doi:10.3969/j.issn.1001-4616.2025.04.008

基于注意力机制与可变卷积神经网络的卫星视频运动目标检测

马洲俊[1],陈锦铭[2],刘浩林[3],张　卡[3]

(1.国网江苏省电力有限公司,江苏 南京 210019)
(2.国网江苏省电力有限公司电力科学研究院,江苏 南京 211103)
(3.南京师范大学地理科学学院,江苏 南京 210023)

[**摘要**] 视频卫星能获得高空间分辨率的视频信息,为运动目标的检测和分析提供有效数据支撑. 然而,由于卫星视频图像中目标像素比例低、纹理细节不清晰、背景复杂等缺点,从卫星视频中检测运动目标存在很大困难. 为此,本文以 YOLOv8 为骨干网络,提出了一种基于注意力机制与可变卷积神经网络的卫星视频运动目标检测算法. 首先,设计 C2f-DCN 模块替换原模型骨干网络中的 C2f 模块,以提高模型对不同尺度目标的特征提取能力. 其次,在检测头前添加 Shuffle Attention 轻量级注意力机制,在保证模型计算速度的前提下增强重要特征,加强通道间信息沟通提高模型特征融合能力. 最后,为了提高模型的学习能力和推理效率,采用 Inner-CIoU 损失函数,并引入辅助边界框概念来解决卫星视频图像中目标像素比例小的问题. 利用 SAT-MTB 卫星视频影像数据集进行对比实验,实验结果表明本文算法的精确度、召回率、mAP50:95 和 $F1$ 分数分别为 75.3%、62.8%、34.9%和 68.48,相较于原始 YOLOv8n 网络,上述指标分别提高了 11.6%、4.2%、3.0%和 7.44,验证了本文方法的有效性和优越性.

[**关键词**] 卫星视频,YOLOv8,轻量级注意力机制,可变形卷积,辅助边框回归
[**中图分类号**]P237;TP751　[**文献标志码**]A　[**文章编号**]1001-4616(2025)04-0078-09

Satellite Video Moving Object Detection Based on Deformable Convolutional Neural Network and Shuffle Attention

Ma Zhoujun[1],Chen Jinming[2],Liu Haolin[3],Zhang Ka[3]

(1.State Grid Jiangsu Electric Power Co.,Ltd.,Nanjing 210019,China)
(2.Electric Power Research Institute,State Grid Jiangsu Electric Power Co.,Ltd.,Nanjing 211103,China)
(3.School of Geography,Nanjing Normal University,Nanjing 210023,China)

Abstract:Video satellites can obtain high spatial resolution video information,providing effective data support for the detection and analysis of moving targets. However,due to the disadvantages of low target pixel proportion,unclear texture details,and complex background in satellite video images,there are significant difficulties in detecting moving targets from satellite videos. Thus,on the basis of the backbone network of YOLOv8,this paper proposes a new detection method of motion targets in satellite video based on deformable convolutional neural network and Shuffle Attention. Firstly,a C2f-DCN module is designed to replace the C2f module in the original model backbone network for improving the model's ability of extracting features from targets with different types and different scales. Secondly,a lightweight Shuffle Attention mechanism is added in front of the detection head to strengthen important features while ensuring the computational speed of the model,enhancing information communication between channels,and improving the model's feature fusion ability. Finally,to improve the learning ability and inference efficiency of the model,the Inner-CIoU loss function is adopted,and the concept of auxiliary bounding boxes is introduced for solving the problem of small proportion of target pixels in satellite video images. Comparative experiments are conducted using the SAT-MTB satellite video image dataset,and the experimental results show that the accuracy,recall,mAP50:95,and $F1$ scores of the algorithm are 75.3%,62.8%,34.9%,and 68.48,respectively. Compared with the original YOLOv8n network,above indexes are improved by 11.6%,4.2%,3.0%,and 7.44.

收稿日期:2024-10-11.
基金项目:国网江苏省电力有限公司科技项目(J2023121).
通讯作者:张卡,博士,教授,研究方向:摄影测量与遥感. E-mail:zhangka81@126.com

Thus,the effectiveness and superiority of the proposed method is verified.

Key words:satellite video,YOLOv8,lightweight attention mechanism,deformable convolution,auxiliary border regression

近年来随着多种对地观测技术的迅速发展,以高时间分辨率、高光谱分辨率和高空间分辨率为特点的遥感新时代已经到来. 在各国高分卫星建设规划的推动下,航天遥感技术取得了巨大进步,为卫星视频的拍摄提供了可能. 卫星视频与固定摄像机和无人机飞行器平台所获得的视频数据不同,其可以从鸟瞰角度持续关注特定区域,获得地面目标的连续观测视频,并具有更大的视场和更丰富的信息,为运动目标检测提供了独特的优势. 视频卫星在各个领域都发挥着重要作用,从地面资源的动态观测到自然灾害的全过程监测,再到社会经济参数的动态反演和军事行动的保障,卫星视频已成为大数据基建的关键组成部分,并为国家安全提供了重要保障[1].

在基于视频的运动目标检测方面,所采用的研究方法主要分为以下四类:帧间差分法、背景差分法、光流法和基于深度学习的方法. 帧间差分法[2]基于视频时域变化差异检测运动目标,具有思路简单、计算量小的特点. 在此基础上,研究者们为应对环境噪声敏感的问题提出了三帧间差法、多帧差法等改进算法[3-4]. 背景差分法首先建立背景模型,通过比较原始像素与背景模型进行前景分割与运动目标检测,并根据检测结果动态更新背景模型,其具有较好的背景变化适应能力,非常适用于静止摄像机拍摄的地面视频数据,但在复杂背景环境中的背景更新策略仍需要深入研究[5-6]. 光流法将物体运动时像素灰度的瞬时变化率的集合作为光流场,根据光流场的运动矢量来区分运动前景和背景[7]. 受限于光流方程的较为复杂解算过程导致其难以满足实时目标检测的应用需求. 基于深度学习的运动目标检测方法则可归类为二阶段(基于候选区)方法和一阶段(基于回归)方法. 二阶段目标检测方法首先对图像通过选择性搜索算法或区域候选网络等提取候选区域,并基于得到的候选区域进行分类和回归获得目标的类别与位置,这类方法的主要代表是 R-CNN 系列,如 Faster R-CNN、Mask R-CNN 等[8]. 一阶段方法则通过建立模型直接获得目标的类别与位置,代表模型是 SSD 系列与 YOLO 系列[9],这类方法与二阶段目标检测方法相比精度较低但计算速度较快,适用于实时检测任务,因此在本研究中选择一阶段方法利用其在处理高时间分辨数据时的优势以保证检测效率和实时性.

地面普通视频和无人机航空视频的运动目标检测取得的显著进展,为卫星视频目标检测提供了有价值的经验[10]. 卫星视频运动目标检测方法大多是在传统算法的基础上,融入目标运动区域提取建议、针对检测目标的多模态特征分析、目标跟踪等策略,以进一步提高运动目标的精度. 目前,卫星视频运动目标检测主要采取以下四种策略:引入道路约束、改进传统检测算法、顾及卫星视频的时空信息及针对现有深度学习模型进行改进. 其中,引入道路约束检测方法的核心原理是利用深度学习模型提取道路信息,抑制背景噪声并采用传统运动目标检测方法进行目标检测[11]. 虽然该方法在一定程度上降低了传统背景差分法对云层以及视频卫星相对运动所造成的伪运动目标的误检,但对所用到的道路分割模型精度要求较高,且会遗留位于道路的伪运动目标. 改进传统检测算法的方法首先使用背景差分方法对卫星视频影像中的运动目标进行初步检测,然后对检测结果使用神经网络进行进一步约束,以提升检测的精确率[12]. 尽管这种策略模型相对简单,但在很大程度上依赖于传统运动目标检测方法的结果,并且这种方法虽然能够使模型在精确率上得到提升,但确无法保证模型召回率. 顾及卫星视频时空信息的检测方法利用单帧信息与相邻多帧图像信息相结合进行目标检测[13],然而该策略的模型结构较为复杂,无法满足检测任务的效率要求. 改进深度学习模型的策略通过对现有模型的改进,结合注意力机制等操作提升模型在卫星视频运动目标检测任务上的表现[14].

基于卫星视频目标检测仍面临着一系列挑战,首先,卫星视频的场景变化在像面的表现为包含平移、旋转、仿射、投影等的混合变换,难以进行准确的变化补偿. 其次,视频中的场景变换具有局部非一致性,易造成虚警误检测. 此外卫星视频中运动目标的尺寸小,且存在大量的非刚体目标,同时目标与背景及邻域的对比度较低,使得前、背景分割和提取困难,基于邻域像素计算或背景建模的传统判别方法容易造成漏检,严重影响检测精度. 针对现有问题,本文基于 YOLOv8 目标检测模型,提出了一种基于注意力机制与可变形卷积神经网络的卫星视频运动目标检测方法,并采用 SAT-MTB 公开卫星视频数据集[15]进行对比试验,以验证本文方法的准确性和可靠性.

1　方法及原理

1.1　整体框架方法总体思想

YOLOv8(You Only Look Once version 8)由 Joseph Redmon 和 Alexey Bochkovskiy 等研究者于 2020 年提出[16],该模型主要由主干特征提取模块(Backbone)、特征加强模块(Neck)、检测模块(Detect)3 个部分构成. 在 YOLO 系列的基础上,YOLOv8 模型采用梯度流更加丰富的 C2f 模块代替 C3 模块,并基于调整通道数策略降低模型参数量,可以实现在一次前向传播中完成对图像中所有目标快速而准确地检测和定位.

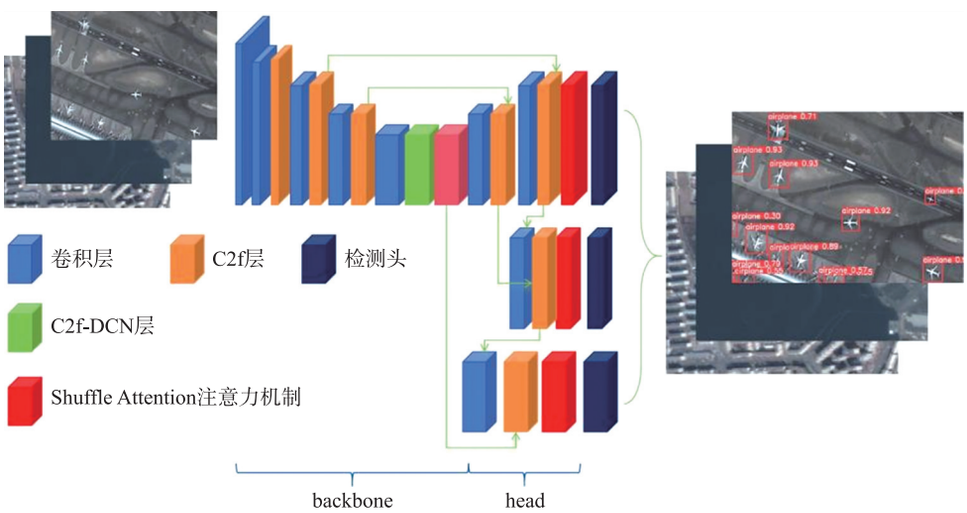


图 1　本文方法的总体技术流程图

Fig. 1　Overall technical flowchat of the proposed method

在卫星视频影像的目标检测任务中,存在着目标所占像素少,受云层光照等背景影响大,同时还存在一部分非刚体目标的特点,现有的 YOLOv8 模型不能满足卫星视频运动目标检测的任务需求. 针对任务特点,本文基于 YOLOv8 目标检测模型,从融合注意力机制、加入可变形卷积模块、更换损失函数等方面对其进行优化,并采用 SAT-MTB 公开卫星视频数据集进行对比验证. 首先,使用可变形卷积替换 C2f 模块中的传统卷积模块,基于其重新设计了 Bottleneck 结构,最终集成为 C2f-DCN 模块用于替换 Backbone 中的 C2f 模块. 其次,在原模型每个检测头前添加一个轻量级高效注意力机制 Shuffle Attention,增强模型的特征提取能力. 最后,引入 Inner-IoU 思想设计 Inner-CIoU 损失函数,增强模型对小目标的检测性能. 本文方法的总体技术流程如图 1 所示.

1.2　C2f-DCN 模块

在针对卫星视频影像的目标检测任务中不仅存在目标所占像素少、分辨率低的问题,还存在有火车等非刚体目标的检测难点. 传统的卷积单元对输入的特征图在固定的位置进行采样,卷积核权重的固定会导致模型在处理不同位置区域的感受野尺寸相同[17]. 由于卫星视频影像中不同的位置可能对应有不同尺寸或者不同形变的物体,如火车目标像素占比往往大于汽车等目标且存在形变,这要求模型中的卷积操作能够实现自动调整尺度或感受野的方法,以提高对非刚体目标和各尺度目标的敏感度.

本文通过引入可变形卷积思想设计 C2f-DCN 模块以替换 Backbone 中的 C2f 模块,通过模型自适应调整卷积核所对像素的偏差,以增强模型对于火车目标的检测能力. 对于输入特征图,将其进行一次卷积核大小为 $n*n$ 卷积操作获得通道维度上大小为 $3*n*n$ 的偏差和权重特征,其中有 $2*n*n$ 通道维度的特征图对应原特征图上每个像素在进行卷积操作时 x 和 y 方向上的偏移量,$n*n$ 通道维度的特征图对应卷积操作时的权重参数. 再对原特征图进行添加偏移量和像素权重的卷积以得到输出特征图. 在模型训练阶段,同步更新模型特征提取卷积核和可变形卷积偏移量卷积核. 可变形卷积的原理示意如图 2 所示.

可变形卷积的数学表达如式(1)所示.

$$y(p)=\sum_{k=1}^{K} w_k\cdot x(p+p_k+\Delta p_k)\cdot\Delta m_k, \tag{1}$$

式中:Δp_k 为自适应学习到的偏移量,Δm_k 为自适应学习到的权重.

本文利用可变形卷积 DCN(Deformable Convolution Network)可以增强对非刚体目标和不同尺度目标的检测效率的优势,重新设计了 Bottleneck 结构,记作 DCN-Bottleneck,将其用于替换 YOLOv8 模型 C2f 模块中原有的 Bottleneck 结构,以此来丰富梯度路径,增强网络学习的效率. 与原有的 Bottleneck 相比,DCN-Bottleneck 保留了残差网络结果,保证了模型的鲁棒性,然后使用可变形卷积 DCN 替换原模型中的第二层传统卷积. 其架构如图 3 所示.

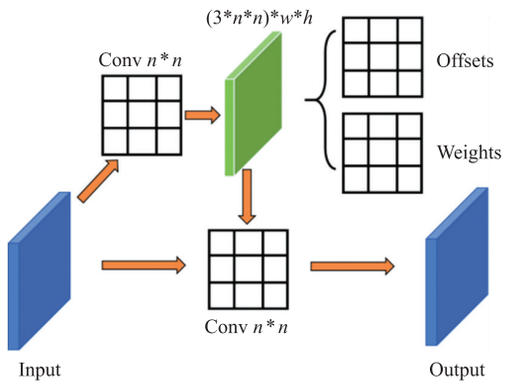


图 2　可变形卷积的原理图

Fig. 2　Schematic diagram of deformable convolution

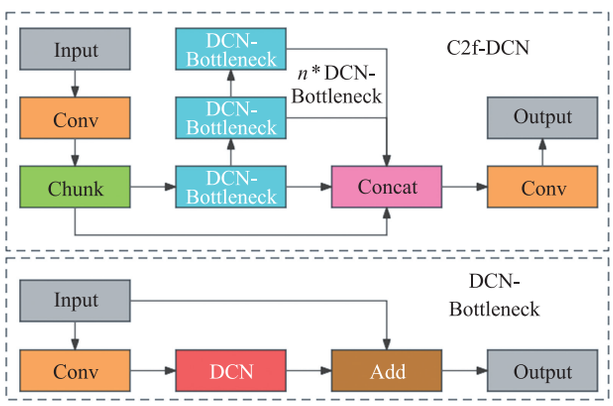


图 3　C2f-DCN 架构

Fig. 3　Structure diagram of C2f-DCN

1.3　Shuffle Attention 注意力机制

由于卫星视频具有更大的视场,其成像区域通常比普通视频大数百倍,这就带来了大量的冗余背景信息和各种背景干扰情况. 在深度学习网络架构中,虽然特征图经过多层卷积和池化操作,但仍有大量的计算资源被用于无目标区域,导致网络训练效率低下. 通过在深度学习中融合注意力机制可以使模型聚焦于关键区域的特征并且抑制不必要的特征,从而提高模型对特征的提取能力. 目前注意力机制按其作用域可分为空间注意力机制、通道注意力机制以及混合域注意力机制,空间注意力机制通过对特征图上每一个像素点赋予不同权重,增强其在空间维度上的特征提取能力,如 STN(Spatial Transformer Networks);通道注意力机制通过在通道维度进行学习,得到各个通道的重要性,如 SE(Squeeze-and-Excitation Networks);混合注意力机制将空间域与通道域融合,如 CBAM(Convolutional Block Attention Module). 本文将高效注意力机制 SA(Shuffle Attention)添加到 YOLOv8 模型中的特征加强模块,以突出卫星视频中的小目标特征增强模型的特征提取能力[18].

对于输入特征图,首先将其按通道维度划分为 N 组生成 N 个子特征图,之后将每个子特征图在通道维度上分成两个分支. 其中一个分支通过计算全局平均池化 GAP 将其嵌入全局信息生成通过到统计数据,最后经过激活函数生成紧凑特征;另一个分支使用组归一化 GN 获得空间统计,经过激活函数后得到空间注意的输出. 然后将两个分支在通道维度上进行连接,并同时将 N 个子特征图连接. 最后采用 channel shuffle 操作使组间信息眼通道维度流动,获得与输入特征图同样大小的输出结果. 其架构如图 4 所示.

在模型的 3 个检测头前嵌入 Shuffle Attention 注意力机制,可以使特征图在进行最终的回归预测操作前,为其同时构建其通道注意力和空间注意力,并实现不同子特征间的信息通信. 可以在保证模型运算速度的前提下,增强模型的特征提取能力,解决卫星视频中目标像素占比少的问题,提高模型的检测性能.

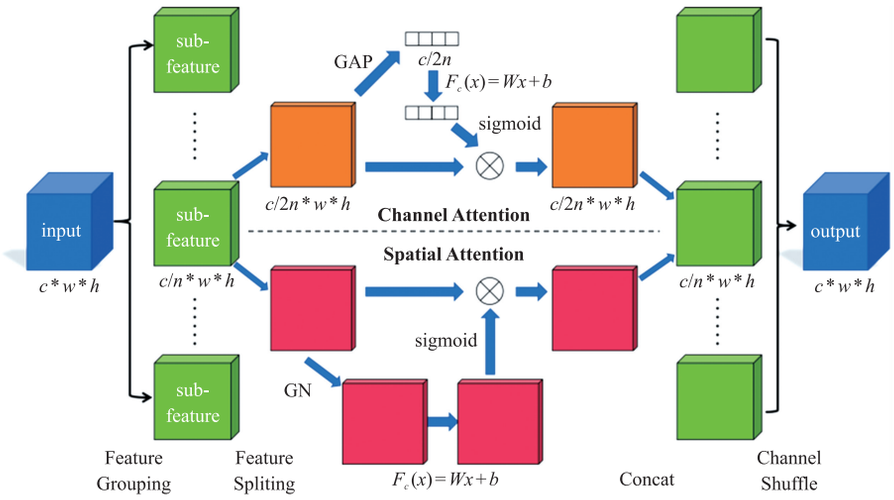


图 4　Shuffle Attention 注意力机制

Fig. 4　Attention mechanism of Shuffle Attention

1.4　Inner-CIoU 优化损失函数

原 YOLOv8 模型中采用 CIoU 和 DFL 损失相结合的方法计算边界框的回归损失,对于卫星视频影像来说,由于其存在目标较小且易发生遮挡等复杂情况,采用 CIoU 计算回归损失时会导致对于小目标来说较小的位置偏移就会使 IoU 急剧下降. 为增强模型对小目标的敏感程度,提升卫星视频影像中像素占比少的小目标检测精度,本文基于 Inner-IoU 思想[19],设计了顾及辅助边框的 Inner-CIoU 优化损失函数.

现有的基于 IoU 的边框回归大多聚焦在加入新的损失项来加速收敛,如 CIoU 在 IoU 基础上考虑了距离与形状损失[20],但现有改进方案忽视了 IoU 损失项其自身的限制. 本文引入 Inner-IoU,采用尺度因子 ratio 控制生成不同尺度的辅助边框用于计算损失,将其应用至现有的 IoU-based 损失函数中以获得更快更为有效的回归效果.

如图 5 所示,GT 框和锚框分别表示为 B^{gt} 和 B,其内部中心点坐标分别为 (x_c^{gt},y_c^{gt}) 和 (x_c,y_c). GT 框的宽度和高度分别表示为 w^{gt} 和 h^{gt},锚框的宽度和高度分别表示为 w 和 h,ratio 为尺度因子. 锚框参数的计算见式(2),交并比的计算见式(3).

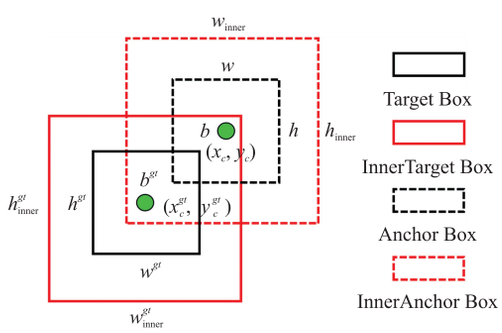


图 5　Inner-IoU 示意图

Fig. 5　Sketch map of Inner-IoU

$$\begin{cases} b_l^{gt}=x_c^{gt}-\dfrac{w^{gt}*\text{ratio}}{2},\ b_r^{gt}=x_c^{gt}-\dfrac{w^{gt}*\text{ratio}}{2} \\ b_t^{gt}=y_c^{gt}-\dfrac{h^{gt}*\text{ratio}}{2},\ b_b^{gt}=y_c^{gt}-\dfrac{h^{gt}*\text{ratio}}{2} \\ b_l=x_c-\dfrac{w*\text{ratio}}{2},\ b_r=x_c-\dfrac{w*\text{ratio}}{2} \\ b_t=y_c-\dfrac{h*\text{ratio}}{2},\ b_b=y_c-\dfrac{h*\text{ratio}}{2} \end{cases} \tag{2}$$

$$\begin{cases} \text{inter}=(\min(b_r^{gt},b_r)-\max(b_l^{gt},b_l))*(\min(b_b^{gt},b_b)-\max(b_t^{gt},b_t)) \\ \text{union}=(w^{gt}*h^{gt})*(\text{ratio})^2+(w*h)*(\text{ratio})^2-\text{inter} \\ \text{IoU}^{\text{inner}}=\dfrac{\text{inter}}{\text{union}} \end{cases} \tag{3}$$

与 IoU 损失相比,当尺度因子 ratio 大于 1 时,辅助边框尺寸大于实际边框,其回归的有效范围大于 IoU 损失,对于低 IoU 的样本的回归有所增益. 由于卫星视频中目标所占像素较小,大部分时间为低 IoU 样本,故本文将尺度因子 ratio 定义为 1.25. 最后将 Inner-IoU 应用至现有的基于 IoU 的边框回归损失函数

中,设计 Inner-CIoU,其损失函数定义如式(4).

$$L_{\text{inner-CIoU}}=L_{\text{CIoU}}+\text{IoU-IoU}^{\text{inner}}. \tag{4}$$

2 实验与讨论

2.1 数据集和评估指标

本文采用中国科学院空间应用工程与技术中心制作的 SAT-MTB 数据集对模型进行训练和评估. 该数据集为目前最多支持任务类型和对象类别的大规模卫星视频数据集,包含飞机、船舶、汽车、火车 4 个类别的 249 段卫星视频,覆盖 5 万余帧和 1 033 511 个注释对象实例. 其中用于目标检测的数据共有 146 段,包含飞机、船舶、火车 3 个类别,本文在各个类别对应视频中采用间隔 10 帧的抽帧策略降低计算开销,并按照 6:2:2 的比例划分训练集、验证集和测试集.

实验软件环境为 Windows10 64 位操作系统,使用 PyCharm 集成开发环境,采用 Python 3.9 编程语言,使用 4.1 版本 OpenCV 库,2.0 版本 Pytorch 深度学习架构. 硬件环境 CPU 为 Inter(R)Core(TM)i7-12700F 处理器,内存为 16G 双通道,GPU 为 NVIDIA GeForce GTX 3060 16G 显存.

本文采用 mAP50:95、$F1$ 分数作为精度检测评估指标,FPS 作为检测速度评估指标. mAP50:95 表示基于 0.05 的步长计算从 0.5 至 0.95 的所有 IoU 阈值下的精测精度平均值. 精度、召回率、$F1$ 分数、mAP、FPS 等定量指标的计算方法见式(5)-式(7).

$$\begin{cases}\text{Precision}=\dfrac{TP}{TP+FP}\\ \text{Recall}=\dfrac{TP}{TP+FN}\\ F1=2*\dfrac{\text{Precision}*\text{Recall}}{\text{Precision}+\text{Recall}}\end{cases} \tag{5}$$

$$\begin{cases}AP=\displaystyle\int_0^1 p(r)\,dr\\ mAP=\dfrac{1}{n}\displaystyle\sum_{i=1}^{n}AP_i\end{cases} \tag{6}$$

$$FPS=1\ 000\ ms/(t_1+t_2+t_3). \tag{7}$$

式中,TP 表示预测正确的样本数,FP 表示将其他负样本预测为正样本的样本数,FN 表示被预测为其他样本的样本数,t_1 为模型预处理所用时间,t_2 为模型推理所用时间,t_3 是模型后处理所用时间.

2.2 消融实验

为验证本文提出的各种改进方案在各项指标上的有效性,使用 SAT-MTB 卫星视频影像数据集以 YOLOv8n 为基线模型进行了消融实验,并在测试集上测试各项改进的检测性能,实验结果如表 1 所示.

表 1 消融实验结果

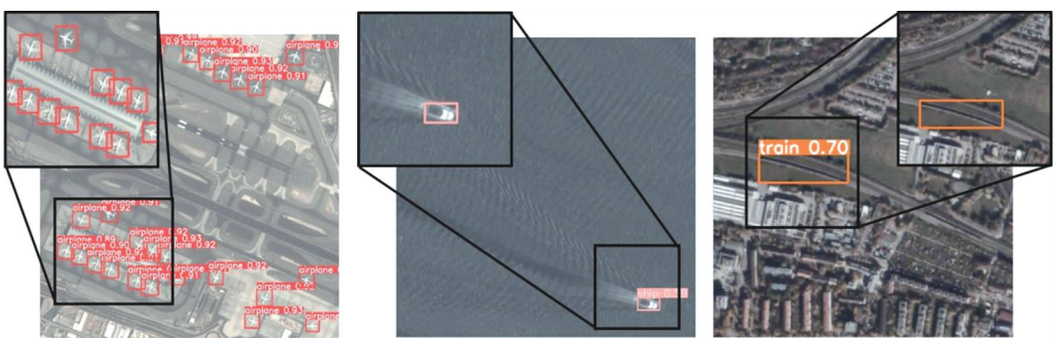
Table 1 Results of ablation experiment

C2f-DCNv2	SA	Inner-CIOU	Precision/%	Recall/%	mAP@0.50-0.95	$F1$	FPS/(frame/s)
×	×	×	63.7	58.6	31.9	61.0	234
✓	×	×	71.3	60.8	32.3	65.6	218
×	✓	×	72.5	60.2	33.9	65.8	227
×	×	✓	72.0	61.1	32.3	66.1	227
✓	✓	✓	75.3	63.1	34.4	67.3	222

由表 1 可见,通过在 Backbone 中添加 C2f-DCN 模块使 mAP50:90 和 $F1$ 分数分别提高 0.4%和 4.6. 在原 YOLOv8 网络结构的 3 个检测头前添加 Shuffle Attention 注意力机制后,可以提升网络的特征提取能力以及对小目标的敏感度,使得 mAP50:90 和 $F1$ 分数相比基线模型分别提高 2.0%和 4.8. 进一步采用 Inner-CIoU 替换原本的 CIoU 损失函数,通过引入辅助边界框的策略适应卫星视频影像的目标检测任务需要,使得 mAP50:90 和 $F1$ 分数分别提高 0.4%和 5.1. 最后将各个改进点加以整合,充分考虑卫星视频影像中目标较小和存在非刚体目标的问题,相比于 YOLOv8n 模型,在 mAP50:90 和 $F1$ 分数两个指标上分别提升

2.5%和 6.3,同时根据计算速度指标 *FPS*,表明本文提出的针对卫星视频影像的目标检测方法在保证模型检测速率的情况下,可以显著提高检测精度.

为了体现本文提出的模型在卫星视频目标检测任务中的检测效果,选取 SAT-MTB 卫星视频影像数据集中的测试集数据进行可视化实验. SAT-MTB 数据集中包含高密度且目标所占像素少的情况,受背景和遮挡严重的情况,以及具有运行中的火车等非刚体目标的情况. 部分实验结果如图 6 所示.



(a) 目标密集情况　　(b) 存在干扰情况　　(c) 非刚体目标情况

图 6　本文卫星视频运动目标检测的可视化实验结果

Fig. 6　Visualization experimental results of the proposed satellite video moving object detection method

在图 6(a)中,本文提出的基于 YOLOv8 的卫星视频目标检测模型能够在目标分布密集且所占像素较少的情况下拥有较高的检测精度. 在图 6(b)中,影像中的船舶目标较小且从颜色上易受云层等环境因素影像,同时还存在遮挡情况,可见使用本文的目标检测模型可以在复杂环境情况下完成卫星视频目标检测任务. 在图 6(c)中,运动中的非刚体变化增加了火车目标检测的难度,本文的方法同样能够实现对其高精度检测.

2.3　对比实验

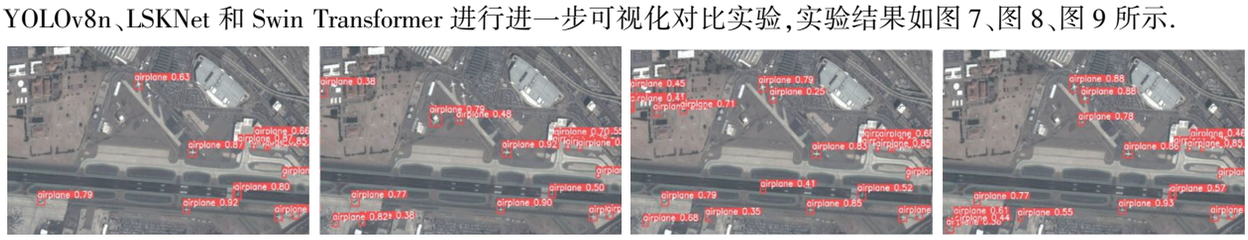
为证明本文改进的模型的有效性和优越性,选取 GOLD-YOLO[21]、LSKNet[22]和 Swin Transformer[23]以及多种经典的目标检测方法,并利用 SAT-MTB 卫星视频影像数据集进行对比实验,具体的定量对比结果列于表 2.

表 2　针对卫星视频运动目标检测的不同方法的对比实验结果

Table 2　Comparative experimental results of different methods for detecting moving objects in satellite videos

	YOLOv8n	DETR	SSD	F-RCNN	GOLD-YOLO	LSKNet	Swin Transformer	Ours
Precision/%	63.7	41.4	61.2	23.9	58.6	61.8	60.7	75.3
Recall/%	58.6	57.7	50.9	63.6	63.4	53.8	59.6	62.8
mAP@0.50−0.95	31.9	20.7	22.8	19.8	30.3	24.4	25.4	34.9
*F*1	61.0	48.2	55.6	34.8	60.9	57.5	60.1	68.5
FPS/(frame/s)	234.0	17.0	36.0	12.0	161.0	17.0	17.0	222.0

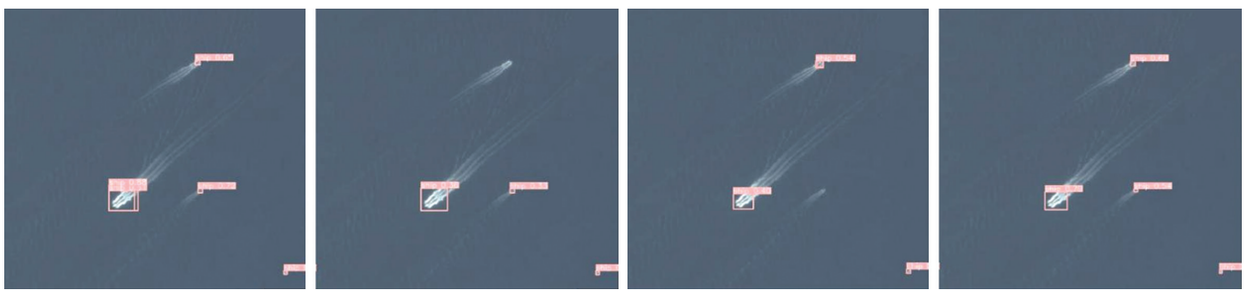
由表 2 可知,本文提出的改进算法在精确率、召回率、mAP50:90 和 *F*1 分数等指标上都高于 SSD 和 F-RCNN 等传统目标检测算法,同时实验与最新的目标检测算法也有较大提升,最后在检测速度指标上本文的改进模型也高于除 YOLOv8n 模型外的其他模型,保证了检测效率. 为直观展示本文算法的优势,选取 YOLOv8n、LSKNet 和 Swin Transformer 进行进一步可视化对比实验,实验结果如图 7、图 8、图 9 所示.



(a) YOLOv8n　　(b) Swin Transformer　　(c) LSKNet　　(d) 本文方法

图 7　不同运动目标检测方法在 SAT-MTB 数据集上对飞机目标的检测结果

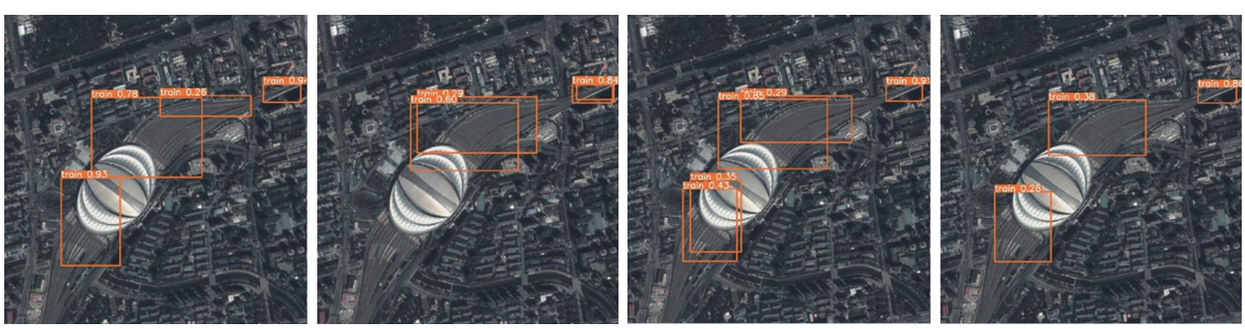
Fig. 7　Detection results of aircraft objects using different moving object detection methods on the SAT-MTB dataset



(a) YOLOv8n (b) Swin Transformer (c) LSKNet (d) 本文方法

图 8 不同运动目标检测方法在 SAT-MTB 数据集上对船舶目标的检测结果

Fig. 8 Detection results of ship objects using different moving object detection methods on the SAT-MTB dataset



(a) YOLOv8n (b) Swin Transformer (c) LSKNet (d) 本文方法

图 9 不同运动目标检测方法在 SAT-MTB 数据集上对火车目标的检测结果

Fig. 9 Detection results of train objects using different moving object detection methods on the SAT-MTB dataset

由图 7 可见,在该场景中 YOLOv8n 模型所得到的检测结果漏检率较高,而 LSKNet 和 Swin Transformer 的检测结果则存在一定的误检率,应用本文的改进算法则实现了较高的精确率和召回率. 由图 8 可见,在第二个场景中存在 3 处船舶目标,其所占像素较少且易受环境影像,导致 LSKNet 和 Swin Transformer 模型的检测结果漏检率较高,而 YOLOv8n 则存在重复检测的问题,本文算法可以适应该场景下的检测任务. 由图 9 可见,在第三个场景中存在运动中的火车目标,且其在颜色上与背景较为相似,YOLOv8n、LSKNet 和 Swin Transformer 模型的检测结果都存在不同程度的重复检测情况,而与其他模型的检测结果相比,本文的改进算法得到的目标框更切合目标实际范围.

3 结论

本文面向基于卫星视频的运动目标检测需求,针对其目标所占影像像素少、背景复杂、存在遮挡情况以及现有模型的误检和漏检率较高等问题,提出了基于改进 YOLOv8 模型的卫星视频运动目标检测方法. 本文方法的主要改进如下:首先,使用重新设计的 C2f-DCN 替换原 C2f 模块,以提高模型对于非刚体目标能取得较好的检测准确率;其次,引入 Shuffle Attention 轻量级注意力机制,以通过空间域和通道域注意力计算结果的融合及混洗操作,放大有用特征而弱化无用特征;最后,通过采用 Inner-CIoU 损失函数优化原有的 CIoU 损失函数,以减少目标像素占比小所带来的影响,进一步提高了模型的检测能力.

在基于 SAT-MTB 数据集的对比实验结果中,本文算法的精确度、召回率、mAP50:95 和 $F1$ 分数分别为 75.3%、62.8%、34.9%和 68.48,相较于原始 YOLOv8n 模型,分别提高了 11.6%、4.2%、3.0%和 7.44. 而且,与 GOLD-YOLO、Swin Transformer 等模型相比也具有一定的优势. 实验结果证实了本文方法具有检测结果受环境影响程度小、检测效率高、能识应不同类型目标检测任务等特点. 因此,对于卫星视频的运动目标检测而言,本文方法具有较好的可行性与优越性. 但是,就信息利用方面,本文方法还未能充分挖掘卫星视频在时序信息上的优越性,在未来的研究中,将继续探讨如何深度利用卫星视频在空间上的连续性,充分利用其时序信息,以进一步提高算法的计算效率与准确率.

[参考文献]

[1] 李贝贝,韩冰,田甜,等. 吉林一号视频卫星应用现状与未来发展[J]. 卫星应用,2018(3):23-27.

[2] 廖一鸣,万剑华,臧文乾,等. 视频卫星数据的运动车辆提取[J]. 测绘科学,2018,43(4):144-149.

[3] 吴昱舟,姚力波,刘勇,等. 基于帧差和背景建模的卫星视频目标检测[J]. 海军航空工程学院学报,2018,33(5):441-446.

[4] SHU M,ZHONG Y,LV P. Small moving vehicle detection via local enhancement fusion for satellite video[J]. International journal of remote sensing,2021,42(19):7189-7214.

[5] 吴佳奇,蒋永华,沈欣,等. 决策树弱分类支持的卫星视频运动检测[J]. 武汉大学学报(信息科学版),2019,44(8):1182-1190.

[6] LEI J,DONG Y,SUI H. Tiny moving vehicle detection in satellite video with constraints of multiple prior information[J]. International journal of remote sensing,2021,42(11),4110-4125.

[7] ZHANG J,JIA X,HU J,et al. Moving vehicle detection for remote sensing video surveillance with nonstationary satellite platform[J]. IEEE transactions on pattern analysis and machine intelligence,2022,44(9):5185-5198.

[8] 魏业文,李梅,解园琳,等. 基于改进 Faster-RCNN 的输电线路巡检图像检测[J]. 电力工程技术,2022,41(2):171-178.

[9] 刘江,关向雨,温跃泉,等. 基于改进 YOLOv4 的 GIS 红外特征识别与温度提取方法[J]. 电力工程技术,2023,42(1):162-168.

[10] 周文青,刘刚. 基于深度学习和无人机图像的架空线路缺陷巡检综述[J]. 电力工程技术,2024,43(2):73-82.

[11] LEI L,GUO D. Multitarget detection and tracking method in remote sensing satellite video[J]. Computational intelligence and neuroscience,2021,2021(1):7381909.

[12] CHEN R,LI X,LI S. A lightweight CNN model for refining moving vehicle detection from satellite videos[J]. IEEE access,2020,8:221897-221917.

[13] FENG J,ZENG D,JIA X,et al. Cross-frame key point-based and spatial motion information-guided networks for moving vehicle detection and tracking in satellite videos[J]. ISPRS journal of photogrammetry and remote sensing,2021,177:116-130.

[14] YU C,FENG Z,WU Z,et al. HB-YOLO:An improved YOLOv7 algorithm for dim-object tracking in satellite remote sensing videos[J]. Remote sensing,2023,15(14):3551.

[15] LI S,ZHOU Z,ZHAO M. A multitask benchmark dataset for satellite video:object detection,tracking,and segmentation[J]. IEEE transactions on geoscience and remote sensing,2023,61:5611021.

[16] RAHMAN S,RONY J H,UDDIN J,et al. A. Real-time obstacle detection with YOLOv8 in a WSN using UAV aerial photography[J]. Journal of imaging,2023,9(10):216.

[17] ZHU X,HU H,LIN S,et al. Deformable convnets V2:More deformable better results[C]//Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition. Long Beach,USA:IEEE,2019:9300-9308.

[18] ZHANG Q L,YANG Y B. SA-Net:Shuffle attention for deep convolutional neural networks[C]//Proceedings of IEEE International Conference on Acoustics,Speech and Signal Processing. Toronto,Canada:IEEE,2021:2235-2239.

[19] ZHANG H,XU C,ZHANG S J. Inner-IoU:more effective intersection over union loss with auxiliary bounding box[J/OL]. arXiv,2023,https://doi.org/10.48550/arXiv.2311.02877.

[20] GEVORGYAN Z. SIoU Loss:more powerful learning for bounding box regression[J/OL]. arXiv,2022,https://doi.org/10.48550/arXiv.2205.12740.

[21] WANG C,HE W,NIE Y,et al. Gold-YOLO:efficient object detector via gather-and-distribute mechanism[J/OL]. arXiv,2023,https://doi.org/10.48550/arXiv.2309.11331.

[22] LI Y,HOU Q,ZHENG Z,et al. Large selective kernel network for remote sensing object detection[J/OL]. arXiv,2023,https://doi.org/10.48550/arXiv.2303.09030.

[23] LIU Z,LIN Y,CAO Y,et al. Swin transformer:hierarchical vision transformer using shifted windows[C]//Proceedings of the IEEE International Conference on Computer Vision. Montreal,Canada:IEEE,2021:9992-10002.

[责任编辑:黄 敏]