

DOI: 10.19650/j.cnki.cjsi.J2513920

一种基于 ByteTrack 的前视声呐多目标跟踪算法 *

陈美龙^{1,2}, 赵新华¹, 叶秀芬¹

(1. 哈尔滨工程大学智能科学与工程学院 哈尔滨 150001; 2. 北京三快在线科技有限公司 北京 100190)

摘要:前视声呐在水下远距离目标检测与跟踪中发挥着重要的作用。然而,前视声呐图像序列帧率较低、目标特征不清晰,容易出现目标丢失的问题。在使用前视声呐进行目标跟踪时,需要对声呐载体旋转和目标遮挡进行补偿,避免目标丢失。为解决以上问题,结合声呐图像序列的特征和目标特征,改进现有的跟踪算法。针对 ByteTrack 算法应用在声呐跟踪上容易出现目标丢失问题,结合前视声呐图像特征,改进关联方式,在第 1 关联提出了一种基于卡尔曼滤波的运动特征和目标外观特征结合的方式作相似性度量,提升了跟踪的准确性。针对前视声呐载体旋转导致目标运动过快的问题,利用声呐姿态数据对 ByteTrack 算法加入旋转补偿,提升了匹配的准确性;最后,通过相似性度量算法对比实验,证明了改进后的关联方式和目标外观特征结合的方式的优越性。对比了 DeepSort、TransTrack 和 ByteTrack 主流目标跟踪算法,改进后的模型跟踪准确度为 76.8%,跟踪召回率为 80.6%;改进后的 ByteTrack 与改进前的 ByteTrack 相比,跟踪精度提升了 9.4%,召回率提升了 10.8%,ID 切换次数降低了 46%。检测与跟踪融合实验表明,改进后的目标检测跟踪融合算法拥有更低的漏检率、误检率,更低的身份切换次数,更能适应前视声呐水下目标的检测和跟踪场景。

关键词:前视声呐;目标跟踪;ByteTrack;图像序列

中图分类号: TH741 TP273 文献标识码: A 国家标准学科分类代码: 510.40

A multi-target tracking algorithm based on ByteTrack for forward-looking sonar

Chen Meilong^{1,2}, Zhao Xinhua¹, Ye Xiufen¹

(1. College of Intelligent Systems Science and Engineering, Harbin Engineering University, Harbin 150001, China;
2. Beijing Sankuai Online Technology Co., Ltd., Beijing 100190, China)

Abstract:Forward looking sonar plays an important role in underwater long-distance target detection and tracking. However, forward-looking sonar image sequences suffer from low frame rates and unclear target features, which can lead to target loss. In addition, effective tracking requires compensating for sonar carrier rotation and handling target occlusion to prevent trajectory discontinuities. To solve the above problems, the characteristics of sonar image sequences and target features were combined to improve existing tracking algorithms in this paper. To mitigate target loss in sonar tracking of ByteTrack algorithm, combined with the features of forward-looking sonar images, an improved correlation method was proposed. In the first correlation, a similarity measurement method based on Kalman filter was proposed by combining motion features and target appearance features, which improved the accuracy of tracking. To address rapid apparent target motion caused by sonar carrier rotation, rotation compensation was added to the ByteTrack algorithm using sonar attitude data to improve the accuracy of matching; Finally, the superiority of the improved association method and the combination of target appearance features were demonstrated through comparative experiments using similarity measurement algorithms. Compared with mainstream target tracking algorithms such as DeepSort, TransTrack, and ByteTrack, the improved model achieved a tracking accuracy of 76.8% and a tracking recall rate of 80.6%. Compared with the original ByteTrack algorithm, the improved ByteTrack has improved tracking accuracy by 9.4%, recall by 10.8%, and ID switching frequency by 46%. The fusion experimental results of detection and tracking show that the improved target detection tracking fusion algorithm has lower miss rate, lower false alarm rate, lower identity

switching times, making it well-suited for underwater target detection and tracking with forward-looking sonar.

Keywords: forward-looking sonar; target tracking; ByteTrack; image sequence

0 引言

前视声呐图像在完成目标检测后,将每帧图像检测结果关联,从而实现目标跟踪,多目标跟踪需要结合目标的位置和类别信息,为每个目标分配身份编号。然而,多目标跟踪方法是为无人驾驶领域的光学图像而设计的,当声呐载体旋转时,直接采用这些算法对前视声呐目标进行跟踪,由于声呐远端的目标运动速度快,目标检测置信度低,将导致身份编号切换频繁的问题。

传统的目标跟踪算法基于第1帧中已知的目标框位置,对后续帧中的目标进行定位,这种跟踪方式难以处理复杂运动问题且需要根据不同场景调参,跟踪效果不理想^[1]。光流法利用像素级别的运动信息进行跟踪,卡尔曼滤波和粒子滤波则通过状态估计和预测实现目标位置的跟踪^[2]。均值漂移方法基于目标颜色直方图的密度估计进行跟踪,适用于目标外观变化较大的情况。随着目标跟踪技术的发展,核相关滤波成为一种常用的技术,其结合了传统方法和核技巧,利用相关滤波器在处理跟踪速度和准确率方面取得了显著进展。Arthanari等^[3]提出一种自适应时域调整目标属性感知背景抑制的深度校正滤波器,改善跟踪性能。逯万春等^[4]提出一种能够识别高速列车轮轨接触几何状态的核相关滤波-哈希-模板匹配算法,用于复杂运营环境下高速列车车轮横向异常晃动的监测。Huang等^[5]也提出了有效的解决方案,如结构化支持向量机,以应对多峰值检测的复杂的目標遮挡情况。Zhao等^[6]为解决线性模型漂移问题,提出了增量核主成分分析的卡尔曼跟踪算法,在长时间跟踪场景中更稳定。

随着深度学习技术在目标跟踪领域的广泛应用,涌现出大量基于深度学习的多目标跟踪算法,这些算法可大致分为2种。一种为端到端的网络,直接得到检测框位置以及目标关联信息,另一种为基于目标检测的跟踪。端到端的算法能够直接输入视频,得到位置和关联情况,然而这种算法标注复杂,训练成本高,且不适用于数据量匮乏的任务。基于检测的跟踪算法,将检测和目标关联解耦,在获取目标的检测结果后,使用跟踪器来追踪目标,并通过关联算法来连接不同帧间的目标,从而生成目标的运动轨迹线。基于检测的跟踪方法可以方便地对检测和关联算法优化,成为多目标跟踪领域的主流。Bewley等^[7]提出了简单在线实时跟踪(simple online and realtime tracking, SORT)算法,首次结合卡尔曼滤波与匈牙利匹配的方法框架,有效地关联目标并增强了算法的

实时性,但对目标遮挡问题处理效果不理想。Kapania等^[8]提出了DeepSORT算法,在SORT算法基础上引入级联匹配,提升了目标遮挡情况下的鲁棒性。最近几年随着Transformer的火热,基于Transformer的多目标跟踪算法相继问世。Cui等^[9]简化了之前跟踪算法的跟踪步骤,通过Transformer模型将特征提取、目标信息融合和目标关联过程统一起来。Zhou等^[10]提出了Transformer全局帧关联跟踪算法。Ma等^[11]提出将单目标与多目标联合的多任务Transformer跟踪算法,推动跟踪领域朝着更加多元化的发展。Zhang等^[12]提出了一种简单有效的数据关联方法,并不是简单的去掉低分检测结果,而是在二次关联中利用它。Huang等^[13]提出了一种迭代扩展的交并比(intersection over union, IoU)和深度特征关联方法,用于特定场景中的多目标跟踪,效果虽好,但实时性较差。

卷积神经网络在前视声呐目标跟踪领域也展露出其强大能力。利用卷积神经网络进行特征提取,提高目标识别和跟踪的鲁棒性,相较传统方法,更准确、更易扩展且对环境变化更敏感^[14]。Ye等^[15]设计了一种相位估计旋转角度的目标跟踪算法,补偿跟踪模型,增加了跟踪的精度,但来自旋转估计的补偿并不是真实的旋转补偿,并不能取得最佳效果。Zhang等^[16]提出了基于高斯粒子滤波器的跟踪方法,解决在杂乱的环境中多目标跟踪问题。崔杰等^[17]基于Mean Shift算法改进,提出了一种自适应目标框能够随着物体运动自动地调整大小。陈美龙等^[18]基于YOLOv8(you only look once version8)算法改进,提出了一种前视声呐图像目标检测算法,有效提升了声纳图像的目标检测精度。骈志康^[19]为提升模型的泛化能力,使用生成对抗网络做扩增,改进了孪生网络的注意力机制,增强了分类回归网络,实现更准确的目标跟踪,虽然速度快,但对多目标跟踪效果不理想。王文胜等^[20]针对无人机目标小、目标不显著的情况以及目标被遮挡后的再跟踪问题,提出一种将改进YOLO和改进KCF模型融合的无人机识别跟踪算法。霍旭等^[21]针对现有多目标跟踪算法关联过程中,外观和几何信息利用不充分,同时跟踪对象的邻域间信息交互不足的问题,提出了一种基于多阶段关联的多目标跟踪算法。

目前主流的多目标跟踪方法多数是针对光学图像的,采用这些算法对前视声呐目标进行跟踪时,因为声呐图像序列帧率较低、目标特征不清晰,容易出现目标丢失问题。声呐载体旋转时,声呐远端的目标运动速度快,导致跟踪不精准、身份编号切换频繁的问题。因此,需要结合声呐图像序列的特征和目标特征,改进现有的跟踪算

法。本研究将优化基于检测的多目标跟踪方法,以实现声呐多目标的跟踪。针对 ByteTrack 不能应对声呐载体旋转和目标遮挡等导致跟踪效果不好的问题,分别提出优化方法。针对前视声呐载体旋转导致远端目标运动过快问题,利用声呐姿态数据对 ByteTrack 算法加入旋转补偿,提升了跟踪的准确性。针对 ByteTrack 算法应用在声呐跟踪上容易出现目标丢失问题,根据前视声呐图像特征,改进关联方式,在第 1 关联提出了一种基于卡尔曼滤波的运动特征和基于 Re-ID(re-identification)目标外观特征结合的方式作相似性度量,提升了跟踪精度。最后,通过与先进跟踪算法的对比试验,证明了本文提出方法的有效性。

1 相关工作

DeepSORT 是一种结合深度学习和卡尔曼滤波的多目标跟踪算法,TransTrack 是一种基于 Transformer 架构的多目标跟踪算法。

1.1 DeepSORT

DeepSORT 是基于深度学习的二阶段多目标跟踪算法,用于视频实时的多目标跟踪^[22-23]。其是基于 SORT 目标跟踪的改进,结合卷积神经网络提取目标外观特征,采用递归的卡尔曼滤波预测运动特征,解决了 SORT 存在的目标遮挡和 ID 频繁跳变问题。DeepSORT 算法流程如图 1 所示,神经网络检测到目标后,将目标检测框与预测框通过 IoU 匹配,分为未匹配的检测框、匹配的检测框、未匹配的轨迹线 3 种状态。

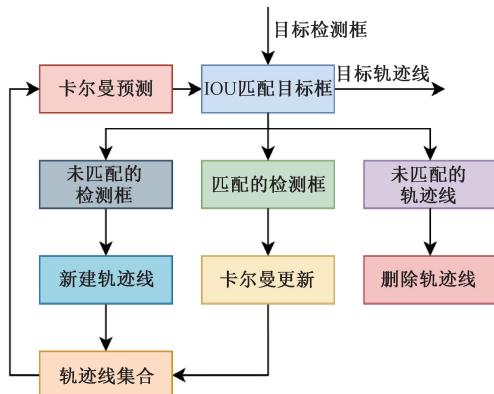


图 1 DeepSORT 算法流程

Fig. 1 DeepSORT algorithm process

未匹配的检测框即当前目标位置未存在轨迹,判定为需要跟踪的新目标,所以要新建轨迹线,当连续 3 帧成功匹配新轨迹线时,再加入轨迹线组,确认为正式轨迹线。成功匹配的检测框,则需要更新卡尔曼预测轨迹算法,便于预测下一帧轨迹线目标位置。存在未匹配成功

的轨迹线时,且在 3 帧内没有匹配上,则判定目标消失,需要从轨迹线组中删除。

1.2 TransTrack

TransTrack 是第 1 次将 Transformer 应用到多目标跟踪领域的算法,TransTrack 是基于 Query-Key 机制的多目标跟踪框架^[24]。Query-Key 的机制是基于 Transformer 设计的,利用 Query-Key 机制,并将学习到的目标查询引入特征提取网络中,以检测新的对象。这种机制在单目标跟踪孪生网络中取得了显著效果,将目标作为查询图像区域的关键,这与传统的联合分布估计范式框架不同,利用 Query-Key 的匹配关系来跟踪当前帧和已存在的目标。

TransTrack 算法的主要步骤如图 2 所示,先用同一个主干网络获取图片的特征,然后将特征分别输入检测分支和跟踪分支,前者经过解码之后变为当前帧的检测框,后者经过解码之后变为跟踪框,分别得到检测结果和跟踪结果,最后将这两个框进行 IoU 匹配,匹配失败的框会作为新目标被添加。总之,TransTrack 算法需要图像进行高效的特征提取,再根据准确的目标检测和匹配算法来保证跟踪的准确性和稳定性。

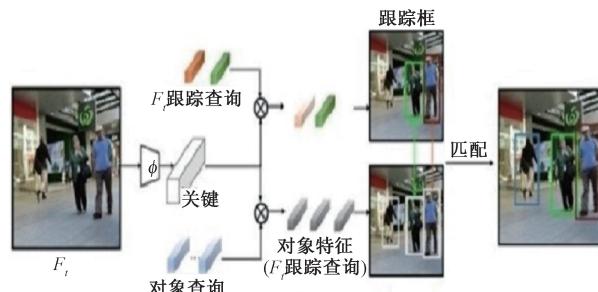


图 2 TransTrack 框架

Fig. 2 Frame of TransTrack

2 本文的方法

2.1 ByteTrack 算法

ByteTrack 是一种先进的二阶段多目标跟踪算法,结合了目标检测器输出的边界框信息,通过卡尔曼滤波器进行目标位置预测,利用匈牙利算法实现多目标跟踪^[12]。它致力于实现目标与轨迹之间的精准匹配,在处理低置信度情况下的二次匹配做出创新,低置信度目标框尽管特征不明显,直接丢弃仍会对跟踪性能造成影响。该算法解决运动模糊和遮挡导致的目标 ID 转换频繁问题,为目标跟踪领域带来新的突破。图 3 展示了 ByteTrack 二次匹配过程。

ByteTrack 算法的主要流程为通过跟踪轨迹来实现目标匹配和持续跟踪。在算法中,目标被分为多种状态,

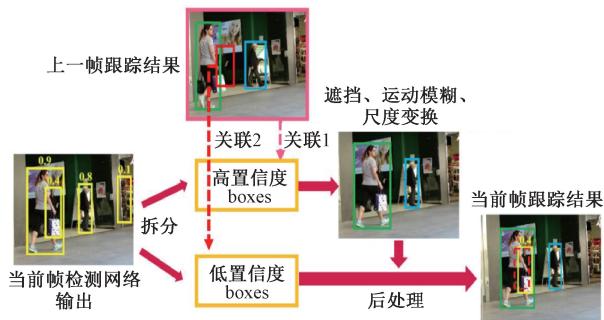


图3 ByteTrack 算法二次匹配过程

Fig. 3 ByteTrack algorithm secondary matching process

包括激活状态、非激活状态、新生轨迹、已追踪轨迹、失踪轨迹和已删除轨迹。每个目标在当前帧，算法通过边界框和评分进行分类，首帧图像用于初始化轨迹对象。通过匹配不同帧之间的目标，实现目标的持续跟踪，确保目标在视频序列中的连续性和稳定性。

ByteTrack 算法目标跟踪流程如图 4 所示，视频流的每一帧通过目标检测算法得到目标框，根据目标检测的置信度分为 3 种情况，大于 0.6 判定为高置信度目标框，用于第 1 关联，大于 0.1 小于 0.6 判定为低置信度目标框，用于第 2 关联，小于 0.1 判定为错误检测结果，直接丢弃。第 1 关联采用卡尔曼滤波预测结果结合 Re-ID 运动特征与当前检测结果计算 IoU，再采用匈牙利算法实现目标块与轨迹线的匹配。在第 2 关联进行低分匹配，更新轨迹信息，这个过程会检测到受到遮挡导致检测分数下降的目标和模糊目标。在第 1 关联未匹配到轨迹线的高分框，会新建轨迹线，未被匹配的轨迹线会保留 30 帧再删除，以解决目标遮挡再出现问题。最终，系统返回所有已跟踪的轨迹线，每条轨迹均具有独一无二的编号，这些编号可作为逐帧跟踪的结果。

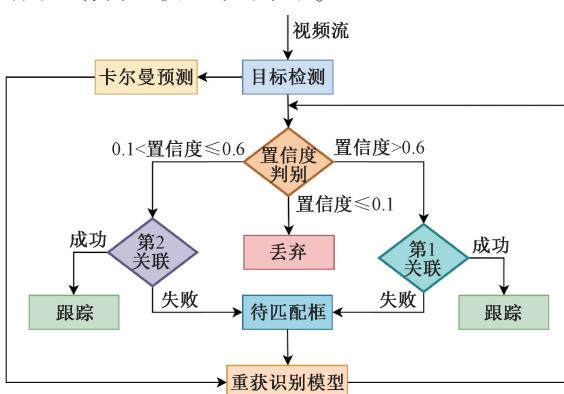


图4 ByteTrack 算法目标跟踪流程

Fig. 4 Target tracking flowchart of ByteTrack algorithm

ByteTrack 算法使用简单有效的关联方式，充分利用了低置信度目标框，将多目标跟踪领域提升到了一个新

的水平。然而这种算法设计之初是为了解决光学图像的目标跟踪问题，直接将 ByteTrack 算法应用在前视声呐图像序列跟踪并不能取得最优效果，需要结合声呐图像序列的特征进行改进。

2.2 改进的 ByteTrack 多目标跟踪算法

1) 目标位置映射算法改进

在水下环境中，声呐系统的性能和跟踪准确性受到多种因素的影响，除了声呐载体旋转导致远端目标跟踪困难外，水下环境的不稳定性如水流、水温变化和复杂声学环境也会影响声呐信号传播的稳定性，进而影响声呐成像的准确性。此外，声呐系统的定位精度和稳定性受航行器运动和姿态变化的影响。这些问题会导致声呐远端的目标运动速度快，容易造成声呐远端目标的跟踪丢失。

如图 5 所示， xoy 为以 O 为原点， x 为横轴， y 为纵轴的世界坐标系。在图 5(a) 中，扇形虚线为 $t-1$ 时刻的声呐图像，其中有 a、b 两个目标，空心圆为目标 a，实心圆为目标 b，这两个目标属同一类目标，即目标特征差异并不明显。在图 5(b) 中，扇形虚线为 t 时刻的声呐图像，其中有 c、d 两个目标，空心圆为目标 c，实心圆为目标 d。在世界坐标系中， $t-1$ 时刻和 t 时刻的目标位置对应关系：a 对应 c，b 对应 d。图 5(c) 为原始的 ByteTrack 算法关联结果，因为在 $t-1$ 和 t 时刻目标几乎没动，声呐的载体姿态发生了变换，导致 $t-1 \sim t$ 时刻得到的声呐图像不一致。ByteTrack 算法的关联是由目标特征和目标运动特征共同作用下的结果，所以会错误地将目标 a 和 d 关联到一起，因为它们属于声呐图像的同一位置，目标 c 和 d 会在关联失败而出现目标丢失的问题。针对这种情况，Brown 等^[25] 改进了目标跟踪运动估计算法。本文利用声呐姿态数据对 ByteTrack 算法加入旋转补偿，使用 ByteTrack 算法更能适应声呐场景下的目标跟踪。

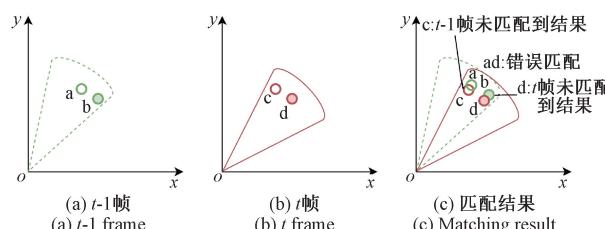


图5 前视声呐旋转造成的错误匹配示意图

Fig. 5 Schematic diagram of incorrect matching caused by forward sonar rotation

在声呐二进制数据解析过程中，可以得到声呐航向角，得到 $t-1$ 时刻的航向角 h_1 和 t 时刻的航向角 h_2 ， h_1-h_2 为两帧的航向角差值 θ ，如图 6 所示，在 xoy 坐标系下， (i, j) 为 $t-1$ 时刻的目标位置， (i', j') 为 t 时刻目标位置，它们的选择角度即为航向角差值 θ 。

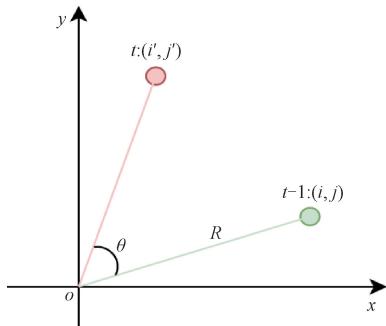


图 6 目标旋转前后在世界坐标系下的位置

Fig. 6 The position of the target in the world coordinate system before and after rotation

坐标的转换关系为：

$$i' = i\cos\theta - j\sin\theta \quad (1)$$

$$j' = i\sin\theta + j\cos\theta \quad (2)$$

目标检测完成之后,将检测到的目标按照式(1)和(2)

进行旋转补偿,再使用 Byte 算法关联,如图 7 所示。得到关联结果如图 8 所示,将它们映射到世界坐标系下,由于 t 时刻下 c 映射到 $t-1$ 时刻的世界坐标系位置接近,所以能正确匹配,b 和 d 目标同理。经过旋转补偿,两帧的目标映射在同一世界坐标系下,而不是在各自的声呐图像坐标系下跟踪,这种改进方式能有效提升跟踪的准确性。

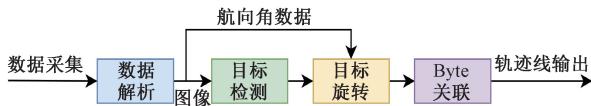


图 7 旋转补偿处理流程

Fig. 7 Rotation compensation processing flow

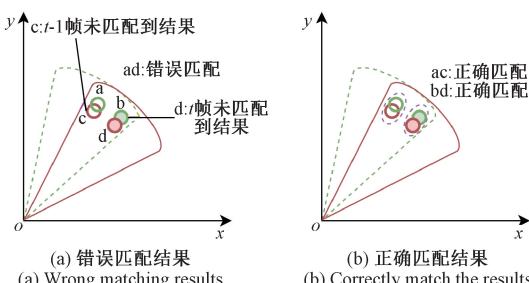


图 8 改进后的关联情况

Fig. 8 Improved correlation situation

2) 相似性度量算法优化

在目标跟踪中,相似性度量是指衡量两个目标之间相似程度的方法。它的主要作用是判断目标在不同帧之间的连续性,以提高目标跟踪的准确性和稳定性。具体来说,相似性度量通常采用特征描述来表示目标,它们可以对目标的外观、运动等特征进行描述,帮助系统对目标

进行关联,即在不同帧中将同一目标进行匹配,使得系统能够在不同帧中识别和跟踪目标。通过相似性度量,系统可以计算目标之间的相似度,从而判断目标在不同帧中是否是同一个目标,实现目标的持续跟踪和识别。

ByteTrack 算法分为两次关联。在第 1 次关联中的相似度量方式为 Re-ID,裁剪检测的图像区域,并将它们馈送到 Re-ID 网络以提取图像特征,其中包括姿态、颜色、纹理和整体外观等,如图 9 所示。然后,基于 Re-ID 特征计算轨迹与检测之间的相似度来关联。

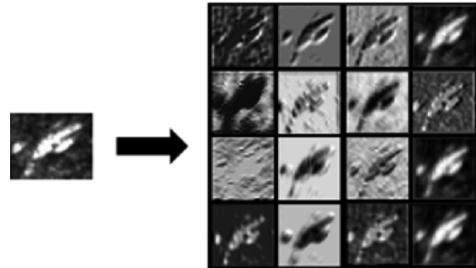


图 9 Re-ID 特征图

Fig. 9 Re-ID feature map

Re-ID 的特征提取网络是独立于目标检测的模块,它专用于提取对应检测框中的特征。这个模块应用一个包含 128 个核的卷积层来提取检测框的 Re-ID 特征。假设提取以 (x, y) 为检测框中心的 Re-ID 特征 $E_{x,y} \subseteq R^{(128HW)}$, H 和 W 表示高和宽。通过全连接层和 Softmax 操作将其映射到分布向量 $R_{x,y}$, 如式(3)所示。

$$R_{x,y} = m(E_{x,y}), P \in [0,1] \quad (3)$$

在第 2 次关联中的相似度量方式为卡尔曼滤波器,预测新帧中轨迹的位置,然后计算检测框和预测框之间的 IoU 作为相似性,如式(4)所示。

$$IoU_{x,y} = \frac{D_{x,y} \cap K(t)}{D_{x,y} \cup K(t)} \quad (4)$$

其中, $D_{x,y}$ 低分目标框, K 为卡尔曼滤波器, t 为轨迹线。因为低分数检测框通常包含严重的遮挡或运动模糊,并且外观特征不可靠,所以第 2 个关联中没有采用外观相似性。

这种相似性度量方式在无人驾驶领域能取得优异的效果,直接应用在声呐图像中却容易出现目标丢失问题,因为声呐图像的成像不清晰、目标特征的可识别性差。因此,直接使用 Re-ID 作为第 1 关联相似度量算法并非最佳做法。在本研究中,将 Re-ID 外观特征和 IoU 运动特征合并,在第 1 关联不只考虑 Re-ID 外观特征,还兼顾了运动特征,提供了一个更全面的相似性度量算法。本小节在第 1 关联提出了一种卡尔曼滤波和目标特征结合的方式作相似性度量算法,如式(5)所示,通过微调 Re-ID 和 IoU 之间的权重,得到加权平均后的度量算法。

$$M_{x,y} = w \times R_{x,y} + (1 - w) \times IoU_{x,y} \quad (5)$$

其中, w 是 Re-ID 的权重, $1-w$ 是 IoU 的权重。

3 实验细节

3.1 数据集及实验细节

在检测部分本次实验使用的 500 张图像,主要包括鱼、潜水员。按照 8:1:1 的比例分为训练集、验证集、测试集,使用 YOLOv8 自带的数据增强,包括翻转、剪裁、缩放等。跟踪部分使用了 6 个声呐视频序列,包括为 4 个鱼类视频序列,2 个蛙人视频序列。本文进行实验的硬件和软件参数如表 1 所示。使用 PyTorch 深度学习框架 1.9.0、CUDA11.1 和 Python3.8.18 实现,并使用 Nvidia3090 显卡进行训练。

表 1 目标检测跟踪实验环境参数

Table 1 Experimental environment parameters of target detection and tracking

环境名称	配置参数
CPU	Intel i9-10900k CPU @ 3.70 GHZ
GPU	NVIDIA GeForce RTX 3090 GPU
内存	SAMSUNG DDR4 3200 MHz 64 GB
操作系统	Windows 11
编译语言	Python 3.8.18
集成开发环境	PyCharm 2021.3
CUDA	11.1
CUDNN	8.0.5
Torch	1.9.0
Torchvision	0.10.0
Numpy	1.24.4
Pandas	1.3.3

3.2 实验部署

声呐设备采集的含有目标的视频经过 YOLOv8s 目标检测网络,得到目标检测框,一方面根据目标检测的置信度进行分类,采用 ByteTrack 进行旋转补偿和目标跟踪;另一方面对目标检测结果进行卡尔曼滤波预测,将预测结果与第 1 关联和第 2 关联失败的待匹配框一起重获识别模型,之后进行 ByteTrack 目标跟踪。

在整个的实验过程中,声呐设备随着水下机器人运动,采集动态视频,采集的数据经过 NVIDIA Jetson Xavier 系统并实时处理,输出跟踪结果。

3.3 评价指标

1) 身份切换(identity switch, IDs)是一种用于评估多目标跟踪算法性能的指标,它表示目标 ID 分配的切换次

数。在多目标跟踪算法中,每个被检测到的目标都会被分配一个独特的 ID,以便在不同帧之间跟踪该目标。通过为每个目标分配一个唯一的 ID,使算法可以准确地识别和跟踪每个目标的运动轨迹,从而实现多目标跟踪任务。IDs 的取值为 ≥ 0 的整数,越接近 0 表示算法越准确,如式(6)所示。

$$IDs = \sum_{t=1}^F N_t \quad (6)$$

其中, N_t 表示第 t 帧中发生 ID 错误的目标数。

2) 多目标跟踪精度(multiple object tracking accuracy, MOTA)代表多目标跟踪准确性的度量,通过比较跟踪器输出与真实目标轨迹之间的匹配情况,计算在跟踪中的误差积累情况。

$$MOTA = 1 - \frac{FN + FP + IDs}{GT} \quad (7)$$

其中, FN 为整个图像序列漏报数量之和, FP 为整个图像序列误报数量之和, GT 为整个图像序列的正确跟踪数量之和。

3) IDF1(integrated detection and F1)代表多目标跟踪的 F1 分数,是结合了检测和跟踪精度来评估跟踪器性能的综合指标。IDF1 是正确识别的检测与平均真实数和检测数之比,计算公式为:

$$IDF1 = \frac{2IDTP}{2IDTP + IDFP + IDFN} \quad (8)$$

其中, $IDTP$ 表示真正 ID 数, $IDFP$ 表示假正 ID 数, $IDFN$ 代表假负 ID 数。

3.4 相似性度量算法对比实验

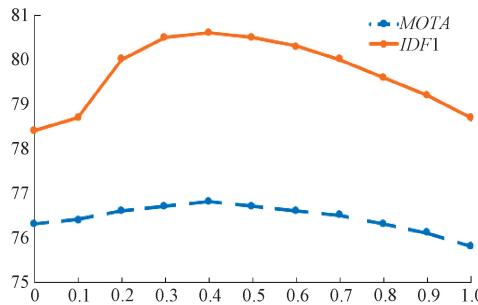
在 Byte 的第 1 关联和第 2 关联的相似性度量上,选择了 Re-ID 和 IoU 两种度量方式。结果如表 2 所示,对比第 1、2 组实验可以发现,在第 1 关联的相似性度量上替换为 IoU 的方式能取得更好的 MOTA 和 IDs,这是因为前视声呐图像分辨率低和目标特征少导致可识别性差,在第 1 关联采用基于运动特征的卡尔曼滤波器更加可靠。第 3 组与其他组实验对比可以发现,第 3 组实验评价指标效果均低于其他实验,这说明在第 2 关联使用 IoU 作为相似度量方式很重要,因为低分检测框通常包含严重的遮挡或运动模糊,这些都会导致 Re-ID 特征不可靠。对比第 4 组与前 2 组实验,可以发现,在第 1 关联采用 Re-ID 和 IoU 结合的方式能取得更优秀的效果,这是因为高分检测框的目标在轮廓和姿态等外观特征上更可靠,但由于声呐数据采集本身帧率较低、目标外观特征少,结合基于运动特征的卡尔曼滤波器能提高跟踪效果。

式(5)的 w 是一个敏感的超参数,决定第 1 关联参考 Re-ID 外观特征和 IoU 运动特征的权重。从 0~1.0 调整 w ,比较 MOTA、IDF1 指标在不同权重的值,结果如图 10 所示。可以看出, w 取值为 0.4 时,取得最优的

表 2 第 1、2 次关联中使用不同类型相似性度量的比较

Table 2 Comparison of different types of similarity measures used in the first and second associations

Similarity#1	Similarity#2	MOTA ↑	IDF1 ↑	IDs ↓
Re-Id	IoU	75.8	78.7	73
IoU	IoU	76.3	78.4	64
Re-Id	Re-Id+IoU	75.2	77.5	105
Re-Id+IoU	IoU	76.6	80.5	59

图 10 第 1 次关联 w 超参数在不同分数阈值下的性能比较Fig. 10 Performance comparison of the first correlation w hyperparameter under different score thresholds

MOTA 和 IDF1, 说明前视声呐图像高分检测框目标的运动特征比轮廓、轮廓和姿态等外观特征的可靠性更强。

3.5 多跟踪算法对比实验

根据上述的理论方法, 使用 YOLOv8 作为基线检测算法, 将改进后的 ByteTrack 跟踪算法与其他主流跟踪算法进行了对比测试。表 3 显示了本文改进算法与先进跟踪算法 DeepSort、TransTrack 和 ByteTrack 的之间的 MOTA、IDF1 和 IDs 的比较。

表 3 多目标跟踪算法对比

Table 3 Comparison of Multi target Tracking Algorithms

跟踪算法	MOTA ↑	IDF1 ↑	IDs ↓
DeepSort ^[22]	66.3	69.2	145
TransTrack ^[24]	67.8	70.1	131
ByteTrack ^[13]	70.2	73.4	103
本文	76.8	80.6	55

分析表 3 可以发现, 改进后的模型跟踪准确度为 76.8%, 跟踪召回率为 80.6%。与改进前的 ByteTrack 算法相比, 改进后的算法与主流跟踪算法 DeepSort 相比跟踪精度提升了 10.5 个百分点, 召回率提升了 11.4 个百分点, ID 切换次数降低了 90 次, 证明 ByteTrack 的跟踪效果优于 DeepSort。改进后的 ByteTrack 与改进前的 ByteTrack 相比, 跟踪精度提升了 9.4%, 召回率提升了

10.8%, ID 切换次数降低了 46%。对于同一个目标, DeepSort 算法在跟踪中出现频繁的 ID 切换, 跟踪效果不佳, ByteTrack 跟踪算法充分利用了低分数检测框, 跟踪效果优于 DeepSort 和 TransTrack。而改进的 ByteTrack 算法通过旋转补偿和改进关联相似度计算, 减少了 ID 的切换次数, 提升了跟踪的准确度。

为了进一步验证改进算法的有效性, 给出基线检测模型下 DeepSort、TransTrack、ByteTrack 以及本文改进的 ByteTrack 算法在声呐图像序列的跟踪结果, 如图 11 和 12 所示。

图 11 图像序列 A 包含 4 个鱼目标在 63、64、74 和 84 帧的情况, 其中在 63~64 帧出现声呐载体旋转, 64~84 帧为声呐目标消失再出现的情况。从图 11 跟踪结果可以看出, 在 63~64 帧出现声呐载体旋转时, DeepSort 在旋转后将所有跟踪失败目标新建轨迹线, 而 TransTrack、ByteTrack 对旋转不够敏感, 近处较大的目标能正确跟踪, 而远处较小的目标跟踪失败, 在改进后加入旋转补偿得到了改善。图 12 图像序列 B 包含两个蛙人目标在 57、77 和 97 帧情况, 在整个过程中, 左侧蛙人转向发生严重形变。从图 12 跟踪结果可以看出, DeepSort、TransTrack 的 ID 增加迅速, 说明出现较多的关联失败导致碎片轨迹线陡增, 这些主流跟踪算法均无法对左侧蛙人形变作出有效的关联, 在进行算法改进后得到改善, 在 57~97 帧并未发生 ID 切换。第 1 关联采用运动特征敏感的 IoU 和对外观特征敏感的 Re-Id 的关联方式, 其中目标的运动特征比外观特征的可靠性更强。实验结果表明, 改进后的目标跟踪算法在提高 MOTA 和 IDF1 的同时, 降低了 ID 切换次数和碎片轨迹线。

3.6 检测与跟踪融合实验

为证明本文提出的检测算法和跟踪算法的可靠性, 使用降噪后的图像序列完成改进前后检测跟踪算法多组消融实验, 分别为 YOLOv8-s 结合 ByteTrack、DP-YOLOv8-s 结合 ByteTrack、YOLOv8-s 结合改进的 ByteTrack 的 3 组检测跟踪算法实验。输入多组声呐图像序列得到跟踪结果, 如表 4 所示。

表 4 改进前后检测跟踪算法对比

Table 4 Comparison of detection tracking algorithms before and after improvement

方法	MOTA	IDF1	IDs ↓	FPS ↓
	/% ↑	/% ↑	↓	↓
YOLOv8-s+ByteTrack	72.7	75.6	83	45
DP-YOLOv8-s+ByteTrack	74.1	77.9	92	39
YOLOv8-s+改进 ByteTrack	78.4	82.4	54	42



图 11 声呐图像序列 A 跟踪效果对比

Fig. 11 Comparison of tracking effects of sonar image sequence A

分析表 4 的前 2 组实验可以发现,保持跟踪算法不变,使用改进后的 YOLOv8-s 算法跟踪相比于改进前的 MOTA 和 IDF1 均得到了提升,而 IDs 会增加,这是因为 DP-YOLOv8-s 是针对小目标的改进,能检测更多的小目标,所以 MOTA 和 IDF1 得到提升,但由于 ByteTrack 无法有效地跟踪导致 IDs 的增加。对比第 1 组和第 3 组实验不难发现,只改进跟踪算法对评价指标有更大的提升,由于针对声呐特征的跟踪算法改进,利用声呐姿态数据对跟踪算法添加旋转补偿,解决了载体旋转导致的跟踪失

败的问题,在第 1 关联使用基于卡尔曼滤波的目标运动特征和基于 Re-ID 目标外观特征结合的方式作相似性度量算法。YOLO8s+ByteTrack 算法有 45 帧率,是因为使用的 YOLO8s 模型小,比官方给的基于 YOLOX 检测算法的 ByteTrack 算法的 FPS 更低。检测跟踪的改进均会增加算法复杂度影响 FPS。

为了进一步验证本文提出算法的可靠性,在测试数据集的声呐图像序列中选取了部分典型的场景,如图 13 所示。

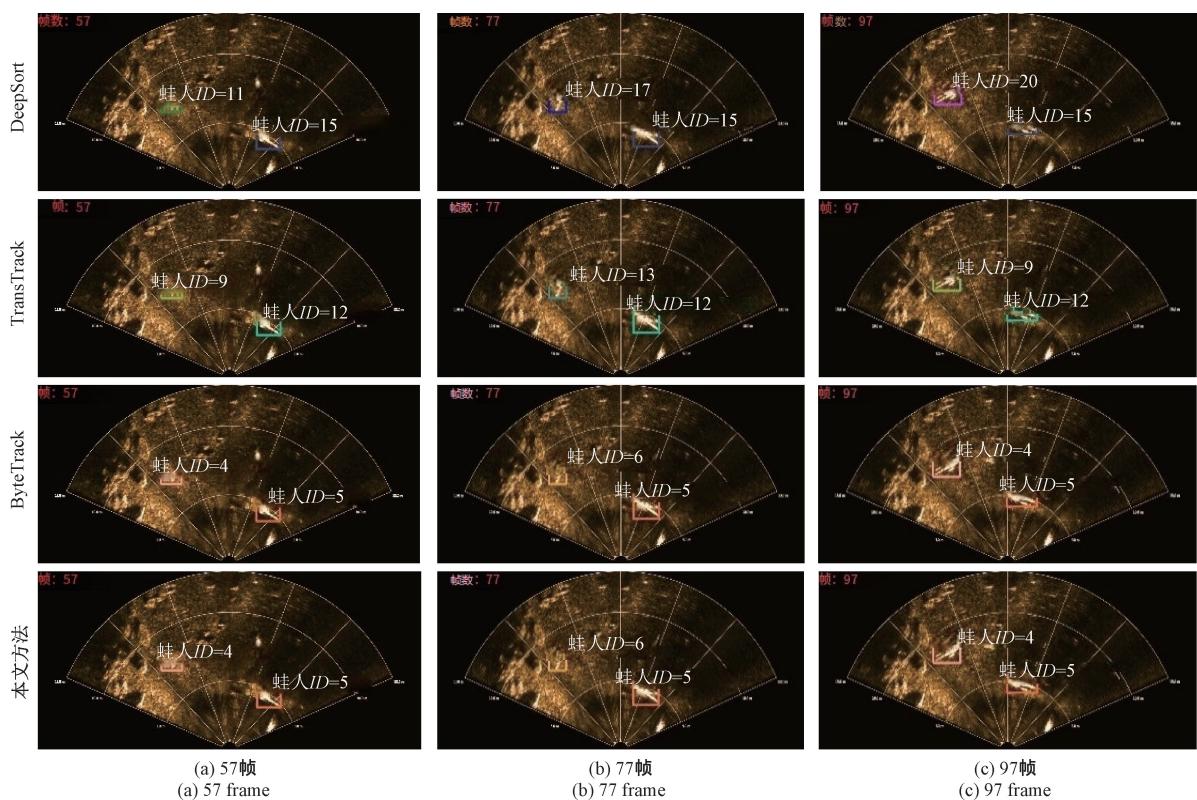


图 12 图像序列 B 跟踪效果对比

Fig. 12 Comparison of tracking effects of image sequence B

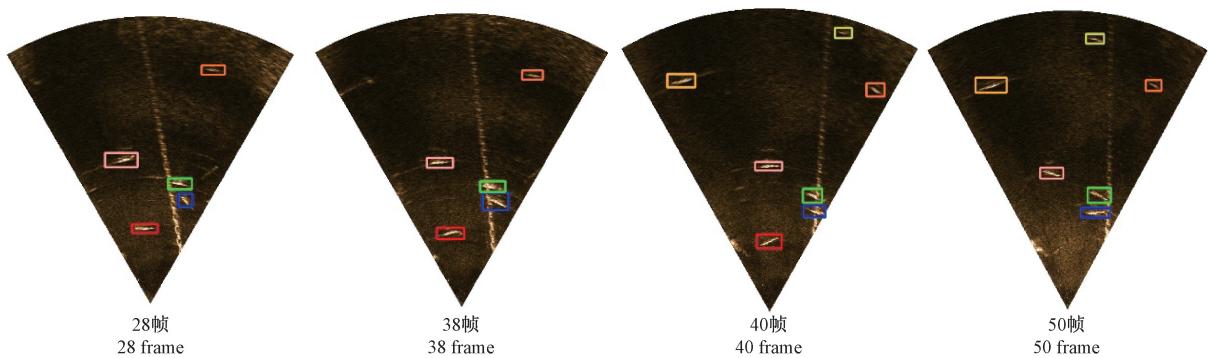
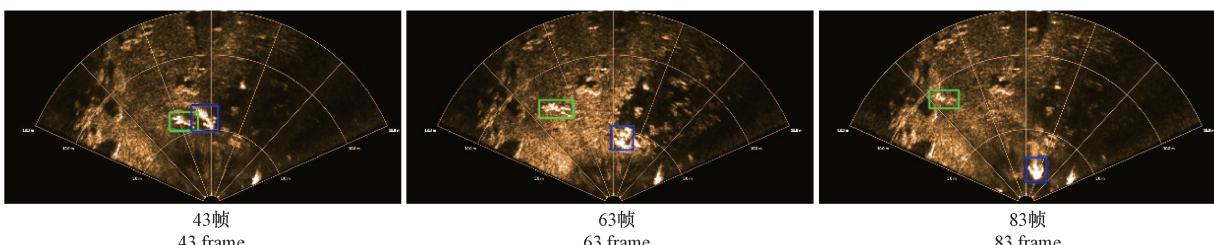
(a) 视频序列A（载体旋转、目标偏小、目标形变、目标遮挡）
(a) Video sequence A(carrier rotation, small target, target deformation, target occlusion)(b) 视频序列B（目标消失再出现、目标模糊）
(b) Video sequence B (target disappearance and reappearance, target blur)

图 13 视频序列跟踪典型场景

Fig. 13 Typical scenarios of video sequence tracking

图13(a)为多个鱼类目标在不同帧的位置情况,在28~50帧为目标被障碍物部分遮挡再出现的情况,38~40帧为声呐载体旋转导致声呐图像偏转的情况,40~50帧目标出现形变,且远端的目标存在目标偏小的问题。图13(b)为多个蛙人目标在不同帧的位置情况,43~63帧为目标由未

遮挡到海底岩石块完全遮挡导致目标消失的情况,在83帧又重新出现,83帧的目标变得模糊。将这两个序列分别使用YOLOv8-s结合ByteTrack、DP-YOLOv8-s结合ByteTrack、YOLOv8-s结合改进的ByteTrack3组算法进行目标检测和跟踪,得到跟踪结果如图14和15所示。

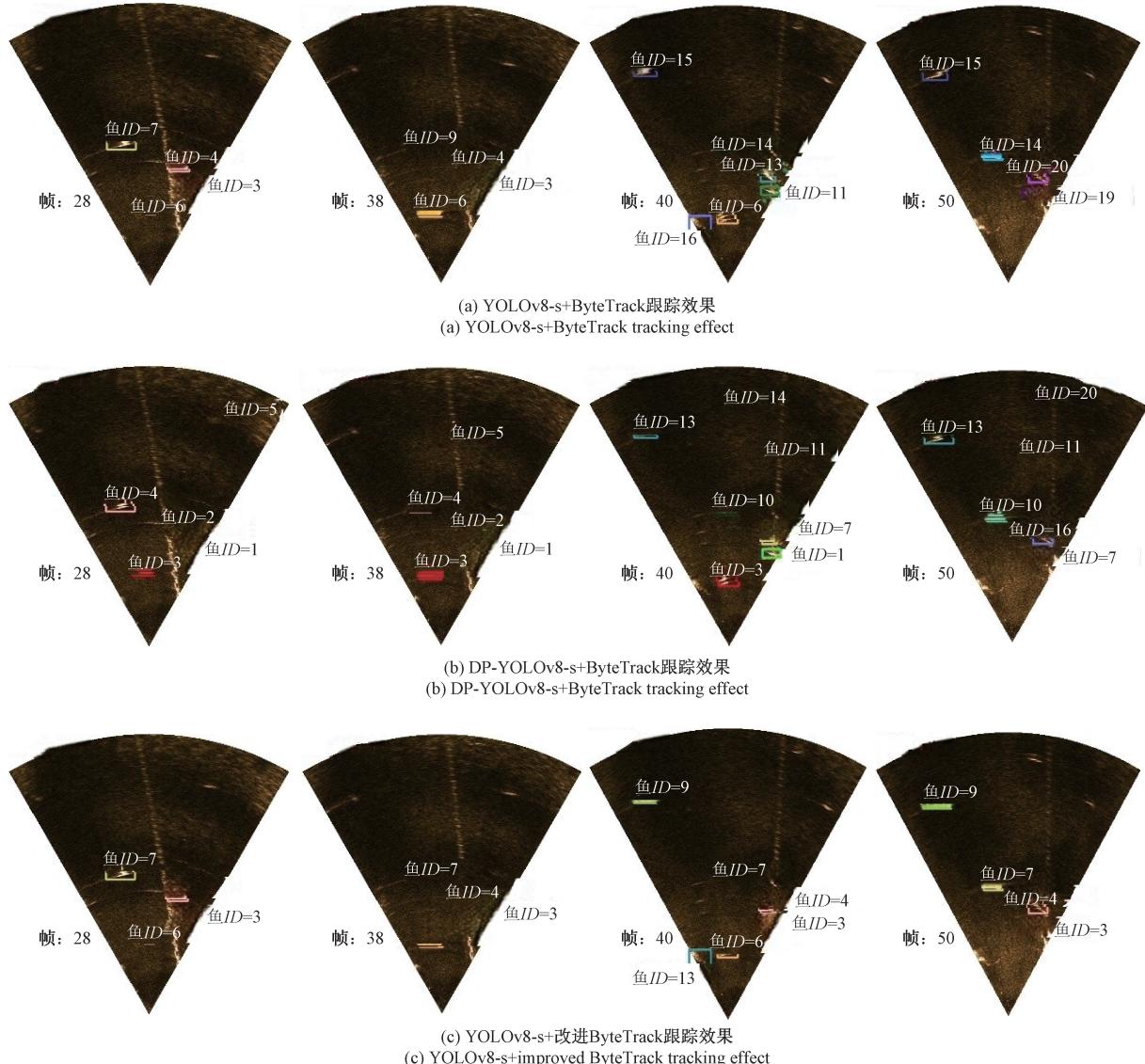


图14 视频序列A 检测跟踪效果对比

Fig. 14 Comparison of detection and tracking effects for video sequence A

对比图14(a)和(b)可以看出,原始的YOLOv8-s算法对声呐图像远端的小目标不够敏感,容易出现漏检,对40帧左下角的干扰物出现误检,DP-YOLOv8-s针对小目标设计的浅层特征检测头和改进的C2N特征提取模块,提高了小目标的检测敏感度和模型表达能力,有效地解决了目标漏检和误检问题。对比图14(a)、(c)可见,在38~40帧由于声呐载体旋转,导致IDs大量切换,目标关

联利用声呐姿态数据对ByteTrack算法加入旋转补偿,有效提升了目标跟踪的准确性。在28~50帧目标右下角两个目标受障碍物的部分遮挡,导致IDs切换,这在改进后得到解决,因为ByteTrack在第1关联使用目标外观特征,而声呐图像目标外观特征并不十分可靠,本文提出在第1关联使用了运动特征和外观特征结合的方式作相似性度量,提升了跟踪的准确性。

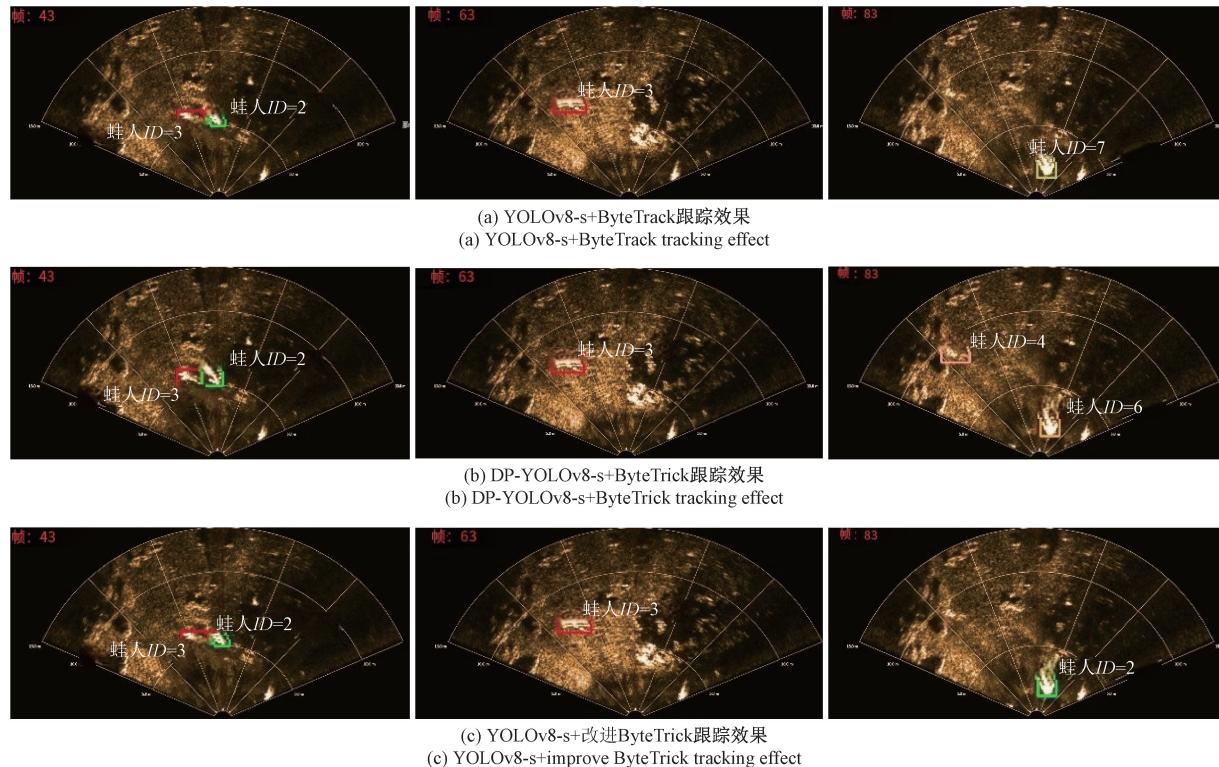


图 15 视频序列 B 检测跟踪效果对比

Fig. 15 Comparison of detection and tracking effects for video sequence B

对比另一个图像序列如图 15 所示,改进后的 DP-YOLOv8-s 算法能检测到 83 帧的模糊目标,因为改进后 C2N 结构更深,具有更强的特征提取能力。改进后的 ByteTrack 成功关联右侧消失再出现的目标,这是因为改进后的 ByteTrack 的第 1 关联相似性度量方式更加依赖于目标的运动特征。实验结果表明,本文提出的改进检测跟踪算法在目标偏小、载体旋转、目标短暂消失、目标遮挡、目标形变的典型场景中拥有不错的效果,改进后的目标检测跟踪融合算法拥有更低的漏检率、误检率,更低的身份切换次数,更能适应前视声呐水下目标的检测和跟踪场景。

4 结 论

本研究对传统多目标跟踪方法和现代多目标跟踪方法进行了研究和分析,包括 DeepSort、TransTrack 和 ByteTrack 跟踪算法。结合前视声呐目标跟踪的特点,对 ByteTrack 算法做了两点改进:针对前视声呐载体旋转导致目标运动过快问题,本文利用声呐姿态数据对 ByteTrack 算法加入旋转补偿,提升了跟踪的准确性;针对 ByteTrack 算法应用在声呐跟踪上容易出现目标丢失的问题,根据前视声呐图像特征,改进关联方式,在第 1 关联提出了一种基于卡尔曼滤波的运动特征和目标外观

特征结合的方式作相似性度量,提升了跟踪的准确性。最后,通过相似性度量算法对比实验,证明了改进后的在第 1 关联使用一种基于卡尔曼滤波的运动特征和目标外观特征结合的方式的优越性;对比了 DeepSort、TransTrack 和 ByteTrack 主流目标跟踪算法,改进后的 ByteTrack 算法在 MOTA、IDF1 和 IDs 评价指标下都取得了最优,证明了本文改进方法的有效性。

参考文献

- [1] CHONG C. An overview of machine learning methods for multiple target tracking [C]. 2021 IEEE 24th International Conference on Information Fusion, 2021: 1-9.
- [2] PAN L, LI W, ZHU J S, et al. Image tracking of fire extinguishing jet drop point based on improved ENet and robust adaptive cubature Kalman filtering [J]. IEEE Transactions on Instrumentation and Measurement, 2024, 73: 7507512.
- [3] ARTHANARI S, MOORTH Y S, JEONG J H, et al. Adaptive spatially regularized target attribute-aware background suppressed deep correlation filter for object tracking [J]. Signal Processing: Image Communication,

- 2025, 136: 117305.
- [4] 逯万春, 姜培斌, 凌亮, 等. 基于 KCF-Hash-Match 目标跟踪算法的高速列车车轮横向晃动识别方法 [J]. 机械工程学报, 2023, 59(24): 223-230.
- LU W CH, JIANG P B, LING L, et al. KCF-Hash-Match target tracking algorithm for identifying wheel lateral sway of high-speed train [J]. Journal of Mechanical Engineering, 2023, 59(24): 223-230.
- [5] HUANG SH J, CAI N G, PACHECO P P, et al. Applications of support vector machine (SVM) learning in cancer genomics [J]. Cancer Genomics & Proteomics, 2018, 15(1): 41-51.
- [6] ZHAO F, HUI K D, WANG T T, et al. A KCF-based incremental target tracking method with constant update speed [J]. IEEE Access, 2021, 9: 73544-73560.
- [7] BEWLE Y A, GE Z Y, OTT L, et al. Simple online and real time tracking [C]. 2016 IEEE International Conference on Image Processing, 2016: 3464-3468.
- [8] KAPANIA S, SAINI D, GOYAL S, et al. Multi object tracking with UAVs using deep SORT and YOLOv3 RetinaNet detection framework [C]. AIMS'20: The 1st ACM Workshop on Autonomous and Intelligent Mobile Systems, 2020: 1-6.
- [9] CUI Y T, ZENG CH K, ZHAO X Y, et al. SportsMOT: A large multi-object tracking dataset in multiple sports scenes [C]. 2023 IEEE/CVF International Conference on Computer Vision, 2023: 9887-9897.
- [10] ZHOU X Y, YIN T W, KOLTUN V, et al. Global tracking transformers [C]. 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2022: 8761-8770.
- [11] MA F, SHOU M Z, ZHU L CH, et al. Unified transformer tracker for object tracking [C]. 2022 IEEE/CVF Conference on Computer Vision And Pattern Recognition, 2022: 8771-8780.
- [12] ZHANG Y F, SUN P Z, JIANG Y, et al. Bytetrack: Multi-object tracking by associating every detection box [C]. Computer Vision-Eccv 2022, 2022: 1-21.
- [13] HUANG H W, YANG CH Y, SUN J CH, et al. Iterative scale-up expansionIoU and deep features association for multi-object tracking in sports [C]. 2024 IEEE/CVF Winter Conference on Applications of Computer Vision, 2024: 163-172.
- [14] YANG H, CHANG F L, HUANG Y H, et al. Multi-object tracking using DeepSORT and modified CenterNet in cotton seedling counting [J]. Computers and Electronics in Agriculture, 2022, 202: 107-339.
- [15] YE X F, MA X L. Improved multi-object tracking algorithm for forward looking sonar based on rotation estimation [C]. Intelligent Robotics and Applications, 2019: 171-183.
- [16] ZHANG T D, LIU SH W, HE X, et al. Underwater target tracking using forward-looking sonar for autonomous underwater vehicles [J]. Sensors, 2019, 20(1): 102-129.
- [17] 崔杰, 胡长青, 徐海东, 等. 改进的 Mean Shift 算法在前视声呐运动目标跟踪中的应用 [J]. 声学技术, 2020, 39(3): 279-283.
- CUI J, HU CH Q, XU H D, et al. Application of improved mean shift algorithm in moving target tracking of forward-looking sonar [J]. Technical Acoustics, 2020, 39(3): 279-283.
- [18] 陈美龙, 赵新华, 叶秀芬. BM3D-YOLOv8-s: 前视声呐图像目标检测算法 [J]. 仪器仪表学报, 2025, 46(2): 234-246.
- CHEN M L, ZHAO X H, YE X F. BM3D-YOLOv8-s: Forward-looking sonar image target detection algorithm [J]. Chinese Journal of Scientific Instrument, 2025, 46(2): 234-246.
- [19] 骈志康. 基于深度学习的前视声纳目标检测跟踪方法 [D]. 哈尔滨: 哈尔滨工程大学, 2020.
- PIAN ZH K. Forward-looking sonar underwater target detection and tracking technology based on deep learning [D]. Harbin: Harbin Engineering university, 2020.
- [20] 王文胜, 何君尧, 黄民, 等. 融合梯度改进 YOLO 和 KCF 模型的无人机目标识别跟踪算法 [J]. 仪器仪表学报, 2025, 46(2): 221-233.
- WANG W SH, HE J Y, HUANG M, et al. Fusion gradient improved YOLO and KCF models for UAV target recognition and tracking algorithm [J]. Chinese Journal of Scientific Instrument, 2025, 46(2): 221-233.
- [21] 霍旭, 盖绍彦, 洪濡, 等. 基于多阶段关联的多目标跟踪算法 [J]. 仪器仪表学报, 2023, 44(11): 205-214.
- HUO X, GAI SH Y, HONG R, et al. Multi-object

- tracking algorithm based on multi-stage association [J]. Chinese Journal of Scientific Instrument, 2023, 44(11): 205-214.
- [22] DU Y H, ZHAO ZH CH, SONG Y, et al. Strongsort: Make deepsort great again [J]. IEEE Transactions on Multimedia, 2023, 25: 13.
- [23] 陈新华, 张龙恩, 郑恩明, 等. 深海海底反射拖船噪声伪目标判别 [J]. 仪器仪表学报, 2024, 45(10): 272-283.
- CHEN X H, ZHANG L EN, ZHENG EN M, et al. False target discrimination of tugboat noise reflected by deep seabed [J]. Chinese Journal of Scientific Instrument, 2024, 45(10): 272-283.
- [24] SUN P Z, CAO J K, JIANG Y, et al. Transtrack: Multiple object tracking with transformer [J]. ArXiv preprint arXiv:2012.15460, 2020.
- [25] BROWN D C, GERC I D, BLANFORD T E. Interpolation kernels for synthetic aperture sonar along-track motion estimation [J]. IEEE Journal of Oceanic Engineering, 2019, 45(4): 1497-1505.

作者简介



陈美龙, 2021 年于哈尔滨学院获得学士学位, 2024 年于哈尔滨工程大学获得硕士学位, 现就职于美团, 任前端研发工程师, 主要研究方向为图像识别和 AIGC。

E-mail: mileschen666@163.com

Chen Meilong received his B. Sc. degree from Harbin College in 2021, and received M. Sc. degree from Harbin Engineering University in 2024. Now he is working in Meituan as front-end developer. His main research interests include image recognition and AIGC.



赵新华(通信作者), 2005 年和 2008 年于哈尔滨工程大学获得硕士、博士学位, 现为哈尔滨工程大学副教授, 主要研究方向为人工智能、声呐图像识别与检测。

E-mail: zhaoxinhua@hrbeu.edu.com

Zhao Xinhua (Corresponding author) received her M. Sc. and Ph. D. degrees from Harbin Engineering University in 2005 and 2008, respectively. Now she is an associate professor at Harbin Engineering University. Her main research interests include artificial intelligence and sonar image recognition and detection.