引文格式：余添添,吴松,唐芝青,等. 利用铁塔视频图像和改进 YOLOv5 的违规施工监测[J]. 地理空间信息, 2024, 22(4): 45-48. doi:10. 3969/j.issn. 1672-4623.2024.04.011

利用铁塔视频图像和改进 YOLOv5 的违规施工监测

余添添[1]，吴　松[1]，唐芝青[2]，屈伟军[2]，肖德勋[1]，李　琪[1]
（1. 北京吉威数源信息技术有限公司，北京 100000；2. 湖南省第二测绘院，湖南 长沙 410000）

摘　要：针对铁塔摄像头视频中挖掘机等工程机械目标极为稀疏、小目标居多、背景复杂、目标遮挡等问题，提出一种改进的 YOLOv5s 算法，首先将数据增强方式改为 mosaic-9，有效增加了小目标样本量；其次，优化了特征提取与特征融合过程，增加了一个检测尺度，提高了小目标的检测精度；最后，引入了 SKNet 注意力模块，使模型更加关注感兴趣区域，抑制无用信息和复杂背景的干扰，提高算法的检测准确率。实验结果表明，改进后算法 *AP* 值达到了 79.63%，相对 YOLOv5s 的 *AP* 值提升了 6.8%，同时小目标检测精度更高，漏检情况更少。

关键词：工程机械检测；YOLOv5s；小目标检测；注意力机制
中图分类号：P258　　**文献标志码：**B　　**文章编号：**1672-4623（2024）04-0045-04

Illegal Construction Monitoring Using Tower-based Video Image and Improved YOLOv5

YU Tiantian[1], WU Song[1], TANG Zhiqing[2], QU Weijun[2], XIAO Dexun[1], LI Qi[1]
(1. Beijing GEOWAY Information Technology Co., Ltd., Beijing 100000, China;
2. Hunan Second Institute of Surveying and Mapping, Changsha 410000, China)

Abstract: The construction equipment targets such as excavators in the tower-based video are extremely sparse, mostly small and with complex backgrounds, and part of them is occluded. So, we proposed an improved detection algorithm based on YOLOv5s. Firstly, we changed the data augmentation mode to Mosain-9, to increase the number of small target. Then, we optimized the process of feature extraction and feature fusion, and added a detection scale to improve the detection ability of small targets and feature fusion. Finally, we introduced an attention mechanism to pay more attention to the region of interest, suppress useless information and background noise, and improve the accuracy of target detection. The experimental results show that AP value of the improved algorithm reaches 79.63%, which is 6.8% higher than that of YOLOv5s. At the same time, the detection accuracy of small targets is higher, and the missing target is less.

Key words: construction equipment detection, YOLOv5s, small target, attention mechanism

铁塔视频是一种近地、实时的对地监测手段，监测频率高、监测粒度细，不易受阴雨天气影响，可实现问题的“早发现、早制止”。以挖掘机、推土机、打桩机等建设施工设备为目标，基于深度学习目标检测算法可及时发现疑似违规/违法施工行为，实现对自然资源高频率、精粒度的监测[1]。铁塔视频数据具有体量大、价值密度低、背景繁多复杂等特点，需配以快速高效的算法模型实现数据的价值“提纯”，因此研究出目标检测速度快、准度高的网络模型是实现铁塔视频违规施工监测的必要前提。YOLO 模型自 2015 年提取以来已历经多次更新改造，其中 YOLOv5 推理速度快、检测精度高，是目前较为广泛使用的目标检测网络，适用于实时视频目标检测[2-4]。

由于铁塔摄像头焦距缩放范围大，挖掘机、推土机、打桩机等建设施工所用到的工程机械设备目标在实际视频中呈现出极大的尺度差异，小目标居多，且背景复杂。而随着网络加深、感受野缓慢增长，容易被大目标主导，小目标提取的特征少，容易被忽略[5-6]，因此本文提出一种改进的 YOLOv5s 算法，提高小目标检测率，以实现对违规/违法施工行为快速、准确的识别。

1　YOLOv5s 算法

YOLOv5s 是 YOLOv5 的轻量化版本，更适合基于视频的目标实时检测。YOLOv5s 模型网络结构由 Input 部分、Backbone 部分、Neck 部分和 Head 部分组成[7]，如图 1 所示。

1.1　Input 部分

Input 部分主要包含数据增强、自适应锚框、自适应图片缩放等。增强方式包括缩放，色彩空间调整和

收稿日期：2023-01-30。
第一作者简介：余添添（1991—），硕士，工程师，主要研究方向为遥感技术与应用、深度学习遥感解译，E-mail: 13618650845_1@163.com。

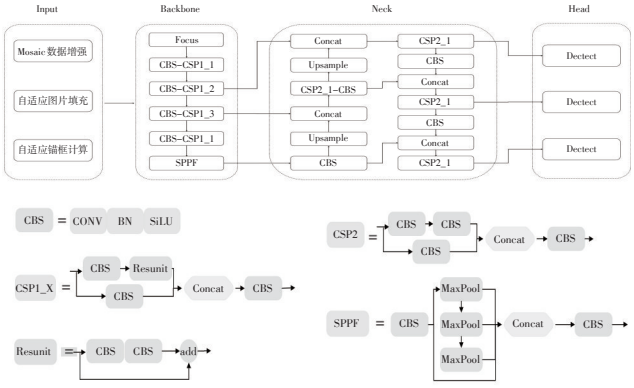


图1 YOLOv5s结构图

Mosaic，其中Mosaic增强是通过将4张图像进行随机缩放、随机裁剪、随机分布方式进行拼接，可使小目标的检测效果得到提升[8]。自适应锚框计算是指YOLOv5会自动按照新的数据集的labels自动学习anchors的尺寸，采用k均值和遗传学习算法对自定义数据集进行分析，获得适合自定义数据集中对象边界框预测的预设锚定框。自适应图片缩放是指YOLOv5通过计算图片最适宜的尺寸，使缺边补充范围达到最小，很好地保留了图片的特征[9]。

1.2 Backbone部分

Backbone部分包括focus结构和CSP结构。其中Focus结构使用切片并concat的方式替换池化操作，完成下采样的同时不损失特征。由于减少了卷积层，所以可以减少计算量，同时不损失精度[10]。

YOLOv5也使用CSPDarknet作为BackBone主干网络，其设计了2种CSP结构，CSP1_X应用于Backbone主干网络，另一种CSP2_X结构则应用于Neck中。CSP结构解决了其他大型卷积网络结构中的重复梯度问题，减少模型参数和检测速度，既保证了推理速度和准确率，又减小了模型尺寸[11]。在Backbone最后采用了SPPF将不同感受也的特征图融合，提高特征图的表达能力，同时速度快于SPPNet[12]。

1.3 Neck部分

Neck部分YOLOv5采用了与YOLOv4一样的FPN+PAN结构，不同点在于YOLOv4的Neck结构中采用的都是普通的卷积操作，而YOLOv5的Neck结构中采用CSP2_X结构，加强网络特征融合的能力[13]。

1.4 Head部分

Head部分算法采用GIOU_Loss作为Bounding box的损失函数，GIOU_Loss损失函数在重叠的目标检测中效果较好，同时YOLOv5算法合理利用freebie和specials进行调优处理，平均精度和检测速度都有一定的提升，可以获得最优目标。

2 改进YOLOv5算法

2.1 Mosaic-9数据增强

一个成功的、可工程化应用的深度学习网络大都需要基于大量的样本数据训练，而实际中样本本身数量少，加之在海量视频中寻找包含目标的数据难度大，训练样本十分缺乏，因此需要使用数据增强方法对样本数据进行增广扩充，最大程度上提高模型的泛化能力。

Mosaic图像增强是在现有样本数据中随机选择4个图像经过裁剪拼接形成新的样本数据，可以丰富样本多样性，增强小目标检测，降低模型对于多样性学习的难度。由于本文选用了挖掘机、推土机、打桩机等目标，小目标多、类别不均衡，因此使用Mosaic-9进行数据增强，具体方法如图2所示，将随机选取的9个图像进行随机裁剪并拼接，形成新的样本数据，其优势在于进一步增加了数据量，同时增加了小目标数量，在归一化处理时，会计算9张图片，不依赖处理参数，减轻了计算量。



图2 Mosaic-9数据增强

2.2 优化特征提取

随着网络加深，网络所提取的特征语音信息丰富，但位置精度会随之降低，尤其是小目标，其影响更为显著，因此文本在CSP1_1部分增加一个输出特征图，接入NECK部分。输入数据大小为640×640时，此部分输出的特征图大小为160×160，每个特征对应的感受野为4×4，可更好地检测到小目标，也可在特征融合过程中提供更多信息。

2.3 引入注意力机制

SKNet是通过一种非线性的多核信息聚合方法实现神经元的自适应感受野大小的算法，是一个基于通道域注意力的轻量级嵌入式模块。SKNet模块可以对特征通道进行重要性加权，使最终得到的特征图更加关注对检测识别等任务有用的特征通道，并且该模块会自动选择调整最优的卷积算子，因此在特征提取后嵌入SKNet，可对特征通道进行重要性加权，再送入目标检测层，在理论上可以改善检测识别的性能[14]。

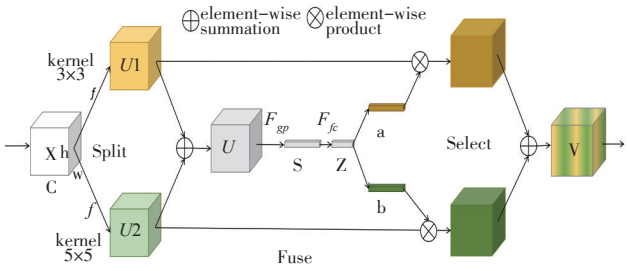


图 3　SKNet 网络结构

如图 3 所示，SKNet 由 Split、Fuse 及 Select 部分组成。Split 是对原特征图经过不同大小的卷积核进行卷积得到 $\tilde{U}$和$\hat{U}$，为了进一步提高效率，将 5O5 核替换为 3O3 核、步距为 2 的膨胀卷积[15]。Fuse 部分基本思想是使用门来控制信息从携带不同信息尺度的多个分支流入下一层的神经元，使得神经元根据刺激信号自适应调整感受野大小。为了实现这个目的，需要整合来自所有分支信息，将两部分的特征求和 $U=\tilde{U}+\hat{U}$。F_{gp} 为全局平均池化操作，F_{fc} 为先降维再升维的 2 层全连接层，输出的 2 个矩阵 $\boldsymbol{a}$ 和 $\boldsymbol{b}$，其中矩阵 $\boldsymbol{b}$ 为冗余矩阵，在如图 2 个分支的情况下 $\boldsymbol{b}=1-\boldsymbol{a}$。Select 是使用跨通道的软注意自适应地选择不同的信息空间尺度。按照信道的方向使用 Softmax 函数计算每个卷积核的权重。使用 $\boldsymbol{a}$ 和 $\boldsymbol{b}$ 两个权重矩阵对 $\tilde{U}$和$\hat{U}$ 进行加权操作，再进行相加得到输出 V[15-16]。经过以上 3 部分改造，得到改进的 YOLOv5s 网络结构，如图 4 所示。

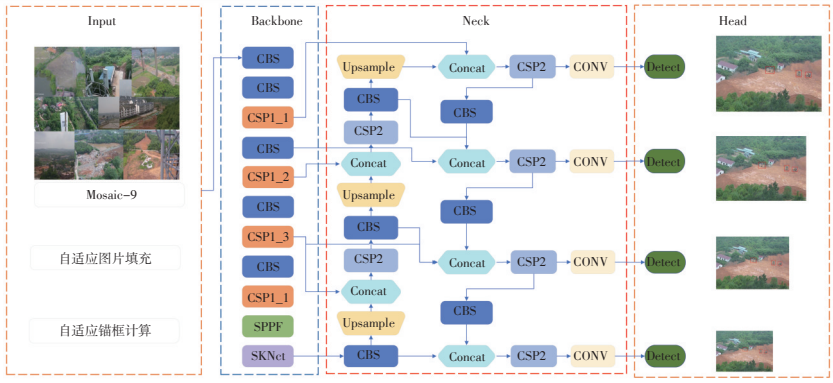


图 4　改进 yolov5s 网络结构图

3　实　验

3.1　实验数据集

实验所用训练样本数据为使用湖南省第二测绘院所建立的铁塔检测系统内 1 779 个摄像头的视频帧数据，共计 5 628 张图片，标注后包含目标 7 873 个，并按照 7∶2∶1 的比例划分训练集和测试集和验证集。

3.2　实验配置与精度评价

实验采用的硬件设备为 Intel Core i7-9700k@3.6 GHz，显卡为 NVIDIA GeoForce RTX 2080 Ti，显存为 11 G，配以 CUDA10.0 及对应的 cudnn 以提高计算效率。模型训练批量大小为 16，epoch 为 100，学习率为 0.001，采用余弦退火策略降低学习率，衰减系数设置为 0.000 5。

目标检测常用的评价指标包括平均精度 AP、均值平均精度 mAP、准确率（Precision，P）和召回率（Recall，R），表达式分别为式（1）、（2）、（3）、（4）。

$$P=\frac{TP}{TP+FP} \tag{1}$$

$$R=\frac{TP}{TP+FN} \tag{2}$$

$$AP=\int_0^1 P(R)\mathrm{d}R \tag{3}$$

$$\mathrm{m}AP=\frac{1}{n}\sum_{i=1}^{n}AP_i \tag{4}$$

式中，TP 为被正确检测的正样本；FP 为被错误检测的正样本；FN 为被检测为目标的负样本；n 为目标类别数，本文所检测目标只有工程机械一类，因此 AP=mAP。

3.3　实验结果分析

为了验证改进算法的有效性，本文对比了 YOLOv5s 改进前后的检测效果，改进后 AP 值达到了 79.63%，而 YOLOv5s 的 AP 值为 74.59%，提升了 6.8%。为了进一步证明改进算法的有效性，对比了使用 YOLOv5s 改进前后算法的检测结果，如图 5~6 所示。其中左侧为改进前算法检测效果，右侧为算法改进后检测效果，改进前算法对于背景复杂、隐蔽的小目标、被遮挡的目标等存在漏检情况，改进后的算法弥补了小目标提取效果不好这一缺点的同时，增加了算法的鲁棒性和抗干扰能力。

为了更直观地展示改进方法的优越性，本文进行了简单的消融实验，实验结果如表 1 所示。由表中数据可知，各组改进均可发挥一定提升作用，其中使用 Mosaic-9 数据增强后，AP 提升了 0.6%，验证了 Mosaic-9 数据增强可提升模型目标检测精度和模型的泛化能



改进前 改进后



改进前 改进后

图5 背景复杂的隐蔽目标



改进前 改进后

图6 目标被遮挡

力；进一步在YOLOv5s中引入4倍下采样，优化特征提取与融合部分后，*AP*提升了3.4%，提升较为显著，尤其是小目标部分，优化较为明显。在此基础上，引入SKNet注意力模块，*AP*提升了6.7%，注意力模块对无用信息进行抑制，使模型更加关注目标，进一步优化了模型。

表1 消融实验结果

改进名称	Mosaic-9优化	特征提取优化	注意力机制	AP@0.5/%	单张检测速度/ms
YOLOv5s	×	×	×	74.59	14
改进一	√	×	×	75.07	14
改进二	√	√	×	77.12	16

4 结 论

选用挖掘机、推土机、打桩机等工程机械设备为目标，对违规/违法占用/破坏自然资源现象进行监测。针对目前监测算法中对小目标漏检情况，提出一种改进的YOLOv5s算法。首先采用mosaic-9对数据增强方式进行了改进，增加小目标样本数量，提高了小目标特征信息量；改进网络结构，增加了一个检测尺度，优化了特征提取与特征融合过程，提高了小目标的检测精度。另外在主干网络末尾处增加了注意力模块，通过注意力机制，使模型更加关注感兴趣区域，抑制无用信息和复杂背景的干扰，提高算法的检测准确率。经实验验证，改进后的YOLOv5s算法在没有增加过多的参数和体量的情况下，保证检测效率不降低的同时，对工程机械目标的检测精度有了一定的提高，可满足工程化生产使用。今后可进一步扩充样本数据集，并对模型的网络结构做进一步的调整，以达到更优的检测效果。

参考文献

[1] 冯笑雨. 基于塔基监控图像的建设施工用地识别与空间定位方法研究[D]. 南京:南京师范大学,2019

[2] 王文杰,何小海,卿粼波,等. 改进YOLOv5的船舶检测算法及嵌入式实现[J]. 无线电工程,2022,52(12):2 116-2 123

[3] 翟宏亮. 基于轻型卷积神经网络的无人机多目标检测系统研究[J]. 地理空间信息,2022,20(12):81-83

[4] 罗茜,赵睿,庄慧珊,等. YOLOv5与Deep-SORT联合优化的无人机多目标跟踪算法[J]. 信号处理,2022,38(12):2 628-2 638

[5] 王一旭,肖小玲,王鹏飞,等. 改进YOLOv5s的小目标烟雾火焰检测算法[J]. 计算机工程与应用,2023,59(1):72-81

[6] 程顺生,覃驭楚,吕炎杰. 城市环境下基于双目视觉的移动目标检测[J]. 地理空间信息,2022,20(3):7-11

[7] YoloV5源码部分注释解读(ultralytics版本)(train.py)[EB/OL]. (2023-06-19) [2023-07-23]. https://github. com/ultralytics/yolov5

[8] 涂海清. 基于yolov5s-se和数据增强的夜间车辆目标检测[D]. 广州:华南理工大学,2021

[9] 陈禹蒲,马晓川,李璇. 基于YOLOv3锚框优化的侧扫声呐图像目标检测[J]. 信号处理,2022,38(11):2 359-2 371

[10] YOLOv5解读[EB/OL]. (2021-12-09)[2023-10-20]. https://blog. csdn. net/zjl892209143/article/details/121833051

[11] Wang C Y, Liao H Y M, Wu Y H, et al. CSPNet:A New Backbone that Can Enhance Learning Capability of CNN[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops, 2020:390-391

[12] He K, Zhang X, Ren S, 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

[13] 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,2020:10 781-10 790

[14] 王战涛,张策,王晓田. 基于YOLOV3的改进目标检测识别算法[J]. 上海航天(中英文),2021,38(6):60-70

[15] Hu J, Shen L, Sun G. Squeeze-and-excitation Networks[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Washington D. C. USA:IEEE Press,2018(7):132-141

[16] 赵梦,于红,李海清,等. 融合SKNet与YOLOv5深度学习的养殖鱼群检测[J]. 大连海洋大学学报,2022,37(2):312-319