基于深度学习的轨条砦障碍毁伤等级识别研究

熊建武，万文乾，姜　波，朱会杰，史俊超
（陆军研究院五所，江苏 无锡 214035）

摘　要：针对传统毁伤评估方法需要大量试验数据修正毁伤评估模型，且需要对炸点位置进行测量并结合弹目交汇条件才能计算毁伤效果的问题，提出了一种基于图像识别的端到端轨条砦障碍毁伤等级识别方法，进而快速对毁伤后轨条砦进行毁伤等级判定。该方案基于改进的YOLO5算法实现轨条砦图像检测、毁伤特征识别和毁伤等级评估。首先，由专家对轨条砦数据集进行标注，利用改进的马赛克拼接提高小目标检测能力；然后，使用卷积块注意力模块代替部分卷积模块，提升模型整体性能；最后，针对轨条砦的尺寸特点优化检测端结构，移除大目标尺度特征以减少推理计算量。实验结果表明：该方法可以在保证检测实时性的同时，毁伤评估精度达到98%。

关键词：轨条砦；神经网络；目标检测；毁伤评估
中图分类号：TJ410　　**文献标志码：**A　　**文章编号：**2097-2741(2024)04-034-007

Rail Obstacles Damage Level Recognition Based on Deep Learning

XIONG Jianwu, WAN Wenqian, JIANG Bo,
ZHU Huijie, SHI Junchao
(The Fifth Institute of Army Academe, Wuxi 214035, Jiangsu)

ABSTRACT: In traditional damage assessment methods, the modification of the damage assessment model requires a large amount of experimental data, while the damage effect could not be calculated without the measurement of the location of the explosion point combined with the intersection conditions of the projectile and target. To solve the problem, an end-to-end rail obstacles damage level recognition method based on image recognition was proposed, by which the damage level of damaged rail obstacles could be determined quickly. Based on YOLO5 algorithm, the proposed method could realize rail obstacle images detection, damage characteristics recognition and damage level assessment. Firstly, the rail obstacle dataset was marked by experts, thereby the detection ability of small targets was improved by modified mosaic stitching. Then, convolutional block attention modules were used to replace some convolutional modules to improve the overall performance of the model. Finally, the detection end structure was optimized based on size characteristics of rail obstacles, and large target scale features were removed to reduce the load of the inference computation. Experimental results on the test set showed that the proposed method could realize real-time detection with a damage assessment accuracy of 98%.

KEYWORDS: rail obstacles; neural network; target detection; damage assessment

收稿日期：2023-10-11

第一作者：熊建武（1969—），男，副研究员，主要从事作战效能仿真评估研究。E-mail：dishiren10@126.com

引用格式：熊建武，万文乾，姜波，等.基于深度学习的轨条砦障碍毁伤等级识别研究[J].火箭军工程大学学报，2024，38（4）：34-39，46.

轨条砦障碍是一种由钢轨和混凝土基座组成的岸防工事，成阵列状排布于海岸，用于阻滞登陆，是登陆作战重点打击目标。快速有效地识别并破除各类轨条砦障碍，特别是对打击后的轨条砦障碍进行毁伤评估、标注通路，是确保登陆成功和向纵深进攻发展的前提。轨条砦具有目标小、排列密集，毁伤后形态各异的特点，因此针对轨条砦目标的探测、识别和毁伤评估具有一定困难。

随着深度学习热潮的到来，许多基于深度学习的方法，尤其是卷积神经网络（Convolutional Neural Network，CNN），在计算机视觉领域展现了出色的性能。越来越多的研究人员倾向于使用CNN来解决各种计算机视觉任务，例如目标检测[1-3]和语义分割[4-6]等。CNN可以在训练过程中学习丰富有效的特征表示，自动提取更深层次的高级语义特征，相比传统手工设计的特征，这些学习到的特征具有更好的鲁棒性和泛化性。基于CNN的目标检测器包括基于候选框的两阶段检测器和一阶段检测器：两阶段检测方法主要是基于区域卷积神经网络（Region with CNN Features，R-CNN）的一系列方法；主流的一阶段检测器包括单次多边框检测（Single Shot MultiBox Detector，SSD）和单阶段快速目标检测算法（You Only Look Once，YOLO）等。其中，YOLO5检测与识别模型广泛使用在学术研究和工业等实际应用中，在精度和速度上都有着显著的优势[7-11]。

目前，国内外缺乏利用图像技术对轨条砦毁伤等级进行识别的相关研究，一般仅针对完整的轨条砦进行图像识别[12]。本文提出一种基于改进YOLO5的轨条砦实时毁伤效果等级评估算法，利用改进的马赛克拼接扩充样本集，在多尺度特征融合网络中删除大目标尺度特征图，提高模型对小目标的检测精度，同时降低计算复杂度。进一步将自注意力模块替换YOLO5中的部分卷积模块和跨阶段局部模块（Cross Stage Partial，CSP），以提升模型整体性能。

1 技术方案

1.1 YOLO5基础架构

根据网络深度和网络宽度的不同，YOLO5可分为4个模型，分别为：small，middle，large和xlarge，模型参数依次增加，目标检测能力逐渐提升。YOLO5模型由特征提取主干网BackBone、特征融合Neck和检测头Head这3部分构成，其中BackBone部分主要由Focus、Conv（卷积）、CSP和空间金字塔池化（Spatial Pyramid Pooling，SPP）等模块组成，如图1所示。

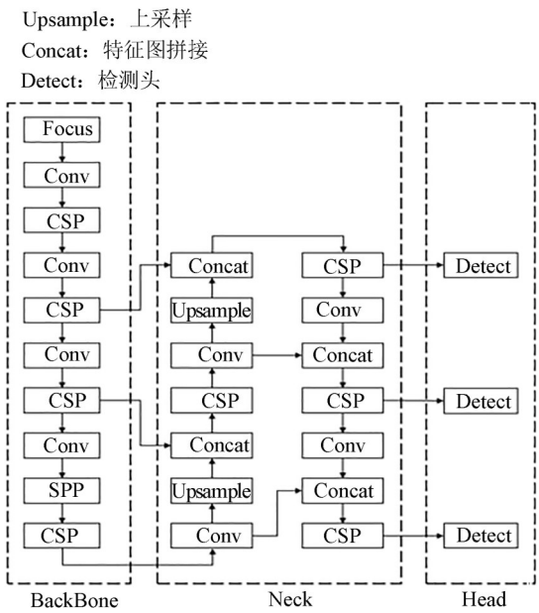

图1 YOLO5基础架构

Fig. 1 Basic structure of YOLO5

1.2 YOLO5模型优化

在YOLO5作为基线模型的基础上，需要进一步针对性的优化。具体措施为小目标改进（改进的马赛克数据增强）、卷积块注意力模块（Convolutional Block Attention Module，CBAM）、移除大目标检测特征图。

1.2.1 改进的马赛克数据增强

采用原始马赛克拼接的数据增强手段，将多张训练图片按一定比例拼接成一张，变相增加每一个训练步骤中的图片数量。不同的拼接方法生成的样本增强了模型的鲁棒性。由于在拼接过程中会对原始图像进行裁剪，大尺度目标被裁剪后可能只保留部分特征，而小尺度目标受裁剪的影响较小，一般被完整保留，因此，马赛克拼接能鼓励模型学习小尺度目标的特征，如图2所示。

每次随机选取4幅不同的训练样本图像，进行随机的裁剪、平移、尺度缩放，然后拼接成一幅图像作为训练输入。该方式能够增强神经网络模型对多尺度目标的学习能力，尤其是提高对小尺度目标的识别精度。在该基础上，本文随机加入小范围的旋转，进一步增强训练样本的广泛性，如图3所示。

图 2　马赛克数据增强图像

Fig. 2　Mosaic data enhancement images

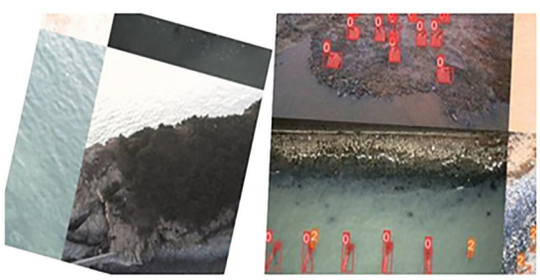

图 3　改进的马赛克数据增强图像

Fig. 3　Improved mosaic data enhancement images

1.2.2　CBAM 模块

CBAM 是一个有效的、轻量的卷积注意力模块，可以插入到大多数主流的 CNN 框架中，而且可以以端到端的形式进行训练。给定一个中间特征映射，该模块沿着通道和空间域这 2 个不同的维度按顺序输入 attention 映射，然后将 attention 映射乘以输入特征映射以进行自适应特征细化，可以得到自适应的精炼特征。通过将 CBAM 融入到不同的模型中，替换 YOLO5 中的部分卷积模块和 CSP 模块，可以使模型的性能得到提升。使用 CBAM 可以帮助 YOLO5 提取受关注区域，减少易混淆信息并关注有用的目标信息。

1.2.3　移除大目标检测特征图

轨条砦图像输入尺度为 768 像素×768 像素，PANet 输出为 96 像素×96 像素、48 像素×48 像素、24 像素×24 像素的 3 个分辨率特征图，检测端在这 3 个尺度上识别毁伤效果。其中，小目标的毁伤检测主要由分辨率较大的特征图负责，分辨率较小的特征图主要负责大目标的毁伤检测。在轨条砦图像中，轨条砦所占的像素数通常较少，大多属于中小目标，YOLO5 模型中负责大目标检测的特征图在此应用场景下往往效果不佳，甚至还会对中小目标的检测产生一定的干扰。

针对轨条砦数据集的特点，本文对原模型检测端的结构进行优化，仅采用检测中、小目标的特征图，移除了 24 像素×24 像素大目标特征图。此外，大目标检测图的移除还能够减少不必要的计算量，提高模型检测速度。

2　实验

2.1　总体流程

实验总体流程如图 4 所示，分为训练步骤、推理识别步骤。

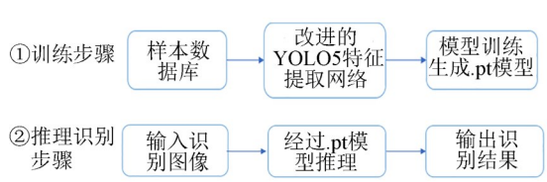

图 4　实验总体流程

Fig. 4　Overall process of the experiment

训练步骤：包括样本数据库、改进的 YOLO5 特征提取网络、模型训练生成 .pt 模型。其中，样本数据库用于对原始数据进行预处理后得到样本数据，同时通过数据增强方式增加样本多样性。改进的 YOLO5 特征提取网络通过深度神经网络，从样本数据中提取特征，做目标检测与分割任务。

模型训练通过梯度下降方法，在训练集上训练模型，生成.pt模型。

推理识别步骤：输入识别数据，经过.pt模型推理识别，输出识别结果。

2.2　训练步骤

2.2.1　样本数据库

原始数据采集通过无人机等载有高清相机的设备，获取原始图像数据，并将原始图像数据传输至样本数据库。

通过抽帧、筛选等方式对原始数据进行预处理，得到样本数据，同时通过数据增强方式，例如图像旋转、平移等增广方式、Cutout、Mixup等组合增强方式，提高样本数据集的多样性。

本方案中，使用某些实际场景的轨条砦数据集。轨条砦的毁伤程度从登陆破障行动是否形成有效通路的视角分为3个等级（类别），分别是无损伤（类别0）、一般损伤（类别1）、重度损伤（类别2），如图5所示。处于破障弹药威力范围外完好的轨条砦定义为无损伤，仍影响通路形成的轨条砦残障定义为一般损伤，不再影响通路形成的轨条砦残障定义为重度损伤。

本文在该轨条砦数据集上进行模型训练与测试实际场景的轨条砦数据集图像文件的数量量级为10^3。每一幅图像文件包含多个类别，标注后的样本数据统计如图6所示。

2.2.2　模型训练

本文实验配置如下：计算机操作系统为Ubuntu18.04、GPU采用NVIDIA RTX2070 SUPER，深度学习框架采用由Ultralytics维护的YOLO5。训练中，输入图像的大小为768像素×

（a）无损伤轨条砦（类别0）

（b）一般损伤轨条砦（类别1）

（c）重度损伤轨条砦（类别2）

图5　轨条砦毁伤等级划分

Fig. 5　Rail obstacles damage level classification

768像素，epoch为50，batch_size为20，初始学习率为0.01，动量因子为0.937，权重衰减系数为0.000 5，且在数据预处理阶段采用马赛克数据增强、自适应锚框计算等操作。

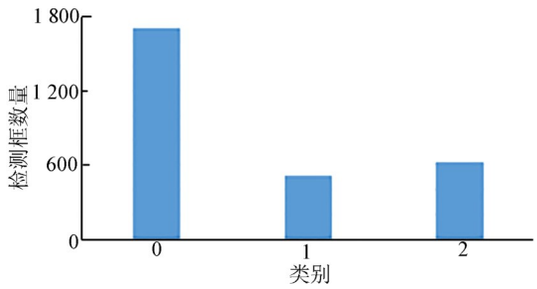

图 6　标注后的样本数据

Fig. 6　Marked sample data

通过梯度下降方法，在训练集上训练模型，生成.pt模型。该模型也可用于其他图像的推理识别预测。

选取70%的样本图像作为训练集，30%作为测试集。训练集经过标注后，可以得出目标的尺度分布，可知目标的大小在一个较大的尺度范围内稠密分布，因此，此任务对模型跨尺度的检测能力提出了更高的要求。采用损失函数box_loss、seg_loss、obj_loss、cls_loss以及度量精度的指标准确率、召回率、mAP_0.5、mAP_0.5∶0.95来表征模型的性能，结果如图7所示。其中：mAP_0.5是指交并比阈值为0.5时的平均准确度（mean Average Precision，mAP）；mAP_0.5∶0.95是指交并比阈值从50%到95%间隔5%时10个mAP值的平均值。

可以看出，模型训练过程中，在训练集上，loss曲线随着迭代次数的增加而迅速下降，准确率、召回率、mAP_0.5、mAP_0.5∶0.95随着迭代次数增加而迅速上升，最后达到良好的训练结果。

测试了原始YOLO5模型基线方法的训练精度，结果为：mAP_0.5为0.88，mAP_0.5∶0.95为0.73。而优化策略的mAP_0.5、mAP_0.5∶0.95分别为0.95和0.79，比原始YOLO5模型提高了8.0%、8.2%。

2.3　推理识别步骤及可视化结果

推理识别步骤的耗时较短，单张图像的耗时一般在80 ms以内。

在测试集上，3个类别的目标数量的真值为746个。推理识别结果为：正确识别的目标数量为731个，错漏识别的目标数量为15个，识别精度为98%。分析错漏识别的图像，发现其色域比较昏暗，背景与目标的区分难度较大，因此，导致了部分目标的漏检和错检。测试集上部分结果如图8所示。

可以看出，模型能够较好地适应复杂环境以及小目标检测任务，将目标进行精确矩形框定位与多边形框分割，可得到可靠的分类。

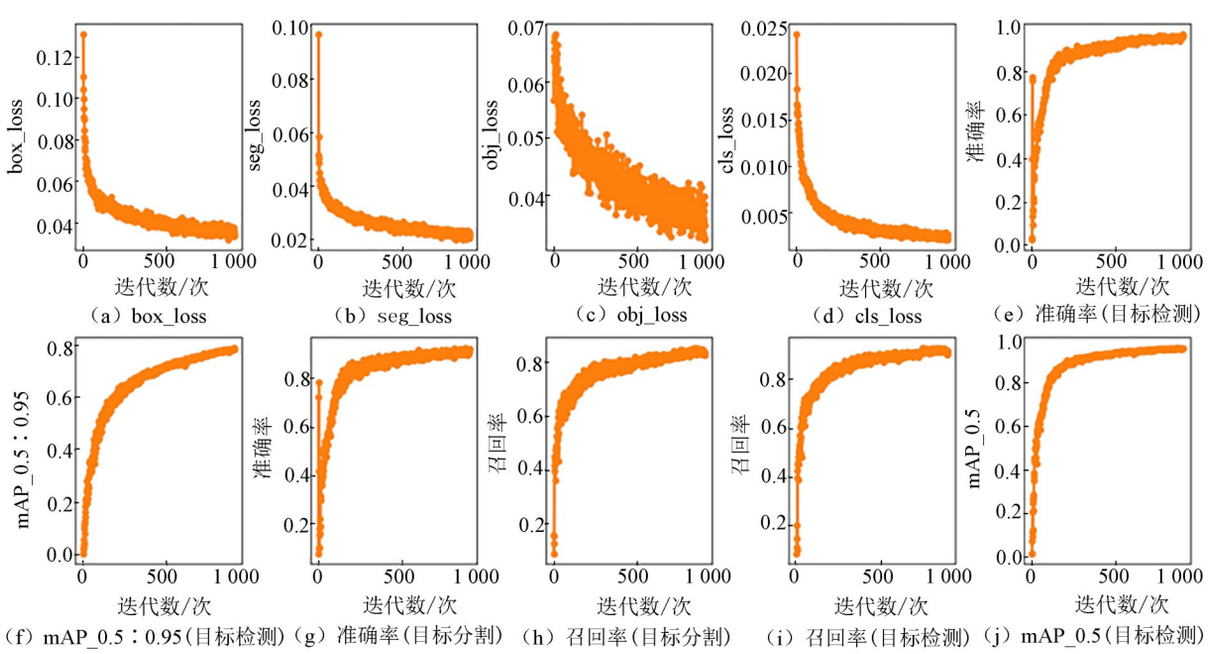

图 7　模型训练结果

Fig. 7　Training results of the model

图 8 测试集上部分识别结果

Fig. 8 Partial recognition results on the test set

3 结论

利用优化YOLO5深度神经网络，建立了一种轨条砦障碍毁伤等级识别方法，试验结果表明，在测试集上测试的识别精度高。分析漏检的图像发现，其色域比较昏暗，背景与目标的区分难度较大，导致了部分目标的漏检。因此，在复杂背景环境的情况下，本文算法还有一定的提高空间。

参考文献：

[1] LI W T，CHEN Y J，HU K X，et al. Oriented RepPoints for aerial object detection[C]//2022 IEEE / CVF Conference on Computer Vision and Pattern Recognition（CVPR）. Piscataway，NJ，USA：IEEE，2022：1829-1838.

[2] LI W Y，LIU X Y，YUAN Y X. Sigma：Semantic-complete graph matching for domain adaptive object detection[C]//2022 IEEE / CVF Conference on Computer Vision and Pattern Recognition（CVPR）. Piscataway，NJ，USA：IEEE，2022：5291-5300.

[3] ZHENG Z H，YE R G，WANG P，et al. Localization distillation for dense object detection[C]//2022 IEEE / CVF Conference on Computer Vision and Pattern Recognition（CVPR）. Piscataway，NJ，USA：IEEE，2022：9407-9416.

[4] LANG C B，CHENG G，TU B F，et al. Learning what not to segment：A new perspective on few-shot segmentation[C]//2022 IEEE / CVF conference on computer vision and pattern recognition（CVPR）. Piscataway，NJ，USA：IEEE，2022：8057-8067.

[5] ATIGH M G，SCHOEP J，ACAR E，et al.

Hyperbolic image segmentation[C]//2022 IEEE / CVF Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway, NJ, USA: IEEE, 2022: 4453-4462.

[6] ZHOU T F, ZHANG M J, ZHAO F, et al. Regional semantic contrast and aggregation for weakly supervised semantic segmentation[C]//2022 IEEE / CVF Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway, NJ, USA: IEEE, 2022: 4299-4309.

[7] HINTON G E, OSINDERO S, TEH Y W. A fast learning algorithm for deep belief nets[J]. Neural Computation, 2006, 18(7): 1527-1554.

[8] PUJOL P, POL S, NADEU C, et al. Comparison and combination of features in a hybrid HMM/MLP and a HMM/GMM speech recognition system [J]. IEEE Transactions on Speech and Audio Processing, 2005, 13(1): 14-22.

[9] DENG L, LI J Y, HUANG J T, et al. Recent advances in deep learning for speech research at Microsoft[C]//2013 IEEE International Conference on Acoustics, Speech and Signal Processing. Piscataway, NJ, USA: IEEE, 2013: 8604-8608.

[10] HORI T, HORI C, MINAMI Y, et al. Efficient WFST-based one-pass decoding with on-the-fly hypothesis rescoring in extremely large vocabulary continuous speech recognition[J]. IEEE Transactions on Audio, Speech, and Language Processing, 2007, 15(4): 1352-1365.

[11] MOHAMED A R, SAINATH T N, DAHL G, et al. Deep belief networks using discriminative features for phone recognition[C]//2011 IEEE International Conference on Acoustics, Speech And Signal Processing (ICASSP). Piscataway, NJ, USA: IEEE, 2011: 5060-5063.

[12] 李宝奇，任露露，陈发，等. 基于前视三维声呐的轨条砦识别方法[J]. 水下无人系统学报，2022，30(6)：747-753.
LI B Q, REN L L, CHEN F, et al. A method of erect rail barricade recognition based on forward-looking 3D sonar[J]. Journal of Unmanned Undersea Systems, 2022, 30(6): 747-753.