

doi: 10.7690/bgzdh.2025.04.005

一种智慧矿山场景下的目标检测算法

姚珊珊, 王静宇, 郝斌, 张飞, 高鹭, 任晓颖

(内蒙古科技大学信息工程学院, 内蒙古 包头 014000)

摘要: 针对露天矿生产场景中存在着目标像素低、小目标众多、背景复杂等问题, 在 YOLOv5s 的基础上提出一种多尺度和超分辨率网络(multiscale and super-resolution network, MS_Net)。在特征融合模块, 将 PANet 的三尺度检测升级为四尺度检测, 提高网络的多尺度学习能力, 并使用子像素卷积作为上采样方法; 提出一种多层次融合(multi layer fusion, MLF)模块, 融合了 PANet 3 个输出层的特征, 得到一个具有丰富语义信息和空间信息的特征图; 在预测层中, 使用 SIoU 作为定位损失函数, 优化模型的参数。实验结果表明: MS_Net 网络在 PASCAL VOC 数据集上 mAP 为 79.4%, FPS 为 59; 在矿山数据集上 mAP 为 80.2%, FPS 为 64.5, 模型可快速、准确、高效地对露天矿中的目标进行识别检测。

关键词: 智慧矿山; 目标检测; YOLOv5s; 多层特征融合; 子像素卷积

中图分类号: TP391 文献标志码: A

A Target Detection Algorithm in Smart Mine Scene

Yao Shanshan, Wang Jingyu, Hao Bin, Zhang Fei, Gao Lu, Ren Xiaoying

(School of Information Engineering, Inner Mongolia University of Science and Technology, Baotou 014000, China)

Abstract: Aiming at the problems of low target pixels, numerous small targets and complex background in the production scene of open-pit mine, an multiscale and super-resolution network (MS_Net) is proposed based on YOLOv5s. In the feature fusion module, the three-scale detection of PANet is upgraded to four-scale detection to improve the multi-scale learning ability of the network, and sub-pixel convolution is used as an up-sampling method; A multi layer fusion (MLF) module is proposed to fuse the features of three output layers of PANet, and a feature map with rich semantic information and spatial information is obtained. In the prediction layer, SIoU is used as the localization loss function to optimize the parameters of the model. The experimental results show that the mAP of MS_Net is 79.4% and the FPS is 59 on PASCAL VOC data set, and the mAP is 80.2% and the FPS is 64.5 on mine data set, and the model can identify and detect the target in the open-pit mine quickly, accurately and efficiently.

Keywords: smart mine; target detection; YOLOv5s; multi-layer feature fusion; sub-pixel convolution

0 引言

在计算机视觉技术的支持下, 煤矿的信息化建设已经从数字矿山建设逐渐转向了智慧矿山建设^[1-2]。在露天矿中, 车辆和人员的秩序是安全、有序、高效开采的重要环节。为了安全作业, 运输车、炸药车、油罐车、巴士等都有自己的行动轨迹及规则。人工对车辆秩序管理存在效率低、成本高、有效性差等问题, 因此, 对矿山车辆和人员进行快速、准确、高效地智能识别至关重要。

随着计算机视觉技术的快速发展, 基于视频的目标检测算法已成为学者们的研究热点^[3]。目标检测算法能检测出图像中的目标类别及位置信息。目前, 主流的目标检测算法主要分为端到端和非端到端^[4]。其中非端到端的方法具有准确性优势, 主要

包括 R-CNN^[5]、Faster R-CNN^[6]、Mask R-CNN^[7]等。端到端的方法具有速度性优势, 主要包括 SSD^[8]、YOLOv1~YOLOv5s^[9-13]、RetinaNet^[14]等。目前目标检测算法在智慧矿山中已有广泛应用, 但智慧矿山生产场景中的目标主要使用激光云台对车辆秩序进行监控, 存在着光影响大、车辆在图像中的尺寸小且多姿态、像素低等问题。

单鹏飞等^[15]在 Faster R-CNN 网络中进行改进, 在 ResNet50 骨干网络中加入 CBAM 注意力机制, 使得特征提取过程重点关注煤研运动状态。该方法在正确率上有一定的提升效果, 但检测速度无法满足实时性要求。阮顺领等^[16]在 RetinaNet 网络的基础上进行改进, 使用 RepVGG 作为骨干网络, 并且使用双向金字塔作为特征融合, 提高了模型的检测

收稿日期: 2024-08-07; 修回日期: 2024-09-14

基金项目: 内蒙古自治区科技计划项目(2021GG0046; 2021GG0048)

第一作者: 姚珊珊(1997—), 女, 四川人, 硕士。

通信作者: 郝斌(1984—), 男, 内蒙古人, 博士。

速度。张保俊等^[17]设计了 Darknet-31 特征提取网络，在提取大目标的语义特征的同时还能关注到小目标的细节信息。王滢暄等^[18]在 YOLOv4 网络上进行改进，将多幅图像拼接后进行车辆检测，在一定程度上解决了目标尺寸小的问题。以上 3 篇文献在检测速度和小目标检测效果上表现较好，但面对背景复杂、多尺度且多姿态时，还需改进。Mandal 等^[19]提出了一种单阶段车辆检测网络 (AVDNet)，在多个尺度上引入了 ConvRes 残差块，以缓解由于包含更深的卷积层而导致的较小对象特征消失的问题。这些残差块与放大的输出特征图一起，确保了小尺寸对象显著特征的稳健表示。但面对分辨率低、背景复杂的目标时还存在漏检和误检情况。

笔者主要针对智慧矿山场景下检测目标背景复杂、尺度多样化、分辨率低等问题，提出了一种多尺度和超分辨率网络 (MS_Net)。MS_Net 网络模型主要分 3 部分：1) Backbone 部分主要负责目标特征的提取，主要使用 CBM、CBMC3、SPPF 提取目标特征。2) Neck 负责不同大小特征图的融合，在上采样时使用子像素卷积，提高 Backbone 操作后特征图的分辨率，使特征在融合过程中减少小目标信息的缺失。为提高网络模型对多尺度且多姿态、目标较小的图像检测正确率，笔者提出多层融合模块 (MLF)，将具有丰富位置信息的浅层特征图、具有丰富位置信息的中层特征图和具有丰富语义信息的深层特征图进行融合。3) Head 层主要负责目标检测，具有 4 个检测头：160×160、80×80、40×40、

20×20，分别用于检测极小、小、中、大 4 种不同尺度的特征图，提高模型的多尺度学习能力。并且使用 SIoU 损失函数进行目标框的定位，提高模型的准确率。

1 所提的方法

1.1 网络模型

图 1 为笔者所提 MS_Net 网络的模型结构。在特征提取阶段，主要使用 CBM 和 CBMC3 模块进行特征的提取，得到所提模型中 F_1 、 F_2 、 F_3 、 F_4 4 个不同空间信息和语义信息的特征图。在特征融合阶段，使用 PANet 网络，采用自底向上和自顶向下 2 条路径融合不同尺度的特征，但由于 PANet 还存在着小目标漏检率和误检率高、多尺度多姿态目标检测效果不佳等问题，笔者提出用 MLF 模块来提高模型的多尺度学习能力，使用四尺度检测。MLF 模块融合了具有丰富位置信息的浅层特征图 F_1 、具有丰富位置信息和语义信息的中层特征图 P_3 和具有丰富语义信息的深层特征图 P_4 ，使得模型在多尺度检测上表现效果较好。最终得到 P_1 、 P_2 、 P_3 、 P_4 4 个特征图，用于后续不同尺度目标的检测工作。Head 层主要用于预测，图像在训练时，对预测值和真实值的位置和类别进行回归和分类，并使用 SIoU 损失函数计算预测框的损失值，然后通过反向传播更新梯度。在推理阶段，对预测器得到的分类定位结果进行阈值滤波和非极大值抑制，消除混叠或异常检测框，得到最终检测结果。

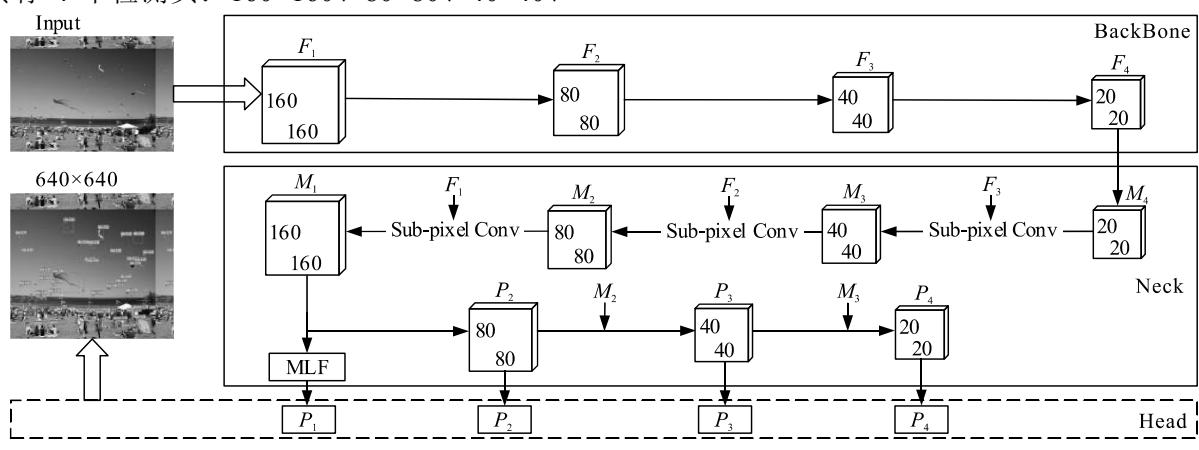


图 1 MS_Net 网络模型

1.2 Backbone 特征提取层

作为 MS_Net 主干网络部分，Backbone 采用 CSPDarknet 作为主干特征提取网络。跨阶段局部网络 (cross stage partial network, CSPNet)^[20]通过将输入图像切分为 2 部分，然后通过跨阶段层次结构

再将切分的 2 部分进行结合，在保持原有正确率的同时还提高了模型的检测速度。

如图 2 所示，Backbone 中，图像输入后先经过 CBM 结构，CBM 结构由 Conv、BN、Mish 3 部分组成。在初步提取特征信息后，再经过 CBMC3 结

构进一步提取细粒度特征，得到具有强定位信息的 F_1 特征图。经过 3 组 CBMC3 结构提取更多更丰富

的语义信息，得到 F_2 、 F_3 、 F_4 3 个具有不同语义信息和空间信息的特征图。

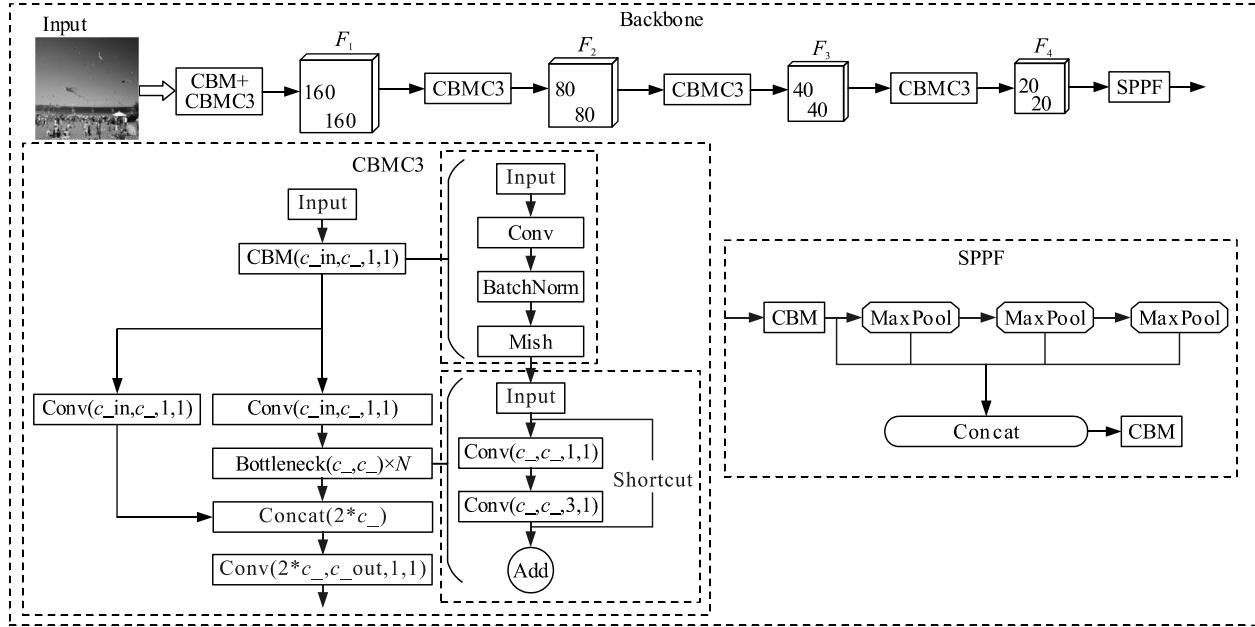


图 2 Backbone 结构

1.3 Neck 特征融合层

在 Neck 层主要使用 PANet^[21]采用自底向上和自顶向下的双向特征融合，使用浅层特征信息对检测目标进行定位功能。PANet 还在浅层和深层之间添加残差连接，缩短层之间的路径。由于 PANet 经过多层的卷积和残差结构会导致小目标信息的丢失以及特征图模糊等情况。经实验表明，PANet 提出的三尺度检测会使模型在面对多尺度多样且目标多姿态时检测效果不佳。如图 1 的 Neck 部分所示，笔者所提模型将 PANet 原来的插值上采样方法替换为子像素卷积，使用超分辨率的方法来提高特征图的分辨率。将 PANet 的三尺度检测升级为四尺度检测，提高模型的多尺度适应能力。

1.3.1 子像素卷积

子像素卷积 (sub-pixel convolution)^[22] 又叫做像素洗牌 (pixel shuffle)，是一种图像及特征图上采样的方法。常见的上采样方法有反卷积、反池化、直接上采样、双线性插值法等。子像素卷积作为一种图像或视频超分辨率方法，是数字图像处理中非常感兴趣的主题，被广泛应用于 HDTV、医学成像、卫星成像、人脸识别和监控领域^[22]。子像素卷积的插值算法是隐含在卷积层当中通过学习的方式得来的，其计算结果更具准确性和鲁棒性。子像素卷积的操作是在小分辨率图像上进行的，因此上采样效率也很高。小目标难以检测的一个重要原因是因为特征

图模糊，在卷积过程中容易被当成背景信息忽略掉；因此，将模糊的特征图通过超分辨率重构的方法提高其分辨率，从而将小目标的特征表达转化为和大、中目标一样或近似的特征表达，是解决小目标检测效果不佳的一个重要的方法。

如图 3 所示，子像素卷积在通道维度上分散像素，在宽度和高度维度上增加像素，从而提高卷积效率。如果要将低分辨率的特征图变成原来 r 倍大小的高分辨率图像，需要生成 r^2 个同等大小的特征图，然后根据前面的卷积层自动学习插值函数，重新排列得到特征维度为 $C \times rH \times rW$ 的特征图，就达到了上采样的效果，如式 1 所示：

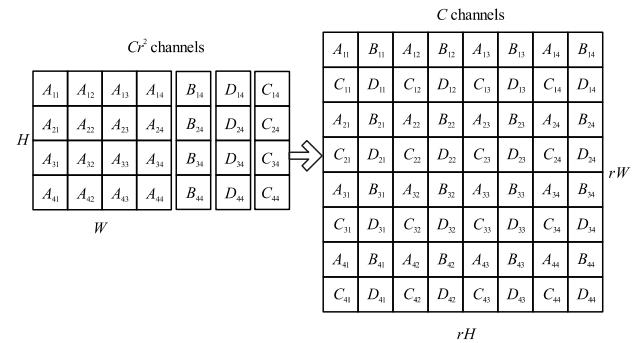


图 3 子像素卷积

$$\text{SUB}(T)_{x,y,c} = T \lfloor x/r \rfloor, \lfloor y/r \rfloor, C \cdot r \cdot \text{mod}(y, r) + C \cdot r \cdot \text{mod}(x, r) + c. \quad (1)$$

式中： $\text{SUB}(T)_{x,y,c}$ 为子像素卷积后的输出结果； r 为特征图要放大的倍数，本文中特征图需要放大 2 倍，

因此 $r=2$; x 、 y 为特征图的长和宽; c 为通道数, C 为输入特征图。

1.3.2 MLF

在 PANet 中使用三尺度对上下文特征融合, 提高了模型的检测正确率。当检测目标背景复杂、尺度多样且多姿态、检测目标较小时, 检测效果不理想。多尺度特征图的融合有利于小目标的检测。此外, 小目标的有效空间信息通常存在于特征图 F_1 中。笔者提出 MLF 模块, 融合浅层具有丰富位置信息的特征图、中间层丰富语义信息和位置信息的特征图以及深层包含小目标语义信息的特征图, 生成语义和空间信息丰富的高分辨率特征图, 从而提高小目标检测能力。

如图 4 所示, P_3 特征图经过 CBM 结构提取细节信息后使用子像素卷积进行上采样操作得到 P'_4 , 如式(2)所示。

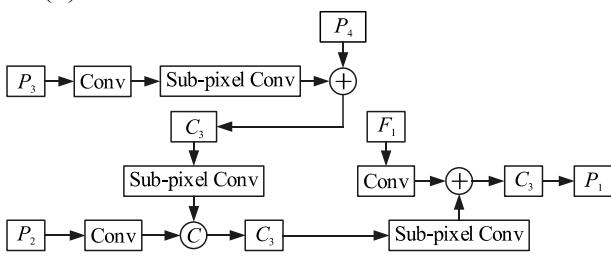


图 4 MLF 结构

$$P'_4 = \text{CBM}(\text{SUB}(P_4)); \quad (2)$$

式中 SUB 为子像素上采样。 P'_4 特征图的通道数和图像大小跟 P_3 特征图一样, 方便与 P_3 特征图进行融合。 P_3 特征图和 P'_4 进行逐元素相加, 融合 P_3 特征图的空间信息和语义信息, 然后经过 CBMC3 模块消除特征图相加可能造成的混叠效应。为了使得特征图能与 P_2 特征图进行融合, 通过子像素卷积扩大特征图的分辨率得到特征图 P'_3 , 如式(3)所示:

$$P'_3 = \text{SUB}(\text{CBMC3}(\text{ADD}(P'_4, P_3))). \quad (3)$$

P_2 特征图通过 CBM 结构进一步提取特征并调整通道数, 并于 P'_3 进一步融合特征, 为使得融合了丰富语义信息的特征图能有效使用浅层特征的强定位信息, 将得到的特征图进行上采样操作, 最终得到特征图 P'_2 , 如式(4)所示:

$$P'_2 = \text{SUB}(\text{CBMC3}(\text{Cat}(\text{CBM}(P_2), P'_3))). \quad (4)$$

式中 Cat 为拼接操作。 F_1 特征图通过 CBM 模块调节通道数与 P'_2 逐元素相加, 使得缺乏位置信息的深层特征图能够使用浅层的位置信息, 以提高小目标的检测效果, 并生成 160×160 大小的特征图 P_1 用于

检测小目标信息, 提高模型的多尺度学习能力, 如式(5)所示:

$$P_1 = \text{CBMC3}(\text{ADD}(P'_2, \text{CBM}(F_1))). \quad (5)$$

1.4 预测层

MS_Net 的预测层由 4 个检测头组成, 分别为 160×160 、 80×80 、 40×40 、 20×20 大小, 由此来实现模型的多尺度学习能力。在训练阶段, MS_Net 通过反向传播不断优化损失 Loss, 更新网络权值, 直到模型收敛。在测试阶段, MS_Net 不执行反向传播。它直接对预测器得到的分类定位结果进行置信度阈值筛选、非极大值抑制处理等操作, 得到最终检测结果^[23]。

根据已有的研究^[23-25], MS_Net 中专门用于分类定位的损失函数主要包括 3 个组成部分: 类别损失函数(classification loss)、置信度损失函数(confidence loss)和定位损失函数(localization loss)。

分类损失 L_{cls} 是检测目标的预测类别与真实类别之间的误差, 如式(6)所示:

$$L_{\text{cls}} = -\sum_{i=0}^{S^2} \sum_{j=0}^B I_{ij}^{\text{obj}} \sum_{C \in \text{cls}} \left[\hat{p}_{ij}^C \log(p_{ij}^C) + (1 - \hat{p}_{ij}^C) \log(1 - p_{ij}^C) \right]. \quad (6)$$

式中: S 为特征图的边长; B 为特征图中每个单元的锚点数量; I_{ij}^{obj} 为第 i 个单元中的第 j 个锚是否对落入该区域的边界盒的概率负责, 若负责 $I_{ij}^{\text{obj}}=1$, 否则=0; p_{ij}^C 为第 i 个单元中第 j 个锚的预测类别; \hat{p}_{ij}^C 为真实的类别。

置信度损失 L_{conf} 是检测目标置信度得分预测的相对误差, 如式(7)所示:

$$L_{\text{conf}} = -\sum_{i=0}^{S^2} \sum_{j=0}^B I_{ij}^{\text{obj}} \log(C_{ij}) - \lambda_{\text{noobj}} \sum_{i=0}^{S^2} \sum_{j=0}^B I_{ij}^{\text{noobj}} \log(1 - C_{ij}). \quad (7)$$

式中: I_{ij}^{noobj} 的定义和值与 I_{ij}^{obj} 相反; C_{ij} 为所提出的第 i 个单元中的第 j 个锚的置信度得分, $C_{ij} \in [0, 1]$; λ_{noobj} 为平衡系数在训练时被设置为 0.5。

定位损失函数 L_{SIOU} 是定位框的预测值和真实值之间的误差。 SIOU ^[26] 损失函数为预测值和实际值之差的和的一半。 SIOU 损失函数由 4 部分组成: Angle cost、Distance cost、Shape cost、IoU cost。 SIOU 损失函数如式(8)所示:

$$\Lambda = 1 - 2 * \sin^2(\arcsin(x) - \pi/4);$$

$$\Delta = \sum_{t=x,y} (1 - e^{-\gamma_{\text{pt}}}), \text{ 其中 } \gamma = 2 - \Lambda;$$

$$\mathcal{Q} = \sum_{t=w,h} (1 - e^{\alpha t})^\beta, \quad L_{IoU} = 1 - IoU; \quad (8)$$

$$L_{SIOU} = 1 - IoU + (\Delta + \mathcal{Q})/2. \quad (9)$$

式中: Δ 为 Angle cost, 如果 $\alpha < \pi/4$, 则收敛过程将最小化 α , 否则, 最小化 β ; Δ 为 Distance cost; \mathcal{Q} 为 Shape cost。SIoU 考虑到了回归之间的 Angle cost, 重新定义了损失函数, 有效地减少了损失的总自由度提高了推理的准确性; 因此, 得到 MS_Net 的总损失 Loss 如下:

$$Loss = L_{SIOU} + L_{conf} + L_{cls}. \quad (10)$$

2 实验结果与分析

2.1 实验环境及数据集

2.1.1 实验环境

笔者所提模型在 Ubuntu20.04 上进行训练, 使用的编程语言为 Python 3.8.13、深度学习框架为 Pytorch 1.8.1、GPU 为 GeForce RTX3080Ti, GPU 显存为 12 G, 使用 CUDA11.6 对 GPU 进行加速。

2.1.2 实验数据集

1) PASCAL VOC。

Pascal VOC 挑战赛提供了 VOC2007 和 VOC2012 2 个数据集。数据集分为 20 类, 主要分人、动物、车辆和室内物品 4 大类。本文中训练验证集使用的是 VOC2012, 按照 8:2 进行划分。train(13 700 张图像), val(3 425 张图像), 测试集使用的是 VOC2007 的 test(4 952 张图像)。

2) 矿山数据集。

智慧矿山中的数据采集均使用激光云台进行采集, 将采集到的视频数据转换成图像数据, 并在图像数据中选择所需种类的图像, 使用 labelme 软件进行标注, 主要分自卸车、挖掘机、洒水车、推土机、巴士、休息室、扬尘、油罐车、炸药车、运输车、钻机等 13 类。然后将标注好的文件转换成 txt 格式。其中训练集 3 968 张图像, 测试集 991 张图像。

2.2 实验细节及评价指标

模型训练参数如表 1 所示。

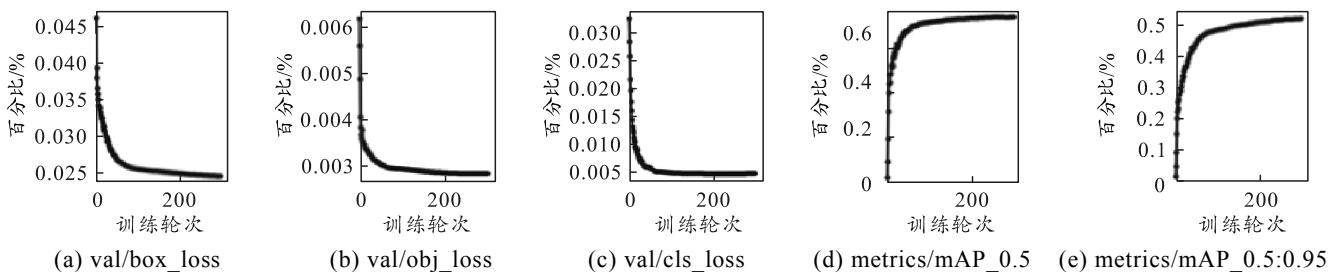


图 5 PASCAL VOC 数据集训练结果

表 1 模型训练参数

参数名称	参数
输入分辨率	640*640
初始学习率(lr0)	0.01
循环学习率(lrf)	0.02
权重衰减系数 (weight_decay)	0.0005
学习率动量 (warmup_momentum)	0.937
训练次数(epoch)	300
批处理量(batchsize)	32
锚框(anchor)	[9, 11, 21, 19, 17, 41] [43, 32, 39, 70, 86, 64] [65, 131, 134, 130, 120, 265] [282, 180, 247, 354, 512, 387]

评估指标是评价目标检测算法方法好坏的重要依据, 笔者采用的评价指标为: 平均准确度 (average precision, AP)、平均准确率均值 (mean average precision, mAP)、交并比 (IoU)、帧率 (frame per second, FPS)^[27]。FPS 为每秒内可以处理的图片数量。mAP@0.5 评价指标为将 IoU 阈值设为 0.5 时, 求每一类的所有图片的精度, 然后所有类别求平均。mAP@.5:.95 表示 IoU 阈值从 0.5 到 0.95, 步长为 0.5 上的平均精度。如下式所示:

$$F_1 = 2 \frac{P \times R}{P + R}, \quad P = \frac{TP}{TP + FP}, \quad R = \frac{TP}{TP + FN}; \quad (11)$$

$$AP = \int_0^1 P(R)d(R); \quad (12)$$

$$mAP = \frac{1}{N} \sum_{i=1}^N AP_i. \quad (13)$$

式中: TP 为正样本被正确识别为正样本; TN 为负样本被正确识别为负样本; FP 为负样本被错误识别为正样本; FN 为正样本被错误识别为负样本。

2.3 实验结果与分析

2.3.1 模型训练

模型训练使用表 1 给出的参数进行训练, 主要在 PASCAL VOC 数据集上和矿山数据集上进行。图 5 为 PASCAL VOC 数据集上的训练结果, 包括损失曲线和正确率曲线。图 6 为矿山数据集的训练结果。

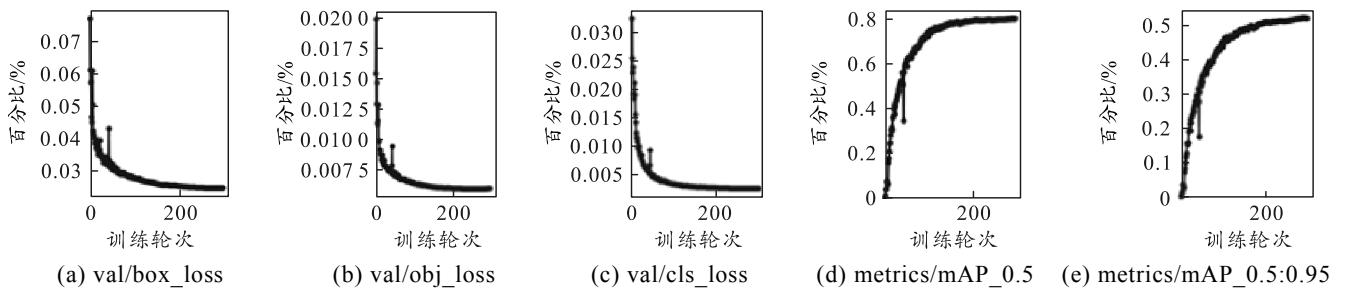


图 6 矿山数据集训练结果

由图 5 和 6 可以看出：MS_Net 网络的损失较小且收敛速度快，正确率稳健上升，矿山数据集正确率稍微有点波动，但最终模型还是有较好的学习能力。

为进一步验证 MS_Net 算法的有效性，笔者与经典的目标检测算法 SSD 和先进的目标检测算法 (YOLOv5s、MSFAN) 进行对比实验。其中 MSFAN^[28] 是一种改进的 MobileNet_YOLOv3 模型，该模型基于 MobileNet_YOLOv3 网络模型，并结合多尺度特征融合等改进措施，在高实时性的同时增强了模型对小目标的检测能力^[28]。结果如表 2 所示。

表 2 不同模型对比 %

数据集	网络模型	特征提取网络	mAP@0.5	mAP@.5:.95
矿山	SSD ^[8]	VGG-16	78.1	50.4
	MSFAN ^[28]	DarkNet-53	76.2	48.6
	YOLOv5s	DarkNet-53	77.4	50.9
	MS_Net	DarkNet-53	80.2	52.1
PASCAL VOC	SSD ^[8]	VGG-16	77.5	51.2
	MSFAN ^[28]	DarkNet-53	75.5	49.3
	YOLOv5s	DarkNet-53	78.1	53.2
	MS_Net	DarkNet-53	79.4	53.9

由表 2 可知，MS_Net 算法较其他算法正确率高。在矿山数据集中，笔者所提模型 mAP@0.5 高于其他模型，达到 80.2%，比 YOLOv5s 模型高 2.8%。在 PASCAL VOC 数据集中，笔者所提模型检测精度较 SSD 提高 1.9%、较 MSFAN 提高 3.9%、较 YOLOv5s 提高 1.3%。实验结果表明，笔者所提方法正确率优于所比较的其他模型。

2.3.2 损失函数实验

为验证在 MS_Net 模型中使用 SIoU 损失函数的有效性，笔者将对比 Alpha^[29]、EIoU^[30]、CIoU^[31]、SIoU^[26] 4 个损失函数，实验结果如表 3 所示。

表 3 MS_Net 模型中损失函数效果 %

损失函数	mAP@0.5	mAP@.5:.95
Alpha	76.1	50.2
EIoU	78.8	50.4
CIoU	79.4	51.4
SIoU	80.2	52.1

根据表 3 所示，SIoU 损失函数运用到 MS_Net

模型中正确率高于 Alpha、EIoU 和 CIoU 损失函数，因此笔者选择 SIoU 损失函数作为回归框定位函数。

2.3.3 多尺度检测实验

为验证 MS_Net 网络对多尺度目标检测的有效性，将检测目标分为 3 类：小目标物体面积小于 32^2 ，中等目标物体面积大于 32^2 小于 96^2 ，大目标物体面积大于 96^2 。此方案可以验证多尺度目标的检测能力，结果如表 4 所示。

表 4 不同大小目标 mAP@.5:.95 检测结果 %

目标	YOLOv5s 算法	本文中算法	数据差值
小目标	21.0	25.4	+4.4
中等目标	46.6	50.2	+3.6
大目标	62.3	63.1	+0.8

从表 4 中可以看出：笔者所提算法小目标、中等目标、大目标的 mAP@.5:.95 都有所提升。多尺度实检测验证明，笔者所提算法提高了模型多尺度目标的检测能力。

2.3.4 消融实验

为验证 MS_Net 的有效性，笔者共同应用在 YOLOv5s 模型框架下对矿山数据集进行消融实验，其结果如表 5 所示。

表 5 消融实验结果 %

分组	A	B	C	mAP@0.5	mAP@.5:.95	FPS
1				78.7	51.9	90.9
3	√			78.6	51.1	76.3
4	√	√		79.4	51.2	64.5
5	√	√	√	80.2	52.1	64.5

A 为使用子像素卷积的 4 尺度检测；B 为使用子像素卷积的 4 尺度检测+MLF；C 为使用子像素卷积的 4 尺度检测+MLF+SIoU 损失函数。其中“√”表示加入所改进的模块。通过表 5 可以看出：当原网络加入四尺度检测后，由于数据集存在大量的大目标和中目标，并且网络复杂度提升可能会使浅层的学习能力下降，导致网络模型正确率有所降低。当模型加入 MLF 模块后，融合了浅层、中层和深层的信息，使模型能充分学习不同层之间的特征，

使得 mAP@0.5 较 YOLOv5s 上升了 0.7 个百分点。加上 SIoU 损失函数后，使得模型具有更快的收敛速度，优化了模型参数，使得模型的正确率再一次提升，达到 80.2%。MS_Net 的 FPS 达到 64.5，满足实时性要求。实验结果证明了 MS_Net 算法具有一定的优越性。

2.4 可视化分析

为进一步证明笔者所提模型的有效性，选取一些存在大量的小目标、背景复杂、图像模糊的图像进行测试。MS_Net 和 YOLOv5s 的检测效果如图 7 所示。

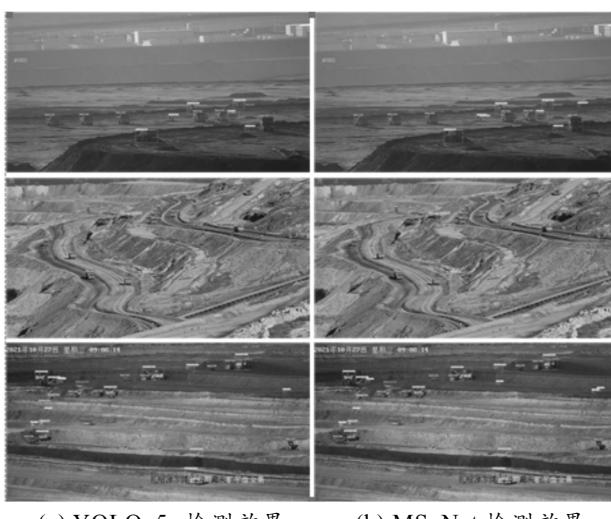


图 7 YOLOv5s 与 MS_Net 检测效果对比

由图 7 可以看出：YOLOv5s 存在部分漏检和误检问题，而 MS_Net 算法能将 YOLOv5s 模型漏检的检测出来，误检较 YOLOv5s 低。检测效果再一次证明，当检测目标较小、背景干扰性强、分辨率低、光影响大的时候，MS_Net 算法的检测效果较 YOLOv5s 查全率更高、误检率更低。检测效果再一次证明，MS_Net 算法的检测效果较 YOLOv5s 好，模型的改进是有效的。

为进一步证明 MS_Net 在小目标检测上的优越性，还将 YOLOv5s 和 MS_Net 网络进行比较，结果如图 8 所示。

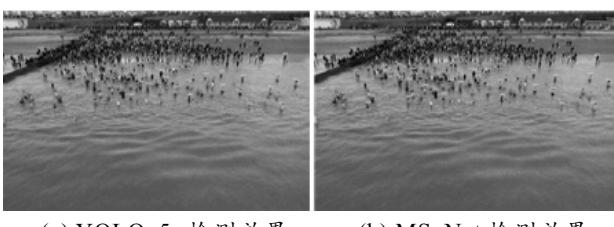


图 8 YOLOv5s 与 MS_Net 小目标检测效果对比

由图 8 可以看出，MS_Net 对小目标的检测效果优于 YOLOv5s 算法。

3 结束语

为满足智慧矿山中检测目标的检测精度和速度要求，笔者设计了一种针对分辨率低、背景干扰性强、小目标众多场景的目标检测网络 MS_Net。实验结果证明：MS_Net 在检测智慧矿山特定场景下的目标时，检测的正确率和效果优于 YOLOv5s 算法，并且检测速度能够满足实时性需求。本文中的矿山数据集训练数据还较少，导致模型的整体正确率不高，后续可以扩充数据集，以满足矿山实际检测需求，保障露天矿开采的安全性和高效性。

参考文献：

- [1] 杜祥龙. 安全系统工程在矿山安全管理中的应用[J]. 世界有色金属, 2021(18): 164–165.
- [2] 李梅, 姜展, 姜龙飞. 三维可视化技术在智慧矿山领域的研究进展[J]. 煤炭科学技术, 2021, 49(2): 153–162.
- [3] 张强, 李嘉锋, 卓力. 车辆识别技术综述[J]. 北京工业大学学报, 2018, 44(3): 382–392.
- [4] 张保俊, 袁广驰, 李志祥. 一种智慧矿山场景下的目标检测方法[J]. 物联网技术, 2022, 12(10): 21–23, 27.
- [5] GIRSHICK R, DONAHUE J, DARRELL T, et al. Rich feature hierarchies for accurate object detection and semantic segmentation[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. IEEE, 2014: 580–587.
- [6] REN S, HE K, GIRSHICK R, et al. Faster R-CNN: Towards real-time object detection with region proposal networks[J]. Advances in neural information processing systems, 2015, 28: 91–99.
- [7] HE K, GKIOXARI G, DOLLÁR P, et al. Mask R-CNN[C]//Proceedings of the IEEE International Conference on Computer Vision. Washington D C, USA: IEEE Press, 2017: 2961–2969.
- [8] LIU W, ANGUELOV D, ERHAN D, et al. SSD: Single Shot MultiBox detector. Proceedings of the 14th European Conference on Computer Vision[C]. Amsterdam: Springer, 2016: 21–37.
- [9] REDMON J, DIVVALA S, GIRSHICK R, et al. You only look once: unified, real-time object detection[Z]. 2016.
- [10] REDMON J, FARHADI A. YOLO9000: Better, faster, stronger [C]//2017 IEEE Conference on Computer Vision and Pattern Recognition(CVPR). IEEE, 2017: 6517–6525.
- [11] REDMON J, FARHADI A. YOLOv3: an incremental improvement[EB/OL]. 2021-03-05. <https://arxiv.org/abs/1804.02767>.
- [12] BOCHKOVSKIY A, WANG C Y, LIAO H Y M. YOLOv4:

- optimal speed and accuracy of object detection[EB/OL]. 2021-03-05. <https://arxiv.org/abs/2004.10934>.
- [13] ZHU X, LYU S, WANG X, et al. TPH-YOLOv5s: Improved YOLOv5s Based on Transformer Prediction Head for Object Detection on Drone-captured Scenarios[J]. IEEE/CVF International Conference on Computer Vision Workshops (ICCVW), 2021: 2778–2788.
- [14] LIN T Y, GOYAL P, GIRSHICK R, et al. Focal loss for dense object detection[C]//Proceedings of the 2017 IEEE International Conference on Computer Vision (ICCV). Venice: IEEE, 2017: 2999–3007.
- [15] 单鹏飞, 孙浩强, 来兴平. 基于改进 Faster R-CNN 的综放煤矸混合放出状态识别方法[J]. 煤炭学报, 2022, 47(3): 1382–1394.
- [16] 阮顺领, 李少博, 顾清华. 基于双向特征融合的露天矿区道路障碍检测[J/OL]. 煤炭学报: 1-13[2023-02-07]. <https://doi.org/10.13225/j.cnki.jccs.2022.0198>.
- [17] 张保俊, 袁广驰, 李志祥. 一种智慧矿山场景下的目标检测方法[J]. 物联网技术, 2022, 12(10): 21–23, 27.
- [18] 王滢暄, 宋焕生, 梁浩翔. 基于改进的 YOLOv4 高速公路车辆目标检测研究[J]. 计算机工程与应用, 2021, 57(13): 218–226.
- [19] MANDAL M, SHAH M, MEENA P, et al. AVDNet: A small-sized vehicle detection network for aerial visual data[J]. IEEE Geoscience and Remote Sensing Letters, 2019, 17(3): 494–498.
- [20] WANG C Y, LIAO H, YEH I H, et al. CSPNet: A new backbone that can enhance learning capability of CNN[C]//2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW). IEEE, 2020.
- [21] LIU S, QI L, QIN H F, et al. Path aggregation network for instance segmentation[C]//In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition
- (CVPR). USA: UT, Salt Lake City, 2018: 8759–8768.
- [22] SHI W, CABALLERO J, HUSZÁR F, et al. Real-time single image and video super-resolution using an efficient sub-pixel convolutional neural network[C]//In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR). USA: NV, Las Vegas, 2016: 1874–1883.
- [23] QI G, ZHANG Y, WANG K, et al. Small object detection method based on adaptive spatial parallel convolution and fast multi-scale fusion[J]. Remote Sensing, 2022, 14(2): 420.
- [24] 何雨, 田军委, 张震, 等. YOLOv5s 目标检测的轻量化研究[J]. 计算机工程与应用, 2023, 59(1): 92–99.
- [25] 毛雨晴, 赵奎. 基于改进 YOLOv5s 的多任务安全人头检测算法[J]. 计算机工程, 2022, 48(8): 136–143.
- [26] GEVORGIAN Z. SIoU loss: More powerful learning for bounding box regression[J]. arXiv preprint arXiv, 2022.
- [27] LI Y, LÜ C. SS-YOLO: an object detection algorithm based on YOLOv3 and ShuffleNet[C]//2020 IEEE 4th Information Technology, Networking, Electronic and Automation Control Conference (ITNEC). IEEE, 2020: 769–772.
- [28] 张陶宁, 陈恩庆, 肖文福. 一种改进 MobileNet—YOLOv3 网络的快速目标检测方法[J]. 小型微型计算机系统, 2021, 42(5): 1008–1014.
- [29] LI H, ZHOU Q, MAO Y, et al. Alpha-SGANet: A multi-attention-scale feature pyramid network combined with lightweight network based on Alpha-IoU loss[J]. Plos one, 2022, 17(10): e0276581.
- [30] ZHANG Y F, REN W, ZHANG Z, et al. Focal and efficient IoU loss for accurate bounding box regression[J]. Neurocomputing, 2022, 506: 146–157.
- [31] ZHENG Z, WANG P, LIU W, et al. Distance-IoU loss: Faster and better learning for bounding box regression[J]. Proceedings of the AAAI conference on artificial intelligence. 2020, 34(7): 12993–13000.
-

(上接第 5 页)

参考文献:

- [1] 汪辉. 基于模糊神经网络滑模变结构火炮伺服系统的控制研究[D]. 南京: 南京理工大学, 2017.
- [2] 胡继辉. 非平衡坦克炮炮控系统研究[D]. 南京: 南京理工大学, 2018.
- [3] 胡近朱. 某侦察车稳定平台伺服系统的控制策略研究[D]. 南京: 南京理工大学, 2021.
- [4] 章钱, 李士勇. 一种新型自适应 RBF 神经网络滑模制导律[J]. 智能系统学报, 2009, 4(4): 339–344.
- [5] 王双霞, 陈丽, 王洪瑞. 基于 RBF 增益调节的机器人滑模控制策略[J]. 机床与液压, 2009, 37(8): 122–124.
- [6] 付天忠, 高国琴, 方志明. 克服链间耦合并联机构双增益自适应滑模控制[J]. 软件导刊, 2022, 21(11): 44–51.
- [7] FENG H, SONG Q Y, MA S L, et al. A new adaptive sliding mode controller based on the RBF neural network for an electro-hydraulic servo system[J]. ISA transactions, 2022, 129(PtA): 472–484.
- [8] 吴梓萌, 茹盼盼, 苏晗. 基于 RBF 神经网络自适应滑模控制技术的某舰炮装填机构控制研究[J]. 火炮发射与控制学报, 2022, 43(3): 56–61.