

doi: 10.6046/zrzyg.2022126

引用格式: 邱磊, 张学志, 郝大为. 基于深度学习的视频 SAR 动目标检测与跟踪算法[J]. 自然资源遥感, 2023, 35(2): 157 – 166. (Qiu L, Zhang X Z, Hao D W. VideoSAR moving target detection and tracking algorithm based on deep learning[J]. Remote Sensing for Natural Resources, 2023, 35(2): 157 – 166.)

基于深度学习的视频 SAR 动目标检测与跟踪算法

邱磊^{1,2}, 张学志², 郝大为²

(1. 海军工程大学兵器工程学院, 武汉 430033; 2. 陆军工程大学军械士官学校雷达系, 武汉 430075)

摘要: 视频合成孔径雷达 (synthetic aperture Radar, SAR) 技术被广泛应用于军事侦查、地质勘探和灾害预测等领域。由于 SAR 视频存在很多的相干斑 (Speckle) 噪声以及镜面反射、叠掩效应等干扰因素, 运动目标容易与背景或其他目标混淆在一起。针对上述问题, 文章提出了一种有效的视频 SAR 目标检测与跟踪算法。首先, 提取视频 SAR 的多个特征用于构造多通道特征图; 然后, 使用改进的轻量 EfficientDet 网络对更深层的特征进行提取, 从而在兼顾算法效率的同时提升 SAR 目标检测的准确度; 最后, 采用基于目标检测框的轨迹关联策略对视频 SAR 中同一目标进行关联。实验表明, 本研究提出的方法针对 SAR 阴影目标检测与跟踪任务取得了较好的效果。

关键词: 视频 SAR; 特征增强; 目标检测; 深度学习; 特征金字塔; 多目标跟踪

中图法分类号: TP 79 **文献标志码:** A **文章编号:** 2097 – 034X(2023)02 – 0157 – 10

0 引言

遥感目标跟踪在军事、环境、资源、气象和农业等多个领域有广泛应用。视频合成孔径雷达 (synthetic aperture Radar, SAR) 通过向地面发射电磁波信号获取地表物体的反射信息, 从而对动目标进行连续成像。相较于传统的可见光影像和其他遥感影像, 视频 SAR 受光照和气候条件影响较小, 特别适合雨、雪、雾和夜间的任务作业。同时, 视频 SAR 还具备一定的穿透效果, 在有云层和植被遮挡的情况下仍然能对目标进行成像。因此, 通过分析 SAR 视频, 对地面运动目标进行检测跟踪, 比一般光学遥感数据有更广泛的适用场景^[1–3]。

在 SAR 目标检测方面, 林旭等^[4]提出了一种基于自适应背景杂波模型的恒虚警率 (constant false – alarm rate, CFAR) 宽幅 SAR 图像舰船检测算法。通过背景窗口的多尺度统计方差判断目标所处的杂波环境, 自适应选择对应的背景杂波分布模型; Huang 等^[5]提出了一种结合目标语义特征的恒定虚警法, 在检测高分辨率 SAR 图像目标时具有更低的虚警率; Kaplan^[6]利用扩展分形 (extended fractal, EF) 特征检测 SAR 图像中的车辆目标, 该特征不仅对目标背景的对比度敏感, 同时也对目标尺寸敏感; 刘冬

等^[7]利用指数小波分形特征检测 SAR 图像目标, 该方法对目标对比度和尺寸敏感, 且考虑了目标边缘特征与区域内部平滑性特征。近年来, 随着深度学习的发展, 基于卷积神经网络 (convolutional neural network, CNN) 的 SAR 图像目标检测方法受到了广泛关注。Fast R – CNN^[8], Faster R – CNN^[9] 和 YOLO^[10] 等目标检测网络被广泛应用于对 SAR 图像的检测。Kang 等^[11]提出了结合上下文的检测方法检测 SAR 图像舰船目标, 利用目标周围的背景信息提高检测结果的可靠性, 减少了虚警; Wang 等^[12]使用最大稳定极值区域 (maximum stability extremal region, MSER) 准则进一步约束目标区域的判定; An 等^[13]用 DRBox – v2 (一种旋转边界框) 取代传统矩形框标记目标, 使得标记选框可以适应不同的目标朝向; 张椰等^[14]将目标检测任务转化为像素分类问题, 利用全卷积神经网络 (fully convolutional neural network, FCN) 对图像进行像素级的分类, 有效避免了背景杂波的干扰, 降低了虚警率。在视频 SAR 目标跟踪方面, Wang 等^[15]研究了目标的阴影对视频 SAR 序列中低雷达散射截面 (Radar cross section, RCS) 目标跟踪的积极影响; Zhang 等^[16]提出了一种基于 CNN 的阴影检测器来提取阴影信息作为定位运动目标的线索; Xu 等^[17]和 Liang 等^[18]利用核化相关滤波器 (kernel correlation filter, KCF) 和

SiamMask 跟踪器对第一帧目标初始化后的阴影进行定位; Zhang 等^[19]采用尺度不变特征变换(scale - invariant feature transform, SIFT)和随机样本一致性(random sample consensus, RANSAC)配准算法来补偿每帧背景的变化,并采用了一种称为最大化 Tallis 熵的阈值分割算法,利用三帧差分法对背景差分进行阴影检测; Henke 等^[20-21]提出了一种基于图像统计和多目标卡尔曼滤波的 SAR 目标跟踪方法; Ding 等^[22]采用改进的基于密度的聚类算法对每一帧检测结果进行过滤,并利用设计的双向长短期记忆(bi - directional long - short term memory, Bi - LSTM)网络抑制检测中的缺失报警; Zhao 等^[23]提出了一种基于时空信息和基于显著性的检测机制来对抗干扰和背景杂波。

近年来,国内对该问题的研究也越来越多。何志华等^[24]利用视频 SAR 帧图像空间域和时间域信息,经过数据配准和相干斑滤波预处理、采用基于改进差分的动目标阴影检测处理和多目标跟踪处理技术,在保证检测率的前提下尽量降低虚警率,实现了视频 SAR 动目标检测与跟踪; 刘雨洁等^[25]融合航迹起始算法与舰船图像特征差异进行长时间间隔序贯图像间的目标匹配,根据速度和加速度设置约束波门,对舰船目标的不同特征设置不同的权值计算目标间的差异,利用差异最小原则筛选航迹,实现目标跟踪; 胡瑶^[26]提出了基于改进的无锚点多目标跟踪框架,引入了注意力机制、修改了损失函数并进行了网络结构的优化,进而提升了跟踪效果。然而,现有工作对于 SAR 图像自身特征的挖掘尚不充分,目标跟踪受到信号强度、障碍遮挡和运动模糊等因素影响较大。这启发本文设计更好的特征表达网络和轨迹关联算法。

针对上述问题,本文提出了基于深度学习的目标检测与跟踪方法。首先,提取出目标的显著特征用于视频 SAR 动目标检测。由于视频 SAR 成像机制与可见光成像不同,视频 SAR 具有其特有的属性。为此,本文分析了视频 SAR 的统计分布特征、极化散射特征、纹理特征和几何特征等多种特征。通过实验发现,增强图像对比度、纹理均值特征、开运算特征、闭运算特征,作为模型的先验输入,能够提升检测效果。其次,设计了更适合于视频 SAR 检测的多尺度特征金字塔^[27]和特征融合策略,使不同尺度特征之间的融合更加高效便捷。本文提出使用轻量级 EfficientDet 网络^[28]对 SAR 图像进行检测,并进一步采用 Bi - LSTM 网络对当前帧上未检测到

的目标进行预测。通过分析不同帧下所有锚点框之间的距离关系,为不同目标设置不同的轨迹信息进行轨迹关联,该轨迹优化策略显著提升了视频 SAR 目标的跟踪效果。

1 多通道特征图目标检测方法

在本小节中,详细介绍了提出的目标检测方法。首先,用多通道特征图构建特征集,对图像进行特征增强; 其次,在深度特征提取阶段,介绍了改进后的多尺度特征金字塔; 接下来,基于上述特征,介绍了采用的轻量 EfficientDet 目标检测模块; 然后,分别介绍了 2 种目标轨迹关联方法,即基于目标框的轨迹关联和基于 Bi - LSTM 的轨迹关联。

1.1 多通道特征图特征增强

在对目标进行提取之前,进行特征集构建。特征集中的特征从 2 个方面来考虑: 一方面是去除图像噪声; 另一方面增强目标与周围地物的差异。对于 Sandia 视频 SAR 数据的单帧图像,在选取合适的去噪滤波后,进行特征计算。本研究选用 Lee 滤波去除噪声,窗口为 5 × 5,在去除噪声后,选取 4 个影像特征和滤波后影像作为特征集。这 4 个特征分别是: 开运算、闭运算、对比度增强、纹理均值特征。其中,对比度增强采用直方图均衡化方法,纹理均值特征使用基于灰度共生矩阵(gray level co - occurrence matrix, GLCM)的纹理特征。特征集构建的流程如图 1 所示。

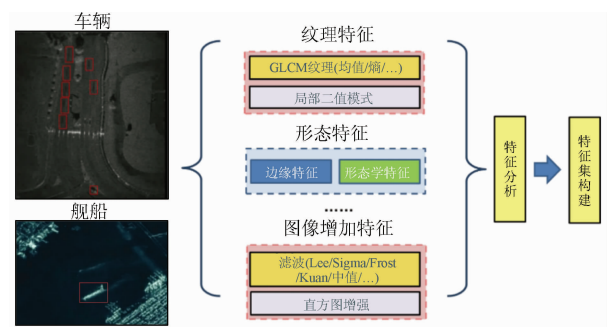


图 1 特征分析流程

Fig. 1 Flowchart of feature analysis

1.2 改进型多尺度特征金字塔

该阶段以自底向上(bottom - up)的多层级语义特征图作为输入,利用上采样和横向连接的方式构建自顶向下(top - down)的多尺度特征金字塔,为多尺度阴影目标检测提供有力基础,目标检测总体结构如图 2 所示。图 2 中 C 表示网络卷积层得到的特征图,P 表示特征金字塔中的特征图。

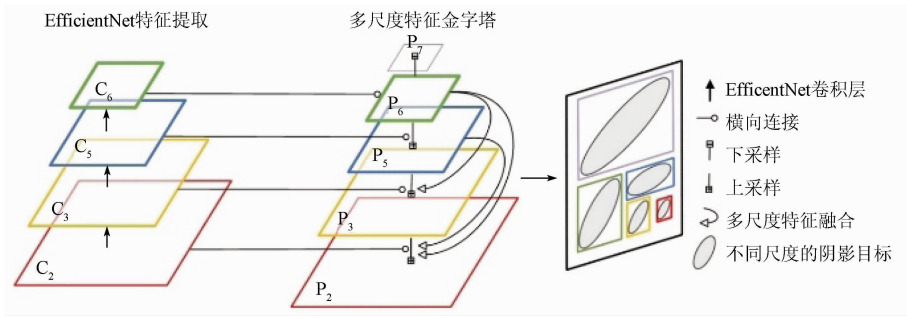


图 2 EfficientNet + 多尺度特征金字塔网络模型结构

Fig. 2 Structure of EfficientNet and the multi-scale feature pyramid

1) 自底向上。该过程由 EfficientNet 主干网络前馈计算而来,其每个阶段最后一层的输出构成金字塔的多个层级,层级间以 2 为缩放比例。

2) 自上向下。由于视频 SAR 中有相当部分的阴影目标尺度较小,且有一定尺度差异,因而对 CNN 网络的尺度适应性也提出更高要求。考虑到在经典特征金字塔网络结构中,小尺度目标的识别主要依赖于金字塔低层中空间分辨率较高的特征图层,因此本文选择在经典特征金字塔网络的基础上,以多尺度特征融合的形式将高维特征图中的语义信息注入至低级特征图中,从而进一步提升网络对小尺度阴影目标的检测能力。多尺度特征融合结构如图 3 所示。

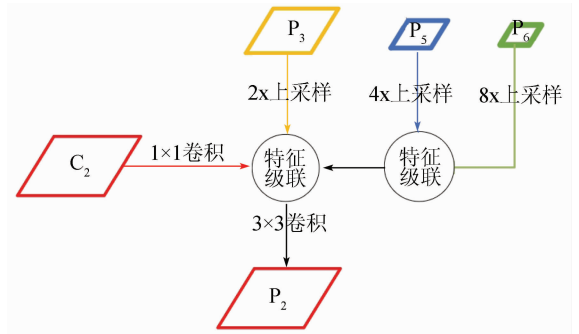


图 3 多尺度特征融合的结构

Fig. 3 Structure of multi-scale feature fusion

具体而言,多尺度特征融合处理中 P_2 和 P_3 的特征可通过式(1)更新,即

$$P_i = \text{Conv}_{3 \times 3} \left(\sum_{j=i+1}^5 \text{Upsample}(P_j) \oplus \text{Conv}_{1 \times 1}(C_i) \right), i \in [2, 3], \quad (1)$$

式中: i 和 j 分别为卷积网络和特征空间金字塔的层号索引; P_i 为对应 C_i 的融合特征映射; $\text{Conv}_{k \times k}(\cdot)$ 为卷积运算; k 为卷积核大小, $\text{Upsample}(\cdot)$ 为最邻近上采样; $\oplus$ 为特征连接操作。本文采用 1×1 卷积对 C_i 进行降维,并采用 3×3 卷积消除上采样带来的混叠效应。

此外,考虑到视频 SAR 中的阴影目标尺度多元,除了大量小尺度目标外,还存在大尺度目标,因此,本文在 P_6 高维特征图的基础上通过取步长为 2 进行下采样得到 P_7 特征图,以在后续目标检测过程中为不同尺度目标提供特定的特征图。为了进一步确保模型效率,令金字塔具体层数与检测目标尺寸相对应。

1.3 基于轻量 EfficientDet 的目标检测

以多尺度特征金字塔为输入,将在各个尺度特征图上以一定步长逐像素生成多种锚点框,该框的生成过程与区域候选网络(region proposal network, RPN)类似,并针对每个锚点框完成 K 分类和框回归任务。

1) 锚点框生成。该过程与 RPN 生成锚点框过程类似。为防止不同尺度目标框之间存在重叠现象且保证效率,为不同尺度特征图依次分配单一的锚点框面积超参数,以促使各特征图能在尺度上具有专一性。对于每个锚点框,若与真实标记框交并比大于 0.5 被视为正样本,小于 0.4 被视为负样本, $[0.4, 0.5)$ 的锚点框则被丢弃,以使模型回归更加准确的锚点框。

2) K 分类和框回归。该过程用于预测 A 个锚点框和类在每个空间位置存在目标的概率及与真实地表之间的偏移量。具体而言,使用若干 3×3 的卷积层对多尺度特征金字塔的逐像素同时进行 $A \times K$ 的类别预测及 $4A$ 的分类回归,2 个任务参数相互独立。在本研究中 A 取 15,分类回归中的 4 代表中心点横纵坐标和锚点框宽高的偏移量。

在训练阶段,本研究采用多任务损失函数,包含检测框回归损失和目标检测损失,以使网络达到端到端的学习,提高分类性能,定义为:

$$L(\{p_i\}, \{t_i\}) = \frac{1}{N_{\text{cls}}} \sum_i L_{\text{cls}}(p_i, p_i^*) + \lambda \frac{1}{N_{\text{reg}}} \sum_i p_i^* L_{\text{reg}}(t_i, t_i^*), \quad (2)$$

$$L_{cls}(p,1) = -\lg p \quad , \quad (3)$$

$$L_{reg}(t_i,t_i^*) = smoothL1(t_i - t_i^*) \quad , \quad (4)$$

式中： L 为损失函数； N_{cls} 和 N_{reg} 分别为目标类别和包围框的个数； i 为每次小批量 (mini - batch) 中的第 i 个锚点框； L_{cls} 为前景和后景的对数损失， p_i 为每一个锚点框属于目标的预测概率， $p_i^* = 1$ 时代表对应样本为正样本， $p_i^* = 0$ 时代表对应样本为负样本。 L_{reg} 为目标锚点框的回归损失； t_i 为对每个样本的预测目标锚点框的坐标参数； t_i^* 为样本对应的真是锚点框的坐标参数； $smoothL1$ 为 $L1$ 范数形式的平滑项损失，定义如 Faster R - CNN^[9]。由式(2)可以看出，当 $p_i^* = 1$ 时，对应样本为正样本时回归框损失才有意义， $p_i^* = 0$ 时，对应样本为负样本时则没有回归框损失。

1.4 基于目标检测框的轨迹关联

以目标检测框为输入，最初通过计算上一帧与当前帧所有目标检测框的交并比来判定轨迹关联信息。然而，研究发现仅通过交并比来判定时，会出现轨迹关联不稳定、轨迹编号迅速变化的现象。通过对轨迹关联进行优化能够有效抑制轨迹信息的快速变化。轨迹关联模块如图 4 所示。

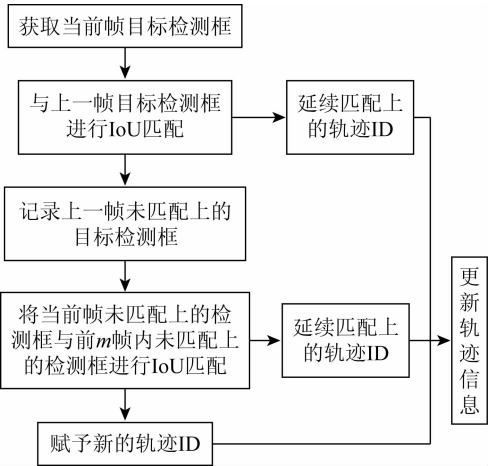


图 4 轨迹关联模块
Fig. 4 Module of track linking

轨迹信息初始化：将每帧中所有正确检测框和误报检测框的中心坐标表示为 $q(x,y)$ ， $x \in [1,M]$ 和 $y \in [1,N]$ ，其中 M 和 N 表示每帧的像素大小。通过测量上一帧与当前帧每个检测框的交并比判定是否为同一轨迹，对同一轨迹的目标设置相同的轨迹关联信息，作为时序上的关联。

存储目标从起始帧到当前帧的轨迹信息，包括轨迹的编号、起始帧、持续帧数和锚点框尺寸。

每检测一帧，则通过计算交并比来判定上下 2 帧距离相近的检测框是否为同一目标轨迹，若当前

帧检测到的与上一帧某一目标检测框交并比大于一定阈值，则判定为同一轨迹，赋予相同轨迹信息。若交并比小于一定阈值，判定为不同目标轨迹，赋予不同轨迹信息。当 SAR 目标比较集中时，可能当前帧存在多个目标与上一帧同一个检测框有较大的重合部分，则在其中选择交并比最大的目标作为轨迹延续。若对于上一帧的某一目标，在当前帧未能找到与其重合部分较大的检测框，说明该目标未被检测到或已离开。若当前帧存在目标未被检测到则在后面的几帧会再次出现，因此需要对这些消失的轨迹进行记录。对于所有中断的轨迹，若连续 8 帧未检测出能够与其连接的目标，则判定该目标已从当前视频中离开，从存储中断轨迹的字典结构中移除。

对于当前帧初始化后新出现的轨迹，与所有被记录的消失轨迹最后一帧进行判定，两两之间距离小于像素阈值且距离最短的判定为同一条轨迹。对于当前帧初始化后新出现的轨迹，若不存在消失轨迹与其距离相近，则判定为新的轨迹。真实阴影目标的运动具有连续性，若轨迹只出现了 1 或 2 帧，即长度小于 3 帧，则判定该轨迹不具有连续性，将其轨迹信息剔除。

1.5 基于 Bi - LSTM 网络的同步目标轨迹预测

Bi - LSTM 是一种双向的长短期记忆网络 (long - short term memory, LSTM)，它同时考虑了时序的正向信息和反向信息。该网络由一个前向时序传播层 $LSTM_f$ 和一个反向时序传播层 $LSTM_b$ 组成。通过正向与反向循环神经网络获得每个时间节点的权重 h_t ，最后由一个全连接层合并所有权重得到输出 Y 。Bi - LSTM 结构如图 5 所示。因此在本项目的框架中，采用 Bi - LSTM 抑制漏报，每检测一帧则对当前帧的检测框同步进行跟踪。

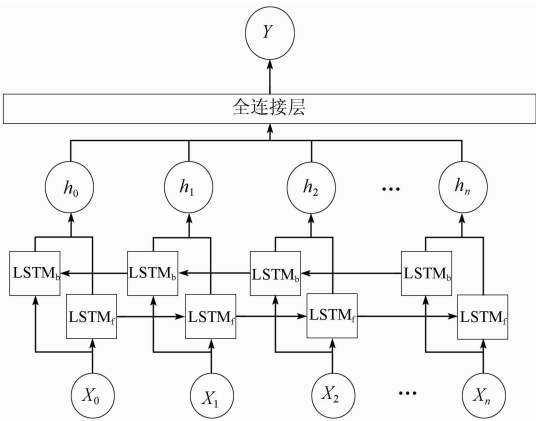


图 5 Bi - LSTM 网络结构
Fig. 5 Structure of Bi - LSTM

本文使用一个包含 $2K + 1$ 个单元的隐藏层，每个单元由 50 个神经元组成，形成从第 $n - k$ 帧到第

n 帧的传播链。而第 n 帧缺失的阴影目标的预测即第 $(K+1)$ 个单元的输出,由前 K 帧到当前帧的 K 个单元对应 K 个中心点坐标并在最后通过全连接层生成缺失目标的预测坐标。在训练阶段,采用预测中心点坐标与真实中心点坐标间的欧氏距离来构建损失函数 $loss$, 定义为:

$$loss = \frac{1}{N} \sum_{m=1}^N (x_m^{\text{pred}} - x'_m)^2 + (y_m^{\text{pred}} - y'_m)^2,$$

(5)

式中: N 为迭代次数; $(x_m^{\text{pred}}, y_m^{\text{pred}})$ 为每次迭代的预测目标中心点坐标; (x'_m, y'_m) 为真实目标中心点坐标。

当第 m 帧有轨迹消失时,根据前 n 帧中心点坐标值,对当前目标中心点坐标进行预测,若第 $m+1$ 帧该轨迹仍未出现,则结合前一帧预测的值继续预测当前帧,连续往后预测 3 帧,若第 $m+2$ 帧目标仍未出现则判定该目标消不再出现。若目标出现则停止预测,若目标连续 3 帧未出现则表明该目标可能已经离开,不再进行预测。

2 实验

本文提出的目标跟踪方法包含检测和轨迹关联 2 个部分,为了验证该方法的实验效果,首先使用单张 SAR 图片对检测功能进行测试,然后再在 Sandia 视频数据上测试跟踪效果。这样做的另一个好处是利用迁移学习的技术弥补了视频 SAR 数据公开样本不足的缺陷。

2.1 样本集的构建

以我国国产高分三号 SAR 数据和 Sentinel-1 SAR 数据为主数据源。高分三号的成像模式是 Strip-Map (UFS), Fine Strip-Map 1 (FSI), Full Polarization 1 (QPSI), Full Polarization 2 (QPSII) 和 Fine Strip-Map 2 (FSII)。这 5 种成像模型的分辨率分别为 3 m, 5 m, 8 m, 25 m 和 10 m。Sentinel-1 的成像模式是条带模式 (S3 和 S6) 和宽幅成像模式, SAR 舰船目标尺寸信息见表 1。共采用 102 景

表 1 SAR 舰船目标尺寸情况

Tab.1 Sizes of ships in SAR images

面积尺寸/像素	舰船目标数量/个
0×0~8×8	269
8×8~16×16	3 231
16×16~32×32	25 302
32×32~64×64	20 842
64×64~128×128	1 209
128×128~n×n	32
合计	50 885

高分三号和 108 景 Sentinel-1 SAR 图像构建高分辨率 SAR 舰船目标深度学习样本库。这些影像拍摄时的入射角大多在 20°~50°之间。目前,该深度学习样本库包含 43 819 舰船切片。

Sandia 视频影像由美国桑迪亚国家实验室提供,本文选择了其中的柯特兰空军基地大门的一段影像。视频显示车辆从各个方向通过大门。沿着道路移动的阴影总是在车辆的实际物理位置。当车辆停止时,车辆的反射能量落在阴影的顶部。一旦车辆继续行驶,影子就再次可见。在屏幕上移动的线是由移动的车辆引起的多普勒频移。该段视频有效帧数为 714,每帧影像分辨率为 640 像素×368 像素,使用 DarkLable 对 Sandia 视频影像车辆标注目标框和 id 信息,总共标记了 48 条车辆轨迹,共 5 020 个包含时序信息的锚点框数据,最长轨迹为 222 帧,最短轨迹为 43 帧。Sandia 阴影目标尺寸信息见表 2。

表 2 Sandia 阴影目标尺寸情况

Tab.2 Sizes of shadow targets in Sandia data

尺寸/像素	阴影目标数量/个
0×0~8×8	0
8×8~16×16	1 374
16×16~16×16	4 041
32×32~16×16	5
64×64~n×n	0
合计	5 020

2.2 参数设置

为确保实验公正性,所有对比试验都采用同样的超参数。根据目前计算机视觉领域对于目标检测任务中 CNN 网络的通用方法,训练前 50k 次学习率为 0.002,后 25k 为 0.000 2,最后 25k 为 0.0000 2。权重衰减和动量系数分别为 0.000 1 和 0.9,选择的优化算法是动量优化。

鉴于遥感影像目标具有朝向随机、大纵横比的特点会导致候选区域交并比 (intersection over union, IoU) 值不稳定,将正样本 IoU 阈值为 0.5,以保证后续的目标检测有充足的正样本。对测试集的定量评估采用准确率、召回率和 F1 指标 3 个指标。

此外,针对不同的数据集,由于目标尺寸分布的不同,将构建不同的特征金字塔结构。本研究针对中科院舰船目标样本集的多尺度特征金字塔构建及锚点框参数设定如表 3 所示。

表 3 中科院舰船目标样本集的多尺度特征金字塔构建

Tab.3 Information of the multi-scale feature pyramid on the ship dataset (像素×像素)

特征金字塔层级	尺寸设定
1	4 ²
2	8 ²

(续表)

特征金字塔层级	尺寸设定
3	16^2
4	32^2
5	64^2
6	128^2

本研究针对 Sandia 阴影目标样本集的多尺度特征金字塔构建及锚点框参数设定如表 4 所示。

表 4 Sandia 阴影目标样本集的多尺度特征金字塔构建

Tab.4 Information of the multi – scale feature pyramid on the Sandia dataset (像素×像素)

特征金字塔层级	尺寸设定
1	10^2
2	20^2
3	40^2

2.3 数据增强

为了去除视频 SAR 影像单帧图像噪声,分别采用 Lee 滤波、增强的 Lee 滤波、中值滤波、Kuan 滤波对 Sandia 视频 SAR 车辆数据单帧数据集进行滤波去噪,其中窗口均为 5×5 。为了评价各种滤波的滤波效果,在阴影附近区域绘制一条曲线,如图 6 所示。记录这条曲线从上至下的原始影像和各种滤波的的像素值,结果如图 7 所示。从结果可以看出,经滤波后图像像素值变化与原始数据相比呈现出更加平缓的变换趋势,有效地抵御了噪声的干扰。

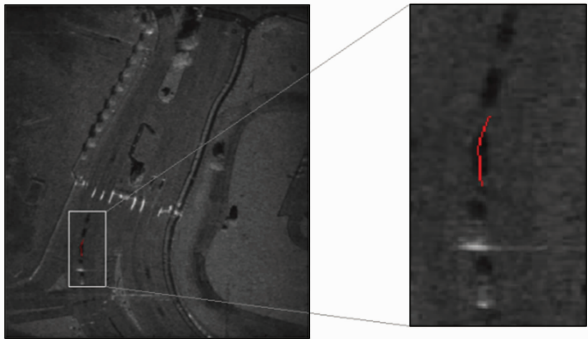


图 6 样本曲线
Fig.6 The curve of the samples

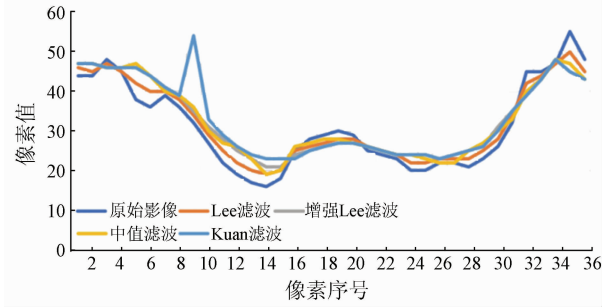


图 7 滤波效果对比
Fig.7 Comparison results of different filters

图 8 为 Sandia 数据特征分析。从图 8 中可以看出,滤波后图像更加平滑,噪声减小,开运算特征结果中亮度较高的物体明显减小,图像更加平滑;闭

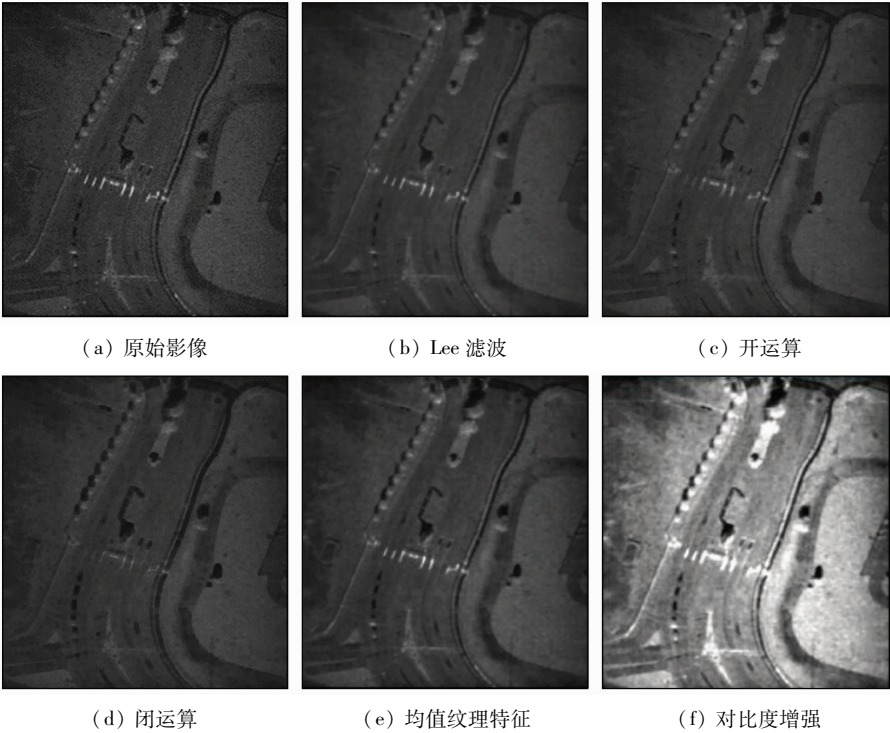


图 8 Sandia 数据特征分析

Fig.8 Analysis of feature on the Sandia dataset

运算特征结果中,与原图相比车辆阴影区域明显变大,特别是位于图像中下方的 2 个车辆阴影更加明

显;均值纹理特征中,车辆阴影特征区域比原图明显增加,阴影特征结果突出。

数据增强是计算机视觉领域用于增强模型鲁棒性和泛化性的重要手段,本研究主要采用平移、亮度调节(亮、暗)、随机噪声、中值滤波、水平垂直翻转、旋转(45° , 90° , 135° , 180° , 225° , 270° , 315°), 约将整个样本库增强 14 倍。

2.4 中科院 SAR 舰船目标检测

此公开样本集主要被用于确定设计的 Effcient-Det 较主流 Faster R – CNN 具有更强的适应性,并探究不同的热门特征提取网络之间的差异性。在评估时,本文设定预测概率大于 0.75, $IoU > 0.5$ 的目标为模型正确检测目标。本文方法及其他常用对比方法的定量评价检测结果如表 5 所示。

表 5 目标检测定量评价结果比较

Tab.5 Comparison of the detection results (%)

网络模型	准确率	召回率	F1 指标
FasterR – CNN	95.68	81.82	88.21
特征金字塔网络	95.70	86.46	90.85
RetinaNet	96.98	88.12	92.34
轻量 EffcientDet	98.08	90.23	93.99
多特征输入 + 轻量 EffcientDet	98.35	91.08	94.58

本研究方法在准确率、召回率和 $F1$ 这 3 项定量指标上分别达到 98.35% ,91.08% 和 94.58% 的最优精度。具体表现如图 9 所示,从测试集中选取多种场景切片进行目视解译,其中包含像素数量小于 102 的小尺度目标,包含像素数量大于 1 282 的大尺度目标,介于其间的中尺度目标,及复杂背景下的船舶目标(建筑物、港口等虚警目标影响)。红色框表示船舶检测结果。

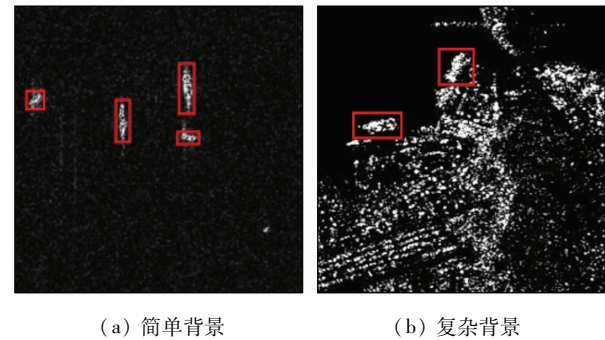


图 9 SAR 舰船目标检测可视化结果

Fig.9 Visualization of the ship detection results

2.5 Sandia 视频 SAR 阴影目标检测

通过中科院 SAR 舰船目标样本集的对比实验,确定了本研究设计的 EffcientDet 模型对复杂场景下的多尺度目标具有较好的检测性能。根据 Sandia 官方的说明,Sandia 视频 SAR 影像上沿着道路移动的阴影总是在车辆的实际物理位置。基于以上结论,将此模型应用到 Sandia 视频 SAR 影像上进行阴影目标检测,并探索 SAR 特征增强后的多特征输入

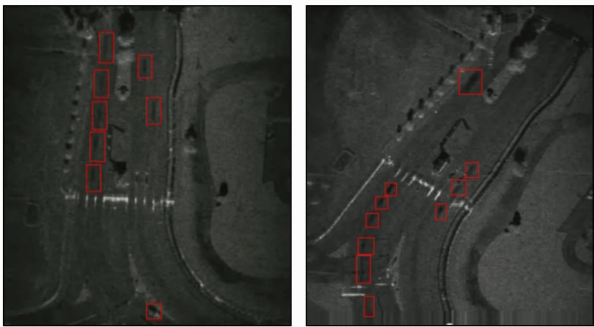
为阴影目标检测带来的促进效果,进而完成最终的视频目标检测及跟踪。与上文类似,在评估时,设定预测概率大于 0.75, $IoU > 0.5$ 的目标为模型正确检测目标。本文方法及其他常用对比方法的定量评价检测结果如表 6 所示。

表 6 阴影目标检测定量评价结果比较

Tab.6 Comparison results of shadow targets detection (%)

网络模型	准确率	召回率	F1 指标
Faster RCNN	90.11	89.65	89.88
特征金字塔网络	92.44	92.26	92.35
RetinaNet	96.99	94.39	95.67
轻量 EffcientDet	97.92	96.37	97.14
多特征输入 + 轻量 EffcientDet	98.85	97.80	98.32

本研究方法多特征输入 + EffcientDet 在准确率、召回率和 $F1$ 这 3 项定量指标上分别达到 98.85% ,97.80% 和 98.32% 的最优结果。具体表现如图 10 所示,红色框表示检测结果。



(a) 时刻 1 检测结果 (b) 时刻 2 检测结果

图 10 在 2 个不同时刻 SAR 车辆目标检测可视化结果

Fig.10 Visualization of car detection results at two timesteps

EfficientDet 模型受利于多尺度特征金字塔,面对多尺度的阴影目标仍然具有较好的适应性,其泛化性能得到进一步证实; SAR 特征增强的多特征输入对已经取得优异精度的模型进一步提供了帮助,实质是阴影特征得到了增强,比如轮廓更加清晰,与背景的差异更大。这较好地改善了单特征输入时,因阴影目标速度过快、目标 SAR 反射强度信号不明显而导致的漏检现象。此外,目标等红绿灯与灯架下方静止时,多特征输入亦能起到促进作用。

2.6 Sandia 视频 SAR 阴影目标跟踪

设计多目标跟踪实验,用于研究跟踪模型在考虑轨迹关联信息的条件下,对多个 SAR 目标同步跟踪的效果。

基于 Bi – LSTM 网络的同步目标轨迹预测结果如图 11 所示,在第 n 帧对漏警目标进行了预测,其中蓝色为正确检测的锚点框,红色为使用 Bi – LSTM

网络通过前 5 帧对第 n 帧漏警目标进行预测的锚点框。从图中可以看出,引入 Bi-LSTM 网络带来 2 个优点: ①对于大部分目标而言,Bi-LSTM 网络预测的结果与实际检测器检测的结果高度重合,这可

以进一步用于辅助判定检测结果的正确性; ②虽然第 n 帧左下角有一个目标被漏检了,但是 Bi-LSTM 网络能正确地预测出该目标的实际位置,对检测器的错误进行了有效的补充。

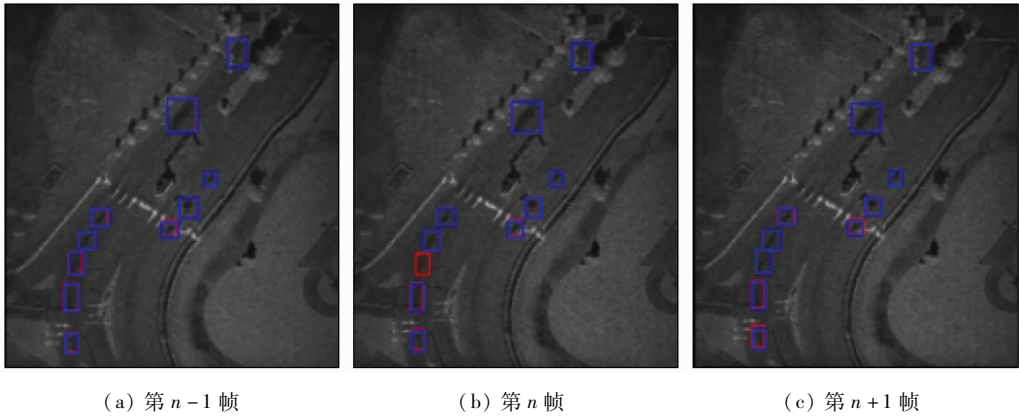


图 11 Bi-LSTM 预测漏警目标示例

Fig. 11 A case in which Bi-LSTM finds the missing target

将 $IoU=0.5$ 作为轨迹初始化时的判定标准,将所有帧下的目标检测与跟踪结果合并为视频,对比仅通过 IoU 判定轨迹关联和添加了跟踪模块后的跟踪结果,可以观察到仅通过 IoU 判定轨迹关联标准没有添加跟踪模块时,轨迹信息会不稳定的快速变化,总共检测到了 1 092 条轨迹,对轨迹信息进行优化后的目标轨迹相对稳定,不会产生过多错误的轨迹关联信息,总共检测到 78 条轨迹,轨迹的优化减少了大量错误的轨迹关联信息。造成这一结果的原因主要是通过引入 Bi-LSTM 网络,降低了目标的漏检率,减少了轨迹的中断次数,进而抑制了目标编号的错误增长。图 12 为轨迹关联优化前(左)和轨迹关联优化后(右)的同一帧图像。

跟踪准确度,结果如表 7 所示。在此仅通过 IoU 判定轨迹关联和通过不同跟踪模型预测轨迹关联的结果。MOTA 用于确定多目标跟踪的准确度,能体现出轨迹 id 的稳定性。IDP, IDR 和 IDF1 用于体现多目标跟踪轨迹信息的正确率。IDP 表示检测到的所有目标中,正确识别到且轨迹信息正确的目标所占比例。IDR 表示所有真值目标中,正确识别到且轨迹信息正确的目标所占比例。IDF1 是 IDP 和 IDR 加权平均得到的结果。这些指标均越大越好。

表 7 多目标跟踪评价结果比较

Tab. 7 Results of multi-target tracking (%)

跟踪模块	MOTA	IDP	IDR	IDF1
$IoU=0.5$	73.3	26.1	25.2	25.5
轨迹优化 + LSTM	90.5	81.9	80.3	81.1
轨迹优化 + Bi-LSTM	91.0	82.1	80.5	81.3

3 结论

本文分别采用多通道特征图对 SAR 图像进行特征增强,同时构建了改进型多尺度特征金字塔对不同尺寸的 SAR 特征进行特征融合,进一步提升了算法对于 SAR 图像目标特征表达的能力。在检测阶段,采用 EfficientDet 网络进行语义特征提取,能够提取到更深层的特征,并引入 Bi-LSTM 网络对目标检测产生的虚警目标进行轨迹关联优化和预测。通过在中科院 SAR 舰船目标样本集下的目标检测实验和美国 Sandia 国家实验室的 SAR 车辆目标样本集上的目标检测与跟踪实验,验证了本研究设计的模型在复杂场景下进行视频 SAR 多尺度目

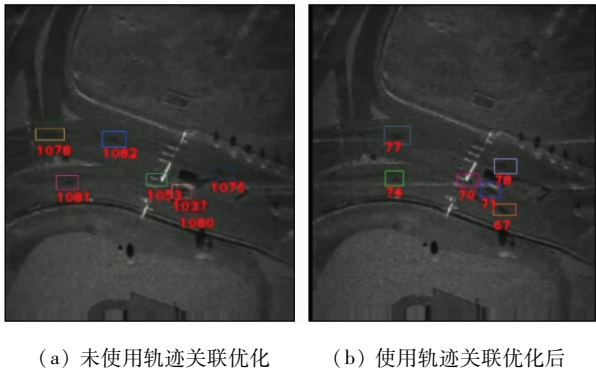


图 12 轨迹关联优化示例

Fig. 12 A sample of our track linking results

通过计算多目标跟踪精度 (multiple object tracking accuracy, MOTA), 识别准确率 (identification precision, IDP) 识别召回率 (identification recall, IDR) 和识别 F1 指标 (IDF1) 这 4 个指标评估多目标

标检测与多目标跟踪时,能有效抑制漏检和误检,同时避免轨迹编号的错误增加,具有更好的效果。

参考文献 (References):

- [1] Wells L, Sorensen K, Doerry A, et al. Developments in SAR and IFSAR systems and technologies at Sandia National Laboratories [C]//Aerospace Conference. IEEE, 2005.
- [2] Wang D, Zhu D Y, Liu R. Video SAR high-speed processing technology based on FPGA [C]//2019 IEEE MTT-S International Microwave Biomedical Conference (IMBioC). IEEE, 2019, 1: 1-4.
- [3] 丁金闪. 视频 SAR 成像与动目标阴影检测技术 [J]. 雷达学报, 2020, 9(2): 14.
Ding J S. Focusing algorithms and moving target detection based on video SAR [J]. Journal of Radars, 2020, 9(2): 14.
- [4] 林旭, 洪峻, 孙显, 等. 一种基于自适应背景杂波模型的宽幅 SAR 图像 CFAR 舰船检测算法 [J]. 遥感技术与应用, 2014, 29(1): 75-81.
Lin X, Hong J, Sun X, et al. New CFAR ship detection algorithm based on aaptive back-ground clutter model in wide swath SAR images [J]. Remote Sensing Technology and Application, 2014, 29(1): 75-81.
- [5] Huang Y, Liu F. Detecting cars in VHR SAR images via semantic CFAR algorithm [J]. IEEE Geoscience and Remote Sensing Letters, 2016, 13(6): 801-805.
- [6] Kaplan L M. Improved SAR target detection via extended fractal features [J]. IEEE Transactions on Aerospace and Electronic Systems, 2001, 37(2): 436-451.
- [7] 刘冬, 张弓. 基于指数小波分形特征的 SAR 图像特定目标检测 [J]. 西安电子科技大学学报, 2010, 37(2): 366-373.
Liu D, Zhang G. Special target detection of the SAR image via exponential wavelet fractal [J]. Journal of Xidian University, 2010, 37(2): 366-373.
- [8] Sommer L, Schmidt N, Schumann A, et al. Search area reduction Fast-RCNN for fast vehicle detection in large aerial imagery [C]//25th IEEE International Conference on Image Processing (ICIP). IEEE, 2018: 3054-3058.
- [9] Li J W, Qu C W, Peng S J. Ship detection in SAR images based on an improved Faster R-CNN [C]//2017 SAR in Big Data Era: Models, Methods and Applications (BIGSAR DATA). IEEE, 2017.
- [10] Jiang S, Zhu M, He Y, et al. Ship detection with SAR based on YOLO [C]//IGARSS 2020 IEEE International Geoscience and Remote Sensing Symposium. IEEE, 2020.
- [11] Kang M, Ji K, Leng X, et al. Contextual region-based convolutional neural network with multilayer fusion for SAR ship detection [J]. Remote Sensing, 2017, 9(8): 860.
- [12] Wang R, Xu F, Pei J, et al. An improved Faster R-CNN based on MSER decision criterion for SAR image ship detection in harbor [C]//IGARSS 2019 IEEE International Geoscience and Remote Sensing Symposium. IEEE, 2019: 1322-1325.
- [13] An Q, Pan Z X, Liu L, et al. DRBox-v2: An improved detector with rotatable boxes for target detection in SAR images [J]. IEEE Transactions on Geoscience and Remote Sensing, 2019, 57(11): 8333-8349.
- [14] 张柳, 朱卫纲, 吴戌. 全卷积神经网络应用于 SAR 目标检测 [J]. 电讯技术, 2018, 58(11): 1244-1251.
Zhang Y, Zhu W G, Wu X. Target detection based on fully convolutional neural network for SAR images [J]. Telecommunication Engineering, 2018, 58(11): 1244-1251.
- [15] Wang H, Chen Z S, Zheng S C. Preliminary research of low-RCS moving target detection based on Ka-band video SAR [J]. IEEE Geoscience and Remote Sensing Letters, 2017, 14(6): 811-815.
- [16] Zhang Y, Yang S Y, Li H B, et al. Shadow tracking of moving target based on CNN for video SAR system [C]//IGARSS 2018 IEEE International Geoscience and Remote Sensing Symposium. IEEE, 2018: 4399-4402.
- [17] Xu Z H, Zhang Y, Li H B, et al. A new shadow tracking method to locate the moving target in SAR imagery based on KCF [C]//International Conference in Communications, Signal Processing, and Systems. Springer, 2017: 2661-2669.
- [18] Liang Z H, Liang C J, Zhang Y, et al. Tracking of moving target based on SiamMask for video SAR system [C]//2019 IEEE International Conference on Signal, Information and Data Processing. IEEE, 2019: 1-4.
- [19] Zhang Y, Zhu D Y, Yu X, et al. Approach to moving targets shadow detection for video SAR [J]. Journal of Electronics and Information Technology 39(9): 2197-2202.
- [20] Henke D, Magnard C, Frioud M, et al. Moving-target tracking in single-channel wide-beam SAR [J]. IEEE Transactions on Geoscience and Remote Sensing, 2012, 50(11): 4735-4747.
- [21] Henke D, Dominguez E M, Small D, et al. Moving target tracking in single- and multichannel SAR [J]. IEEE Transactions on Geoscience and Remote Sensing, 2015, 53(6): 3146-3159.
- [22] Ding J, Wen L, Zhong C, et al. Video SAR moving target indication using deep neural network [J]. IEEE Transactions on Geoscience and Remote Sensing, 2020, 58(10): 7194-7204.
- [23] Zhao B, Han Y, Wang H, et al. Robust shadow tracking for video SAR [J]. IEEE Geoscience and Remote Sensing Letters, 2021, 18(5): 821-825.
- [24] 何志华, 陈兴, 于春锐, 等. 一种稳健的视频 SAR 动目标阴影检测与跟踪处理方法 [J]. 电子与信息学报, 2022, 44: 1-9.
He Z H, Chen X, Yu C R, et al. A robust moving target shadow detection and tracking method for VideoSAR [J]. Journal of Electronics and Information Technology, 2022, 44: 1-9.
- [25] 刘雨洁, 齐向阳. 基于长时间间隔序贯 SAR 图像的运动舰船跟踪方法 [J]. 中国科学院大学学报, 2021, 38(5): 7.
Liu Y J, Qi X Y. Moving ship tracking method based on long time interval sequential SAR images [J]. Journal of University of Chinese Academy of Sciences, 2021, 38(5): 7.
- [26] 胡瑶. 基于阴影的 SAR 多目标跟踪方法研究 [D]. 成都: 电子科技大学, 2021.
Hu Y. Research on shadow-based SAR multi-target tracking method [D]. Chengdu: University of Electronic Science and Technology of China, 2021.
- [27] Lin T Y, Dollar P, Girshick R, et al. Feature pyramid networks for object detection [C]//2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). IEEE, 2017: 936-944.
- [28] Tan M, Pang R, Le Q V. EfficientDet: Scalable and efficient object

VideoSAR moving target detection and tracking
algorithm based on deep learning

QIU Lei^{1,2}, ZHANG Xuezh², HAO Dawei²

(1. Ordnance Enginnering College, Naval University of Engineering, Wuhan 430033, China; 2. Ordnance
NCO Academy, Army Engineering University of PLA, Wuhan 430075, China)

Abstract: The video synthetic aperture radar (VideoSAR) technology is widely used in military reconnaissance, geological exploration, and disaster prediction, among other fields. Owing to multiple interference factors in SAR videos, such as speckle noise, specular reflection, and overlay effect, moving targets are easily mixed with background or other targets. Therefore, this study proposed an effective VideoSAR target detection and tracking algorithm. Firstly, several features of VideoSAR were extracted to construct multichannel feature maps. Then, deeper features were extracted using the improved lightweight EfficientDet network, thus improving the accuracy of SAR target detection while considering algorithm efficiency. Finally, the trajectory association strategy based on bounding boxes was employed to associate the same target in VideoSAR. The experimental results show that the method proposed in this study is effective for SAR shadow target detection and tracking.

Keywords: VideoSAR; feature enhancement; target detection; deep learning; feature pyramid network; multi – target tracking

(责任编辑: 张 仙)