文章编号:1671-4598(2025)09-0318-08 DOI:10.16526/j.cnki.11-4762/tp.2025.09.038 中图分类号:TP75 文献标识码:A

基于多任务联合学习的多舰船目标跟踪方法

陈 旸[1],刘 宇[2],王 敏[2],常晓宇[2],陈金勇[2]
(1.中国电子科技集团公司 第七研究所,广州 510310;
2.中国电子科技集团公司 第五十四研究所,石家庄 050081)

摘要:利用高分辨率卫星视频数据对大范围海上目标进行实时检测和跟踪,在国防军事、公共安全等领域具有重要应用;然而,现有目标跟踪研究主要集中在自然图像领域,且主要针对单目标跟踪问题;针对海面背景复杂、舰船尺寸多样且部分目标占像素较少、目标间的交叉运动容易出现目标混淆和跟踪丢失等问题,提出了一种基于点表示的目标检测和身份重识别的多任务学习模型,在CenterNet网络中使用多层次特征聚合网络来提取多尺度特征,并以目标中心点的识别和定位来实现目标检测,避免了anchor框和非极大值机制的计算开销;此外,采用目标身份分类和多任务学习损失自适应平衡方法,解决了舰船目标检测和身份重识别任务的特征学习目标冲突问题;通过自建目标跟踪数据集,在24段视频的测试下,多目标跟踪系统的目标检测召回率、目标检测准确率、多目标跟踪召回率、多目标跟踪精确率分别达到88.3%、95.9%、90.2%和98.0%,较好地实现了目标跟踪。

关键词:多舰船目标;目标跟踪;目标检测;卫星视频;多层次特征聚合网络

A Multi-ship Target Tracking Method Based on Multi-task Joint Learning

CHEN Yang[1], LIU Yu[2], WANG Min[2], CHANG Xiaoyu[2], CHEN Jinyong[2]
(1. The 7th Research Institution of CETC, Guangzhou 510310, China;
2. The 54th Research Institution of CETC, Shijiazhuang 050081, China)

Abstract: The video data of high-resolution satellites is used to detect and track large-scale maritime targets in real time, which has important applications in national defense, military, public safety, and other fields. However, existing research on object tracking mainly focuses on the field of natural images and mainly single-object tracking. To address complex sea backgrounds, diverse ship sizes, few pixels for some targets, and intersecting movements between targets that easily lead to target confusion and tracking loss, a multi-task learning mode for object detect and identity re-identification based on the point representation is proposed. The model uses a multi-level feature aggregation network in the CenterNet network to extract multi-scale features, and achieves object detection by identifying and locating the target center point, thus avoiding the computational overhead of an anchor box and non-maximum suppression. In addition, the target identity classification and multi-task learning loss adaptive balancing methods are used to solve the conflict between feature learning targets in detecting ship targets and identity re-identification tasks. By building a target tracking dataset and testing it on 24 videos, the target detection recall rate, target detection accuracy rate, multi-target tracking recall rate, and multi-target tracking accuracy rate of the multi-target tracking system reach up to 88.3%, 95.9%, 90.2%, and 98.0%, respectively, achieving a good target tracking performance.

Keywords: multi-ship target; target tracking; target detection; satellite video; multi-level feature aggregation network

0 引言

作为海洋大国,我国在油气资源开发、岛礁主权归属、海域边界界定及战略通道安全等领域面临的海洋权益争端持续升温,强化海洋综合管控能力已成为国家战略的重要课题。随着天基观测体系在高空间分辨率与高时间分辨率上的跨越式发展,依托高分辨率卫星视频数据构建的动态监测系统,对特定海域进行高价值目标检测和跟踪,具有重要的实用价值[1-2]。

针对目标跟踪,国内外研究者展开了初步研究,但

收稿日期:2025-07-07; **修回日期**:2025-08-01。
作者简介:陈 旸(1971-),男,硕士,高级工程师。
引用格式:陈 旸,刘 宇,王 敏,等.基于多任务联合学习的多舰船目标跟踪方法[J].计算机测量与控制,2025,33(9):318-325.

相较于卫星视频，在地面采集的自然视频数据的获取和标注的难度、成本更低，可用的数据量也更多，这也导致目前多目标跟踪问题的研究工作大量集中在自然视频上。多目标跟踪问题最早由文献［3］在 1979 年提出，将多目标跟踪任务建模为多个检测假设之间的关联问题，采用概率模型进行数据关联[4]，通过求解假设集合上的概率最优解得到跟踪轨迹。随着深度学习模型的发展，越来越多的研究人员将深度学习方法引入目标跟踪研究。文献［5］公开了视觉多目标跟踪 benchmark MOT Challenge 2015，数据集利用 Faster R-CNN 等目标检测算法对视频序列中所有目标进行了标注，供研究人员使用。按照多目标跟踪中目标检测和目标关联两部分处理是否在统一模型中，可以将研究方法分为检测跟踪分离的方法和检测跟踪一体的方法。检测跟踪分离的方法中一部分方法仅使用单一的运行线索进行目标关联。例如，SORT 方法首先依据过往历史航线，使用卡尔曼滤波器预测舰船的未来轨迹，然后再预测后续帧图像中的目标位置，最后使用匈牙利算法融合预测结果和检测结果[6]。为了解决跟踪过程轨迹中断的问题，文献［7］提出了一种运动估计网络，学习轨迹的远程特征，将目标检测结果和轨迹进行关联。另一类方法认为目标的视觉外观线索同等重要，利用目标的外观线索实现跟踪，例如最近的一些工作[8-9]将检测结果提供给 ReID（身份重识别）网络[10]，根据 ReID 特征计算轨迹和检测之间的相似性，最终使用匈牙利算法完成匹配。随着多任务学习方法的快速发展，使用单个深度学习网络同时实现目标检测和跟踪的方法开始成为研究热点。如文献［11］提出了一个 Track-RCNN 方法，通过 3D 卷积扩展了 MaskRCNN，将时间信息包含在网络中，并在 MaskRCNN 顶部添加了一个 ReID 头，通过关联头来实现随时间的目标识别和跟踪。文献［12］提出了一种端到端 Chained-Tracker 方法，将目标检测、特征提取和数据关联融合到一个框架中。

与地面摄像机采集的自然视频相比，视频卫星采集的图像通常具有更大的视野、更多的对象和更复杂的背景。此外，卫星视频的空间分辨率只能达到米级，比自然图像要低得多，所以卫星视频上的目标缺乏详细的目标外观和纹理信息。这给依赖外观特征的方法带来了巨大挑战[13]。在卫星视频目标检测方面，现有的方法通常通过利用运动线索或背景建模来实现视频中的运动目标检测[14-15]。帧差分和背景减法是最常用的框架。帧差分方法通过对连续帧之间的差分设置阈值来识别运动对象[16]。帧差分方法通过检测前景中发生的变化来利用视频中的运动线索，进一步分离前景和背景。基于背景建模的方法首先将背景干扰和噪声表示为背景模型，然后使用更新的模型执行差分[17]。随着深度学习方法在各种视觉任务的突破，一些研究人员开始利用深度学习技术从卫星视频中学习时空线索来提升目标检测和跟踪的性能。文献［18］提出了一个两阶段的时空卷积网络框架，第一阶段在卷积架构中组合运动和外观信息，第二阶段通过热力图给出目标的质心位置。然而，由于缺乏大规模的卫星视频公开数据集，如何将深度学习方法应用至该领域仍然面临巨大挑战。

基于以上分析，本文自建了高分辨率卫星图像舰船跟踪数据集，并提出了一种基于目标中心点的舰船检测和基于目标身份重识别跟踪的联合网络，实现检测和跟踪的一体化模型，一方面降低了跟踪系统的计算量，一方面可以通过多任务学习整体提升检测和跟踪的精度。最后，以舰船目标检测和身份重识别联合模型为核心，结合卡尔曼滤波运动估计方法，综合目标外观特征和空间位置特征，实现多舰船目标实时跟踪。

1　多舰船目标检测与身份重识别多任务联合学习模型

1.1　多尺度深度特征聚合网络

卷积神经网络提取的图像特征随着网络深度的变化高层次语义性逐渐增强、低层次的目标结构特点逐渐减弱，同时，随着网络加深，图像的特征图尺度逐渐下降。对于目标检测任务而言，特征图尺寸会随着卷积而下降，小目标的语义信息会因此损失，检测效果和定位精度也会降低，因此设计多尺度的特征提取网络结构能够提升对多尺度目标的深层语义表达能力。对于身份重识别任务而言，低层次的目标结构特征有助于增加模型对目标身份的判别能力，因此设计高层次语义信息和低层次结构特点相融合的特征提取网络能够有效提升目标身份识别的效果。本文采用多层次特征聚合（DLA，deep layer aggregation）网络提取目标特征[19]。

1.1.1　DLA34-base 网络结构

DLA34 网络结构如图 1 所示，由于网络可以包含许多层和连接，模组化设计可以通过分组与复制来克服网络过于复杂的问题。DLA 核心模块有两个：Iterative Deep Aggregation（IDA）和 Hierarchical Deep Aggregation（HDA）。IDA 模块用于融合相邻两个阶段不同深度的特征。HDA 模块按照分辨率将不同层级的特征划分成不同层级，浅层特征具有更多几何信息，深层特征具有更多语义信息。不同层级的跳跃连接能够融和不同尺度和分辨率的语义信息。

1.1.2　可形变卷积

常规卷积操作是在输入的特征图上直接进行采样加权，如公式（1）所示。可形变卷积（Deformable Convolution）通过在标准卷积上学习到每个卷积点的偏置量，然后对偏置位置的像素点进行卷积，如公式（2）所示：

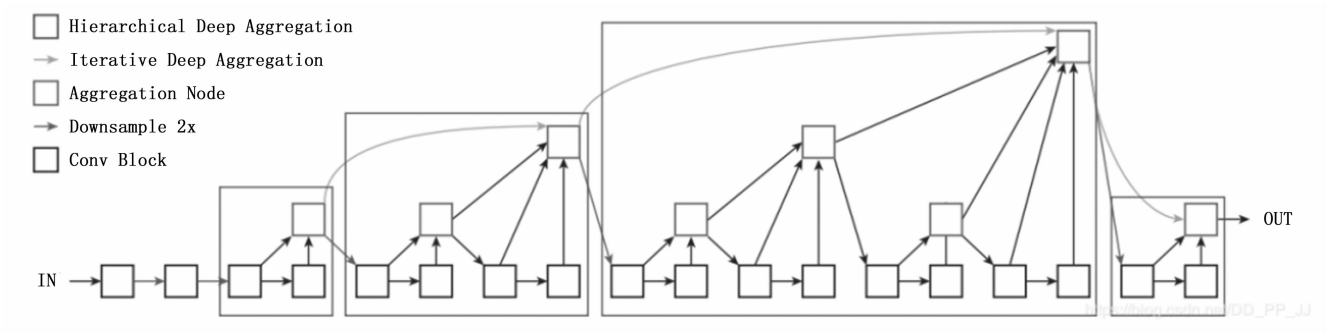

图 1 DLA34-base 网络结构

$$y(p_0)=\sum_{p_n\in R}w(p_n)\cdot x(p_0+p_n)$$

$$R=\{(-1,-1),(-1,0),\cdots,(0,1),(1,1)\} \quad (1)$$

其中：y 表示输出特征图，x 表示输入特征图，p_0 表示像素位置，p_n 表示卷积位置，w 为卷积核权重。

$$y(p_0)=\sum_{p_n\in R}w(p_n)\cdot x(p_0+p_n+\Delta p_n) \quad (2)$$

其中：Δp_n 表示基于 p_n 计算得到的偏置量。

当舰船目标尺寸较小时，标准卷积可能无法聚焦于关键部件，而可形变卷积可以自动使得卷积核向高纹理区域偏移，并以此来提升对小目标的特征表达。

1.1.3 DLA-Seg 网络结构

本文使用的骨干网络是 DLA-Seg，它在 DLA 的基础上使用可形变卷积和 Upsample 层组合进行信息提取，提升了空间分辨率，DLA-Seg 网络结构如图 2 所示。

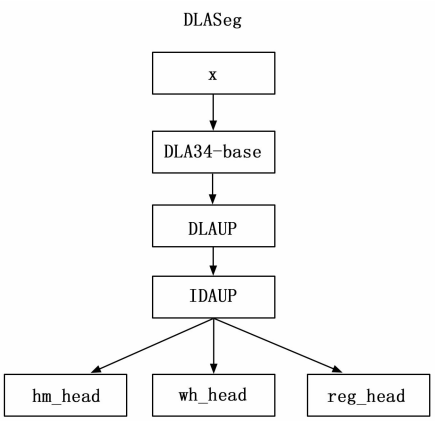

图 2 DLA-Seg 网络结构

DLA-up 对应于图 3 左侧一半的融合模块，IDA-up 对应于右侧一半的融合模块，这两个模块都使用了可形变卷积，并使用反卷积进行上采样。经过融合后的特征保持了输入尺寸 4 倍下采样的分辨率，然后输入到目标检测分支以及多目标关联分支中，针对本文中的小目标舰船有较好的跟踪表现。

1.2 舰船检测网络

检测卫星视频中的舰船目标是指提取出一段视频中每一帧图像内所有舰船目标的位置和大小，以矩形框形

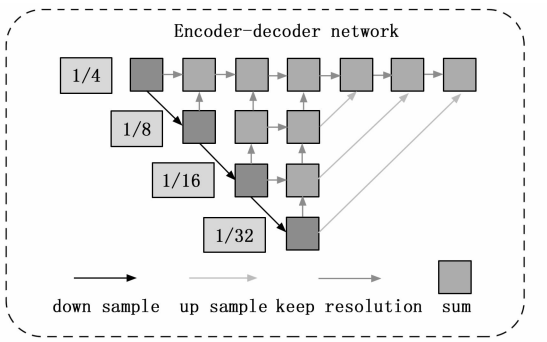

图 3 基于 DLA34base 的 DLA-up 以及 IDA-up 结构图

式表示目标中心位置和大小并输出。卫星视频视角下，舰船目标在视觉外观上有如下特点：(1) 由于舰船目标本身的机动性，目标具有动静之分，动态舰船和静态舰船的自身区域上下文特征有明显区别，相比静态舰船而言，运动舰船目标有较明显的尾迹浪花伴随在船只的末端；(2) 舰船类型涵盖范围广，舰船目标尺寸变化的动态范围较大，从几米长度的游艇到几百米长的航空母舰不等；(3) 由于视频拍摄时舰船目标与卫星镜头间的相对运动造成的运动模糊和薄云雾的遮挡等因素，目标存在模糊和遮挡的问题。

针对当前目标检测算法中非极大值抑制不适用于大尺寸的卫星视频图像问题，本文采用基于点表示的目标检测模型 CenterNet 来实现舰船目标的检测。CenterNet 模型抛弃了 anchor 机制，基于对图像中目标中心点的识别和定位来实现目标检测，避免了密集冗余 anchor 框和 NMS 后处理的计算开销，有效提高目标检测算法的效率。CenterNet 模型架构如图 4 所示。

CenterNet 模型包含两个部分，分别是特征提取模块（使用 DLA 网络）和目标提取模块，特征提取模块提取图像的语义特征，然后将特征图输入到检测模块中，检测模块对特征图进行卷积，分别得到目标位置热力图、目标特征图、目标中心偏移图、目标长宽预测图四张不同意义的图像，位置热力图中响应越大的点表示目标存在的概率越高，在热力图上筛选出高概率的目标位置，然后在目标中心偏移图的对应位置找到目标中心

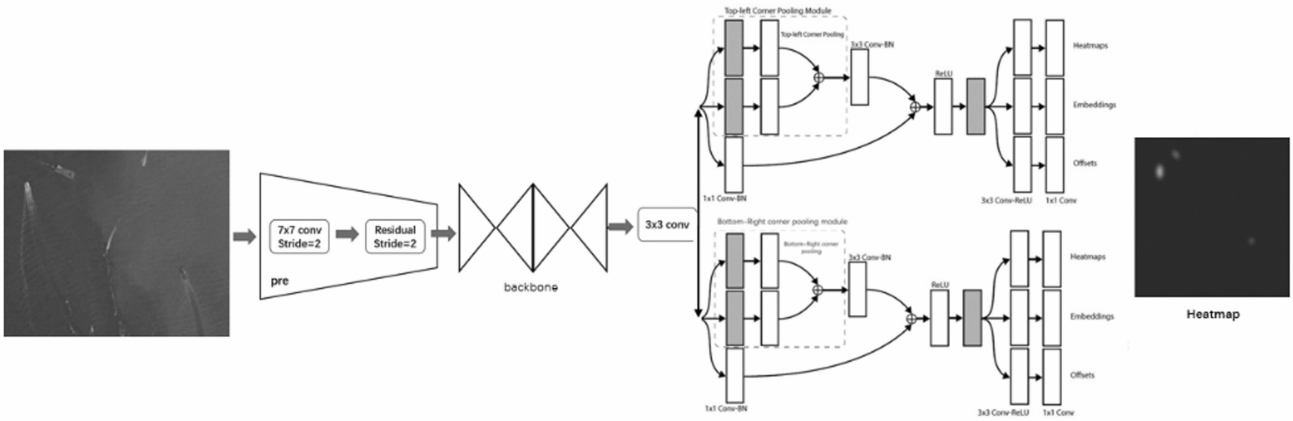

图 4　CenterNet 网络架构

的偏移量，得到精确的目标中心，最后取对应目标位置处目标长宽预测图上的值作为，得到目标的大小，取对应目标位置处的目标特征进行分类，得到舰船目标，排除误检测。获取图像特征图后，需要在目标提取模块网络中对特征图进行进一步的特征表达，得到目标位置热力图、目标特征图、目标中心偏移图和目标长宽预测图。目标位置热力图并不是仅在目标中心响应而其他地方完全抑制，因此，需要提取目标响应图中各响应区域的峰值，以峰值位置作为目标的初始中心位置。具体的做法是将热力图上每个点和其八邻域位置点上的响应值进行比较，如果该点的响应值大于所有八邻域点的响应值，则认为该点为峰值点。考虑到卫星视频图像中可能存在多个目标，取前 200 个峰值点作为目标检测的初始目标位置。假设找到 $P=\{(x_i,y_i)\mid i=1,2,\cdots,200\}$。然后找到每个峰值对应位置处的目标中心偏移图上的中心偏移向量 $O_c=\{(\mathrm{d}x_i,\mathrm{d}y_i)\mid i=1,2,\cdots,200\}$ 和目标长宽预测图上对应的目标长宽值 $S=\{(w_i,h_i)\mid i=1,2,\cdots,200\}$。按如下公式得到目标左上角和目标右下角的坐标：$bbox=\left(x_i+\mathrm{d}x_i-\frac{w_i}{2},y_i+\mathrm{d}y_i-\frac{h_i}{2},x_i+\mathrm{d}x_i+\frac{w_i}{2},y_i+\mathrm{d}y_i+\frac{h_i}{2}\right)$。再将这 200 个候选目标的特征用于分类，得到目标的类别并排除误检测。

1.3　多舰船目标身份重识别网络

本文设计了图 5 所示的舰船目标检测和 ReID 多任务学习模型；检测和 ReID 两个任务在同一个模型中端到端的联合学习，同时实现舰船目标检测和 ReID。模型包含一个 Encoder-Decoder 特征提取器，一个检测分支，一个 ReID 分支。当前帧图像输入特征提取器后，输出图像特征，然后将图像特征输入检测子网络，检测子网络为基于点表示的目标检测网络，输出目标中心热力图和目标中心偏移预测图和目标长宽预测图，并进一步得到目标中心。检测子网络将检测到的目标中心点坐标传入 ReID 子网络，ReID 子网络得到每个目标中心点的特征向量，然后将当前帧所检测到的舰船目标的特征向量与上一帧所检测到的舰船目标的特征向量进行匹配，匹配算法用的是匈牙利算法，经过匹配后，得到当前帧中舰船目标与上一帧中舰船目标的身份对应关系。

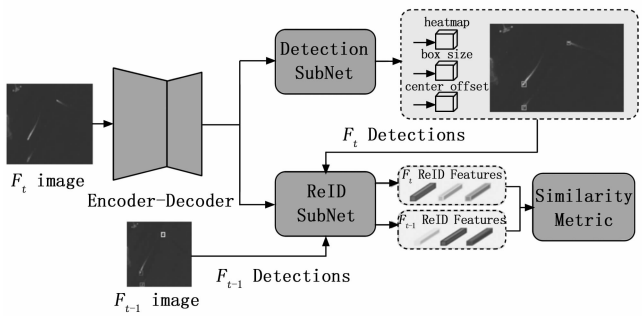

图 5　舰船目标检测与 ReID 多任务学习模型

为了解决多舰船目标检测和跟踪联合框架中检测任务和 ReID 任务学习目标不一致的问题，本文采用多类型损失自适应平衡方法，将两分支的损失比例设计为可学习的；在联合训练的时候，网络能够自动学习到最适合当前多舰船目标跟踪数据的损失比例系数：

$$L_{\mathrm{total}}=\frac{1}{2}\left(\frac{1}{e^{w_1}}L_{\mathrm{det}}+\frac{1}{e^{w_2}}L_{\mathrm{REID}}+w_1+w_2\right)\tag{3}$$

其中：L_{total} 表示网络总损失，L_{det} 和 L_{REID} 分别表示检测分支的损失和 ReID 分支的损失，w_1 和 w_2 是可学习的一组参数，可以调和检测分支损失和 ReID 分支损失的权重；将权重设计为指数形式，是避免负权重破坏优化过程；线性项则可以防止权重无限增大。初始训练时，w_1 和 w_2 均设置为 0，此时 L_{det} 和 L_{REID} 的权重均为 1，即检测与 ReID 损失权重相等。训练过程中两者的更新梯度如下式所示：

$$\begin{cases}\dfrac{\partial L_{\mathrm{total}}}{\partial w_1}=-e^{-w_1}L_{\mathrm{det}}+1\\[2ex]\dfrac{\partial L_{\mathrm{total}}}{\partial w_2}=-e^{-w_2}L_{\mathrm{REID}}+1\end{cases}\tag{4}$$

当某个分支的损失较大时，会使得对应权重的梯度下降，进一步使得权重上升，即提升该任务的重要性；反之，则权重降低。

1.3.1 检测分支损失函数 L_{det}

检测分支输出的是目标中心热力图、目标中心偏移向量图、目标长宽预测图，因此对应的有目标中心热力图损失，目标中心偏移损失和目标长宽预测损失三个损失值。检测分支的损失函数公式如下：

$$L_{\text{det}} = L_{\text{heat}} + \lambda_{\text{size}} L_{\text{size}} + \lambda_{\text{off}} L_{\text{off}} \tag{5}$$

其中：L_{heat}表示热力图预测损失，L_{size}和 L_{off}分别表示目标尺寸预测损失和目标中心位置偏移预测损失，λ_{size}和 λ_{off}用于调和不同损失值之间的权重。

目标热力图损失函数如下：

$$L_{\text{heat}} = \frac{-1}{N} \sum_{xyc} \begin{cases} (1-\hat{Y}_{xy})^{\alpha} \log(\hat{Y}_{xy}) & \text{if} \quad \hat{Y}_{xy} = 1 \\ (1-\hat{Y}_{xy})^{\beta} (\hat{Y}_{xy})^{\alpha} \log(1-\hat{Y}_{xy}) & \text{otherwise} \end{cases} \tag{6}$$

$$\hat{Y}_{xy} = \exp\left(-\frac{(x-\tilde{p}_x)^2 + (y-\tilde{p}_y)^2}{2\sigma_p^2}\right) \tag{7}$$

其中：σ 和 β 是控制正负样本损失比例的参数（采用默认值，分别为 2 和 4），对于易分样本而言，loss 较小，对于难分样本而言，loss 较大，相当于增加了其训练权重。N 是样本总数量，$\hat{Y}_{xy}$ 是以目标中心为中心，标准差为 σ 的二元高斯分布，$\tilde{p}_x$ 和 $\tilde{p}_y$ 分别表示目标中心横纵坐标。

1.3.2 ReID 分支损失函数 L_{ReID}

为了训练 ReID 分支区分不同目标个体的能力，设计了以目标 ID 为类别的分类任务，提高 ReID 特征的可判别性。分类训练的损失函数如下公式所示：

$$L_{\text{ReID}} = -\sum_{i=1}^{C} \sum_{k=1}^{K} L^{i}(k) \log(p(k)) \tag{8}$$

其中：C 表示总目标 ID 数，K 表示训练数据中所有的舰船目标数量，$L^{i}(k)$ 表示第 k 个目标属于第 i 个 ID 的概率，属于为 1，不属于为 0。

1.4 多舰船目标跟踪和轨迹生成

舰船目标检测和 ReID 联合的网络模型只能获取前后帧中的舰船目标位置、大小及其匹配关系。为了得到目标在整个视频序列上的完整轨迹，本文设计了如图 6 所示的多舰船目标跟踪系统。

输入第一帧视频图像时，检测图像中的目标，得到目标中心位置，为每个目标分配一个 ID 号，放入活跃状态轨迹库中；输入后续帧时，检测当前帧图像，得到图像中目标位置和 ReID 特征后，进行目标和轨迹的两阶段匹配。第一阶段使用卡尔曼滤波器估计当前帧中目标的位置，并计算预测的目标位置与检测到的目标的马氏距离；计算上一帧图像和当前帧图像舰船目标的

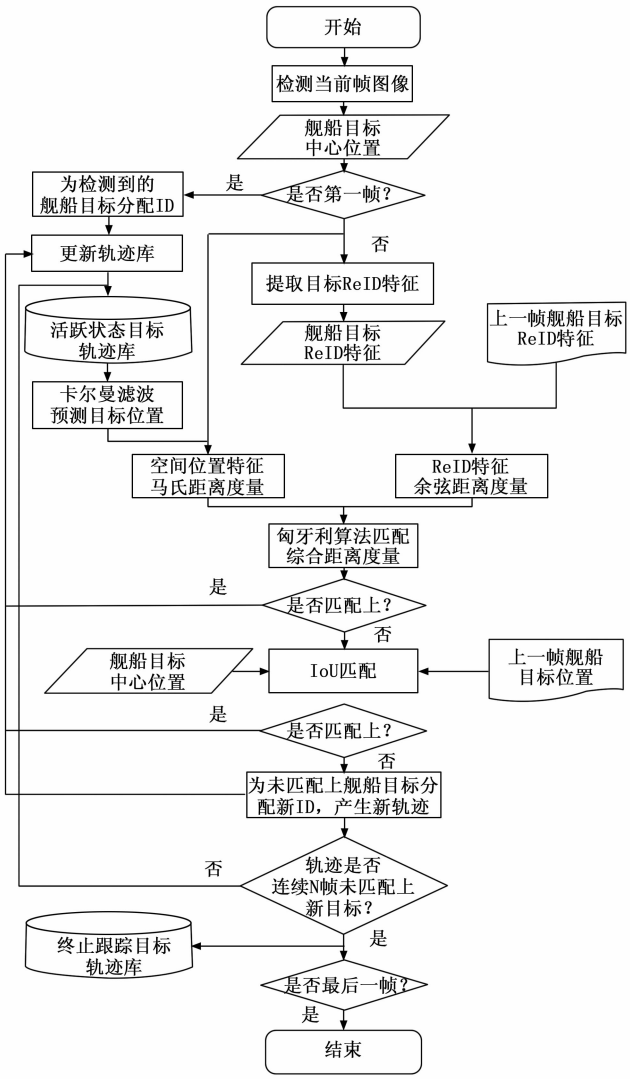

图 6 多舰船目标跟踪流程

ReID 特征余弦距离；然后利用匈牙利匹配算法综合计算位置特征距离和 ReID 距离，以匹配舰船目标。第二阶段，对第一阶段未匹配上的目标采用基于 IoU 距离的匹配；当 F_{t-1} 帧图像因卫星镜头出现抖动目标位置发生较大偏差时，卡尔曼滤波需要一段时间重新收敛，F_t 帧图像的目标位置与 F_{t-1} 帧轨迹的空间位置特征相似性程度不高，若此时 ReID 特征匹配失效，将导致目标跟丢。因此本文加入基于两帧间目标框 IoU 匹配对这种情况进行修正。将前后两帧目标检测框的 IoU 值作为距离度量，采用线性匹配算法得到目标和轨迹的匹配得分。对于在两个阶段均未匹配上的目标，为其分配新的 ID，加入活跃状态轨迹库中。对于在两个阶段均未匹配上的轨迹，判断其是否连续 N 帧未匹配上，如果是，则终止对该轨迹的跟踪，将其从活跃状态轨迹库中删除，放入终止跟踪轨迹库中。如果否，则将该轨迹保留在活跃状态轨迹库中。

2　实验结果

2.1　实验数据

2.1.1　舰船目标检测样本库

在舰船检测数据集中，使用 DOTA 数据集、HRSC 数据集、Airbus Ship 数据集进行扩充，并使用传统方法对图像进行预处理以增强样本库，比如：旋转、缩放、光照变化、模糊处理等等。

2.1.2　目标跟踪样本库

使用第五届“中科星图杯”国际高分遥感图像解译大赛中的目标跟踪数据集。数据源为吉林一号视频卫星，空间分辨率为 1 米，帧率为 10 帧/秒，视频尺寸为 1 920×1 080 像素，共标注了 14 个有效视频片段。

此外，本文还收集了 9 段 12 000×5 000 大小的吉林一号卫星视频数据；由于视频视场大，直接标注难以准确标注目标的边缘。因此对其进行分区标注，将视频裁切成 2 000×1 000 大小的视频片段进行标注，共标注了 83 个有效视频片段。

标注工具为 Darklabel 多目标跟踪标注软件。标注每个舰船目标的编号信息，以此区分视频中多个同类目标，得到贯穿整个视频序列的目标运动轨迹。

2.2　评价指标

2.2.1　目标检测召回率与准确率

在检测目标框与真实目标框的 IoU 达到 0.5 时，认为检测目标框为正确检测目标框，否则，视为错误检测目标框。正确检测目标框的数量计为 D_p，错误检测目标框的数量计为 D_n，测试数据中真实标注的目标框数量计为 D_{gt}。

目标检测召回率计算公式为：

$$recall_{\text{det}} = \frac{D_p}{D_{gt}} \tag{9}$$

目标检测准确率计算公式为：

$$accuracy_{\text{det}} = \frac{D_p}{D_p + D_n} \tag{10}$$

2.2.2　多目标跟踪召回率与准确率

当跟踪目标框与真实目标框的 IoU 达到 0.5 以上且跟踪目标框的 ID 与真实目标框的 ID 相同时，认为跟踪目标框为正确跟踪目标框，否则，视为错误跟踪目标框。正确跟踪目标框的数量计为 T_p，错误跟踪目标框的数量计为 T_n，测试数据中真实标注的目标跟踪框数量计为 T_{gt}。

多目标跟踪召回率计算公式为：

$$recall_{\text{mot}} = \frac{T_p}{T_{gt}} \tag{11}$$

多目标跟踪准确率计算公式为：

$$accuracy_{\text{mot}} = \frac{T_p}{T_p + T_n} \tag{12}$$

2.3　实验结果分析

选取共 24 段视频测试，视频中包含多尺度舰船、海浪、云雾遮挡、相似舰船等复杂场景，测试结果如表 1 所示，本文多目标跟踪系统的目标检测召回率、目标检测准确率、多目标跟踪召回率、多目标跟踪精确率分别为 88.3%、95.9%、90.2%和 98.0%。

表 1　多舰船目标跟踪测试结果

视频	跟踪准确率	跟踪召回率	检测准确率	检测召回率
FS3-4	0.940	0.897	0.955	0.911
FS5-2	0.901	0.928	0.960	0.988
FS3-6	0.694	1.000	0.694	1.000
FS4-2	0.833	0.906	0.918	0.998
NFK1-3	0.833	1.000	0.833	1.000
NFK2-2	0.964	0.970	0.964	0.970
NFK5-3	0.981	0.992	0.981	0.992
HXH5-5	0.948	0.919	0.983	0.953
HXH5-6	0.987	1.000	0.987	1.000
HXH5-4	1.000	0.997	1.000	0.997
PCMS4-3	0.920	1.000	0.920	1.000
PCMS5-4	0.974	0.996	0.976	0.998
PCMS2-3	0.965	1.000	0.965	1.000
BLMD2-4	0.999	1.000	0.999	1.000
SDG4-3	0.679	0.787	0.852	0.988
SDG5-3	0.527	0.626	0.611	0.726
XN1-6	0.601	1.000	0.601	1.000
XN2-6	0.828	0.999	0.828	0.999
ZY2-4	0.925	1.000	0.925	1.000
ZY3-6	0.936	0.996	0.936	0.996
LY2-3	0.958	1.000	0.958	1.000
LY3-2	0.824	1.000	0.824	1.000
LY3-3	0.992	1.000	0.992	1.000
LY1-3	0.980	0.999	0.980	0.999
平均	0.883	0.959	0.902	0.980

不同卫星视频的舰船跟踪结果如图 7 所示。不同的框代表不同舰船连续帧形成的轨迹。图 7（a）中较好地跟踪到了 4 艘船的运动轨迹，(b) 中同样较好地跟踪到了 4 艘船的运动轨迹。

2.4　对比实验

在对比方法方面选择在多舰船目标跟踪方面有较好表现的 ByteTrack 算法，其中目标检测算法使用 YOLOX。

ByteTrack 在目标检测召回率、目标检测准确率、多目标跟踪召回率、多目标跟踪准确率四个指标上均低于本文算法；其中目标检测准确率和召回率分别低于本文算法 1.98%和 2.86%，而多目标跟踪准确率和召回率分别低于本文算法 2.98%和 4.50%。

2.5　消融实验

为了验证不同组件对于目标检测或目标跟踪的有效

(a)

(b)

图 7 多舰船目标跟踪结果

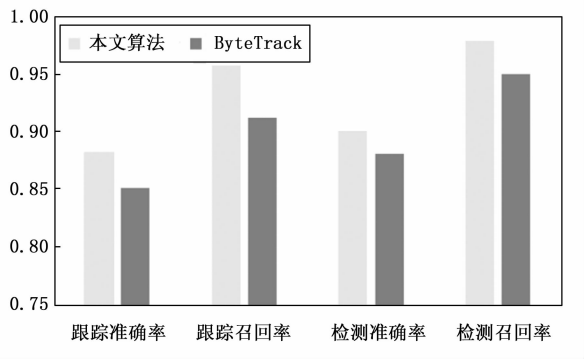

图 8 不同算法的多舰船目标跟踪对比

性，分别在使用 ResNet50 作为骨干网络、使用传统卷积、目标检测和 REID 分支损失权重为 1∶1 三种情况下分别验证 DLA 网络、可变形卷积和多任务学习损失自适应平衡方法的有效性。

由图 9 可知，骨干网络的选择对于目标检测和跟踪影响较大，当选择 ResNet50 作为骨干网络时，目标的检测准确率和跟踪准确率分别较本文算法降低了 3.5% 和 4.8%。可变形卷积也一定程度上提高了目标检测和跟踪效果，但效果没有骨干网络和多任务学习损失自适应平衡方法显著。多任务学习损失自适应平衡方法对于跟踪准确率、跟踪召回率、检测准确率、检测召回率四项指标分别提升了 0.2%、1.3%、1.6%和 1%。

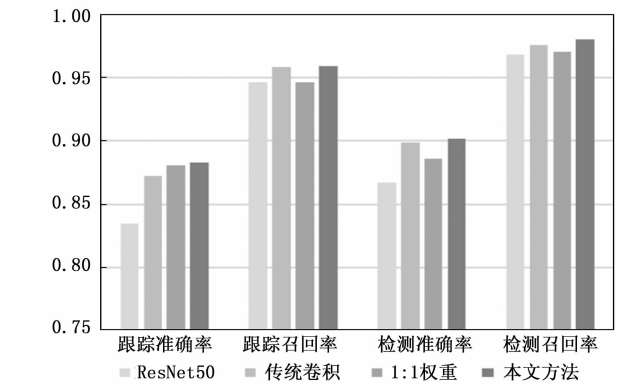

图 9 消融实验对比

3 结束语

本文面向卫星视频的复杂背景下多舰船目标检测和跟踪需求，提出了一个基于点表示的目标检测和身份重识别的多任务学习模型。该模型使用多尺度深度特征聚合网络来实现感知不同尺寸的舰船目标的特征，并使用可形变卷积来聚焦舰船目标的关键特征，以提升舰船目标的特征感知；此外，CenterNet 避免使用 anchor 机制，而是直接将目标抽象为点，从而更好地适应大尺寸图像。为了自动平衡目标检测和跟踪两个任务，提出了多类型损失的自适应平衡方法，实现了两个任务的联合优化，同时实现目标检测与跟踪。最后，自建了包含 97 段舰船视频的数据集，通过 24 段视频的测试，多目标跟踪系统的目标检测召回率、目标检测准确率、多目标跟踪召回率、多目标跟踪精确率分别达到 88.3%、95.9%、90.2%和 98.0%。该方法在军事、海洋资源保护、港口管理等方向有重要应用前景。

参考文献：

[1] 刘 韬. 静止轨道高分辨率光学成像卫星展示海洋监视的重大潜力 [J]. 卫星应用，2014 (12)：70－74.

[2] 周越冬. 基于深度学习的遥感图像舰船多目标跟踪方法研究 [D]. 西安：西安电子科技大学，2021.

[3] REID D. An algorithm for tracking multiple targets [J]. IEEE Transactions on Automatic Control，1979，24 (6)：843－854.

[4] REZATOFIGHI S H，MILAN A，ZHANG Z，et al. Joint probabilistic data association revisited [C] // 2015 IEEE International Conference on Computer Vision (ICCV)，2015：3047－3055.

[5] LEAL-TAIXé L，MILAN A，REID I，et al. MOTChallenge 2015：towards a benchmark for multi-target tracking [J]. ArXiv E-Prints，2015：ArXiv：1504.01942.

[6] BEWLEY A，GE Z，OTT L，et al. Simple online and realtime tracking [C] // 2016 IEEE International Conference on Image Processing (ICIP)，2016：3464－3468.

[7] ZHANG Y，SHENG H，WU Y，et al. Long-term tracking with deep tracklet association [J]. IEEE Transactions on Image Processing，2020，29：6694－6706.

[8] YU F，LI W，LI Q，et al. POI：multiple object tracking with high performance detection and appearance feature [C] // Hua G，Jégou H. Computer Vision-ECCV 2016 Workshops. Cham：Springer International Publishing，2016：36－42.

[9] ZHOU Z，XING J，ZHANG M，et al. Online multi-target

tracking with tensor-based high-order graph matching [C] // 2018 24th International Conference on Pattern Recognition (ICPR), 2018: 1809 - 1814.

[10] LUO H, GU Y, LIAO X, et al. Bag of tricks and a strong baseline for deep person re-identification [C] // 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), 2019: 1487 - 1495.

[11] VOIGTLAENDER P, KRAUSE M, OSEP A, et al. MOTS: multi-object tracking and segmentation [C] // 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2019: 7934 - 7943.

[12] PENG J, WANG C, WAN F, et al. Chained-Tracker: chaining paired attentive regression results for end-to-end joint multiple-object detection and tracking [C] // Computer Vision-ECCV 2020: 16th European Conference, Glasgow, UK, August 23 - 28, 2020, Proceedings, Part IV. Berlin, Heidelberg: Springer-Verlag, 2020: 145 - 161.

[13] 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)：642 - 648.

[14] ZHANG J, JIA X. Improved low rank plus structured sparsity and unstructured sparsity decomposition for moving object detection in satellite videos [C] // IGARSS 2019 - 2019 IEEE International Geoscience and Remote Sensing Symposium, 2019: 5421 - 5424.

[15] ZHANG J, JIA X, HU J. Error bounded foreground and background modeling for moving object detection in satellite videos [J]. IEEE Transactions on Geoscience and Remote Sensing, 2020, 58 (4): 2659 - 2669.

[16] SALEEMI I, SHAH M. Multiframe many-many point correspondence for vehicle tracking in high density wide area aerial videos [J]. International Journal of Computer Vision, 2013, 104 (2): 198 - 219.

[17] ZHOU X, YANG C, YU W. Moving object detection by detecting contiguous outliers in the low-rank representation [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2013, 35 (3): 597 - 610.

[18] LALONDE R, ZHANG D, SHAH M. ClusterNet: detecting small objects in large scenes by exploiting spatio-temporal information [C] // 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2018: 4003 - 4012.

[19] YU F, WANG D, SHELHAMER E, et al. Deep layer aggregation [C] // 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2018: 2403 - 2412.

(上接第 317 页)

[17] POTLAPALLI V, ZAMIR S W, KHAN S H, et al. Promptir: Prompting for all-in-one image restoration [J]. Advances in Neural Information Processing Systems, 2023, 36: 71275 - 71293.

[18] SIFRE L, MALLAT S. Rigid-motion scattering for texture classification [J]. ArXiv Preprint ArXiv: 1403.1687, 2014.

[19] SHEN Z, WANG W, LU X, et al. Human-aware motion deblurring [C] //Proceedings of the IEEE/CVF International Conference on Computer Vision, 2019: 5572 - 5581.

[20] RIM J, LEE H, WON J, et al. Real-world blur dataset for learning and benchmarking deblurring algorithms [C] //Computer Vision-ECCV 2020: 16th European Conference, Glasgow, UK, August 23 - 28, 2020, Proceedings, Part XXV 16. Springer International Publishing, 2020: 184 - 201.

[21] GAO H, TAO X, SHEN X, et al. Dynamic scene deblurring with parameter selective sharing and nested skip connections [C] //Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2019: 3848 - 3856.

[22] SUIN M, PUROHIT K, RAJAGOPALAN A N. Spatially-attentive patch-hierarchical network for adaptive motion deblurring [C] //Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2020: 3606 - 3615.

[23] ZHANG K, LUO W, ZHONG Y, et al. Deblurring by realistic blurring [C] //Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2020: 2737 - 2746.

[24] XU Y, ZHU Y, QUAN Y, et al. Attentive deep network for blind motion deblurring on dynamic scenes [J]. Computer Vision and Image Understanding, 2021, 205: 103169.

[25] PARK D, KANG D U, KIM J, et al. Multi-temporal recurrent neural networks for progressive non-uniform single image deblurring with incremental temporal training [C] //European Conference on Computer Vision, Cham: Springer International Publishing, 2020: 327 - 343.

[26] KIM K, LEE S, CHO S. Mssnet: Multi-scale-stage network for single image deblurring [C] //European Conference on Computer Vision, Cham: Springer Nature Switzerland, 2022: 524 - 539.

[27] ZHANG B, SUN J, SUN F, et al. Image deblurring method based on self-attention and residual wavelet transform [J]. Expert Systems with Applications, 2024, 244: 123005.