

中文引用格式:李刚,杜亚波,杨庆贺,等. 基于多尺度特征融合和注意力机制的矿区道路障碍检测[J]. 中国安全科学学报, 2024, 34(9): 87-98.

英文引用格式:LI Gang, DU Yabo, YANG Qinghe, et al. Obstacle detection on mining roads based on multi-scale feature fusion and attention mechanism[J]. China Safety Science Journal, 2024, 34(9): 87-98.

基于多尺度特征融合和注意力机制的 矿区道路障碍检测*

李刚 教授, 杜亚波, 杨庆贺, 毛梦影, 贾冬平
(辽宁工程技术大学 矿业学院, 辽宁 阜新 123000)

中图分类号: X948

文献标志码: A

DOI: 10.16265/j.cnki.issn1003-3033.2024.09.1631

【摘要】 为解决复杂露天矿区背景下的行车障碍检测问题,提出一种基于改进跨尺度特征融合的矿区道路障碍检测算法。首先,针对原始矿山数据集中小目标样本类别不平衡的问题,采用一种基于几何变换和加权泊松融合的数据增强方法扩大样本数量;其次,在特征提取阶段提出更适用于障碍检测的跨阶段连接网络,以增大检测尺度,提高算法对小目标特征的学习能力;然后,在特征融合阶段提出基于3D无参注意力(SimAM)和去权重的双向特征融合金字塔网络(Bi-FPN),通过扩大预测特征图和特征感受野,提升多尺度检测性能;最后,针对训练中样本不平衡和障碍物边界框定位不精准问题,引入质量焦点损失函数(QFL)和可扩展的交并比损失函数(SIoU),将分类得分与位置的质量预测结合,提高对密集遮挡目标的定位精度。结果表明:改进方法能有效识别复杂背景下露天矿区非结构化道路障碍物,在实际应用中,检测精度达到91.88%,检测速度达到68.7帧/s,相较于主流检测方法有着更好的小目标和多尺度检测性能,可满足露天矿区无人矿卡行进中的障碍安全检测要求。

【关键词】 多尺度; 特征融合; 注意力机制; 矿区道路; 障碍检测; 数据增强

Obstacle detection on mining roads based on multi-scale feature fusion and attention mechanism

LI Gang, DU Yabo, YANG Qinghe, MAO Mengying, JIA Dongping

(Mining Institute, Liaoning Technical University, Fuxin Liaoning 123000, China)

Abstract: In order to solve the problem of travelling obstacle detection in the context of complex open pit mines, a mining road obstacle detection algorithm based on improved cross-scale feature fusion is proposed. Firstly, to address the problem of unbalanced small target sample categories in the original mine dataset, a data enhancement method based on geometric transformation and weighted Poisson fusion is used to expand the number of samples. Secondly, a cross-stage connectivity network that is more suitable for obstacle detection is proposed in the feature extraction stage to increase the detection scale and improve the algorithm's learning ability of the small target features, and then a 3D parameterless attention (SimAM) and de-weighted Bi-directional feature fusion pyramid network (Bi-FPN) are proposed in the feature fusion stage to improve the multi-scale detection performance by enlarging the predicted feature map and feature receptive field. Finally, to address the problems of sample imbalance and imprecise obstacle bounding box

localisation in the training, the quality focal loss function (QFL) and the scalable Intersection and combination ratio loss function (SIoU), which combines the classification score with the quality prediction of the position to improve the localisation accuracy for dense occlusion targets. The results show that the improved method can effectively identify unstructured road obstacles in open pit mining area under complex background, and in practical application, the detection accuracy reaches 91.88% and the detection speed reaches 68.7 f/s, which has a better performance of small-target and multi-scale detection compared with the mainstream detection methods, and it can satisfy the requirements of obstacle safety detection in the travelling of unmanned mine cards in open pit mining area.

Keywords: multi-scale; feature fusion; attention mechanism; mining roads; obstacle detection; data enhancement

0 引言

随着智能化技术的快速发展,矿区无人驾驶成为智慧矿山建设下的一个重要发展方向,其关键影响因素是露天矿区非结构化道路中存在大量碎石、坑洞等影响行车安全的问题。在露天矿区中,无人矿卡必须能够快速、准确地识别矿区道路中的车辆、行人以及障碍物,为行车道路的安全性提供必要保障^[1]。因此,有必要研究一种具备高精度和高实时性的矿区道路障碍检测方法。

矿山非结构化道路的检测算法大致可分为传统的图像检测和基于深度学习的图像检测 2 种方法^[2]。在传统的图像障碍检测算法中,使用图像获取—特征提取—模式分类这一经典模式识别框架检测行人^[3],张彦斐等^[4]通过提取道路的纹理和颜色等特征信息检测道路,对道路的形状具有良好的鲁棒性,但是单纯的纹理特征容易受到光照、天气等环境因素影响。后来发展了基于图像的边缘检测,如卢才武等^[5]针对矿区道路边缘模糊,提出基于空间图像处理的边缘检测方法,可有效突出道路边缘。郝明等^[6]提出基于改进主动轮廓模型的矿区裂隙检测算法,利用得到的边缘信息检测和定位研究区域内的裂隙。随着深度学习算法的发展与应用,基于卷积神经网络的目标检测技术在准确性方面远远超过了传统方法。CHENG Yong 等^[7]利用深度稀疏卷积网络检测具有多尺度特征的车辆,获得了更高的精度和检测效果。针对道路车辆实时检测遮挡问题,马永杰等^[8]提出了基于改进 YOLOv3^[9]的道路车辆检测方法,提高了模型的特征表达能力,使得模型具有良好的自适性和鲁棒性。张炳力等^[10]基于 RetinaNet 网络,针对车辆特征优化设计网络结构,达到了较好的检测精度和 23 帧/s 的检测速度。以上文献都是利用深度学习算法对结构化道路中障碍

物的检测与跟踪,但是面对道路中坑洞、积水等负向障碍物的检测均不理想。同时深度学习算法在复杂矿区非结构化道路中的实际适用性和准确性仍存在众多挑战。

鉴于以上分析,笔者针对复杂矿区道路中负向障碍漏检多、检测难等问题,提出一种基于深度学习的矿区道路障碍检测算法。该方法在特征提取阶段,以高效的跨阶段连接网络(Cross Stage Partial Darknet, CSPDarknet)作为主干网络,扩大目标感受野,增强模型的多尺度检测性能;在融合阶段,使用去权重的双向特征网络融合底层与高层的特征信息,引入注意力机制加强对小目标的特征提取能力;通过质量焦点损失函数和可扩展的交并比损失函数作为分类和边界框损失函数,加快模型的收敛速度并提高精度,以期为无人驾驶车辆在露天矿区非结构化道路中的行驶提供一种高效率、高可靠性的障碍检测方法。

1 露天矿区道路障碍检测模型

矿区道路与城市结构化道路不同的是,露天矿区的行车道路通常是临时修建的。矿区道路背景信息复杂多变,主要体现在:①车辆组成单一。矿区道路具有一定的特殊性,多独立存在,不构成网络。因此,车辆组成也较为单一,多以重载卡车、工程车和小汽车为主。②负障碍尺度跨度大。负障碍检测模型特征信息如图 1 所示。路面的负障碍因雨水冲刷导致积水与坑洞形状多变,具有障碍边界模糊、小目标特征存在遮挡等特点。

在露天矿区的复杂道路背景下,原始 YOLOv5 算法中用来检测小目标的最小尺度为 80×80,最小检测为 8×8 像素,无法满足矿山道路小目标检测的需求。为提升露天矿区障碍检测模型在小目标和多尺度检测方面的能力,一种有效的策略就是增加网

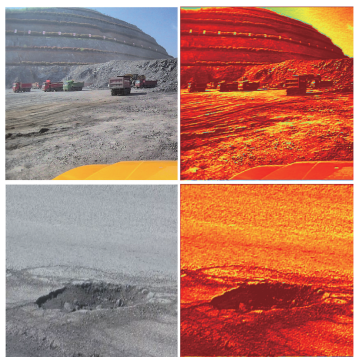


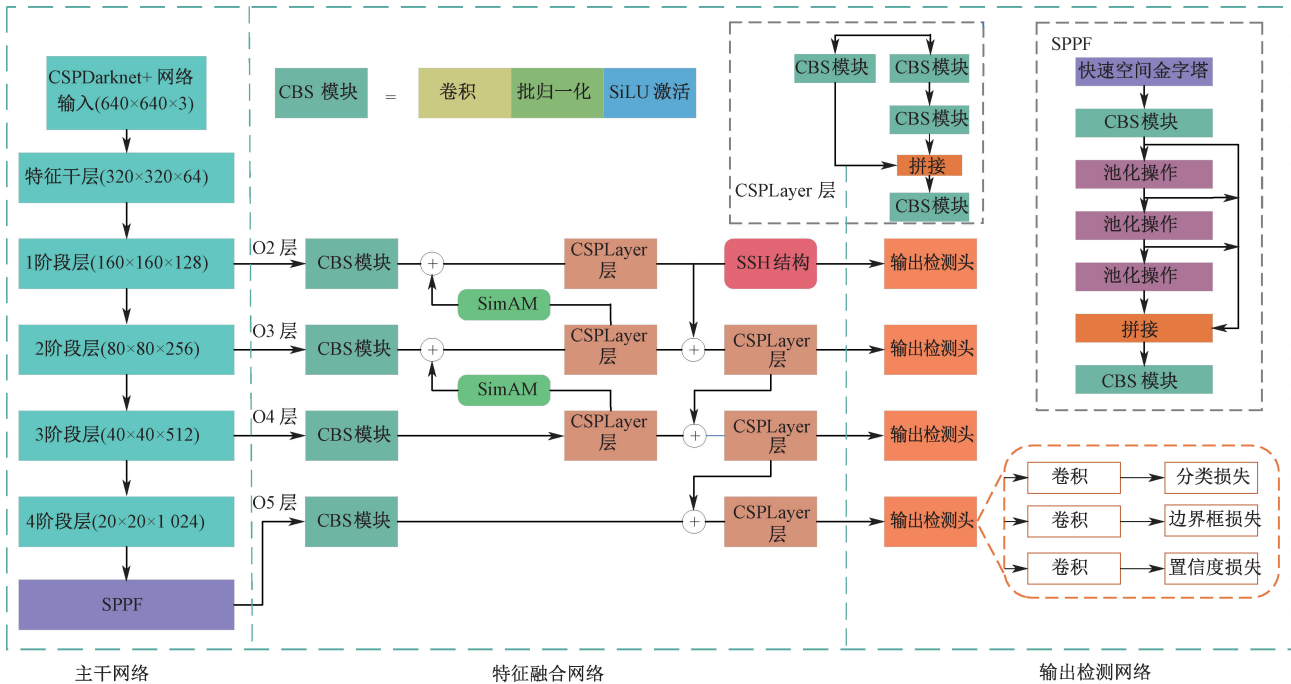
图 1 负障碍检测模型特征信息

Fig. 1 Open pit obstacles and their characteristics

网络中的特征图尺寸,使网络在进行分类预测时获得更丰富的特征信息,从而提高对小目标的识别准确性。因此,提出以改进的 CSPDarknet+作为骨干特征提取网络,保证模型的检测速度和精度。在原主

干网络的基础上添加一个 4 倍下采样的检测头,在 3 尺度主干网络的基础上引入一个 160×160 的检测层来满足更小目标的检测需求。这种多尺度检测架构利用低层次的特征图,为模型提供了具有区分性的特征信息,从而显著提升了对小型目标的检测和识别能力。

CSPDarknet+的网络结构和 ResNet 类似,整体结构分为 1 个特征干层和 4 个阶段层。特征干层是一个 6×6 卷积核的卷积模块,比 YOLOv5 算法中的焦点模块更加高效。O2、O3 和 O4 分别是主干网络 1 阶段层、2 阶段层和 3 阶段层的经过 3×3 卷积后的输出特征图,O5 是经过 4 阶段层的卷积后,再经过多个池化处理的特征层。其中,增加的 O2 层特征尺寸为 160×160 ,有助于在更小的尺度上获取更全面的特征信息。其结构如图 2 所示。



注:快速空间金字塔池(Spatial Pyramid Pooling Fast, SPPF);卷积批归一化激活(Conv Batch Normalization SiLU, CBS);3D 无参注意力(3D Parameterless Attention, SimAM);跨阶段连接层(Cross Stage Partial Layer, CSPLayer);单阶段无头检测结构(Single Stage Headless, SSH)。

图 2 双向特征融合的矿区障碍检测模型

Fig. 2 Bidirectional feature fusion for mining obstacle detection modeling

最后一个阶段层增加了 SPPF 结构,如图 3 所示。空间金字塔池化(Spatial Pyramid Pooling, SPP)^[11]分别指定 3 个不同大小的池化操作,将不同尺寸的特征图映射为固定长度的特征向量,使模型具有处理不同尺度图像的能力。SPPF 结构是将输入串行通过多个 5×5 大小的池化层,在计算上合并这些池化操作,只进行一次卷积和池化。在设计

上更加考虑效率,减少了计算量,同时,采用改进的池化策略,有助于更好地保留重要信息,并减少噪声,而且,与 SPP 模块效果相同,但速度更快。

1.1 基于注意力的去权双向特征融合模块

原始的加权双向特征金字塔网络(Bidirectional Feature Pyramid Network, BiFPN)通过引入可学习的权值来调整不同输入特征的贡献,同时,根据需要

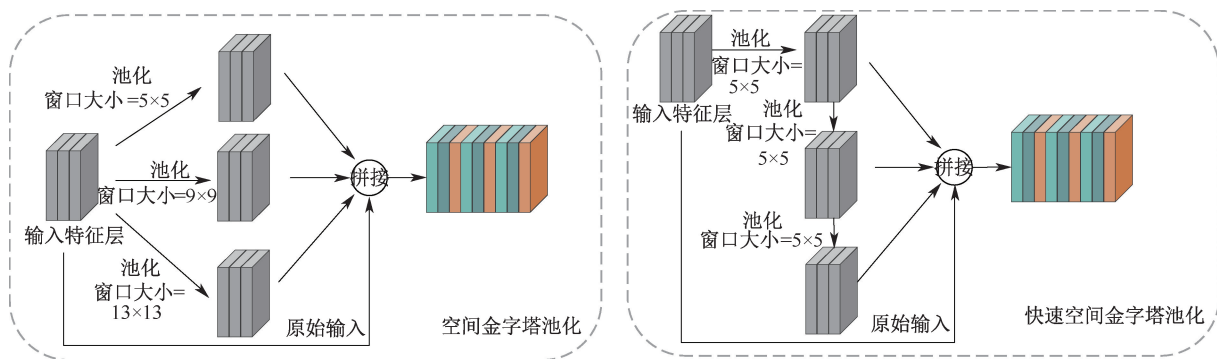


图3 SPP 和 SPPF

Fig. 3 Spatial pyramid pooling and spatial pyramid pooling fast

整个结构循环使用以加强特征融合的能力。BiFPN 的带权特征融合公式为：

$$Q = \sum_i \frac{w_i}{\varepsilon + \sum_j w_j} \cdot I_i \quad (1)$$

式中： Q 为融合后的特征输出； ε 为一个非常小的正数，防止分母为 0； w_i 为 i 个输入特征图的权重； j 为特征融合操作中所有输入特征的索引； I_i 为第 i 个输入特征。

引入带权重的 BiFPN 会增加过多的参数量和计算量，导致训练时间过长且不容易拟合。因此，提出改进的去权重 Bi-FPN 网络，只借鉴其特征融合方法，这样既可以提升网络的多尺度检测性能，也能获得更加抽象高效的语义特征。基于 SimAM 与去权双向特征融合模块如图 4 所示。在自下而上的融合

过程中，从主干网络的 O4 层输出开始，经过卷积 (Conv)、批量归一化 (Batch Normalization, BN) 和 SiLU 激活函数后，进入 CSPLayer 的处理，再经过上采样后的 O4 层通过特征增强与 O3 特征层拼接。重复上述操作，从而获取主干网络各层丰富的特征信息，减少原有噪声信息的影响。最后，采用 SSH^[12] 在 O2 层后进一步拓展特征图的感受野，进而提升对微小目标的检测性能。

为减少矿山复杂背景的影响，传统检测往往会采用霍夫变换等方法进一步划分矿山道路区域，但是该方法处理效果低下且分割效果并不理想。为提升模型的特征表示能力，在特征金字塔网络中引入 SimAM 空间与通道注意力机制^[13]。SimAM 能够更全面高效地评估特征权重。通过增强小目标特征、减少矿山背景干扰，提升模型的干扰抑制能力。相

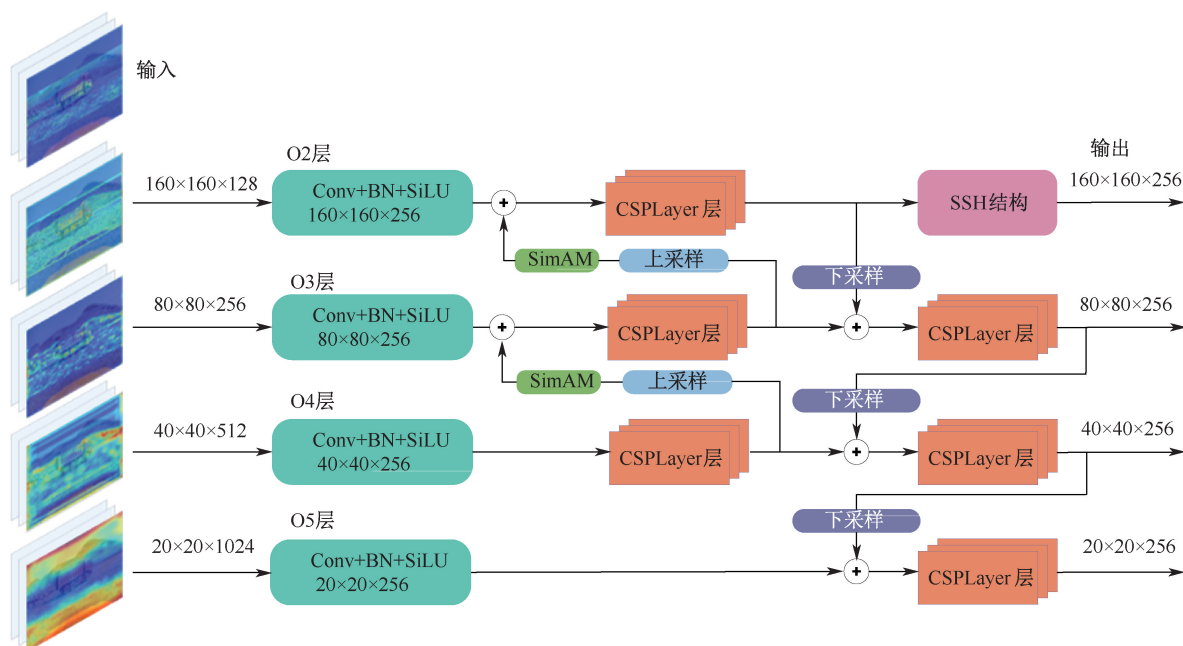


图4 基于 SimAM 注意力与去权双向特征融合模块

Fig. 4 SimAM-based attention and de-weighting bidirectional feature fusion module

较于压缩与激励注意力 (Squeeze-and-Excitation Networks, SENet)^[14] 和混合注意力 (Convolutional Block Attention Module, CBAM)^[15], 它们侧重于通过池化和全连接层手工设计注意力模块, 而 SimAM 基于神经科学理论中的能量函数评估各处特征的重要性, 因此, 更具可解释性且无需引入可学习参数。此外, SimAM 作为一种 3D 注意力机制, 与 SENet 关注于通道维度的重要性, CBAM 侧重于依次关注通道和空间维度的重要性相比, 它直接评估了各个独立神经元的重要性。在保持运算高效性的同时, 能够更全面地评估各神经元的重要性并进行加权处理。

SimAM 注意力机制的作用原理通过能量函数实现, 而能量函数的定义与神经科学领域中的空域抑制现象有关。即活跃的神经元会抑制周围信息匮乏神经元的表现能力。因此, 神经元的重要程度可通过能量函数表达。能量函数的表达见下式:

$$e_i(w_i, b_i, v, x_k) = \frac{1}{M-1} \sum_{i=1}^{M-1} [-1 - (w_i x_k + b_i)]^2 + [1 - (w_i x_k + b_i)]^2 + \lambda w_i^2 \quad (2)$$

式中: w_i 和 b_i 分别为对应的权重和偏置量; v 为计算目标神经元和其他神经元区分度的标签; t, x_k 分别为输入特征图中目标神经元和其他神经元; k 为空间维度上的索引; M 为该通道上所有神经元的数量; λ 为正则化系数; 通过计算出 w_i, b_i 的解析解, 得到最小能量 e_i^* 为:

$$e_i^* = \frac{4(\hat{\sigma}_2 + \lambda)}{(t - \hat{u})^2 + 2\hat{\sigma}_2 + 2\lambda} \quad (3)$$

$$\text{其中, } \hat{u} = \frac{1}{M} \sum_{i=1}^M x_k, \hat{\sigma}_2 = \frac{1}{M-1} \sum_{i=1}^M (x_k - \hat{u})^2.$$

由式(3)可知: 能量越低, t 神经元和其他神经元的区别越大, 也就越重要。SimAM 模块最终优化为:

$$\tilde{x} = \text{Sigmoid}\left(\frac{1}{E}\right) \odot X \quad (4)$$

式中: E 为 e_i^* 在所有通道和空间维度的和; \odot 为逐元素相乘操作; X 为输入的特征张量。

SimAM 注意力机制结构如图 5 所示。 C 为特征图的通道数; H 和 W 分别为特征图的高度和宽度; Sigmoid 为激活函数, 用来限制 E 中较大的值, 不会影响每个神经元的相对重要性。因此, 网络模型使用 SimAM 注意力机制, 在没有引入额外参数的情况下, 能够更好地聚焦在矿山小目标上, 提取出更具信息量的特征。

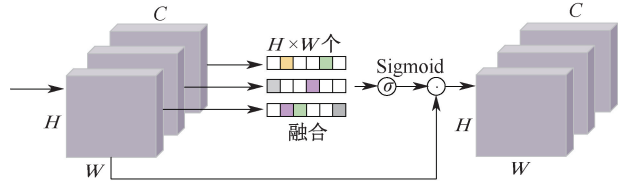


图 5 SimAM 注意力机制结构

Fig. 5 SimAM attention mechanism structure diagram

1.2 损失函数优化

YOLOv5 模型使用完全交并比损失函数 (Complete Intersection Over Union loss, CIoU) 来衡量边界框的损失。但是, CIoU 没有考虑到如何在回归时能够更好地描述重叠信息, 因为边界框可能在训练过程中“徘徊”, 最终导致收敛速度较慢且效率不高。针对上述问题, 提出使用可扩展的交并比损失函数 (Scalable Intersection over Union Loss, SIoU) 替代 CIoU, 考虑到边界框重叠面积、中心点距离、长宽比和所需回归之间的向量角度。SIoU 包含 4 个部分: 角度损失 Λ , 距离损失 Δ 、形状损失 Ω 以及交并比损失 U 。其中, 角度损失公式如下:

$$\Lambda = 1 - 2 * \sin^2\left(\arcsin(x) - \frac{\pi}{4}\right) \quad (5)$$

其中

$$x = \frac{C_h}{\mu} = \sin(\alpha) \quad (6)$$

$$\mu = \sqrt{(b_x^g - b_x)^2 + ((b_y^g - b_y)^2)} \quad (7)$$

$$C_h = \max(b_x^g, b_y) - \min(b_x^g, b_y) \quad (8)$$

损失函数中的角度损失如图 6 所示。其中, 点 B 和点 B^g 分别为预测框中心点及真实框中心点; g 为真实数据的标签; C_w, C_h 分别为 B 和 B^g 点的横坐标与纵坐标之差; w 和 h 为真实框的宽和高; α 为 2 点连线与水平线的夹角; (b_x, b_y) 为预测框中心坐

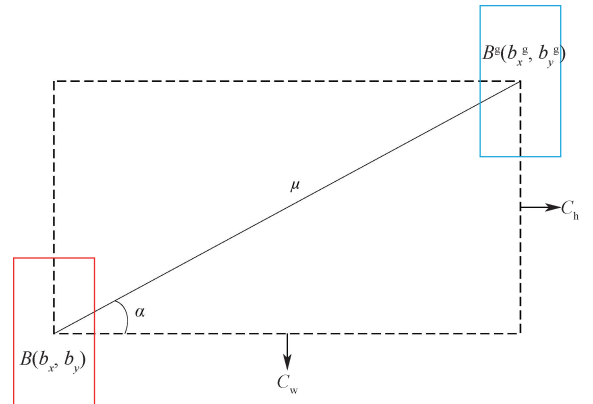


图 6 损失函数中的角度损失

Fig. 6 Angular loss in loss function

标, (b_x^g, b_y^g) 为真实框中心点坐标, x 和 y 为中心点的横坐标和纵坐标; μ 为真实框和预测框中心点距离。根据角度损失 Λ , 定义距离损失 Δ 计算公式如下:

$$\Delta = \sum_{t=x,y} (1 - \exp(-\gamma\rho_t)) \quad (9)$$

其中

$$\rho_x = \left(\frac{b_x^g - b_x}{C_w} \right)^2, \rho_y = \left(\frac{b_y^g - b_y}{C_h} \right)^2, \quad \gamma = 2 - \Lambda \quad (10)$$

形状损失 Ω 定义如下:

$$\Omega = \sum_{t=w,h} (1 - \exp(-\omega_t))^\theta \quad (11)$$

其中

$$\omega_w = \frac{|w - w^g|}{\max(w, w^g)}, \omega_h = \frac{|h - h^g|}{\max(h, h^g)} \quad (12)$$

式中: θ 为网络控制对形状损失的关注程度, 取值范围为 2~6; w^g 和 h^g 为预测框的宽和高; SIoU 损失 L 的定义为:

$$L = 1 - U + \frac{\Delta + \Omega}{2} \quad (13)$$

YOLOv5 的置信度与分类损失使用二值交叉熵损失函数, 二值交叉熵损失函数通常用于检测易于区分的样本, 但是对于矿山小目标来说, 存在着大量负样本, 类别极其不均衡, 且占比过大容易对训练结果产生影响, 从而影响模型检测目标的精度。因此, 引入质量焦点损失函数 (Quality Focal Loss, QFL)^[16], 传统的焦点损失函数 (Focal Loss, FL) 只支持 $\{0, 1\}$ 离散标签, 利用 QFL 优化分类、质量联合分数连续值标签。首先, 将交叉熵部分扩展到完全形式, 然后, 将每个样本的缩放因子泛化为预测值和实际值之间的绝对值, 最后, 扩展后的损失函数即为 QFL。函数定义如下:

$$QFL(\hat{p}) = - |d - \hat{p}|^\beta ((1 - d) \lg(1 - \hat{p}) + d \lg(\hat{p})) \quad (14)$$

式中: d 为实际的质量评分, 范围为 $[0, 1]$; \hat{p} 为模型的预测分数; β 为调节因子, 用于控制损失对难分类样本的敏感度; 在复杂的道路背景下, 当遇到远距离小目标聚集且互相存在遮挡的场景时, 目标之间的重叠和交叉现象就会有所增加, 因此, 存在于训练和推理之间的误差也会变大, 定位精度也会受到影响。通过引入 QFL 损失函数, 缩小训练和测试之间的差距, 使分类和定位质量得到更简单、联合和有效地表示, 提高了对矿山遮挡小目标的多尺度检测能力。

2 矿区道路障碍数据集构建

矿山数据来源于 2022 年 10 月间的某大型露天矿作业开采现场, 使用数据采集车采集。在不同时间段和天气环境下, 共采集 4 909 张图像数据, 以此构建露天矿山行车道路数据集, 包括矿山运载卡车、行人、挖机、小汽车、推土机和坑洞共 6 类目标。由于露天矿区视野空旷, 远距离采集的图像中矿卡和行人目标尺寸都比较小, 这导致数据集中目标空间分布极不均衡。同时, 在道路中的坑洞等目标形状各异且尺度变化较大, 这也导致矿区道路数据集存在大量多尺度特征和小目标物体。数据集按照 8 : 2 的比例划分训练集和验证集, 划分后的训练集仅有 3 927 张图片, 对于通用的大型目标检测模型来说, 容易出现过拟合的现象, 因此, 对数据集做一定的数据增强处理。按照通用物体数据集 (Common Objects in Context, COCO) 数据格式标准分析矿山数据集, 数据增强前后样本分布见表 1。卡车类在数据量方面与其他 5 类目标相差过大, 各目标类别尺寸分布也存在不均衡的问题, 小目标占比都比较少。

表 1 数据增强前后样本分布

Table 1 Sample distribution before and after data enhancement

类别	样本数量/个						
	卡车	挖机	推土机	行人	汽车	坑洞	总图片数
训练集数据增强前	7 739	2 786	779	387	997	331	4 909
训练集数据增强后	38 695	13 930	3 895	1 935	4 985	1 655	24 545
训练集数据均衡后	9 240	4 779	4 887	8 605	7 385	3 906	24 204

为解决目标尺寸分布不均衡和样本多样性不足的问题, 对原始图片数据集进行随机水平翻转、随机对比度增强、添加高斯噪声和随机擦除 4 种数据增强操作来模拟不同的环境变化, 使样本数量扩增为原来的 5 倍。数据增强操作创造了额外的训练样本, 从而扩大训练集的规模和多样性, 但是也可能导致模型对目标的某种特征陷入过拟合状态。因此, 提出一种改进的数据均衡方法, 即在原始数据集上抽取 50% 的图像, 分别做 4 种数据增强处理, 然后对数据集中数量过少的某些类别添加基于公开数据集的图像数据和小目标增强的数据。数据均衡方法在数据增强的基础上进一步解决了类别不平衡的问题。

题,使不同类别的样本数量接近平衡,改善了模型性能。解决了矿区道路数据集中尺寸分布不均衡的问题,显著增加了数据集的样本数量,扩充了小目标样本,从而降低了过拟合的风险

由于坑洞积水等障碍目标在图像中占据很少的像素比例,且处于车辆等大型目标的周围,为此,提出一种加权泊松融合^[17]的方法将小目标图像与矿山场景相结合进行数据扩增,其具体流程为:①选择

源图像和目标图像的对应区域。②计算源图像和目标图像对应区域的梯度信息。③根据蒙版和权重设置源图像和目标图像的加权值。④在目标图像上通过解泊松方程来计算差异项,使得差异项的梯度等于源图像对应区域的梯度,并考虑权重的影响。⑤将目标图像和计算得到的差异项相加,得到使用加权泊松融合方法将小目标图片与原始图片相融合后的图像(图7)。



图7 数据增强图像

Fig.7 Data enhanced images

3 试验设计与分析

3.1 试验设计及评价指标

试验平台系统为 ubuntu 22.04,网络模型基于 PyTorch 深度学习框架搭建。试验训练各项超参数配置见表2。

表2 试验各项参数配置

Table 2 Configuration of experimental parameters

训练数据	预训练数据	文中数据
图像输入尺寸	640×640×3	640×640×3
冻结训练 batchsize	—	16
冻结训练 Epoch	—	20
冻结训练学习率	—	0.001 25
解冻训练 batchsize	16	16
解冻训练 Epoch	50	280
解冻训练学习率	0.005	0.01
动量参数	0.937	0.937
优化方法	SGD	SGD
NMS 阈值	0.65	0.65

为加速和改善新任务的学习过程,采用迁移学

习的方式,使用基于 COCO2017 数据训练的数据集权重,修改数据集分类分支以符合文中数据集的类别数目。同时防止权值被破坏,在刚开始训练阶段冻结主干网络的权重,使特征提取网络在训练初期不发生改变。在训练 20 个 epoch 后,解冻主干网络权重并进行权值更新,继续完成后续的训练。

为全面评估检测算法的性能,采用召回率 R 、精确率 P 、平均精度(Average Precision, AP)和均值平均精度(mean Average Precision, mAP)作为评价指标,其计算公式如下:

$$P = \frac{TP}{TP + FP} \quad (15)$$

$$R = \frac{TP}{TP + FN} \quad (16)$$

$$AP = \frac{\sum_1^n P \times R}{N} \quad (17)$$

$$mAP = \frac{\sum_1^n AP}{n} \quad (18)$$

式中:TP 为真正例,即正确预测为正类别的样本数;FP 为假正例,即错误预测为正类别的样本数;FN 为假负例,即错误预测为负类别的样本数; n 为类别总数, N 为检测到的类别数。mAP 评估的是整个模型对矿山目标的检测效果。

模型的复杂程度主要用网络的参数数量和计算量作为评估标准。参数数量是指网络模型中需要学习的权重和偏差的总数,计算量指的是模型在进行一次前向推理时所需的浮点运算的总次数。模型的推理速度通常用每秒帧数(Frames Per Second,FPS)来衡量。

$$FPS = \frac{1}{T} \quad (19)$$

式中 T 为模型处理一张图像所需的时间,s。

3.2 模型试验与性能分析

不同网络模型精度对比见表 3。模型均采用视觉对象数据集(PASCAL Visual Object Classes,

表 3 不同网络模型对比

Table 3 Comparison of different network models

模型	输入尺寸	精确率/%	速度/(帧·s ⁻¹)	参数量/MB	计算量/GB	坑洞精度
SSD	300×300	61.37	176.3	24.41	30.68	41.74
Faster-RCNN	800×800	80.36	48.0	41.37	134.09	61.57
RetinaNet	600×600	84.50	38.1	36.43	82.45	65.54
EfficientDet-d3	896×896	78.72	42.5	18.44	108.20	58.70
RepVGG-A2	512×512	83.99	52.4	41.12	57.66	67.39
YOLOX-m	640×640	82.04	95.6	25.28	36.76	61.68
YOLOX-tiny	640×640	77.54	133.0	5.03	7.58	56.88
Our Model	640×640	91.88	68.7	17.80	20.77	76.51

文中提出的检测模型在矿山多种复杂路况和天气条件下,无论何种距离都能准确可靠地识别出道路间的目标,显示出了优异的性能。矿区道路障碍物检测结果如图 8 所示,在检测可能存在密集遮挡的场景时,该模型能准确判定目标框边界,不存在边界框重合问题,置信度高达 0.9。此外,它能精准地识别多尺度特征的坑洞等负向障碍小目标,满足了露天矿区行进行道路间的障碍检测要求。

3.3 障碍检测模型预训练影响因素分析

文中设计的主干网络作为特征提取模型的上游任务,为后续分类和回归任务提供了强有力的特征表示,在预训练阶段使用 COCO2017 大规模图像分类数据集进行训练,作为迁移学习的基础权重。通过加载预训练权重模型和权重随机初始化 2 种场景,分别使用原始的 CSPDarknet 网络、改进后的 CSPDarknet+主干网络在 VOC 数据集上进行对比训

练,验证迁移学习的必要性和改进主干网络的有效性。不同主干网络训练结果如图 9 所示。CSPDarknet_pre 和 CSPDarknet+_pre 分别为使用了预训练权重的原始主干网络模型和改进的主干网络模型;CSPDarknet_nopre 为未使用预训练权重的原始主干网络模型。CSPDarknet_pre 为加载了预训练权重的改进矿区障碍检测模型。RepVGG0_pre 和 ResNet50_pre 分别为加载了预训练权重的原始 RepVGG 模型和 ResNet50 模型。由于加载了预训练权重的 CSPDarknet 主干网络从一开始就能提取到比较有效的特征,损失下降的更快,也更容易收敛,相较于权重随机初始化的 CSPDarknet 主干网络精度提高了 3%,最终达到 87.82%的精度。

VOC) 作为预训练数据集,文中模型与主流模型对比,使用 COCO 数据集训练的预训练权重。文中模型达到 91.88%的 mAP,取得较好的检测效果,而参数量、计算量和检测速度分别达到 17.80 MB、20.77GB 和 68.7 帧/s。

YOLOX 是基于 anchor free 的 2 阶段算法,其中,轻量级的模型 YOLOX-tiny 参数量仅有 5.03 MB,速度达到 133.0 帧/s,但是对于坑洞的精度为 56.88%,未达到很好的检测效果。EfficientDet 虽然使用大量基于深度可分离卷积的 BiFPN 作为 Neck 网络,但是其输入分辨率过大,模型较为复杂,精度仅为 78.72%。RetinaNet 和 RepVGG-A2 由于网络更深,所以,能提取到更多的语义信息,模型精度分别达到了 84.50%和 83.99%。SSD 和 Faster-RCNN 等传统检测算法在识别坑洞等小目标方面表现不太理想,但是 SSD 达到了 176.3 帧/s 的检测速度。

Table 3 Comparison of different network models

练,验证迁移学习的必要性和改进主干网络的有效性。不同主干网络训练结果如图 9 所示。CSPDarknet_pre 和 CSPDarknet+_pre 分别为使用了预训练权重的原始主干网络模型和改进的主干网络模型;CSPDarknet_nopre 为未使用预训练权重的原始主干网络模型。CSPDarknet_pre 为加载了预训练权重的改进矿区障碍检测模型。RepVGG0_pre 和 ResNet50_pre 分别为加载了预训练权重的原始 RepVGG 模型和 ResNet50 模型。由于加载了预训练权重的 CSPDarknet 主干网络从一开始就能提取到比较有效的特征,损失下降的更快,也更容易收敛,相较于权重随机初始化的 CSPDarknet 主干网络精度提高了 3%,最终达到 87.82%的精度。

3.4 矿区道路障碍检测模型有效性分析

3.4.1 主干网络试验分析

为验证 CSPDarknet+主干网络选择的有效性



图 8 矿区道路障碍物检测结果

Fig. 8 Detection results of road obstacles in mining area

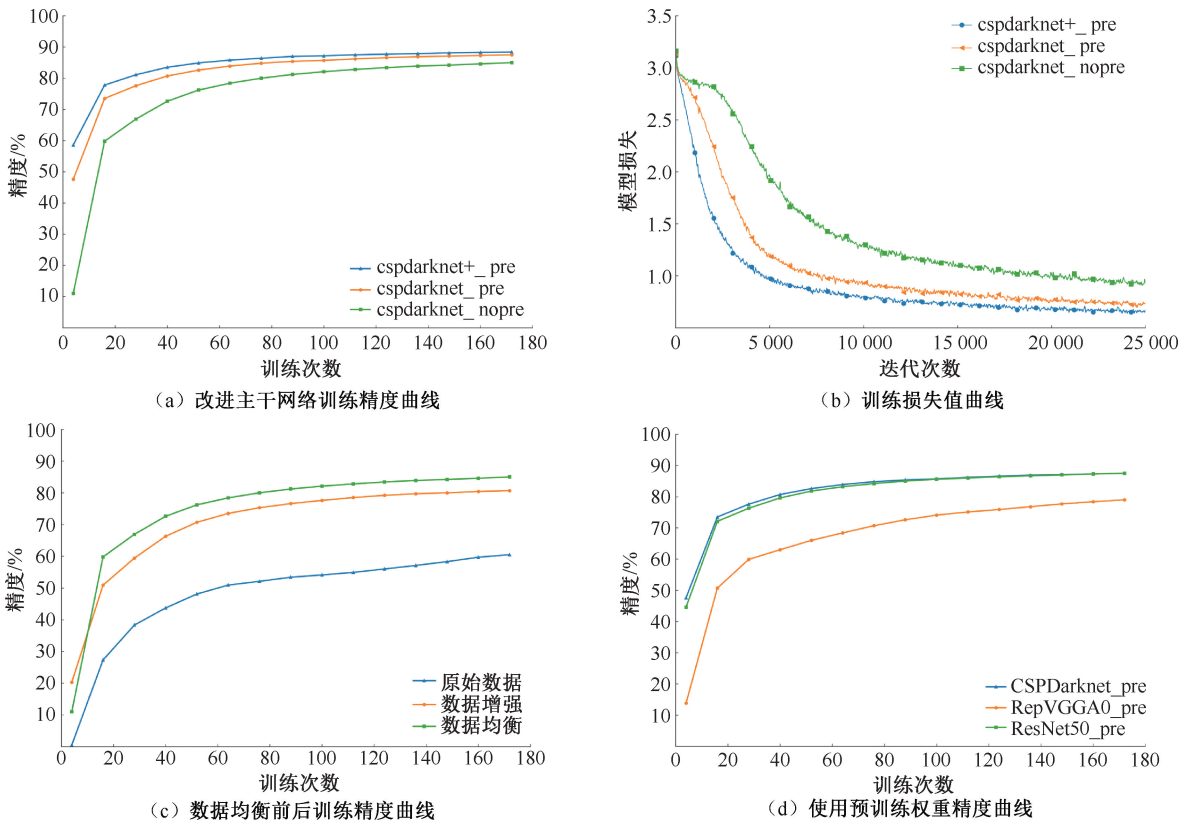


图 9 不同主干网络训练结果

Fig. 9 Training results of different backbone networks

和合理性,分别在 VOC 数据集和矿区道路数据集上训练模型,试验结果见表 4。

由表 4 可知:ResNet50 由于有着更深的网络深度,有更强的特征提取能力,模型精度达到最高的 88.61%,但其参数量大,无法满足实时部署要求。

RepVGG A0 虽然具有更少的参数量,因其提取不到深层次的特征信息,精度也较低。文中针对矿区道路实际的作业场景,提出的改进的 CSPDarknet+主干网络,虽然增加了少许参数量,但是检测精度达到 88.49%,对于坑洞等小目标检测精度也达到 73.30%。

表 4 不同主干网络对模型性能影响

Table 4 Impact of different backbone networks on model performance

主干网络	VOC 精度/%	矿山精度/%	参数量/MB	坑洞精度/%
Base	88.71	87.82	7.06	72.69
ResNet50	91.47	88.61	25.60	73.01
RepVGG A0	74.33	77.70	7.03	56.81
CSPDarknet+	88.90	88.49	7.79	73.30

3.4.2 改进注意力的试验分析

为使模型更关注于矿区道路中的小目标特征,在主干网络输出特征层后引入 SimAM 注意力机制。不同注意力机制对模型性能影响见表 5。CBAM 和坐标注意力(Coordinate Attention, CA)^[18] 由于计算复杂度更高,未能有理想的效果。而压缩与激励注意力 SENet、高效通道注意力(Efficient Channel Attention, ECA)^[19] 和无参注意力 SimAM,对于小目标的特征提取都有明显的增强。基于梯度加权映射(Gradient weight Class Activation Mapping, GradCAM)^[20] 的不同

表 5 不同注意力机制对模型性能影响

Table 5 Impact of different attention mechanisms on model performance

注意力模块	VOC 精度/%	矿山精度/%	坑洞精度/%	参数量/MB	计算量/MB
Baseline	89.30	88.89	73.65	+0	+0
SENet	90.95	89.39	74.80	+0.021	3.072
CBAM	90.78	88.54	72.67	+0.003	6.144
CA	90.87	88.98	73.56	+0.012	7.168
ECA	90.97	89.40	75.05	+0	2.048
SimAM	90.89	89.40	74.92	+0	+0

表 6 消融试验

Table 6 Ablation experiment

数据扩增	数据均衡	多尺度	Bi-FPN	SimAM	QFL	SIoU	VOC 数据精度/%	矿山数据精度/%	坑洞精度/%
—	—	—	—	—	—	—	88.71	73.67	34.24
√	—	—	—	—	—	—	88.71	83.69	65.52
—	√	—	—	—	—	—	88.71	87.82	72.69
—	√	√	—	—	—	—	88.90	88.49	73.30
—	√	√	√	—	—	—	89.30	88.89	73.65
—	√	√	√	√	—	—	90.89	89.40	74.98
—	√	√	√	√	√	—	92.13	90.56	75.66
—	√	√	√	√	√	√	93.21	91.88	77.12

注:√表示引入该方法,—表示未引入该方法。

提出改进的 4 尺度 CSPDarknet+ 主干网络。为兼顾推理速度与检测精度,重新设计 Neck 网络,提出去权重的双向特征融合 Bi-FPN 结构,同时,添加 3D 无参的 SimAM 注意力机制,显著提高了模型

注意力机制特征如图 10 所示,为了更直观展现改进注意力对模型检测效果的影响,采用 GradCAM 方法可视化不同注意力机制在特征提取阶段的特征层。

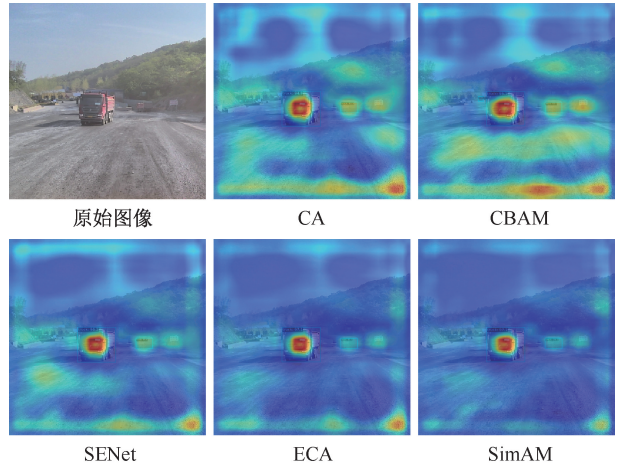


图 10 基于 GradCAM 的不同注意力机制特征
Fig. 10 Characterization of different attention mechanisms based on GradCAM

虽然 ECA 注意力有着更高的精度提升,但是 SimAM 注意力在矿区数据上精度仅略低于 ECA,同时在不增加参数和计算量的基础上,注意力分布也更加集中和清晰,更专注于小目标的细节特征。

3.4.3 消融试验与分析

为进一步证明文中提出的改进方法对矿区道路目标检测任务的有效性,设计消融试验,试验结果见表 6。由于初始数据集处于欠拟合状态,针对矿区道路数据集样本容量少以及类别尺度不均衡的问题进行数据均衡处理,均衡化可大幅度提升模型精度。

对特征的学习能力。进一步优化分类损失,采用 QFL 损失函数对分类和定位更精准的表示,并使用 SIoU 损失函数替代 CIoU 损失函数,能够更好地衡量边界框的损失。文中提出的改进模型在 VOC 数

据集和矿区道路数据集上分别达到 93.21% 和 91.88% 的精度检测效果。

4 结 论

1) 首次将数据均衡的处理方法运用到图像数据增强中,该方法能有效地模拟不同环境下的噪声数据,证明数据均衡操作能明显提升模型的检测性能。

2) 改进的障碍检测模型在多种矿区道路场景下,均有良好的检测性能,达到 91.88% 的精度和

68.7 帧/s 的检测速度,满足露天矿区道路实时检测的部署要求,总体性能优于目前主流的目标检测网络。将深度学习的方法应用于露天矿山的非结构化道路检测中来,取得较好的效果,验证了障碍检测算法对矿区无人驾驶的有效性。

3) 由于露天矿区数据采集难度大,风险系数高,因此,只进行了行进间的道路障碍检测,下一步考虑更多场景下图像数据与其他传感器数据融合的方法,应用到矿区道路的目标检测任务中来。

参 考 文 献

- [1] 付恩三,刘光伟,白润才,等. 大数据时代矿山协同监管监察新思路[J]. 中国安全科学学报, 2023, 33(8): 39-44.
FU Ensan, LIU Guangwei, BAI Runcai, et al. Exploration and research of mine multi-element cooperative supervision new thinking in era of big data[J]. China Safety Science Journal, 2023, 33(8): 39-44.
- [2] 范冰倩,董秉聿,王彪,等. 基于深度学习的地铁施工作业人员不安全行为识别与应用[J]. 中国安全科学学报, 2023, 33(1): 41-47.
FAN Bingqian, DONG Bingyu, WANG Biao, et al. Identification and application of unsafe behaviors of subway construction workers based on deep learning[J]. China Safety Science Journal, 2023, 33(1): 41-47.
- [3] 陈志发,余贵珍,张传莹,等. 露天矿区无人驾驶行车风险评估及防控策略仿真研究[J]. 煤炭学报, 2023, 48(4): 1782-1797.
CHEN Zhifa, YU Guizhen, ZHANG Chuanying, et al. Driving risk assessment and prevention strategies for autonomous vehicle in open-pits[J]. Journal of China Coal Society, 2023, 48(4): 1782-1797.
- [4] 张彦斐,封子晗,张嘉恒,等. 基于特征融合的果园非结构化道路识别方法[J]. 农业机械学报, 2023, 54(7): 35-44, 67.
ZHANG Yanfei, FENG Zihan, ZHANG Jiaheng, et al. Recognition method of orchard unstructured road based on feature fusion[J]. Transactions of the Chinese Society for Agricultural Machinery, 2023, 54(7): 35-44, 67.
- [5] 卢才武,马玲,阮顺领. 矿区复杂道路边缘检测[J]. 测绘科学, 2020, 45(9): 123-131.
LU Caiwu, MA Ling, RUAN Shunling. Mining area complex road edge detection based on HSV space[J]. Science of Surveying and Mapping, 2020, 45(9): 123-131.
- [6] 郝明,林惠晶,高彦彦. 基于改进主动轮廓模型的无人机影像矿区地裂缝提取[J]. 地球信息科学学报, 2022, 24(12): 2448-2457.
HAO Ming, LIN Huijing, GAO Yanyan. Ground fissure extraction method based on improved active contour model for UAV images in mining areas[J]. Journal of Geo-information Science, 2022, 24(12): 2448-2457.
- [7] CHENG Yong, WANG Wei, ZHANG Wenjie, et al. A multi-feature fusion and attention network for multi-scale object detection in remote sensing images[J]. Remote Sensing, 2023, 15(8): 2096-2115.
- [8] 马永杰,马芸婷,程时升,等. 基于改进 YOLOv3 模型与 Deep-SORT 算法的道路车辆检测方法[J]. 交通运输工程学报, 2021, 21(2): 222-231.
MA Yongjie, MA Yunting, CHENG Shisheng, et al. Road vehicle detection method based on improved YOLOv3 model and Deep-SORT algorithm[J]. Journal of Traffic and Transportation Engineering, 2021, 21(2): 222-231.
- [9] REDMON J, FARHADI A. Yolov3: an incremental improvement[C]. IEEE Conference on Computer Vision and Pattern Recognition, 2018: 89-95.
- [10] 张炳力,秦浩然,江尚,等. 基于 RetinaNet 及优化损失函数的夜间车辆检测方法[J]. 汽车工程, 2021, 43(8): 1195-1202.
ZHANG Bingli, QIN Haoran, JIANG Shang, et al. A method of vehicle detection at night based on RetinaNet and

- optimized loss functions[J]. *Automotive Engineering*, 2021, 43(8): 1 195–1 202.
- [11] HE Kaiming, ZHANG Xiangyu, REN Saoqing, et al. Spatial pyramid pooling in deep convolutional networks for visual recognition[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2015, 37(9): 1 904–1 916.
- [12] NAJIBI M, SAMANGOUEI P, CHELLAPPA R, et al. Ssh: single stage headless face detector[C]. *Proceedings of the IEEE International Conference on Computer Vision*, 2017: 4 875–4 884.
- [13] YANG Lingxiao, ZHANG Ruyuan, LI Lida, et al. SimAM: a simple, parameter-free attention module for convolutional neural networks[C]. *International Conference on Machine Learning*. PMLR, 2021: 11 863–11 874.
- [14] HU Jie, SHEN LI, SUN Gang. Squeeze-and-excitation networks[C]. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2018: 7 132–7 141.
- [15] WOO S, PARK J, LEE J Y, et al. Cbam: convolutional block attention module[C]. *Proceedings of the European Conference on Computer Vision (ECCV)*, 2018: 3–19.
- [16] LI Xiang, WANG Wenhai, WU Lijun, et al. Generalized focal loss: learning qualified and distributed bounding boxes for dense object detection[J]. *Advances in Neural Information Processing Systems*, 2020, 33: 21 002–21 012.
- [17] 石祥滨, 贾贺林. 基于控制点网格优化的航拍图像拼接算法[J]. *计算机应用与软件*, 2020, 37(5): 303–308.
- SHI Xiangbin, JIA Helin. Aerial image stitching algorithm based on control point grid optimization[J]. *Computer Applications and Software*, 2020, 37(5): 303–308.
- [18] HAO Qibin, ZHOU Daquan, FENG Jiashi. Coordinate attention for efficient mobile network design[C]. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2021: 13 713–13 722.
- [19] WANG Qilong, WU Banggu, ZHU Pengfei, et al. ECA-Net: efficient channel attention for deep convolutional neural networks[C]. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2020: 11 534–11 542.
- [20] SELVARAJU R R, COGSWELL M, DAS A, et al. Grad-cam: visual explanations from deep networks via gradient-based localization[C]. *Proceedings of the IEEE International Conference on Computer Vision*, 2017: 618–626.



作者简介：李刚（1979—），男，吉林德惠人，博士，教授，主要从事矿山压力及巷道围岩控制和智慧矿山等方面的研究。E-mail: ligang@lntu.edu.cn。