了一些研究。刘家银等<sup>[1]</sup>提出一种基于多激光雷达与组合特征的方法,设计了一种具有互补能力的多激光雷达安装方式。其次,提出了基于幅向局部凸性和后沿壁局部密集特征的64线雷达负障碍物特征点对检测方法,以及基于径向距离跳变和后沿壁局部密集特征的32线雷达负障碍物特征点对检测方法。汪佩等<sup>[2]</sup>提出一种基于单线激光雷达和单目视觉的负障碍检测算法,对检测到的负障碍区域在摄像机画面中进行跟踪,结合跟踪结果对负障碍区域做进一步判别。蔡云飞等<sup>[3]</sup>把双多线激光雷达数据映射到多尺度栅格,统计栅格点云密度与相对高度等特征并标记,对其特征点进行聚类并过滤之后识别出负障碍。笔者<sup>[4]</sup>对露天矿区道路负障碍特征进行深入分析,构建了基于机器视觉的轻量化目标检测模型,其召回率、平均精确度均值、检测精度均明显高于当时的主流网络。LI等<sup>[5]</sup>通过自定义算式判断负障碍距离,将3D点云转变成二维高程栅格图像,基于图像滤波技术和点云去噪技术提出了一种基于高程栅格图像的几何特征估计方法,以此来判断负障碍信息。

露天矿非结构化道路因重型卡车碾压容易出现路面下陷的情况。例如露天矿装载区,不停的改变道路结构使路面易出现坑洼,尤其是雨天过后,露天矿

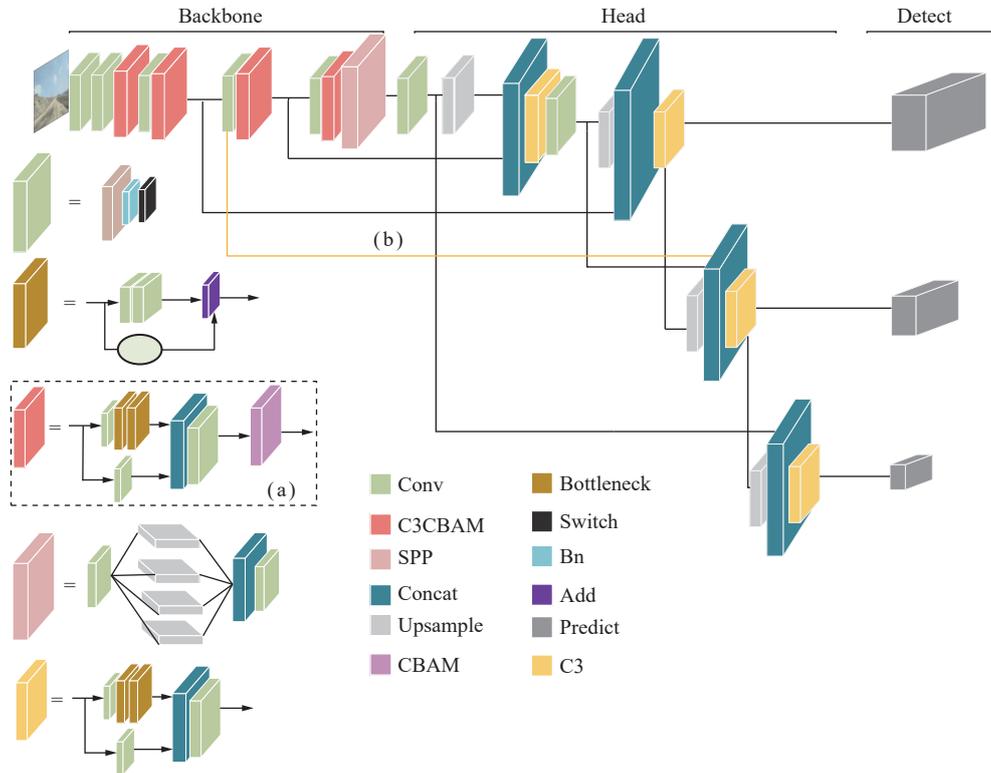
非结构化道路变的泥泞,在碾压过后的道路出现坑洞、水坑等负障碍易使卡车出现轮胎下陷,甚至出现翻车的可能性。笔者基于以上论文分析并结合露天矿非结构化道路特点,提出了一种基于改进YOLOv5(You Only Look Once v5)多特种融合的负障碍检测方法,使露天矿区非结构化道路负障碍检测更加高效、精准和智能。为了使负障碍检测模型更注重路面负障碍信息,笔者改进了原模型的特征提取模块,在每个特征提取模块后加入CBAM<sup>[6]</sup>卷积注意力机制,提高网络对图像重要特征的提取。为了检测到更小更精细的负障碍,笔者修改FPN<sup>[7]</sup>的连接方式,将池化金字塔修改为加权双向池化金字塔BiFPN<sup>[8]</sup>(Bidirectional Feature Pyramid Network),增加网络对小目标检测的能力。选择SGD为优化器,SIOU<sup>[9]</sup>作为边界回归损失函数并且使用Kmeans++<sup>[10]</sup>方法对笔者数据集进行聚类生成新的Anchor,通过动态的学习率防止模型拟合,提升算法准确度,使用遗传算法<sup>[11]</sup>为数据找到更适配的超参数,最终实现对复杂环境下露天矿区道路负障碍的快速、精准和智能识别。

## 1 矿区道路负障碍检测优化模型

露天矿区非结构化道路负障碍尺寸跨度、形状不规则,具有多尺度特征、障碍边界不清晰、多小目标等特点。为了解决无人车在矿区前进的安全问题,并结合YOLOv5<sup>[12-15]</sup>在露天矿山非结构化道路小型负障碍检测的不足,提出了基于改进YOLOv5的矿区负障碍检测模型。露天矿区负障碍道路检测模型如图1所示。其中,针对矿区负障碍小目标检测不到的问题,笔者提出增加BiFPN小目标检测模块。在第3个预测层中加入第6层的特征图,并在网络获得不同尺度的特征图后增大对小特征的权重,提高网络对负障碍

小目标的检测。其次针对矿区负障碍边缘信息不明显, 容易与正常道路搞混淆的问题, 笔者在原特征提取模块后引入 CBAM(Convolutional Block Attention Module) 机制。通过引入通道注意力机制和空间注意力机制提高网络模型对负障碍边缘特征的学习, 加权重要的负障碍特征, 减少不必要冗余, 模型参数变少且

检测精度更高。而针对负障碍检测模型梯度收敛较慢问题提出引入 SIOU 作为新的边框损失回归函数, 加快网络学习速率, 进一步提升网络精度。笔者为使检测框更适配矿区负障碍, 提出修改原网络 Anchor, 使其更适配被检测目标, 提高模型平均精确度均值。最后对原模型的学习率和超参数做出调整, 优化模型学习过程。



注: (a) 优化为 C3CBAM 模块; (b) 引入 BiFPN 优化模块。

图 1 负障碍检测模型结构

Fig.1 Negative barrier detection model structure

### 1.1 小尺度负障碍检测模块

露天矿区非结构化道路负障碍存在外形尺寸不规则、多小坑等特点, 尤其使距离卡车较远时的小型负障碍, 很难被及时的被检测出来, 其特征点不易被察觉且包含的图片语义信息过少。在负障碍检测网络训练的过程中负障碍的尺寸不一, 特征点不同, 导致在对负障碍特征信息提取的过程中更大的负障碍目标可以保留更多的特征, 而较小的负障碍目标则会丢失语义信息甚至检测不到小型负障碍。针对小型的负障碍检测问题, 笔者在 YOLOv5 的 FPN 结构处引入 BiFPN 的特征融合方式。提高模型对多个小目标特征的融合能力, 增加小目标的相关权重, 使形状较小的目标被检测的精度和准确度更高。FPN 等结构如图 2 所示。

传统的 FPN 结构只有自上而下的单项的信息流, 而 YOLOv5 结合了 PAnet<sup>[16]</sup>网络的特性。在 FPN 的基础上额外增加了一条自底向上的路径进行信息的

增强虽然能将不同特征层进行有效融合, 但其本质上仍是将不同特征进行简单相加。然而由于在不同的图像中被检测目标大小不一, 因此训练中会产生不同分辨率尺寸的特征。在路径聚合网络中仍将它们简单相加, 这将导致同一类型的不同尺寸特征对融合后输出的特征产生不平等的权重。大尺寸的特征被更多的融入网络而小尺寸特征贡献较小。相较于 FPN-PAN 结构来说, BiFPN 是加权双向特征金字塔结构。针对融合的不同尺寸的特征图增加了独特的权重, 动态学习调节每个尺度的贡献, 使网络将其更好的融合。同时增加残差连接的模块以增强其特征的表达能力以更好的读取语义信息。BiFPN 的计算公式为

$$\begin{cases} P_6^{td} = \text{Conv} \left[ \frac{\omega_1 P_6^{in} + \omega_2 \text{Resize}(P_7^{in})}{\omega_1 + \omega_2 + \varepsilon} \right] \\ P_6^{out} = \text{Conv} \left[ \frac{\omega'_1 P_6^{in} + \omega'_2 P_6^{td} + \omega'_3 \text{Resize}(P_5^{out})}{\omega'_1 + \omega'_2 + \omega'_3 + \varepsilon} \right] \end{cases} \quad (1)$$

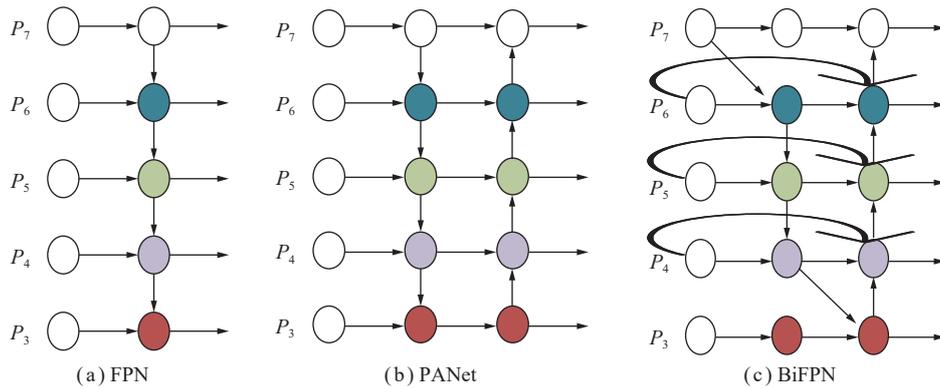


图 2 特征融合方式结构

Fig.2 Structure diagram of feature fusion mode

式中,  $P_6^d$  为自上而下路径中第 6 层的中间特性;  $P_6^{out}$  为自下而上路径中第 6 层的输出特性;  $P_5^{out}$ 、 $P_6^{in}$ 、 $P_7^{in}$ 、 $\varepsilon$ 、 $\text{Resize}$  分别为第 5 层特征的输出、第 6 层的输入、第 7 层的输入、无穷小量和一个上下采样的操作;  $\omega_1$ 、 $\omega_2$ 、 $\omega'_1$ 、 $\omega'_2$ 、 $\omega'_3$  为不同的权重。

### 1.2 负障碍特征提取模块优化

露天矿区道路负障碍的颜色与无障碍道路并无很大区别,且负障碍的边缘处并没有明显界限,这使检测模型在露天矿区负障碍近处难以进行高效精准的识别。针对露天矿区负障碍信息边缘特征不明显,检测模型难以捕捉到特点信息的问题,笔者提出了一种融合了空间和通道注意力机制的新模块 C3CBAM,使模型在检测过程中对负障碍的信息变得更敏感。负障碍检测模型在多重卷积融合后的特征图处添加通道注意力机制和空间注意力机制。让网络更好的学习图片中重要特征的权重,减少非必要参数。在为模型带来剪枝操作的同时增加其精度、平均精确度均值等。C3CBAM 结构主要图解如图 3 所示。

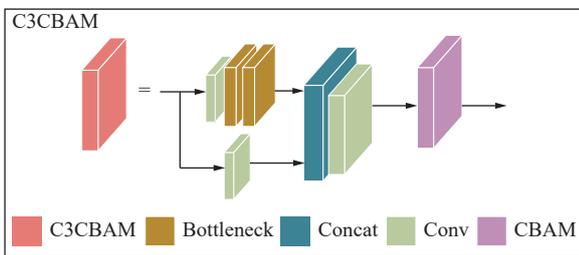


图 3 特征提取优化 C3CBAM 结构

Fig.3 Feature extraction optimization C3CBAM structure diagram

其中 CBAM 表示卷积模块的注意力机制模块,是一种结合了空间和通道的注意力机制模块。如图 4 为 CBAM 模块的图解。CBAM 主要由 CAM 和 SAM 两个模块组成,分别代表了通道注意力模块和空间注意力模块。输入的负障碍特征图会先进入通

道注意力和空间注意力,之后输出新的特征图。其中  $\otimes$  表示元素乘法。

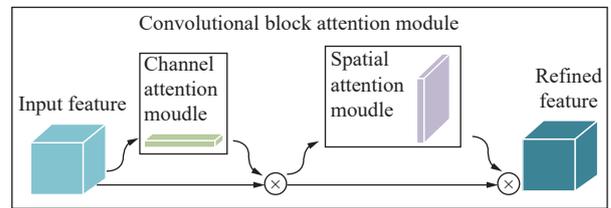


图 4 CBAM 注意力机制结构

Fig.4 CBAM attention mechanism structure chart

通道注意力模块让通道维度不变,压缩空间维度,结构如图 5 所示。该模块关注负障碍图片中的负障碍边界等有独特特征意义的信息。模块将输入的负障碍特征图经过 2 个并行的最大池化层和平均池化层,将特征图的高和宽变成  $1 \times 1$  的大小,然后经过 Share MLP 模块。在该模块中,它先将通道数压缩为原来的  $1/r$  ( $r$  为减少率) 倍,再扩张到原通道数。经过 ReLU 激活函数得到 2 个激活后的结果。将这 2 个输出结果进行逐元素相加,再通过一个 sigmoid 激活函数得到通道注意力的输出结果。再将这个输出结果乘原图,变回  $C \times H \times W$  的大小 ( $C$  为 Channel,  $H$  为 Height,  $W$  为 Wide)。其通道注意力公式为

$$\begin{cases} M_c(F) = \text{sigmoid}(\text{MLP}[\text{AvgPool}(F)] + \text{MLP}[\text{MaxPool}(F)]) \\ M_c(F) = \text{sigmoid}(W_1[W_0(F_{\text{avg}}^c)] + W_1[W_0(F_{\text{max}}^c)]) \end{cases} \quad (2)$$

式中,  $F$  为输入的特征图;  $\text{sigmoid}$  为损失函数;  $\text{AvgPool}(F)$  和  $\text{MaxPool}(F)$  为特征图经过平均池化和最大池化;

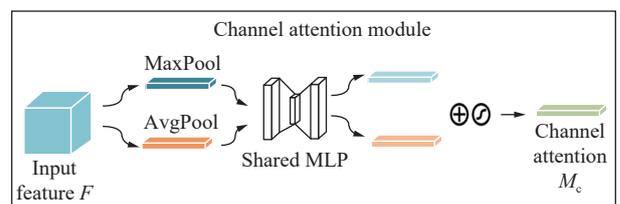


图 5 CAM 通道注意力模块结构

Fig.5 CAM channel attention module structure

$W_0$ 、 $W_1$ 为权重;  $M_c$ 为经过通道的注意力;  $F_{avg}^c$ 为经过通道注意力机制并采取平均池化的特征图;  $F_{max}^c$ 为经过通道注意力机制并采取最大池化的特征图;  $M_c(F)$ 为让特征图  $F$  进入通道注意力机制; MLP 为多层感知器。

空间注意力模块让空间维度不变, 压缩通道维度, 结构如图 6 所示。该模块更关注负障碍相对于正常道路的位置信息。将通道注意力的输出结果通过最大池化和平均池化得到 2 个  $1 \times H \times W$  的特征图。之后对 2 个特征图进行拼接。通过  $7 \times 7$  卷积变为 1 通道的特征图, 之后经过 sigmoid 函数得到空间注意力的特征图。最后将输出结果乘原图变回  $C \times H \times W$  大小。其空间注意力表示为

$$\begin{cases} M_s(F) = \text{sigmoid}(f^{7 \times 7} [\text{AvgPool}(F); \text{MaxPool}(F)]) \\ M_s(F) = \text{sigmoid}(f^{7 \times 7} [F_{avg}^s; F_{max}^s]) \end{cases} \quad (3)$$

式中,  $f^{7 \times 7}$  为  $7 \times 7$  的特征图;  $s$  为空间;  $M_s$  为基于空间的注意力;  $M_s(F)$  为特征图经过空间注意力机制;  $[F_{avg}^s; F_{max}^s]$  为空间注意力中经过平均池化和最大池化的特征图。

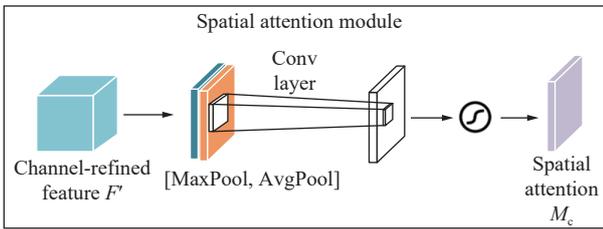


图 6 SAM 空间注意力模块结构

Fig.6 SAM spatial attention module structure diagram

### 1.3 检测模型损失函数优化

为了使露天矿区负障碍检测模型训练加速并提高模型检测准确率, 使模型更快的学习到负障碍特征, 故对边框回归损失函数进行优化。损失函数通过对候选框做边框回归, 在多个候选预测框中选择精度最高的, 使最终检测到的目标定位更加接近真实值, 提高定位准确率, 提升检测精确度。原模型使用 DIoU (Distance-IoU)<sup>[17]</sup> 作为损失函数, 其通过对预测框纵横比添加惩罚因子, 再判断不同预测框和真实框的欧氏距离来确定最终预测框。但当其纵横比值为 1 时惩罚因子失效, 定位精度下降, 模型准确率会因此降低。而 SIoU<sup>[9]</sup> 通过重新定义角度、距离、预测框与真实框的位置关系, 解决了当预测框的长宽一样时出现的精度和训练速度下降的问题。提高了模型训练速度和推理的准确性。针对露天矿区负障碍, 可以使模型更快学习到负障碍信息。SIoU 主要由 4 个 Cost 函数组

成分别为 Angle cost、Distance cost、Shape cost 和 IoU cost。其中 Angle cost 对角度进行了重新定义, 其角度 cost 公式为

$$\begin{cases} \Lambda = 1 - 2\sin^2\left(\arcsin x - \frac{\pi}{4}\right) \\ x = \frac{c_h}{\sigma} = \sin \alpha \end{cases} \quad (4)$$

$$\sigma = \sqrt{(b_{c_x}^{gt} - b_{c_x})^2 + (b_{c_y}^{gt} - b_{c_y})^2} \quad (5)$$

$$c_h = \max(b_{c_x}^{gt}, b_{c_x}) - \min(b_{c_x}^{gt}, b_{c_x}) \quad (6)$$

其中,  $\sigma$  为 2 个框的中心点之间  $x$  和  $y$  的距离;  $c_h$ 、 $c_w$  为根据 2 个框中心点距离形成的矩形的高和宽;  $\Lambda$  为角度数值;  $\alpha$  为检测框  $B$  和真实框  $B^{GT}$  之间的水平夹角;  $b$  为检测框  $B$  的中心点;  $b_{c_x}$ 、 $b_{c_y}$  为  $B$  中心点  $x$  值和  $y$  值;  $b^{gt}$  为真实框  $B^{GT}$  中心点;  $c_h$  为两中心点之间形成矩形的高。考虑到角度的更新, 笔者也将距离进行了重新定义:

$$\begin{cases} \Delta = (1 - e^{\gamma\rho_x}) + (1 - e^{\gamma\rho_y}) \\ \rho_x = \frac{b_{c_x}^{gt} - b_{c_x}^2}{c_w}, \rho_y = \frac{b_{c_y}^{gt} - b_{c_y}^2}{c_h}, \gamma = 2 - \Lambda \end{cases} \quad (7)$$

其中,  $e$  为自然数;  $\gamma$  为随着角度的增加, 距离值具有给定的时间优先级数值;  $\Delta$  为距离损失, 其越大, 距离成本越高, 检测框精度越低;  $\rho_x$  为  $B$  和  $B^{GT}$  两框中心点  $x$  值与两中心点形成的矩形宽之比,  $\rho_y$  为  $B$  和  $B^{GT}$  两框中心点  $y$  值与两中心点形成的矩形高之比;  $c_w$  为两中心点之间形成矩形的宽。

$\Omega$  名为 shape cost, 其作为优化检测的依据数值是唯一的。式 (8) 中的  $\theta$  数值控制着对 Shape cost 的关注度, 其数值设为 1, 它将立即优化一个 Shape。  $\omega_w$ 、 $\omega_h$  为 2 个相交预测框不重叠部分的宽高比, 对于 Shape cost 和  $\omega_w$ 、 $\omega_h$  的定义为

$$\begin{cases} \Omega = (1 - e^{-\omega_w})^\theta + (1 - e^{-\omega_h})^\theta \\ \omega_w = \frac{|w - w^{gt}|}{\max(w, w^{gt})}, \omega_h = \frac{|h - h^{gt}|}{\max(h, h^{gt})} \end{cases} \quad (8)$$

整个 IoU Cost 的定义为

$$L_{IoU\ Cost} = 1 - IoU \quad (9)$$

$$IoU = \frac{|B \cap B^{GT}|}{|B \cup B^{GT}|} \quad (10)$$

其中, IoU 为真实框和预测框的交集;  $L_{IoU\ Cost}$  为预测框和真实框的交并比;  $w$ 、 $h$  为真实框的宽和高;  $w^{gt}$ 、 $h^{gt}$  为预测框的宽和高。最后, 整个回归损失函数和总损失函数为

$$L_{\text{box}} = 1 - \text{IoU} + \frac{\Delta + \Omega}{2} \quad (11)$$

$$L = W_{\text{box}} L_{\text{box}} + W_{\text{cls}} + L_{\text{cls}} \quad (12)$$

式中,  $L_{\text{box}}$  为预测框的损失值;  $L$  为全局损失值;  $L_{\text{cls}}$ 、 $W_{\text{cls}}$ 、 $W_{\text{box}}$  分别为全局损失、框和分类损失权重。 $W_{\text{cls}}$ 、 $W_{\text{box}}$  是用遗传算法<sup>[11]</sup>解析出来的。

#### 1.4 负障碍目标定位模块优化

露天矿区非结构化道路负障碍具有尺寸跨度大、形状不规则等特点,像位于拍摄位置远近不一的负障碍,其大小和形状都不相同。适配负障碍图像的先验框可以提升模型学习速度、提高模型的准确率。在原算法中,模型使用默认锚框作为先验框,之后和真实框进行对比得到预测框。原模型基于 coco 数据集设定了 3 个层面,一共 9 组默认锚框。考虑到路面负障碍中各个目标大小相差较大,为获得更适用于矿区道路负障碍图像的锚框,笔者使用  $k$ -means++ 方法对锚框进行聚类。其具体方法为:

(1) 将被检测负障碍数据的宽和高作为一个数组,将整个训练集中所有数组组成集合  $A$  并指定  $k$  个聚类中心。

(2) 从集合  $A$  中任选一组宽和高作为初始聚类点  $c_1$ 。计算  $A$  中所有样本与已选定聚类点之间的最短距离记为  $D(x)$ 。计算每一个点成为下一个聚类中心的概率  $P(x)$ , 其计算方式为

$$P(x_q) = \frac{D(x_q)^2}{\sum_{q=1}^n D(x_q)^2} \quad (13)$$

式中,  $n$  为集合  $A$  中的所有点个数;  $q$  为集合  $A$  中的任意一点,取值为 1 到  $n$ 。

(3) 重复步骤 (2) 直到找到  $k$  个聚类中心。

(4) 对于集合  $A$  中的每个数据  $x_q$ , 计算其到  $k$  个聚类中心的 IoU 距离,并将其划分至距离最小的聚类中心所属类别。

(5) 重新选择聚类中心  $c_1$ , 并重复步骤 (2)~(4) 直至聚类中心不再发生变化。

通过以上方式获得的锚框会更好的贴合露天矿区负障碍的轮廓。 $k$ -means++ 聚类后得到的新锚框见表 1。它可以对负障碍做出更加精准判断,让预测框更加贴合被检测物,小幅度的提升检测精度。

表 1 Anchor 框分配

Table 1 Anchor box allocation

特征图像素尺寸	感受野	锚框像素尺寸		
		大障碍	中型障碍	小障碍
20×20	大	72×134	155×252	382×382
40×40	中	3×420	12×869	332×204
160×160	小	3×171	8×234	210×117

#### 1.5 检测模型学习率动态优化

露天矿区复杂的道路环境,已经采集的负样本数据比较有限,不能对矿区的整体道路环境进行概括。为使修改模型更加适配露天矿区负障碍检测,提出利用遗传算法 (GA) 对超参数进行优化。图 7 为优化后的部分超参数,分别代表了学习率、余弦退火学习率参数、学习率动量。其中学习率代表为模型初始学习率。余弦退火学习率参数代表模型阶段性学习种对学习率重置后的大小,学习率动量代表模型使用的优化器的配比数值。图中点越密集,颜色越偏向浅黄色的地方代表数据集更适应这些数值的超参数。它们会被选取作为新的超参数数值。

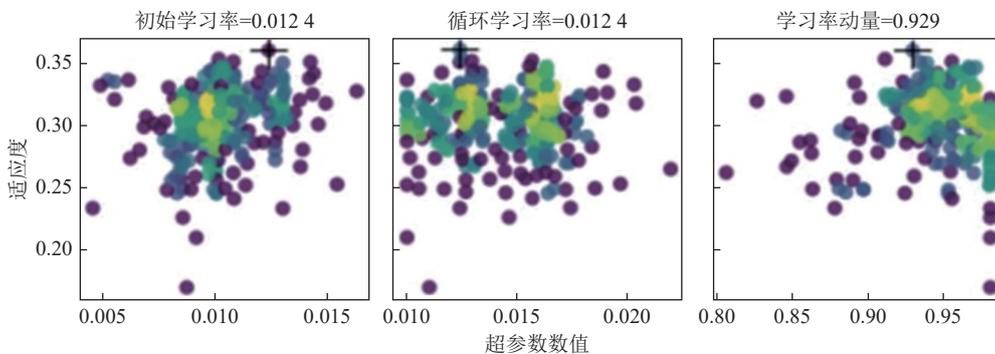


图 7 动态优化详细参数

Fig.7 Detailed parameter diagram of dynamic optimization

其中余弦退火学习率<sup>[18]</sup>优化较为重要。它主要通过动态调整不同阶段的学习率对模型学习的过程进行优化。初始阶段设置较大学习率,在模型多次训

练后因训练参数减少逐渐减小学习率,在训练多次后再增大至初始学习率使模型冲出马鞍面。这种下降模式能和学习率配合,以一种较为有效的计算方式来

产生很好的效果。余弦退火公式为

$$\eta_t = \eta_{\min}^i + \frac{1}{2}(\eta_{\max}^i - \eta_{\min}^i) \left[ 1 + \cos \left( \frac{T_{\text{cur}}}{T_i} \pi \right) \right] \quad (14)$$

式中,  $i$  为第几次训练的索引值;  $\eta_{\max}^i, \eta_{\min}^i$  分别为学习率的最大值和最小值;  $T_i$  为当前已经执行了多少轮数的训练数目;  $\eta_t$  为余弦退火学习率在等  $t$  次迭代时的值;  $T_{\text{cur}}$  为当前执行训练的轮数。

## 2 实验与结果分析

### 2.1 数据集构建及实验准备

实验数据来源于 3 个大型露天矿区, 包括河南洛阳某大型露天矿装载区道路图像、新疆某大型露天矿主干运输道路图像、内蒙古某大型露天煤矿道路图像。数据均使用 canonEOS80d 数字相机, 采集不同的光照

条件和不同的拍摄角度下的负障碍信息图像共计 814 张, 图片像素为 1 080×1 080。路面的负向障碍因雨水冲刷导致的积水与路面坑洞有着较大的特征差异。所以本数据集将负障碍分为坑洞与积水 2 类, 在训练过程中背景作为单独一个类别参与训练。因此本模型实际上是一个三分类目标检测问题。为了提升露天矿区负障碍检测的精度以及实用性, 提高负障碍模型的泛化能力, 文章对初始数据进行了 5 种方式的数据增强, 包括水平和垂直方向上的翻转、镜像、改变 HSV 空间、加入高斯噪声、锐化。通过数据扩充的方式可以增加负障碍图像的可变性, 从而提高负障碍检测的精度并提升检测模型的鲁棒性, 使检测模型适配于环境不一的露天矿区。不同方式的增强效果如图 8 所示。



图 8 露天矿区负障碍数据增强

Fig.8 Enhancement of negative obstacle data in Open-pit Mining Area

将增强后得到的 4 070 张矿区道路负障碍图像尺寸固定为 1 080×1 080, 以便模型通过原始大小图学习相关特征。将数据增强的图像与原图像混合后得到 4 884 张图像, 之后将数据集按照 8 : 2 的比例分成训练集和测试集, 数据集的具体划分细节见表 2。

本实验的环境配置部署在云端服务器上, 其操作系统为 Linux, 显卡型号为 NVIDIA RTX3090, Python 版本为 3.9, Pytorch 版本为 1.8。网络训练的输入尺寸设置为 640×640, 批处理大小设置为 8, 优化器选择为 SGD, 设置最大迭代次数为 300。笔者使用精度

表 2 数据增强前后样本分析

Table 2 Sample analysis before and after data enhancement

类别	坑洞/张	积水/张	照片数量/张
训练数据增强前	298	353	651
训练数据增强后	1 788	2 118	3 906
测试数据增强前	74	89	163
测试数据增强后	444	534	978

(Precision,  $P$ )、召回率 (Recall,  $R$ )、平均正确率 (Average Precision,  $P_{\text{av}}$ )、所以平均准确度均值 (Mean Average Precision,  $P_{\text{mav}}$ ) 作为主要评判手段

$$\left\{ \begin{array}{l} P = \frac{T_p}{T_p + F_p} \\ R = \frac{T_p}{T_p + F_n} \\ P_{av} = \frac{1}{S} \sum_1^n PR \\ P_{mav} = \frac{1}{n} \sum_1^n P_{av} \end{array} \right. \quad (15)$$

式中,  $T_p$  指预测值和真实值一样, 预测值为正样本 (真实值为正样本);  $F_p$  指预测值和真实值不一样, 预测值为正样本 (真实值为负样本);  $F_n$  指预测值和真实值不一样, 预测值为负样本 (真实值为正样本);  $S$  为检测到的障碍数量;  $n$  为检测的种类数。

## 2.2 负障碍检测模型有效性分析

为了验证改进模型的性能, 笔者使用现在的主流网络与笔者模型对比。实验的详细结果见表 3。其中 Pre 为检测精度, 指模型给出的所有预测结果中, 正确预测的结果所占的比例, 其数值越大说明正确检测的类别准确度约稿。R 为召回率, 指模型在预测正确

的结果中, 其原始的标签与检测出标签的交并比例, 其数值越大说明检测出来的检测框越贴合实际数值。mAP@0.5 指在 IoU 为 0.5 时计算每一类的所有图片的  $P_{av}$ , 之后对所有类别求平均的结果, 其数值越大代表检测精度越高。速度是指在模型实时运行过程中每秒可检测的帧数大小, 其单位为 fps/s, 其数值越大代表 1 s 中可检测的图片数量越多。GFLOPs(Giga Floating-point Operations Per Second) 意为 10 亿次 1 s 的浮点运算数, 代表了模型运行时需要的运算量, 其数值越大说明模型结构所需的运算数量越大, 代表其所需参数量更多且模型结构越复杂。考虑到不同模型的实际效果不同, 笔者对 5 种模型进行负障碍应用实验。详细对比如图 8 所示。其中 YOLOv7<sup>[19]</sup> 为原模型, YOLOv6<sup>[20]</sup> 为 s 模型。它们和改进后的 YOLOv5s 模型均训练了 300 轮次。SSD<sup>[21]</sup> 的主干网络由 VGG16<sup>[22]</sup> 实现, 训练 400 轮取其最好的权重文件。Fast-rcnn<sup>[23]</sup> 的主干网络由 Resnet50<sup>[24]</sup> 实现, 训练 100 轮取其最好的权重文件。

表 3 主流检测网络模型性能对比

Table 3 Performance comparison of mainstream detection network models

模型	P/%	R/%	$P_{mav}/%$	模型大小/MB	速度/(fps · s <sup>-1</sup> )	GFLOPs
YOLOv5-s	93.6	89.1	93.7	13.7	67.0	16.5
YOLOv6-s	93.7	89.1	93.4	36.2	56.0	45.3
YOLOv6-n	93.2	86.4	92.7	9.3	74.0	11.4
YOLOv7	91.0	81.7	85.4	74.8	53.0	103.3
YOLOv7-x	88.2	83.7	86.2	142.1	41.0	188.2
SSD-Vgg16	94.3	63.1	85.8	91.1	25.5	110.6
Fsat-rcnn-rsnet	86.2	79.2	75.4	108.0	8.9	139.8
Ours	96.9	89.9	95.3	12.7	70.0	14.2

通过上述实验可知, 笔者改进模型在统一化的评判指标下出色的完成了任务。在同样的数据集训练的情况下, 改进过后的 YOLOv5 模型拥有更高的准确性、更少的参数和更小的模型。相较于单阶段目标检测的 YOLOv6 和 YOLOv7。它们模型参数量和模型太大, 并不能很好的适应于实时检测的矿山负障碍检测上。相较于 3 阶段目标检测的 SSD、Fast-rcnn。它们模型的精度高, 但是模型的体积较大。并且 SSD 和 Fast-rcnn 作为 3 阶段目标检测网络模型速度要比单目标检测模型慢, 无法实现矿山无人车在行进过程中的实时检测要求。

如图 9 所示为不同的 5 个模型对应的对矿区负障碍检测图, 其中图 9(a)、(b) 为坑洞图; 图 9(c)、(d)

为水坑图; 图 9(e)、(f) 为复杂环境图; 图 9(g)、(h) 为多目标图; 图 9(i)、(j) 为复杂光源图。由图 9 可知, 在图 9(e)、(f) 中 YOLOv7、SSD 和 YOLOv6-s 出现了漏检。在图 9(g)、(h) 中 SSD 出现了错检, 把石头识别成了水坑, YOLOv6-s 将坑洞识别成了水坑, 而 Fast-rcnn 出现了个别漏检。在图 9(i)、(j) 中 Fast-rcnn 出现了错检, 其他 3 种模型均出现了漏检的情况。综合数据和实测后判断, 在同等数据集训练的情况下, 笔者对于 YOLOv5 的改进模型效果最好, 且模型体积最小, 检测速度最快, 检测精度最高。

## 2.3 负障碍检测模型性能优化分析

为了验证模型的改进的有效性, 笔者对所有改进模块进行消融实验, 总体模块的消融实验具体效果见

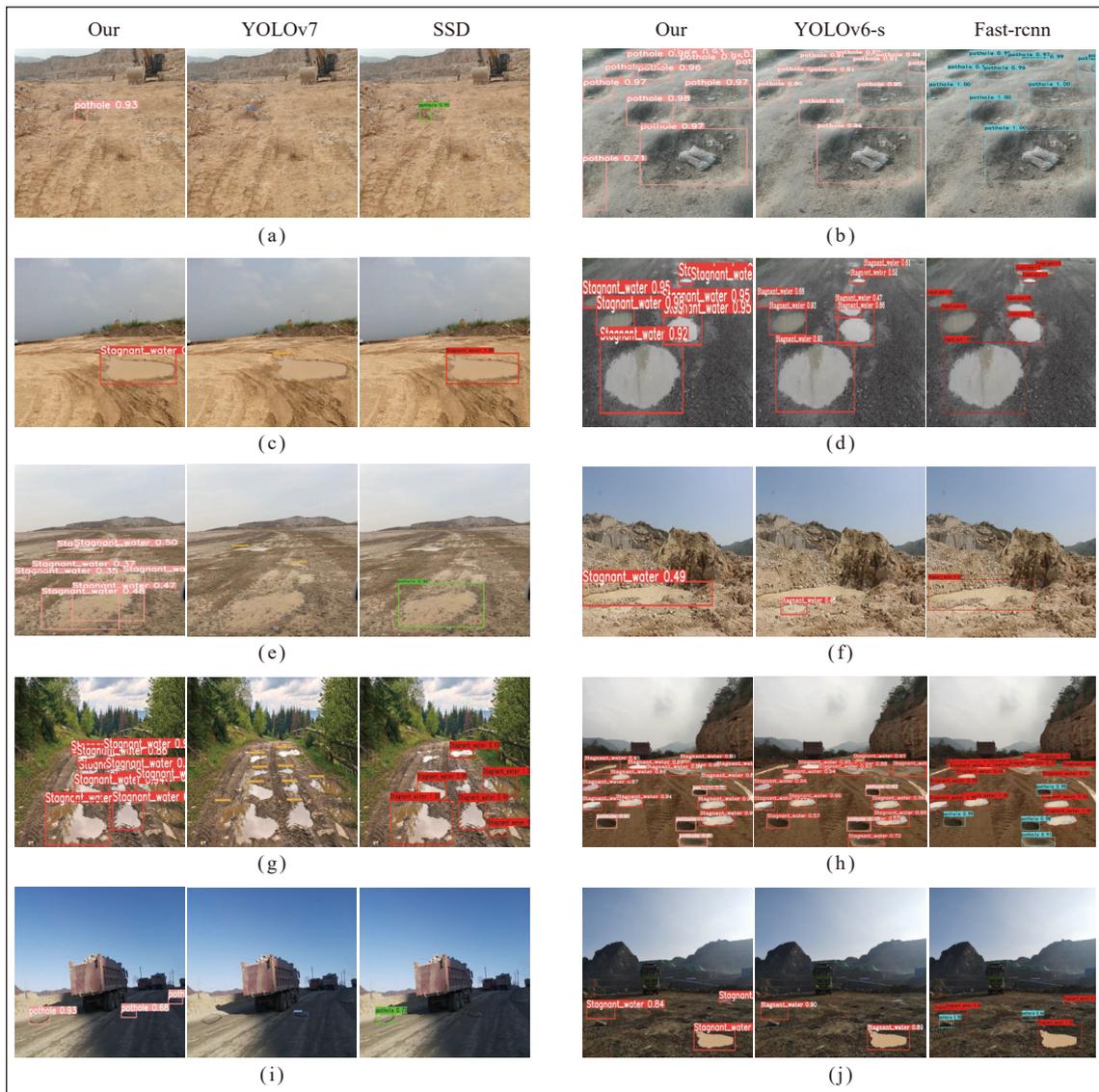


图 9 多模型矿区负障碍检测

Fig.9 Detection of negative obstacles in multi model mining area

表 4。

序号 1 为没有任何改动的 YOLOv5s 模型, 通过迁移学习 YOLOv5s.pt 文件的权重进行学习。笔者最终对模型的修改提升了 3.3% 的精确度, 召回率

表 4 总体模型消融实验表

Table 4 Overall model ablation experiment table

序号	New-hyp	Anchor	Siou	Bifpn	C3CBAM	P/%	R/%
1						93.6	89.1
2	√					93.9	89.7
3	√	√				94.6	86.5
4	√	√	√			95.7	88.2
5	√	√	√	√		94.8	90.5
6	√	√	√	√	√	96.9	89.9

仅上升 2.8%。前面针对边框损失函数、Anchor 等方式的修改虽然没有改变模型的大小, 却在训练时使其梯度下降得更快, 边框更适合被检测物体。对于增加 BiFPN 的修改相比没有改动的序号 1 精度有所提升, 但相对于序号 4 却有所降低。这是因为在修改模型 FPN 的连接模式之后, BiFPN 所带有的权重进入了各个特征层。让图像在卷积操作后不仅能记住大目标的特征, 加强了权重的小目标特征也会被检测到。虽然精度有所下降, 但是对于露天矿山无人车的安全行进检测来说小目标的检测也是必不可少的。

### 2.3.1 模型主干结构消融实验

为了验证不同主干模块给模型带来的影响, 笔者以 YOLOv5s 模型为基础进行修改, 分别添加相对应模块, 其效果见表 5。

表 5 不同模块性能对比

Table 5 Performance comparison of different modules

模型	P/%	R/%	$P_{max}/%$	模型大小/MB
YOLOv5-s	93.6	89.1	93.7	13.7
YOLOv5-Bifpn	94.1	89.8	93.9	14.1
YOLOv5-C3CBAM	96.0	90.7	95.1	11.6
YOLOv5-C3CBAM-Bifpn	96.3	89.9	95.1	12.7

为使用 YOLOv5 仅修改相关模块后的消融实验,修改 BiFPN 增加模型的大小,但更好地注意了小目标层的语义信息,对其赋予更大的权重。对于 C3CBAM 模块来说,在传统特征提取模块后加入 CBAM 机制,

可以降低其使用参数,并且使特征图注意主要的地方,提高重要特征的权重。不仅可以提升模型检测精度,也可以抛弃部分冗余参数。

图 10 为 YOLOv5 结构的 FPN-PAN 特征融合模块和 BiFPN 特征融合模块对于水坑检测的 Grad-CAM 对比图。从图 10 可以看出原结构对于中心区域的特征比较敏感但对边缘区域的信息热度较低。反观 BiFPN 结构,对于整个水坑的信息热度更高,且边缘区域的热度远高于原模型,对于非目标区域的信息热度较低。结合对于负障碍物的检测精准度和注意力热力图,判断 BiFPN 结构更好的适配于矿区负障碍检测场景。

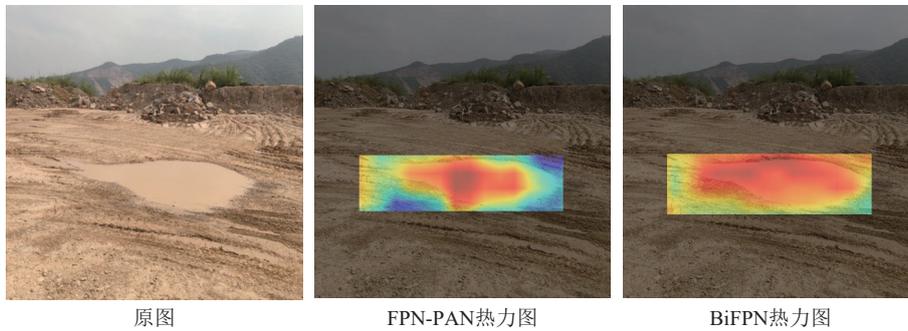


图 10 FPN-PAN 和 BiFPN 热力图对比

Fig.10 Comparison of FPN-PAN and BiFPN heat maps

2.3.2 模型定位消融实验

为了对比非模型改动的模块对于露天矿区负障碍模型的影响,笔者对原模型进行了参数部分模块的消融实验,其具体结果见表 6。

表 6 不同参数优化对比

Table 6 Comparison of optimization of different parameters

模型	P/%	R/%	$P_{max}/%$
YOLOv5-s	93.6	89.1	93.7
YOLOv5-New-hyp	93.9	89.7	94.7
YOLOv5-SIoU	95.5	88.1	95.0
YOLOv5-Anchor	94.3	86.4	93.4
YOLOv5-New-hyp-SIoU	95.6	88.9	94.8
YOLOv5-New-hyp-Anchor	94.7	88.4	94.2
YOLOv5-SIoU-Anchor	95.7	87.3	92.8
YOLOv5-New-hyp-SIoU-Anchor	95.7	88.2	94.7

由表 6 可以看出,在对 YOLOv5 进行参数优化后,准确率都有明显提升。改变超参数的提升并不高,但对于露天矿区非结构化道路的场景来说,其对于 Recall 和 mAP 的提升有利于整体模型的精确度,可以精确负障碍检测的具体定位。而将边框回归损失函数

由 DIoU 变成 SIoU 之后,训练时模型的梯度下降的很快,反应在表中的变化为其精度数值有了大幅度的提升,提高了负障碍信息的分类准确度。改变 Anchor 的准确率提升并不明显,但是在预测图片时,预测框可以更好的贴合所检测的负障碍,提供了更精确的障碍信息。

图 11 为在训练期间,不同的 Anchor 带来的不同的边框损失数值。蓝色代表使用 kmeans++ 聚类而成的新 Anchor,粉色代表了适用于 coco 数据集的 YOLOv5 的 Anchor。修改过后的 box\_loss 数值明显低于修改之前的。且修改后下降速度更快,可以更快的找到全局最小值。

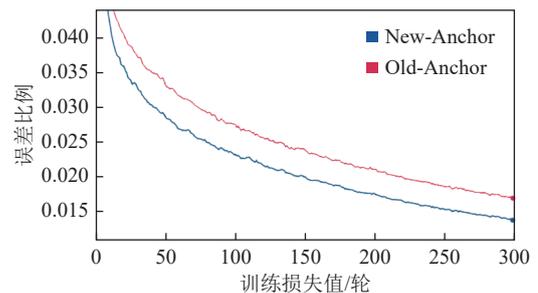


图 11 box\_loss 对比

Fig.11 Box\_loss comparison chart

图 12 所示为不同 Anchor 所形成的检测框。在外围是适用于 coco 数据集的 Anchor, 在内是使用 kmeans++ 聚类过的新 Anchor。从上图中可以发现原模型锚框虽然大但包含了很多不是水坑的区域。相比较之下聚类生成的锚框更精准的反应了水坑的具体位置, 且检测的准确度相比于原模型高出 3 个百分点。以模型训练速度和实际精准度测试判断, 对于锚框的重新聚类可以有效提升模型的整体性能。



图 12 Anchor 对比

Fig.12 Anchor comparison chart

### 2.3.3 模型损失函数消融实验

图 13 为模型在测试集中, SIoU 和 DIoU 对于物体边框的损失图。在测试的起始阶段, 2 种损失函数相差不大, SIoU 略微低于 DIoU。在测试的中后期图线出现了巨大的震荡, 尤其是 DIoU, 被困在局部极小值中出不来。但反观 SIoU, 虽然震荡但是 obj\_loss 与 DIoU 有很大的差别, 相较于 DIoU 来说也更稳定。

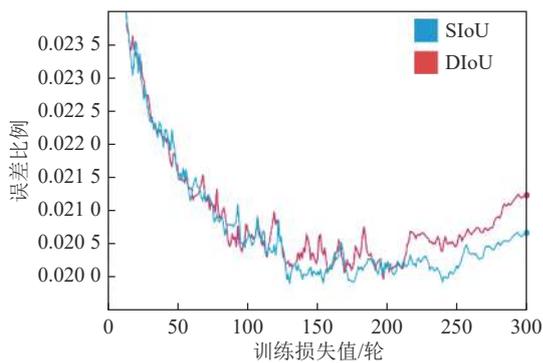


图 13 obj\_loss 对比

Fig.13 Obj\_loss comparison chart

## 3 结论及分析

(1) 笔者针对露天矿区非结构化道路上的负障碍问题, 提出了基于多特征融合的矿区非结构化道路负障碍目标检测网络模型。在露天矿区多种复杂道路情况下依然保持良好的检测性能, 满足矿区无人车在露天矿区复杂光源、复杂环境下对负障碍物快速准确

识别的要求。

(2) 多轮实验结果表明, 针对露天矿区非结构化道路的特殊情景修改的 YOLOv5 模型, 在多种露天矿区场景下具有良好的检测效果。其中平均准确度均值, 精确度和召回率上分别达到了 95.3%、96.9%、89.9%。对比现阶段主流的单阶段目标检测模型, 笔者的模型具有检测效果好、识别精度高的优势。对比现阶段主流的 2 阶段目标检测模型, 笔者的模型具有参数量、计算量小, 检测速度快, 实时性效果更好的优点。

(3) 针对露天矿区的特定场景, 笔者提出了适配于矿区场景的负障碍检测模型骨干, 优化负障碍检测模型在训练过程中的相关函数、参数。方法针对露天矿区非结构化道路的不同问题, 增加了不同的解决方法的模块, 具有问题针对性和多矿区适用性。但论文还存在一定局限性, 模型作为给无人车系统提供感知信息的模块并不参与直接决策, 且由于拍摄设备的原因, 文章无法获得图片的深度信息, 下一步考虑采用双目相机再进行实地取景, 细化负障碍信息的大小、深度等信息, 将仿真实验变成现场实地应用, 为无人车避障提供更精确的数据。

### 参考文献(References):

- [1] 刘家银, 唐振民, 王安东, 等. 基于多激光雷达与组合特征的非结构化环境负障碍物检测[J]. 机器人, 2017, 39(5): 638-651.  
LIU Jiayin, TANG Zhenmin, WANG Andong, et al. Detection of negative obstacles in unstructured environments based on multiple lidars and combined features[J]. Robot, 2017, 39(5): 638-651.
- [2] 汪佩, 郭剑辉, 李伦波, 等. 基于单线激光雷达与视觉融合的负障碍检测算法[J]. 计算机工程, 2017, 43(7): 303-308.  
WANG Pei, GUO Jianhui, LI Lunbo, et al. Negative obstacle detection algorithm based on single line laser radar and vision fusion[J]. Computer Engineering, 2017, 43(7): 303-308.
- [3] 蔡云飞, 石庭敏, 唐振民. 基于双多线激光雷达的非结构化环境负障碍感知技术[J]. 自动化学报, 2018, 44(3): 569-576.  
CAI Yunfei, SHI Tingmin, TANG Zhenmin. Sensing technology of negative obstacles in unstructured environment based on dual multi line laser radar[J]. Journal of Automation, 2018, 44(3): 569-576.
- [4] 阮顺领, 李少博, 卢才武, 等. 多尺度特征融合的露天矿区道路负障碍检测[J]. 煤炭学报, 2021, 46(S2): 1170-1179.  
RUAN Shunling, LI Shaobo, LU Caiwu, et al. Detection of road negative obstacles in open-pit mining area with multi-scale feature fusion[J]. Journal of China Coal Society, 2021, 46(S2): 1170-1179.
- [5] LI X, GAO Z, CHEN X, et al. Research on estimation method of geometric features of structured negative obstacle based on single-frame 3D laser point cloud[J]. Information, 2021, 12(6): 235-257.
- [6] WOO S, PARK J, LEE J, et al. Cbam: Convolutional block attention module[C]// Proceedings of the European conference on computer vision. Springer, 2018: 3-19.
- [7] LIN T Y, DOLLÁR P, GIRSHICK R, et al. Feature pyramid net-

- works for object detection[C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. IEEE, 2017: 2117–2125.
- [8] TAN M, PANG R, LE Q V. Efficientdet: Scalable and efficient object detection[C]// Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. IEEE, 2020: 10781–10790.
- [9] GEVORGYAN Z. SIoU loss: More powerful learning for bounding box regression[EB/OL]. (2022–05–25)[2023–04–28]. <https://arxiv.org/abs/2205.12740v1>.
- [10] ARTHUR D, VASSILVITSKII S. K-Means++: The Advantages of Careful Seeding[C]// Proceedings of the eighteenth annual ACM-SIAM symposium on discrete algorithms, New Orleans, Louisiana. USA: SODA, 2007.
- [11] HOLLAND JOHN. Adaptation in natural and artificial systems[M]. Cambridge, MA: MIT Press. 1975: 148–153.
- [12] REDMON J, DIVVALA S, GIRSHICK R, et al. You only look once: Unified, real-time object detection[C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. IEEE, 2016: 779–788.
- [13] REDMON J, FARHADI A. YOLO9000: better, faster, stronger[C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. IEEE, 2017: 7263–7271.
- [14] FARHADI A, REDMON J. Yolov3: An incremental improvement [C]// Computer Vision and Pattern Recognition. Berlin: Springer Press, 2018: 1804–2767.
- [15] BOCHKOVSKIY A, WANG C Y, LIAO H Y M. Yolov4: Optimal speed and accuracy of object detection[EB/OL]. (2020–04–23) [2023–04–28]. <https://arxiv.org/abs/2004.10934>.
- [16] WANG W, XIE E, SONG X, et al. Efficient and accurate arbitrary-shaped text detection with pixel aggregation network[C]// Proceedings of the IEEE/CVF International Conference on Computer Vision. IEEE, 2019: 8440–8449.
- [17] ZHENG Z, WANG P, LIU W, et al. Distance-IoU loss: Faster and better learning for bounding box regression[C]// Proceedings of the AAAI conference on artificial intelligence. AAAI, 2020, 34(7): 12993–13000.
- [18] LOSHCHILOV I, HUTTER F. Sgdr: Stochastic gradient descent with warm restarts[EB/OL]. (2016–08–13)[2023–04–28]. <https://arxiv.org/abs/1608.03983>.
- [19] WANG C Y, BOCHKOVSKIY A, LIAO H Y M. YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors[C]// Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. IEEE, 2023: 7464–7475.
- [20] LI C, LI L, JIANG H, et al. YOLOv6: a single-stage object detection framework for industrial applications[EB/OL]. (2022–09–27) [2023–04–28]. <https://arxiv.org/abs/2209.02976>.
- [21] LIU W, ANGUELOV D, ERHAN D, et al. Ssd: Single shot multibox detector[C]// European Conference on Computer Vision. Springer, 2016: 21–37.
- [22] SIMONYAN K, ZISSERMAN A. Very deep convolutional networks for large-scale image recognition[EB/OL]. (2014–09–04) [2023–04–28]. <https://arxiv.org/abs/1409.1556>.
- [23] REN S, HE K, 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.
- [24] 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. IEEE, 2016: 770–777.