

光电工程

Opto-Electronic Engineering

中文核心期刊 中国科技核心期刊
Scopus CSCD

面向多类别舰船多目标跟踪的改进CSTrack算法

袁志安, 谷雨, 马淦

引用本文:

袁志安, 谷雨, 马淦. 面向多类别舰船多目标跟踪的改进CSTrack算法[J]. *光电工程*, 2023, **50**(12): 230218.

Yuan Z A, Gu Y, Ma G. Improved CSTrack algorithm for multi-class ship multi-object tracking[J]. *Opto-Electron Eng*, 2023, **50**(12): 230218.

<https://doi.org/10.12086/oe.2023.230218>

收稿日期: 2023-09-01; 修改日期: 2023-12-04; 录用日期: 2023-12-05

相关论文

基于改进YOLOv5s的无人机图像实时目标检测

陈旭, 彭冬亮, 谷雨

光电工程 2022, **49**(3): 210372 doi: [10.12086/oe.2022.210372](https://doi.org/10.12086/oe.2022.210372)

基于双分支多尺度融合网络的毫米波SAR图像多目标语义分割方法

丁俊华, 袁明辉

光电工程 2023, **50**(12): 230242 doi: [10.12086/oe.2023.230242](https://doi.org/10.12086/oe.2023.230242)

时空特征对齐的多目标跟踪算法

程稳, 陈忠碧, 李庆庆, 李美惠, 张建林, 魏宇星

光电工程 2023, **50**(6): 230009 doi: [10.12086/oe.2023.230009](https://doi.org/10.12086/oe.2023.230009)

融合空间掩膜预测与点云投影的多目标跟踪

陆康亮, 薛俊, 陶重犇

光电工程 2022, **49**(9): 220024 doi: [10.12086/oe.2022.220024](https://doi.org/10.12086/oe.2022.220024)

更多相关论文见光电期刊集群网站 



<http://cn.ojournal.org/oe>



OE_Journal



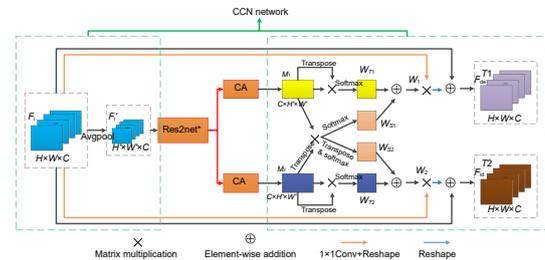
Website

DOI: 10.12086/oe.2023.230218

面向多类别舰船多目标跟踪的改进 CSTrack 算法

袁志安, 谷雨*, 马滢

杭州电子科技大学自动化学院, 浙江 杭州 310018



摘要: 针对海面舰船多目标跟踪过程中图像背景复杂、目标尺度差异大等难点, 提出了一种改进 CSTrack 的舰船多目标跟踪算法。首先, 针对 CSTrack 算法使用暴力解耦分解颈部特征造成目标特征损失的问题, 提出了一种结合 Res2net 模块的改进互相关解耦网络 RES_CCN, 使网络解耦后获得更加细粒度的特征。其次, 为提升对多类别舰船的跟踪性能, 采用检测头网络解耦设计分别预测目标类别、置信度和位置。最后, 采用 MOT2016 数据集进行消融实验, 验证了所提模块的有效性, 在新加坡海事数据集上进行测试, 所提算法的多目标跟踪精度提升了 8.4%, 目标识别准确度提升了 3.1%, 优于 ByteTrack 等算法。本文所提算法具有跟踪精度高、误检率低等优点, 适用于海面舰船多目标跟踪任务。

关键词: 多目标跟踪; 目标重识别; 目标检测; 细粒度特征; 注意力机制

中图分类号: TP391.41

文献标志码: A

袁志安, 谷雨, 马滢. 面向多类别舰船多目标跟踪的改进 CSTrack 算法 [J]. 光电工程, 2023, 50(12): 230218

Yuan Z A, Gu Y, Ma G. Improved CSTrack algorithm for multi-class ship multi-object tracking[J]. *Opto-Electron Eng*, 2023, 50(12): 230218

Improved CSTrack algorithm for multi-class ship multi-object tracking

Yuan Zhian, Gu Yu*, Ma Gan

School of Automation, Hangzhou Dianzi University, Hangzhou, Zhejiang 310018, China

Abstract: Due to the difficulties of complex backgrounds and large-scale differences between objects during the process of ship multi-object tracking in sea-surface scenarios, an improved CSTrack algorithm for ship multi-object tracking is proposed in this paper. Firstly, as violent decoupling is used in the CSTrack algorithm to decompose neck features and cause object feature loss, an improved cross-correlation decoupling network that combines the Res2net module (RES_CCN) is proposed, and thus more fine-grained features can be obtained. Secondly, to improve the tracking performance of multi-class ships, the decoupled design of the detection head network is used to predict the class, confidence, and position of objects, respectively. Finally, the MOT2016 dataset is used for the ablation experiment to verify the effectiveness of the proposed module. When tested on the Singapore maritime dataset, the multiple object tracking accuracy of the proposed algorithm is improved by 8.4% and the identification F1 score is increased by 3.1%, which are better than those of the ByteTrack and other algorithms. The proposed

收稿日期: 2023-09-01; 修回日期: 2023-12-04; 录用日期: 2023-12-05

基金项目: 浙江省自然科学基金资助项目 (LY21F030010, LZ23F030002)

*通信作者: 谷雨, guyu@hdu.edu.cn。

版权所有©2023 中国科学院光电技术研究所

algorithm has the advantages of high tracking accuracy and low error detection rate and is suitable for ship multi-object tracking in sea-surface scenarios.

Keywords: multi-object tracking; re-identification; object detection; fine-grained feature; attention mechanism

1 引言

舰船多目标跟踪作为多目标跟踪 (multi-object tracking, MOT) 的重要应用场景之一, 在海面救援、渔业管理、海面交通疏导以及舰船监测等领域有着重要的现实意义, 其目的是定位多个感兴趣的舰船目标, 并维持每个目标唯一的身份编号 (identification, ID), 记录连续的运动轨迹^[1]。然而, 在复杂海面场景下进行鲁棒跟踪仍存在诸多困难, 主要体现在以下 4 个方面: 1) 跟踪过程中舰船目标被遮挡导致难以被精准定位; 2) 不同舰船目标间可能具有较高的外观相似性, 增加了维持目标 ID 的难度; 3) 舰船目标间交互可能造成跟踪框漂移^[2]; 4) 海面上成像模糊和存在多个类别舰船目标带来的检测挑战。这些问题与行人多目标跟踪类似, 但又存在差别, 因此将行人多目标跟踪的方法迁移到舰船多目标跟踪时需要进行针对性调整。

传统的多目标跟踪算法包含马尔可夫决策、联合概率数据关联、粒子滤波、最近邻标准滤波等^[2], 这些传统方法在预测位置时存在较大的误差, 并且对遮挡或相似物体的干扰鲁棒性不足。随着深度学习理论与技术的飞速发展, 基于深度学习的多目标跟踪方法在性能上已超越传统方法并成为研究的主流方向^[3]。现有基于深度学习的多目标跟踪算法中, 目标检测和基于重识别的特征提取是影响多目标跟踪性能的两个主要因素。2016 年 Xiao 等人首次在一个卷积神经网络里实现同时处理行人检测和行人重识别任务^[4], 2017 年 Bewley 等人提出的 DeepSORT 算法^[5] 利用深度学习网络模型提取重识别 (re-identification, ReID) 特征向量进行数据关联实现行人的鲁棒跟踪。随后, 研究学者在此基础上提出了多种改进算法, 显著提高了多目标跟踪性能^[6-9]。StrongSORT^[6] 算法在文献 [5] 的基础上优化特征提取器和特征更新策略, 并使用相机运动补偿和自适应计算噪声协方差的卡尔曼滤波算法提升了目标跟踪精度; ByteTrack^[7] 算法利用检测框和跟踪轨迹之间的相似性, 在保留高分检测结果的同时, 从低分检测结果中去除背景, 挖掘出模

糊、被遮挡的困难目标样本, 从而降低漏检并提高轨迹的连续性; FairMOT 算法^[8] 则根据文献 [4] 的思想使用 anchor-free 方式减少 anchor 之间的歧义, 通过聚合多层特征生成 ReID 特征向量提高网络对不同尺寸目标的适应能力。然而, 在一个卷积神经网络中尝试同时实现目标检测和目标重识别这两种相互矛盾的任务会导致网络性能不佳。为解决这两种任务之间的矛盾, CStrack 算法^[9] 通过解耦的方式将网络在颈部 (neck) 得到的特征图分解成两部分, 用不同的网络头实现不同的分支任务。这种方式虽然有效地提升了多目标跟踪的性能, 但其暴力解耦的方式会带来目标特征损失的问题, 导致在目标遮挡、小目标或密集目标等情况下跟踪性能下降。

为提高 CStrack 算法对海面舰船多目标跟踪性能, 本文从解耦网络优化设计等方面进行改进, 主要贡献如下:

1) 为减少 Neck 处理部分的目标特征损失, 使解耦后的各分支任务网络头提供更丰富的特征, 提出了一种结合细粒度特征和注意力机制的改进特征解耦网络 (cross-correlation network, CCN)—RES_CCN 网络。

2) 对网络的检测头部分进行解耦设计, 通过降低误检率提升多目标跟踪性能。

3) 针对多类别舰船跟踪任务, 通过增加类别损失, 将原算法扩展到多类别多目标跟踪任务。

4) 采用新加坡海事数据集 (Singapore maritime dataset, SMD)^[10] 和 MOT16^[11] 行人数据集进行测试验证, 验证了所提方法的有效性。

本文的组织结构如下: 第二节综述文献中的相关工作, 第三节介绍本文提出的方法, 第四节给出实验的结果和分析, 第五节对本文进行了总结。

2 相关工作

基于深度学习的多目标跟踪算法根据其网络模型完成目标检测、特征提取和数据关联三项子任务中的一项或几项可分成不同的范式, 分别是分离检测与特征提取的范式 (separate detection and embedding, SDE)、

联合检测与特征提取的范式 (joint detection and embedding, JDE) 和联合检测与跟踪的范式 (joint detection and tracking, JDT)^[2]。SDE 范式将检测和检测器进行数据关联分成两个独立的任务, 通常使用现成的目标检测器进行检测, 使用另一个独立的关联算法进行数据关联, 可以通过更换不同检测器实现性能的提升, 结构灵活, 但因采用两阶段处理, 所以效率低下。JDT 范式在一个网络里完成检测、特征提取以及 ID 分配三种任务, 可以利用网络的反向传播实现全局优化能力, 虽然大多 JDT 范式的主干网络共享特征, 但检测网络通常独立进行, 没有利用跟踪网络的线索和结果特征, 且常见的 ReID 损失与检测损失不兼容也会造成网络的检测性能下降^[12]。JDE 范式在一个网络里完成目标检测和 ReID 特征提取两种任务, 相比 SDE 范式, JDE 范式集成度更高; 相比 JDT 范式, JDE 范式充分考虑检测任务和 ReID 任务之间的关系, 实现了任务之间优化。

本文方法基于舰船跟踪的应用背景, 从 JDE 跟踪范式和 CStrack 跟踪方法出发, 进一步改进了 CStrack 跟踪方法, 以实现更强大的舰船多目标跟踪能力, 与本文方法相关工作的介绍具体介绍如下。

2.1 JDE 范式跟踪方法

JDE 范式跟踪方法又被称为单阶段跟踪方法, 其框架最早由 Xiao 等人修改 Faster RCNN^[13] 检测器提出, 通过引入额外的全连接层, 检测器获得产生 ID 嵌入的能力。随后 Wang 等人提出 JDE 多目标跟踪算法^[14], 然而其效果并不如 SDE 范式的算法, 为了提升 JDE 范式跟踪方法的性能, 研究人员陆续提出了一系列改进算法。除上文提到的 FairMOT^[8]、CStrack^[9]、RelationTrack^[15] 与 CStrack 类似, 都是设计特征解耦模块, 通过将特征解耦成两部分来满足各自任务的需求, 不同的是 CStrack 对来自不同特征层的 ReID 向量进行特征对齐, 设计了具有多尺度学习能力的尺度敏感注意网络 (scale-aware attention network, SAAN), 而 RelationTrack 则考虑全局语义相关性, 设计了引导式变换编码网络学习全局感知 ReID 特征。此外还有研究者利用时空信息来提升多目标跟踪的性能。Wan 等人^[16] 设计了一种多通道时空特征, 将目标的外观和运动特征编码到不同通道中, 通过更丰富的特征来兼顾检测和 ReID 任务。将注意力机制用于提升多目标跟踪性能时, Meng 等人^[17] 使用基于语义特征的空间注意力机制分析历史帧的运动轨迹从而预测目

标当前位置。Guo 等人^[18] 提出了目标注意力机制和干扰物注意力机制, 有效增强模型区分不同目标的能力。

2.2 基于深度学习的舰船多目标跟踪方法

基于深度学习的舰船多目标跟踪方法研究主要涉及两个方面: 舰船数据集的研究和基于深度学习的舰船检测和跟踪方法研究。质量或相关性高的数据集是训练高性能网络模型的先决条件, 当前常用的舰船公开数据集包括: Bloisi 等人在 2015 年发布的 MarDCT 数据集^[19], Prasad 等人在 2017 年发布的 SMD 数据集, Shao 等人在 2018 年发布的 SeaShips 数据集^[20], 以及基于海面浮动摄像头拍摄的数据集 Buoy^[21] 等。基于深度学习的舰船检测和跟踪算法亦有诸多研究^[22-26], 徐安林^[22] 等人提出一种结合层次化搜索与视觉残差网络的光学舰船目标检测方法, 对于光学遥感图像中的舰船目标检测具有良好的抗干扰性能。于国莉等人^[23] 利用相关滤波的方法, 将相关图中最大响应值位置作为舰船目标最新位置, 输出舰船目标实时跟踪结果。Li 等人^[24] 提出基于图匹配的舰船目标跟踪方法, 将检测网络输入的连续帧图像替换为图数据结构, 图中的点和线表示目标之间的位置信息以实现跟踪的目的。周越东^[25] 在特征匹配方面引入多粒度网络提取目标外观信息, 并根据目标出现次数设置优先级实现出色的舰船跟踪。陈庆林^[26] 在 DeepSORT 的基础上提出了一种基于外观余弦距离平均值的改进方法, 减少了舰船跟踪过程中的 ID 切换实现性能提升。

3 本文方法

为详细说明本文方法, 本节内容安排如下: 首先在 3.1 节介绍 CStrack 方法和网络结构, 然后在 3.2 节介绍本文提出的改进 CStrack 多目标跟踪算法, 最后在 3.3 节介绍数据关联算法。

3.1 CStrack 网络结构

CStrack 在 JDE 算法框架的基础上进行改进, 如图 1 所示。图 1(a) 和 1(b) 分别是 JDE 和 CStrack 的算法流程图, 检测任务要求同一类别的物体 (例如行人) 具有相似的语义信息, 类间距离最大, 而 ReID 任务则要求同类目标具有不同的语义信息 (例如区分两个不同的行人), 类内距离最大。CStrack 算法设计时认为采用共享的网络参数和结构同时实现上述两个任务会存在竞争关系, 故利用共性和特性的思想在 JDE 算法基础上设计了解耦网络模块 CCN, 并针对

ReID 任务设计了多尺度的特征增强模块 SAAN, 以减少不同特征层的 ID 向量带来的噪声干扰, 从而提升多目标跟踪性能。

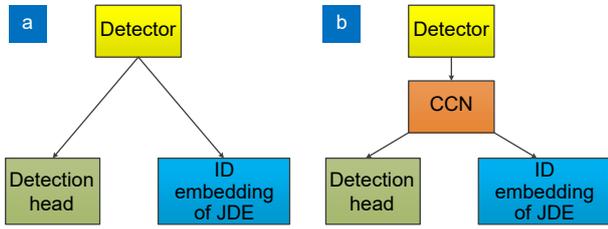


图 1 JDE 和 CStrack 算法流程图。(a) JDE; (b) CStrack
Fig. 1 Flowchart of the JDE and CStrack algorithms.
(a) JDE; (b) CStrack

如图 2(a) 所示, CStrack 算法的网络结构分为特征提取模块 (feature extraction module)、特征融合模块 (feature fusion module)、特征解耦模块 (feature decoupling module)、检测头分支 (detection head branch)、ReID 分支 (ReID branch), 其中 CCN 网络和 SAAN 网络分别如图 2(b) 的绿色虚线框和图 2(c) 所示。在 SAAN 网络中, 为了将网络不同层的输出特征对齐,

分别将 1/16 和 1/32 输入图像大小的特征通过上采样方式都统一到 1/8 图像大小。此外引入空间注意力机制 (spatial attention module, SAM)^[9] 和通道注意力机制 (channel attention module, CAM)^[9], 分别如图 2(d) 和图 2(e) 所示。SAM 网络增强目标相关区域特征并抑制背景噪声, CAM 网络帮助模型更好地关注图像中重要特征通道, 提高模型的准确性和泛化能力。在 SAM 网络中, Maxout 和 Avgout 分别表示输入特征在第一维度上的最大值和均值, [Maxout, Avgout] 表示两种特征向量拼接, 将拼接后的向量经过卷积和激活, 最后得到输入向量的空间权重。在 CAM 网络中, Avgpool 和 Maxpool 分别表示平均池化和最大池化, 将池化后的向量分别经过卷积和全连接 (fully connected, FC), 最后将得到的两个特征使用元素乘法融合并经过激活函数得到输入向量的通道权重。

CStrack 算法通过图 2 所示网络检测目标位置并提取外观特征向量, 然后采用数据关联算法实现 ID 分配。给定一帧图像 X , 首先通过特征提取网络和融合网络得到不同分辨率的特征层 $F_i, i=1,2,3$, 然后将

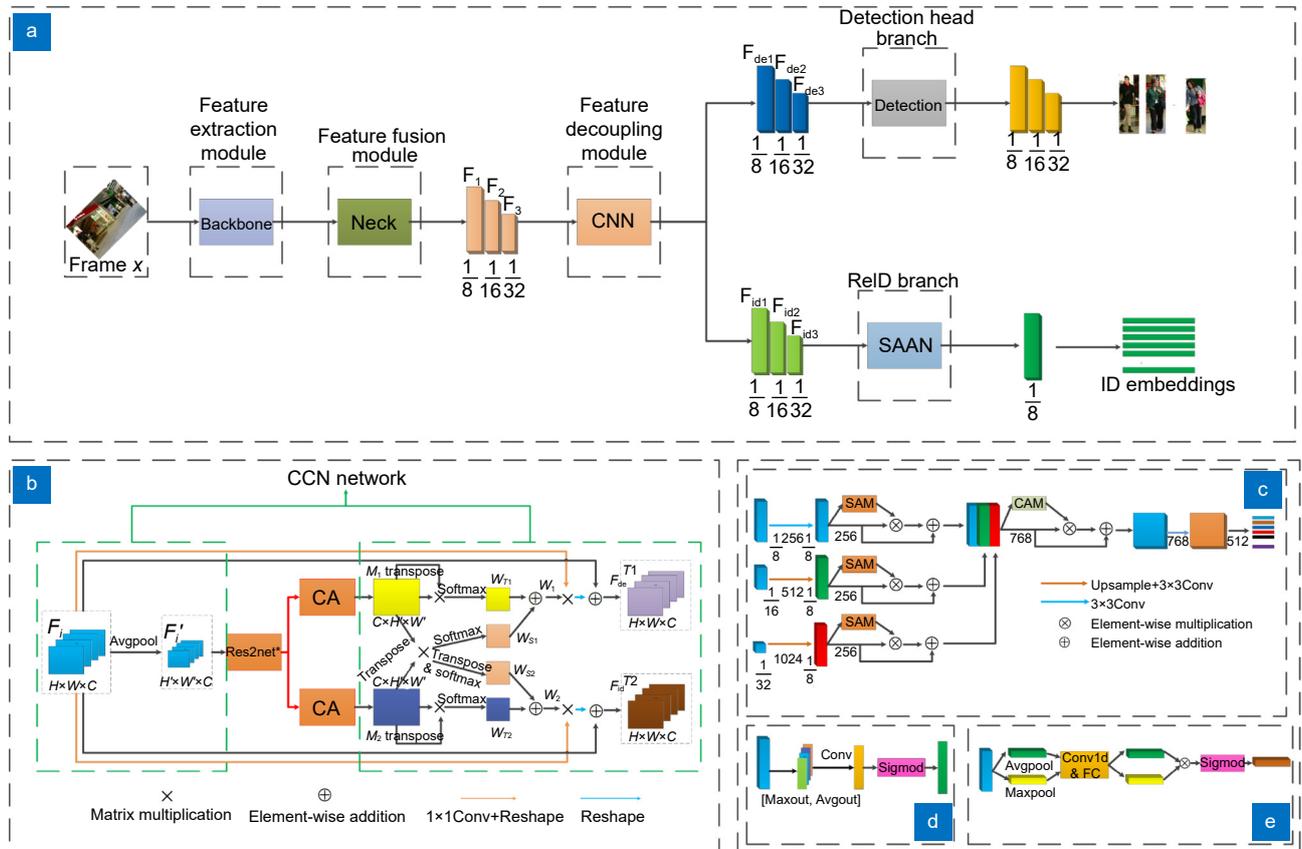


图 2 CStrack 网络结构。(a) 总体框架; (b) CCN 和 RES_CCN 网络; (c) SAAN 网络; (d) SAM 网络; (e) CAM 网络
Fig. 2 Network architecture of the CStrack. (a) Overall framework; (b) CCN and Res_CCN networks;
(c) SAAN network; (d) SAM network; (e) CAM network

$F_i, i = 1, 2, 3$ 特征都分别通过特征解耦网络模块得到适用不同任务的特征层 $F_{\text{det}}/i = 1, 2, 3$ 和 $F_{\text{id}}/i = 1, 2, 3$ 。最后将不同的特征层通过检测头分支和 ReID 分支得到对应预测检测结果以及原始的 ReID 特征, 其中预测检测结果经过非极大值抑制处理得到目标位置候选框。在数据关联阶段, 利用候选框的位置信息从 ReID 向量池中取出对应目标的 ReID 特征向量, 计算当前帧检测目标与前一帧跟踪目标的代价矩阵, 将代价矩阵作为匈牙利算法的输入特征以完成目标的 ID 分配。

3.2 改进 CStrack 多目标跟踪算法

针对 CStrack 多目标跟踪算法存在的问题, 本文从主干网络、特征解耦网络和检测输出网络等几个方面进行改进, 提出的算法总体框图和特征提取网络结构如图 3 所示, 现具体介绍如下。

3.2.1 特征提取网络改进

CStrack 采用了 YOLOv5^[27] 的早期版本作为特征提取和融合的网络基础。为实现更优秀的检测性能, 本文的方法对 CStrack 的特征提取和融合网络进行了优化和改进。本文将 CStrack 中使用的 YOLOv5 第三个版本的特征提取和融合网络升级为第七个版

本的网络。这二者的区别在于: 1) 将激活函数从 LeakyReLU 换成了 SiLU 函数, 保留更多的输入信息; 2) 将 Focus 模块换成了 Conv 模块, 减少模型的复杂度和计算量, 方便模型部署; 3) 将 BottleneckCSP 换成了 C3 模块, 提高模型的精度和泛化能力; 4) 将 Backbone 网络的最后两层 SPP 和 BottleneckCSP 分别换成 C3 和 SPPF 模块, 提高检测速度和准确率。网络结构如图 3 的红色虚线框所示。

3.2.2 结合改进 Res2net 和 CA (coordinate attention) 注意力机制的解耦模块 Res_CCN

Res_CCN 解耦网络是对 CCN 网络的改进, 其中 CCN 的网络结构如图 2(b) 的绿色虚线框所示。特征层首先经过自适应平均池化得到新的不同大小的特征层 F'_i , 接着 CCN 中会将新的特征层通过一对相同的卷积函数得到不同分支任务的特征 M_1 和 M_2 , 最后为了促进两个任务的学习, 利用注意力方式获得自注意力权重图和交叉相关性权重图, 自注意力权重图促进各自任务的学习, 交叉相关性权重图提高两个任务的协同学习。自注意力权重图 W_{T1} 、 W_{T2} 由 M_1 和 M_2 与其对应的转置张量进行矩阵乘法, 并将结果经过

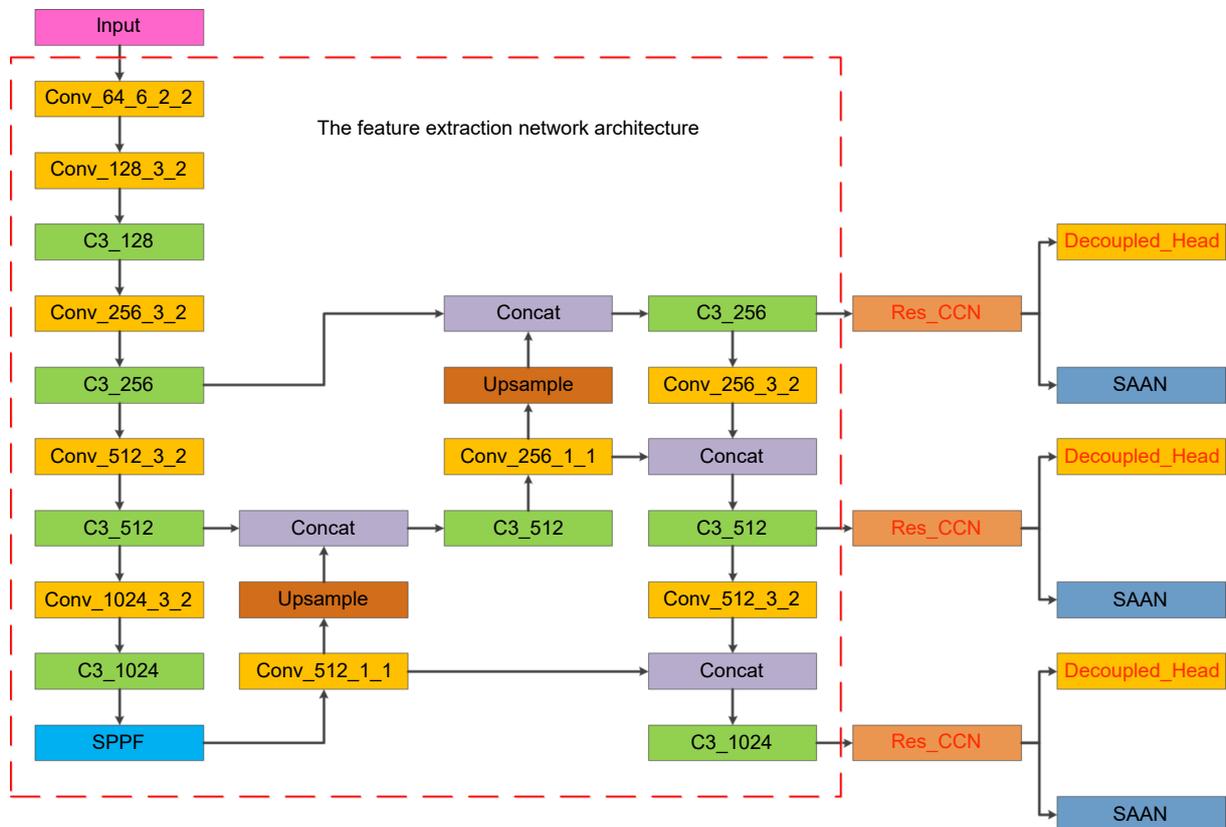


图 3 本文方法总体框图和特征提取网络结构

Fig. 3 Overall framework and feature extraction network architecture of the proposed method

softmax 函数得到, 如式 (1) 所示:

$$w_{Tk}^{ij} = \frac{\exp(m_k^i \cdot m_k^j)}{\sum_{j=1}^C \exp(m_k^i \cdot m_k^j)}, k \in \{1, 2\}, \quad (1)$$

式中: “ \cdot ” 为点乘运算, m_k^i 和 m_k^j 分别表示 M_1 或 M_2 的第 i 行和第 j 行, w_{Tk}^{ij} 表示权重映射图 w_{Tk} 上 (i, j) 位置处的值。

交叉相关性权重图 W_{s1} 由 M_1 和 M_2 的转置进行矩阵乘法之后经过 softmax 函数得到, 交叉相关性权重图 W_{s2} 由 M_1 和 M_2 的转置进行矩阵乘法之后再经过转置和 softmax 函数得到, 如式 (2) 所示:

$$w_{sk}^{ij} = \frac{\exp(m_k^i \cdot m_h^j)}{\sum_{j=1}^C \exp(m_k^i \cdot m_h^j)}, (k, h) \in \{(1, 2), (2, 1)\}, \quad (2)$$

式中: w_{sk}^{ij} 表示 M_1 或 M_2 的第 i 个特征通道对 M_2 或 M_1 的第 j 个特征通道的影响。最后得到的自相关和交叉相关性权重映射图通过训练权重 λ 进行融合得到 W_1 和 W_2 , 如式 (3) 所示:

$$W_k = \lambda_k \times W_{Tk} + (1 - \lambda_k) \times W_{sk}, k \in \{1, 2\}. \quad (3)$$

然后原始的特征权重 F_i 与 W_1 或 W_2 使用矩阵乘法得到增强的表示特征, 表示特征与原始特征图 F_i 使用元素加法进行特征融合, 最后得到属于不同分支任务的特征张量 F_i^{T1} 和 F_i^{T2} 。

对于 CCN 暴力解耦带来的特征损失问题, 本文从提升细粒度特征的角度在 CCN 解耦网络中添加了改进的 Res2net 网络, 改进后的 Res2net 在残差单元中引入更多的分层残差连接, 并结合 CA 注意力机制, 从而不仅能够获得更为精细化的特征, 同时还能让网络聚焦于与任务相关的特征, 为解耦后的检测头分支和 ReID 分支提供更多的细节特征, 接着解耦后每个

分支任务之间的差异性通过 CA 注意力机制进一步调整, 从而实现性能的提升。本文提出的 Res_CCN 解耦网络结构如图 2(b) 所示, 现分别介绍如下:

1) 改进的 Res2net 模块

Res_CCN 网络使用改进的 Res2net block^[28] 模块, 如图 4 所示。输入特征首先经过 1×1 卷积进行通道分组, 通过参数 s 控制通道组, 将输入平均划分到 s 个子集, 定义为 $\{X1, X2, \dots, Xs\}$, 这样每一个特征都有相同的尺度大小, 均是输入特征的 $1/s$, 除了第一个子集其余的子集接着通过不同的 3×3 卷积核, 对应的输出为 $\{Y1, Y2, \dots, Ys\}$, 将得到的子特征重新拼接并通过 1×1 的卷积得到最后的特征。每个子集的 3×3 卷积核都可以潜在地接受所有其左边的特征信息, 这样每一个输出都能增大感受野, 同时 Res2net 里面的分层残差连接使感受野在更细粒度级别上能够捕获细节和全局特性的变化。接着对特征再进行尺度的细化, 将 1×1 卷积得到的特征通过 CA 注意力机制, 这样在保持大范围感受野的同时使用注意力机制让网络更聚焦于任务相关的特征提升性能, 最后经过残差连接后使用 Silu 函数进行激活。

2) CA 注意力机制

注意力机制^[29-32] 已被证明有助于各种计算机视觉任务, 在 Res_CCN 解耦网络中也同样使用注意力机制, 经典的注意力机制有 CA^[29]、SE^[30](squeeze-and-excitation attention)、CBAM^[31](convolutional block attention module)、ECA^[32](efficient channel attention) 等。Res_CCN 网络的构建目标, 一方面是为了消除解耦模块可能导致的特征信息丢失, 另一方面也着重于精确定位各类任务所需的特定目标特征, 以减少不准确目标定位对 ReID 特征向量可能产生的干扰。其中, CA 注意力机制不仅关注通道的信息, 还深入考

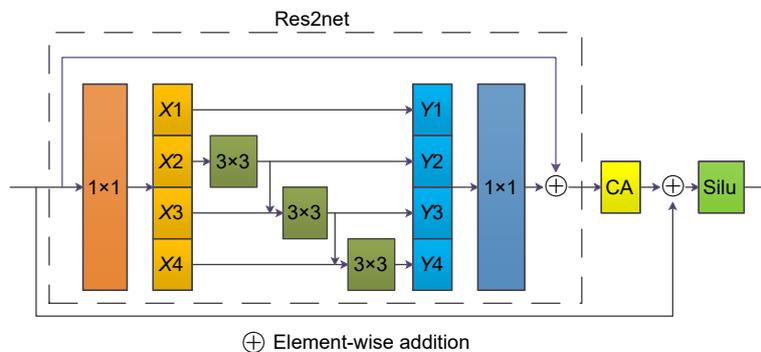


图 4 改进的 Res2net 网络结构
Fig. 4 Network architecture of the improved Res2net

虑了方向相关的地理位置信息, 这种设计使得它能够高效地捕捉到大范围的位置信息, 同时又不会引入过多的计算负担, 从而进一步提升了模型的性能。鉴于 Res_CCN 网络的目标和 CA 注意力机制的独特性质, 本文决定采用 CA 注意力机制来增强网络的特征提取能力。

CA 注意力机制网络结构图如图 5 所示, 输入特征首先分别按照 X 方向和 Y 方向进行池化, 然后进行拼接和二维卷积计算进行降维, 接着将激活后的特征按照 X 和 Y 方向进行切分并分别通过不同的二维卷积和激活函数得到不同方向上的权重, 最后将输入特征与权重相乘得到最后的输出特征。

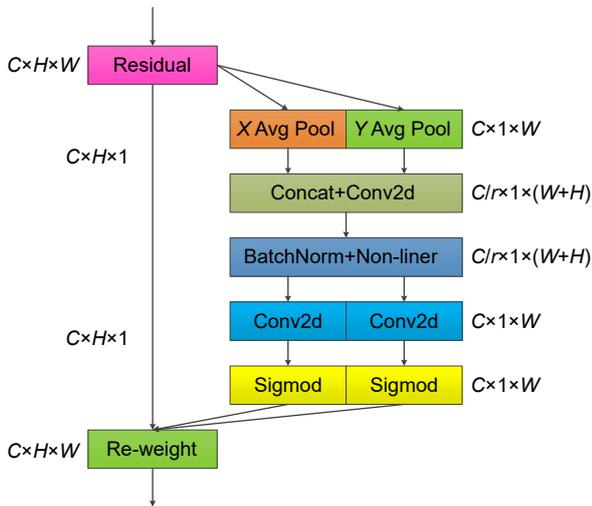


图 5 CA 网络结构
Fig. 5 Network architecture of CA

3.2.3 检测解耦头设计和多类别扩展

与行人的多目标跟踪不同, 舰船的多目标跟踪需要处理包含多种类别、尺度变化较大的目标, 这对算法的目标检测能力提出了更高的要求, 而多目标跟踪

的性能与检测的稳定性紧密相关, 因此, 本文从提升目标检测性能的角度出发, 通过降低误检率进一步提高多目标跟踪的表现。在 YOLOv6^[33]、YOLOX^[34] 以及最新的 YOLOv8 中, 解耦检测头被证明可以提高对目标的检测能力, 因此本文对 CStrack 的检测头进行解耦设计, 解耦检测头的网络设计如图 6 所示, 特征金字塔网络 (feature pyramid network, FPN) 的每层输出特征首先会通过一个内核大小为 1x1 的 Conv 卷积模块, 接着将得到的结果分别通过两个内核大小为 3x3 的 Conv 卷积模块, 最后利用二维卷积得到不同任务的输出结果。

由于 CStrack 仅支持行人单一类别目标的跟踪, 针对 SMD 数据集等包含多个舰船类别目标的跟踪需求, 本文对损失函数进行改进, 添加了分类损失, 使其能够同时检测多个类别的目标。最后总损失函数由定位损失、分类损失、置信度损失和 ReID 损失组成。其中, 定位损失采用 CIOU (complete-IOU) 损失, 而分类损失和置信度损失则采用二元交叉熵损失, 参数设置同 YOLOv5 保持一致, ReID 损失采用交叉熵损失, 其损失函数的定义如式 (4) 所示:

$$\mathcal{L}_{id} = -\frac{1}{N} \sum_{i=1}^N \sum_{c=1}^C Y^i(c) \log(p(c)), \quad (4)$$

其中: N 代表当前图像中目标的数量, $Y^i(c) \in \mathbb{R}^{C \times 1 \times 1}$ 表示真值标签, $P = \{p(c), c \in [1, 2, \dots, C]\}$ 表示 ID 分布向量表。最后多类别扩展的总损失函数定义如式 (5) 所示:

$$\mathcal{L}_{total} = \alpha \mathcal{L}_{cls} + \beta \mathcal{L}_{box} + \mathcal{L}_{obj} + \eta \mathcal{L}_{id}, \quad (5)$$

其中: \mathcal{L}_{cls} 、 \mathcal{L}_{box} 、 \mathcal{L}_{obj} 、 \mathcal{L}_{id} 分别表示分类损失、定位损失、置信度损失和 ReID 损失, α 、 β 、 η 是常数, α 为 0.5, β 为 0.05, η 在程序中设定为 4×10^{-4} 用于平衡目标检测任务和 ReID 任务。

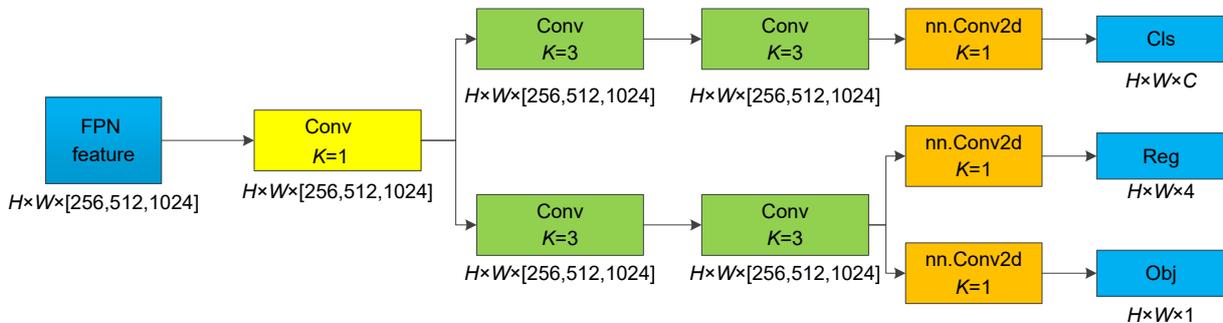


图 6 解耦检测头网络结构
Fig. 6 Network architecture of decoupled head

3.3 数据关联

根据目标检测及提取 ReID 特征的结果进行数据关联, 流程如图 7 所示。首先基于第一帧中检测到的目标进行轨迹起始, 接着后续帧中检测目标和跟踪目标的匹配过程则分为三个阶段, 在第一阶段, 检测目标和前一帧中保留的跟踪目标使用 ReID 向量和位置信息分别计算余弦距离 D_r 和马氏距离 D_m , 然后通过加权的方式融合在一起, 得到代价矩阵 $D_{\text{cost}} = 0.98D_r + 0.02D_m$, 匈牙利算法利用代价矩阵得到第一次的目标分配结果, 成功匹配的目标会更新位置和 ReID 特征信息; 第二阶段使用交并比 (intersection over union, IOU) 距离计算目标的代价矩阵, 然后利用匈牙利算法完成目标和轨迹的第二次匹配; 第三阶段对两次都没匹配上的检测目标初始化为新的跟踪目标, 连续 30 帧未匹配上的跟踪目标则会被删除。

4 实验结果与分析

4.1 数据集和模型评价指标

实验所用的舰船数据集是新加坡海事数据集 SMD, 由 Prasad 等人在 2017 年公开, 分为甲板和岸上两种拍摄环境, 共 40 个视频序列, 包含十类目标。为了更具针对性, 本文选取其中跟海面舰船相关的六类目标进行实验, 分别是 Ferry、Vessel-ship、Speed-Boat、Boat、Kayak、Sail-boat。Moosbauer 等人^[35]分析了 SMD 数据集的分布情况, 提出了 SMD 数据集的训练集、验证集和测试集的划分方法。本文根据其论文的划分方法在进行结果测试时, 发现测试集中存在着类别不均衡的问题: Boat、Kayak、Sail-boat 这三类目标从测试开始到结束都没有出现相应的目标对象, 这对于分析这三类目标的检测和跟踪情况不利。因此, 本文在统计 SMD 不同视频序列中的目标类别及其数量之后, 对文献 [35] 的划分方法进行微调, 将训练集的三个视频序列 MVI_1452、MVI_1587、MVI_

1592 变为测试集, 以此补充测试集中缺少的类别和样本, 另外将三个验证集 MVI_1469、MVI_1578、MVI_0790 调整为训练集, 以此增加训练集的样本。相关视频序列的目标统计结果和划分情况如表 1 所示。

除采用舰船数据集进行算法测试, 本文还采用 MOT16 行人数据集进行消融实验以验证提出各模块的性能。消融实验训练集由 ETH、CityPerson、CalTech、MOT17、MOT15、CUDK-SYSU 和 PRW 构成, 其中 ETH 和 CityPerson 仅提供检测信息, 训练检测分支, 其余数据集同时提供 ID 和检测信息, 训练 ReID 和检测分支。验证和测试数据集由 MOT16 训练集构成, 共 7 个视频序列 5316 帧 79790 个标注检测框及 731 个行人 ID, 挑战主要包括摄像头抖动、视角变化、光照变化、行人密集等。

为了定量地评价多目标跟踪算法的性能, 通常结合多个指标对模型的跟踪性能进行评价, 其中最重要的两个指标是 MOTA (multiple object tracking accuracy) 和 IDF1 (identification F1 score), 此外还有多数跟踪占比 (mostly tracked, MT)、多数丢失占比 (mostly lost, ML) 指标。

MOTA 综合考虑了 FP (false positive)、FN (false negative) 和 IDS (ID switches) 三个指标, 通常该值越高, 多目标跟踪性能越佳。如式 (6) 所示:

$$MOTA = 1 - \frac{\sum(FN + FP + IDS)}{N(GT)}, \quad (6)$$

其中: $N(GT)$ 表示真实值的总数, FN 表示被预测为负样本的正样本数, FP 表示被预测为正样本的负样本数, IDS 表示整个跟踪过程中所有目标的 ID 交换次数, 其值越低越好。

IDF1 表示正确识别的检测与平均真实数和计算检测数之比, 综合考虑了目标 ID 的准确率和召回率, 主要反映数据关联的准确性。如式 (7) 所示:

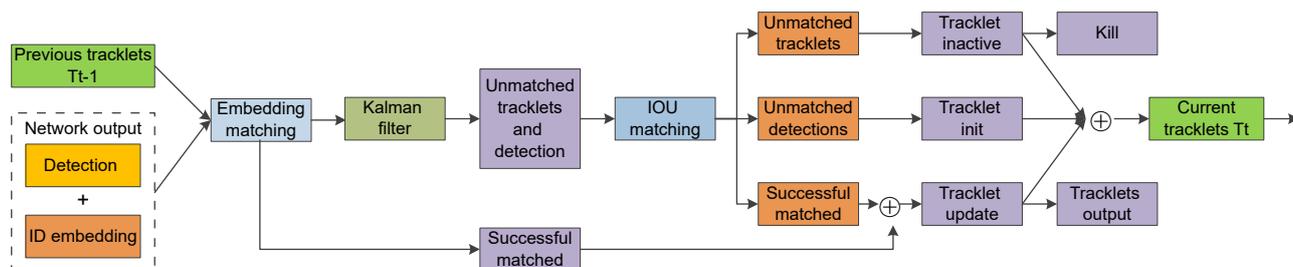


图 7 级联匹配流程图

Fig. 7 Flowchart of matching cascade

表 1 调整的 SMD 数据集视频序列相关参数
Table 1 Adjusted SMD dataset video sequence related parameters

SMD视频序列	视频帧数	Ferry	Vessel-ship	Speed-boat	Boat	Kayak	Sail-boat	调整前	调整后
MVI_1448	600	-	3210	1410	-	-	-	测试集	-
MVI_1474	445	890	3560	-	-	-	-	测试集	-
MVI_1484	600	600	1200	-	-	-	-	测试集	-
MVI_1486	600	1023	4200	-	-	-	-	测试集	-
MVI_1582	540	540	5400	-	-	-	-	测试集	-
MVI_1612	261	165	2349	-	-	-	-	测试集	-
MVI_1626	556	-	2775	-	-	-	-	测试集	-
MVI_1627	600	-	4200	-	-	-	-	测试集	-
MVI_1640	310	-	1677	274	-	-	-	测试集	-
MVI_0797	600	-	767	-	-	-	-	测试集	-
MVI_1587	600	-	7800	-	600	-	-	训练集	测试集
MVI_1592	491	491	2347	-	-	791	-	训练集	测试集
MVI_1452	340	-	1360	-	-	-	340	训练集	测试集
MVI_1469	600	-	3600	941	-	-	-	验证集	训练集
MVI_1578	505	-	3535	-	-	-	-	验证集	训练集
MVI_0790	600	-	70	-	140	-	-	验证集	训练集
MVI_0799	600	-	390	170	-	-	-	-	训练集

$$IDF1 = \frac{IDTP}{IDTP + 0.5IDFP + 0.5IDFN}, \quad (7)$$

其中: $IDTP$ 为正确分配数量, $IDFN$ 为漏分配数量, $IDFP$ 为错误分配数量。

多数跟踪占比表示 80% 以上的轨迹被成功跟踪的目标占比。

多数丢失占比表示 80% 以上的轨迹跟踪失败的目标占比。

4.2 实验环境和训练细节

本文使用 SGD 优化器在单个 NVIDIA GeForce RTX 3090 GPU 上训练网络 30 轮, 使用在 COCO 数据集上训练的 YOLOv5l 模型对网络模型进行初始化。初始学习率设为 5×10^{-4} , 在训练到 20 轮时, 会将学习率衰减到 5×10^{-5} 。在解耦网络中, 输入特征经过平均池化处理后的尺寸为 (H, W) , 在 SMD 数据集和行人数据集上, 其尺寸大小均设置为 (24,40)。对于不同数据集设置的 batch size 也不同, 舰船数据集设置的 batch size 为 4, 行人数据集设置的 batch size 为 16, 其余的参数设置都与 CStrack 算法保持一致。

4.3 消融实验

经本文训练并测试的 CStrack 算法在 MOT16 训练集上的指标如表 2 所示。其中 Baseline 表示

CStrack 方法在 MOT16 上测试得到的指标, Baseline* 表示 CStrack 方法改进特征提取网络后测试得到的指标, Res2net* 表示改进的 Res2net 模块。从表 2 中可以看出, Baseline* 相比 Baseline 在 MOTA 指标上提升了 2.8, Baseline+CA 相比 Baseline 在 MOTA 和 IDF1 指标上分别提升 3.0 和 0.4, Baseline+检测头解耦相比 Baseline 在 MOTA 和 IDF1 指标上分别提升 3.3 和 1.3, 证明这三个改进模块都能提升多目标跟踪性能; 在 Baseline*+Res2net* 和 Baseline*+Res2net 对比实验中, Baseline*+Res2net* 相比 Baseline*+Res2net 方法虽然 MOTA 值由 83.2 下降到 83.1, 但是 IDF1 值由 75.8 上升到 80.8, 目标切换次数由 758 下降到了 536, 证明本文的改进方法 Res2net* 相比 Res2net, 在保持 MOTA 指标相当的情况下在目标关联上有着更好的性能; 此外, 本文方法相比 Baseline 方法, MOTA 值由 79.4 上升到了 84.0, IDF1 由 77.9 上升到了 81.3, 同时 FP、FN、MT、ML、IDS 指标也均表现出不同程度的上升, 表明本文提出的模块结合到 baseline 中是有效的。

为了研究 RES_CCN 网络模块对 Baseline* 的影响, 本文也对 Baseline* 及 RES_CCN 内部使用的网络模块进行了定量分析, 如表 2 所示。Baseline*+Res2net* 相比 Baseline* 在 MOTA 和 IDF1 指标上分别提升 0.9、

表 2 MOT16 数据集上不同模块对跟踪性能的影响

Table 2 Influence of different modules on the tracking performance on MOT16 dataset

模型	MOTA↑	IDF1↑	FP↓	FN↓	MT↑	ML↓	IDS↓
Baseline	79.4	77.9	6235	15584	354	29	876
Baseline+Res2net*	82.8	79.7	4714	13966	390	21	616
Baseline+CA	82.4	78.3	4776	14022	377	21	642
Baseline+检测头解耦	82.7	79.2	4628	14318	375	28	571
Baseline*	82.2	75.4	4927	13801	389	23	875
Baseline*+Res2net	83.2	75.8	4459	13350	398	22	758
Baseline*+Res2net*	83.1	80.8	4413	13720	385	23	536
Baseline*+ Res2net* +CA注意力机制(Baseline*+RES_CCN)	83.4	81.9	4335	13434	393	18	571
Baseline*+ Res2net* +CA注意力机制+检测头解耦	84.0	81.3	4000	13107	400	20	480

5.4, 表明 Baseline*结合改进的 Res2net 模块能提升目标跟踪性能; 将改进的 Res2net 模块与 CA 注意力机制结合 Baseline*在 MOTA 提升 1.2, IDF1 提升 6.5, 同时 FP、FN、MT、ML、IDS 都有不同程度的上升, 表明改进 RES_CCN 网络对 Baseline*的跟踪性能提升同样是有有效的。

此外, 为评估 CA 注意力机制与其他常见注意力机制在网络跟踪性能方面的影响, 本文对四种典型的注意力机制进行了详细的对比测试, 包括 SE、CBAM、ECA 和 CA 注意力机制。采用这些注意力机制的实验结果如表 3 所示。其中 CA 注意力机制相比其他注意力机制拥有最高的 MOTA 值, IDF1、FN、IDS 的值也最好, 因此本文方法选用 CA 注意力机制提升网络的跟踪性能。

由于检测损失和 ReID 损失的权重参数的选择对

跟踪性能有较大影响, 本文通过实验分析其影响。实验中检测损失的权重参数固定, 将 η 从 4×10^{-2} 到 4×10^{-5} 变化, 多目标跟踪性能指标如表 4 所示。当 ReID 权重参数由 4×10^{-2} 下降到 4×10^{-4} 的时候, MOTA、FP、FN、MT、ML、IDS 的指标呈现变好的趋势, 但是 IDF1 值下降, 说明随着 ReID 权重参数的下降能提升算法的目标检测性能, 但是对于数据关联是不利的; 不过随着 ReID 权重参数不断下降, MOTA、FP、FN、MT、ML、IDS 指标也不再上升, 反而呈现下降的趋势。由上述分析可知, ReID 权重参数的下降在一定程度上有利于算法的检测性能提升, 但是不利于关联性能的提升。从表 4 中可以看出, 当权重参数选为 4×10^{-4} 时, MOTA 值最高, 漏检、误检以及目标切换次数最少, 因此本文方法设置为 4×10^{-4} 。

表 3 不同注意力机制对跟踪性能的影响

Table 3 Influence of different attention mechanisms on tracking performance

模型	MOTA↑	IDF1↑	FP↓	FN↓	MT↑	ML↓	IDS↓
SE	83.0	78.6	4624	13557	394	18	589
CBAM	83.6	80.8	4229	13402	391	20	491
ECA	80.5	79.3	3316	17806	351	29	489
CA	84.0	81.3	4000	13107	400	20	480

表 4 ReID 权重参数对跟踪性能的影响

Table 4 Influence of ReID weight parameters on tracking performance

ReID权重参数	MOTA↑	IDF1↑	FP↓	FN↓	MT↑	ML↓	IDS↓
4×10^{-2}	80.1	83.0	4685	14488	374	28	576
4×10^{-3}	81.1	82.6	4416	13687	388	23	530
4×10^{-4}	84.0	81.3	4000	13107	400	20	480
4×10^{-5}	83.5	80.2	4319	13379	396	22	530

4.4 定量分析

为测试和比较本文所提方法与其他先进方法在舰船多目标跟踪方面的性能, 本小节在 SMD 数据集上进行了定量分析。本文训练并测试了几种先进算法在 SMD 数据集上的指标, 如表 5 所示。CSTrack 算法因为使用特征解耦和 ReID 特征对齐所以相比 DeepSORT 算法 MOTA 值提升了 7.4, 提升明显, IDF1 值提升了 0.3, 但是相比 StrongSORT 和 ByteTrack 算法, CSTrack 则性能稍差。本文方法在 CSTrack 算法上进行改进, 通过 RES_CCN 网络保留了更多的细粒度特征以及对检测头的解耦操作, 提升算法的多目标跟踪性能, 相比 StrongSORT 方法, 本文方法的 MOTA 值高出 4.8, IDF1、FP、FN 等指标也都优于 StrongSORT; 相比 ByteTrack 方法, MOTA 值高出 2.1, FP、FN、ML 也都优于 ByteTrack; 相比 CSTrack 方法, MOTA 提升了 8.4, IDF1 提升了 3.1, FP、FN 和 ML 指标也都有不同程度的提升。MOTA 综合了 FP、FN、IDS 指标, 从而衡量算法的综合跟踪性能。从表 5 中可以看出, 相比其他方法, 本文所提方法的 MOTA 指标最优。

4.5 定性分析

将改进特征提取网络后的方法 CSTrack*作为基

准, 通过对实验的结果进行可视化来分析本文方法在目标遮挡、小目标或密集目标场景中等情况下的表现出的效果。由于原数据集的图片过大, 含有的目标比较多, 不方便对比分析, 截取了实验中间比较有代表性的一些跟踪结果来比较效果, 分别如图 8 和图 9 所示。

如图 8(a) 所示, 在 MVI_1627_VIS 的视频序列中, 基准方法在第 397 帧中对尺寸较大船只旁边的小船检测不到, 出现漏检, 而本文方法则可以正确检测, 说明本文方法相比基准算法能保留更多的特征信息; 在 MVI_1484_VIS 视频序列中, 基准方法出现了误检, 将船舱部分检测为一个新的目标, 而本文方法则检测正确, 且置信度较高。如图 8(b) 所示, 在 MVI_1640_VIS 的视频序列中, 基准方法和本文方法对连在一块的两个船只都出现了漏检现象, 而在第 36 帧的时候基准方法出现了 ID 切换, 本文方法则保持原来的 ID。通过以上的定性分析证明本文的改进特征解耦模块和检测解耦头设计在海面舰船跟踪方面发挥了作用。

在图 9(a) 所示的 MOT16-04 视频序列中, 第 001 帧中人流密集的场合基准方法出现了误检, 而本文方法则检测正确; 在 MOT16-13 视频序列的第 367 帧中, 基准方法出现了漏检, 而本文方法则能正确地检测出。

表 5 本文方法与其他先进方法在 SMD 数据集上跟踪表现的对比结果

Table 5 Comparison of tracking performance between the proposed method and other state-of-the-art methods on SMD dataset

算法	MOTA \uparrow	IDF1 \uparrow	FP \downarrow	FN \downarrow	MT \uparrow	ML \downarrow	IDS \downarrow
DeepSORT	31.1	62.3	21678	11082	69	25	224
StrongSORT	42.1	65	13264	17233	63	21	224
ByteTrack	44.8	67.3	9387	17003	57	26	49
CSTrack	38.5	62.6	9760	19617	48	33	109
本文方法	46.9	65.7	6658	16565	43	23	172

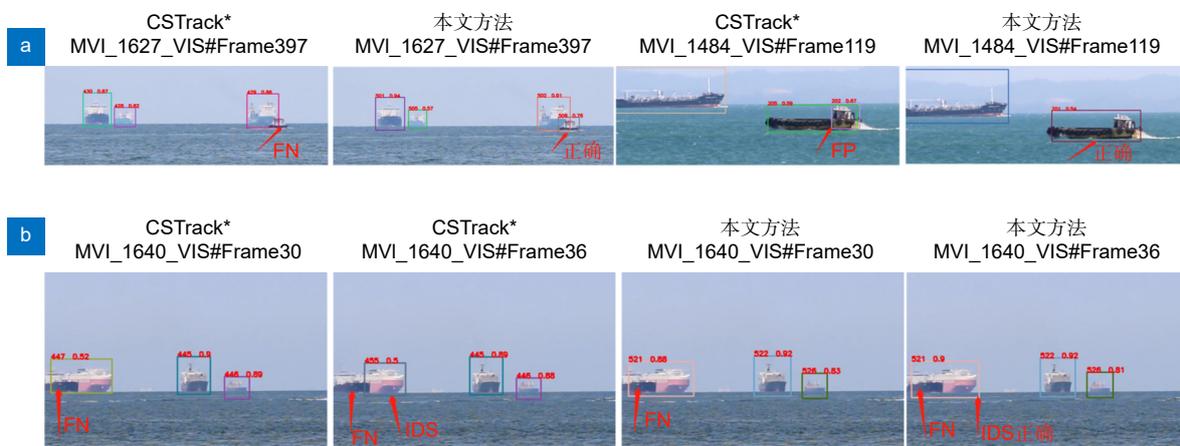


图 8 本文方法与基准方法在 SMD 验证集上的可视化结果对比。(a)漏检和误检; (b) ID 切换和漏检

Fig. 8 Comparison of visualization results between our method and baseline on SMD validation set. (a) FN and FP; (b) ID switch and FN

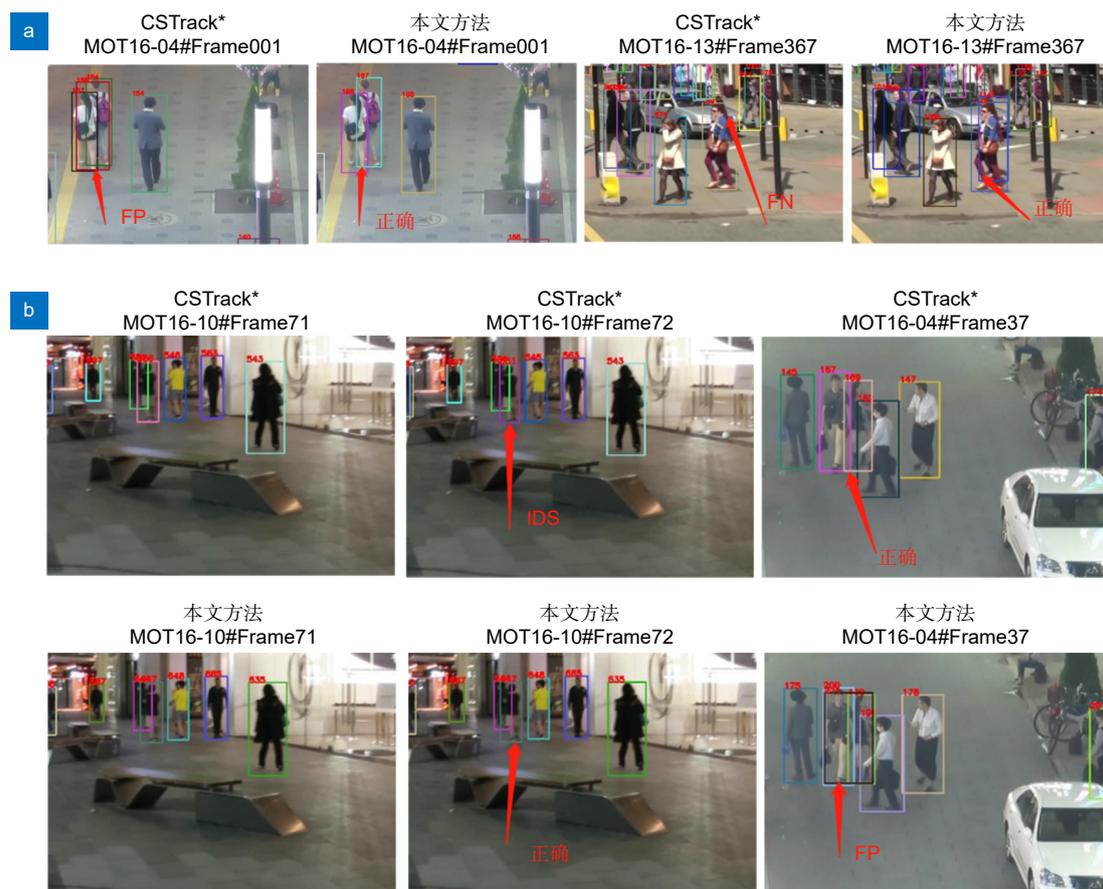


图 9 本文方法与基准方法在 MOT 验证集上的可视化结果对比。(a) 误检和漏检; (b) ID 切换和特定的误检

Fig. 9 Comparison of visualization results between our method and baseline on MOT validation set.

(a) FP and FN; (b) ID switch and special FP

在图 9(b) 所示的 MOT16-10 视频序列中, 基准方法的第 72 帧相比前一帧出现了 ID 切换, 重新生成了一个新的跟踪目标, 由粉红色目标框切换为了紫色的目标框, 而本文方法在人流密集的场所仍能保持 ID。通过上面的定性分析, 证明本文所提方法的有效性。不过本文方法同样存在一定的问题, 如在 MOT16-04 视频序列的 37 帧中, 本文方法相比基准方法出现了误检, 本文方法在分别检测出两个行人的同时将两个行人又作为一个整体的检测对象, 虽然这种误检发生的概率极低, 但仍然影响跟踪器的性能, 可能因为改进的 Res2net 在特征解耦前关注了大范围的位置信息, 带来了多余噪声信息导致出现误检。

5 结论

CSTrack 通过解耦的方式虽然能够缓解目标检测和目标重识别任务之间的矛盾, 提升多目标跟踪的性能, 但这种暴力解耦的方法会导致目标特征损失, 从

而降低算法跟踪性能。为提高 CSTrack 算法的海面舰船多目标跟踪性能, 本文在算法多个方面进行改进。在网络的解耦部分, 提出了一种结合细粒度特征和注意力机制的改进特征解耦网络 RES_CCN, 其通过增加感受野并在残差单元中插入更多带层级的残差连接结构, 获得了更加细粒度的特征, 提升了多目标跟踪的性能。为满足多类别舰船多目标跟踪的需求, 采用检测头解耦网络及改进损失函数进一步提升了跟踪性能。最后, 本文在 SMD 和 MOT16 数据集上进行了测试和验证, 实验结果证明所提网络方法的有效性。但是, 本文方法也存在局限性, RES_CCN 中的 Res2net 网络部分在特征解耦之前关注了大范围的位置信息, 可能引入了多余的噪声信息, 从而影响了跟踪性能。因此, 下一步的研究重点是尽量减少网络引入的噪声, 以避免误检现象对跟踪性能的影响。

利益冲突: 所有作者声明无利益冲突

参考文献

- [1] Ciaparrone G, Sánchez F L, Tabik S, et al. Deep learning in video multi-object tracking: a survey[J]. *Neurocomputing*, 2020, **381**: 61–88.
- [2] Wu H, Lie J H, Zhang Z W, et al. Deep learning-based visual multiple object tracking: a review[J]. *Comput Sci*, 2023, **50**(4): 77–87.
伍瀚, 聂佳浩, 张照妮, 等. 基于深度学习的视觉多目标跟踪研究综述[J]. *计算机科学*, 2023, **50**(4): 77–87.
- [3] Wang G A, Song M L, Hwang J N. Recent advances in embedding methods for multi-object tracking: a survey[Z]. arXiv: 2205.10766, 2022. <https://doi.org/10.48550/arXiv.2205.10766>.
- [4] Xiao T, Li S, Wang B C, et al. Joint detection and identification feature learning for person search[C]//*Proceedings of 2017 IEEE Conference on Computer Vision and Pattern Recognition*, Honolulu, 2017: 3376–3385. <https://doi.org/10.1109/CVPR.2017.360>.
- [5] Wojke N, Bewley A, Paulus D. Simple online and realtime tracking with a deep association metric[C]//*Proceedings of 2017 IEEE International Conference on Image Processing*, Beijing, 2017: 3645–3649. <https://doi.org/10.1109/ICIP.2017.8296962>.
- [6] Du Y H, Zhao Z C, Song Y, et al. StrongSORT: make deepSORT great again[Z]. arXiv: 2202.13514, 2023. <https://doi.org/10.48550/arXiv.2202.13514>.
- [7] Zhang Y F, Sun P Z, Jiang Y, et al. ByteTrack: multi-object tracking by associating every detection box[C]//*Proceedings of the 17th European Conference on Computer Vision*, Tel Aviv, 2022: 1–21. https://doi.org/10.1007/978-3-031-20047-2_1.
- [8] Zhang Y F, Wang C Y, Wang X G, et al. FairMOT: on the fairness of detection and re-identification in multiple object tracking[J]. *Int J Comput Vis*, 2021, **129**(11): 3069–3087.
- [9] Liang C, Zhang Z P, Zhou X, et al. Rethinking the competition between detection and ReID in multiobject tracking[J]. *IEEE Trans Image Process*, 2022, **31**: 3182–3196.
- [10] Prasad D K, Rajan D, Rachmawati L, et al. Video processing from electro-optical sensors for object detection and tracking in a maritime environment: a survey[J]. *IEEE Trans Intell Transp Syst*, 2017, **18**(8): 1993–2016.
- [11] Milan A, Leal-Taixé L, Reid I, et al. MOT16: a benchmark for multi-object tracking[Z]. arXiv: 1603.00831, 2016. <https://doi.org/10.48550/arXiv.1603.00831>.
- [12] Wu J L, Cao J L, Song L C, et al. Track to detect and segment: an online multi-object tracker[C]//*Proceedings of 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, Nashville, 2021: 12347–12356. <https://doi.org/10.1109/CVPR46437.2021.01217>.
- [13] Ren S Q, He K M, Girshick R, et al. Faster R-CNN: towards real-time object detection with region proposal networks[J]. *IEEE Trans Pattern Anal Mach Intell*, 2017, **39**(6): 1137–1149.
- [14] Wang Z D, Zheng L, Liu Y X, et al. Towards real-time multi-object tracking[C]//*Proceedings of the 16th European Conference on Computer Vision*, Glasgow, 2020: 107–122. https://doi.org/10.1007/978-3-030-58621-8_7.
- [15] Yu E, Li Z L, Han S D, et al. RelationTrack: relation-aware multiple object tracking with decoupled representation[J]. *IEEE Trans Multimedia*, 2022, **25**: 2686–2697.
- [16] Wan X Y, Zhou S P, Wang J J, et al. Multiple object tracking by trajectory map regression with temporal priors embedding[C]//*Proceedings of the 29th ACM International Conference on Multimedia*, 2021: 1377–1386. <https://doi.org/10.1145/3474085.3475304>.
- [17] Meng F J, Wang X Q, Wang D, et al. Spatial-semantic and temporal attention mechanism-based online multi-object tracking[J]. *Sensors*, 2020, **20**(6): 1653.
- [18] Guo S, Wang J Y, Wang X C, et al. Online multiple object tracking with cross-task synergy[C]//*Proceedings of 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, Nashville, 2021: 8132–8141. <https://doi.org/10.1109/CVPR46437.2021.00804>.
- [19] Bloisi D D, Iocchi L, Pennisi A, et al. ARGOS-Venice boat classification[C]//*Proceedings of the 2015 12th IEEE International Conference on Advanced Video and Signal Based Surveillance*, Karlsruhe, 2015: 1–6. <https://doi.org/10.1109/AVSS.2015.7301727>.
- [20] Shao Z F, Wu W J, Wang Z Y, et al. SeaShips: a large-scale precisely annotated dataset for ship detection[J]. *IEEE Trans Multimedia*, 2018, **20**(10): 2593–2604.
- [21] Ribeiro R, Cruz G, Matos J, et al. A data set for airborne maritime surveillance environments[J]. *IEEE Trans Circuits Syst Video Technol*, 2017, **29**(9): 2720–2732.
- [22] Xu A L, Du D, Wang H H, et al. Optical ship target detection method combining hierarchical search and visual residual network[J]. *Opto-Electron Eng*, 2021, **48**(4): 200249.
徐安林, 杜丹, 王海红, 等. 结合层次化搜索与视觉残差网络的光学舰船目标检测方法[J]. *光电工程*, 2021, **48**(4): 200249.
- [23] Yu G L, Sang J G, Li J R. Ship real-time target tracking and recognition technology based on improved convolutional neural network[J]. *Ship Sci Technol*, 2022, **44**(21): 152–155.
于国莉, 桑金歌, 李俊荣. 基于改进卷积神经网络的舰船实时目标跟踪识别技术[J]. *舰船科学技术*, 2022, **44**(21): 152–155.
- [24] Li G Y, Qiao Y L. A ship target detection and tracking algorithm based on graph matching[J]. *J Phys Conf Ser*, 2021, **1873**: 012056.
- [25] Zhou Y D. Research on ship multiple object tracking in remote sensing image based on deep learning[D]. Xi'an: Xidian University, 2021. <https://doi.org/10.27389/d.cnki.gxadu.2021.000391>.
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. <https://doi.org/10.27389/d.cnki.gxadu.2021.000391>.
- [26] Chen Q L. Research on automatic annotation and multi-target tracking algorithm for ship video target detection[D]. Hangzhou: Hangzhou Dianzi University, 2021. <https://doi.org/10.27075/d.cnki.ghzdc.2021.000349>.
陈庆林. 面向舰船视频目标检测的标注与多目标跟踪算法研究[D]. 杭州: 杭州电子科技大学, 2021. <https://doi.org/10.27075/d.cnki.ghzdc.2021.000349>.
- [27] Chen X, Peng D L, Gu Y. Real-time object detection for UAV images based on improved YOLOv5s[J]. *Opto-Electron Eng*, 2022, **49**(3): 210372.
陈旭, 彭冬亮, 谷雨. 基于改进YOLOv5s的无人机图像实时目标检测[J]. *光电工程*, 2022, **49**(3): 210372.
- [28] Gao S H, Cheng M M, Zhao K, et al. Res2Net: a new multi-scale backbone architecture[J]. *IEEE Trans Pattern Anal Mach Intell*, 2019, **43**(2): 652–662.
- [29] Hou Q B, Zhou D Q, Feng J S. Coordinate attention for efficient mobile network design[C]//*Proceedings of 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, Nashville, 2021: 13708–13717. <https://doi.org/10.1109/CVPR46437.2021.01350>.

- [30] Hu J, Shen L, Sun G. Squeeze-and-excitation networks[C]//*Proceedings of 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, Salt Lake City, 2018: 7132–7141. <https://doi.org/10.1109/CVPR.2018.00745>.
- [31] Woo S, Park J, Lee J Y, et al. CBAM: convolutional block attention module[C]//*Proceedings of the 15th European Conference on Computer Vision*, Munich, 2018: 3–19. https://doi.org/10.1007/978-3-030-01234-2_1.
- [32] Wang Q L, Wu B G, Zhu P F, et al. ECA-Net: efficient channel attention for deep convolutional neural networks[C]//*Proceedings of 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, Seattle, 2020: 11531–11539. <https://doi.org/10.1109/CVPR42600.2020.01155>.
- [33] Li C Y, Li L L, Jiang H L, et al. YOLOv6: a single-stage object detection framework for industrial applications[Z]. arXiv: 2209.02976, 2022. <https://doi.org/10.48550/arXiv.2209.02976>.
- [34] Ge Z, Liu S T, Wang F, et al. YOLOX: exceeding YOLO series in 2021[Z]. arXiv: 2107.08430, 2021. <https://doi.org/10.48550/arXiv.2107.08430>.
- [35] Moosbauer S, König D, Jäkel J, et al. A benchmark for deep learning based object detection in maritime environments[C]//*Proceedings of 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, Long Beach, 2019: 916–925. <https://doi.org/10.1109/CVPRW.2019.00121>.

作者简介



袁志安(1998-), 男, 硕士研究生, 主要从事图像目标识别、多目标跟踪的研究。

E-mail: 1426771675@qq.com



【通信作者】谷雨(1982-), 男, 博士, 副教授, 主要从事多源信息融合、遥感图像处理、目标检测、识别与跟踪等方面的研究。

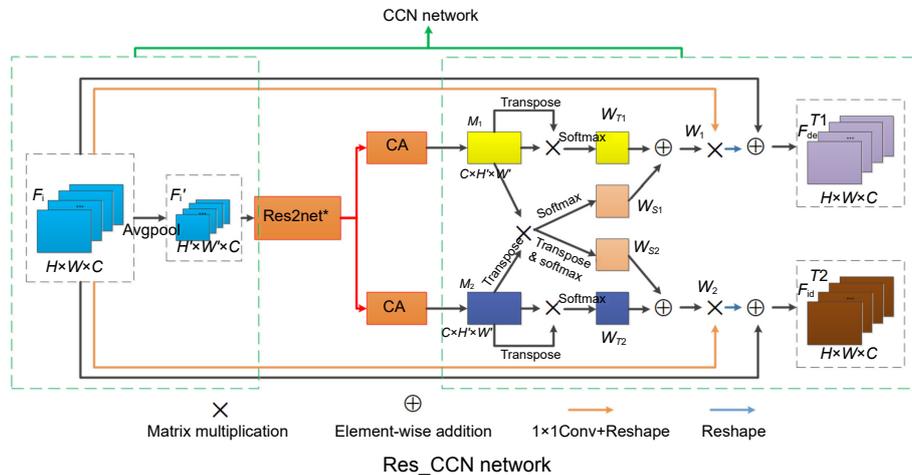
E-mail: guyu@hdu.edu.cn



扫描二维码, 获取PDF全文

Improved CTrack algorithm for multi-class ship multi-object tracking

Yuan Zhian, Gu Yu*, Ma Gan



Overview: Ship multi-object tracking is an important application scenario in the field of multi-object tracking (MOT), and can be widely applied in both military and civilian fields. The objective of MOT is to locate multiple ship objects and maintain a unique identification (ID) number for each ship object, and record its continuous trajectory. The difficulty of MOT lies in the uncertainty of false positives, false negatives, ID switches, and object numbers. The feature maps obtained by the neck part of the network in CTrack multi-object tracking algorithm are decomposed into two different feature vectors by decoupling, and are as the input of object detection and Re-identification networks respectively to alleviate the contradiction between these two tasks and improve the performance of multi-object tracking. However, this kind of violent decoupling will bring about the problem of object feature loss, which leads to the deterioration of tracking performance in the case of object occlusion, small objects, or dense objects. To solve this issue, an improved cross-correlation network (CCN) named RES_CCN which can extract fine-grained features is proposed in this paper. This network is composed of an improved Res2net network, coordinate attention, and CCN network, and is inserted between the neck and head modules of the network, so that more fine-grained features can be obtained by increasing the receptive field and inserting more hierarchical residual connection structures into the residual unit before feature decoupling. To meet the requirements of multi-class ship multi-object tracking and improve the detection performance of the algorithm, the decoupled design of the detection head network is used to predict class, confidence, and position of objects, respectively, and binary cross-entropy is used as class loss function and added to the total loss function. Finally, the ablation experimental results on the MOT2016 dataset show that the multiple object tracking accuracy (MOTA) of the proposed algorithm has an improvement of 4.6 compared with that of the original algorithm, and the identification F1 score (IDF1) is increased by 3.4. When tested on the Singapore maritime dataset, the MOTA of the proposed algorithm is improved by 8.4 compared with that of the original CTrack, and IDF1 is increased by 3.1, which are better than the performance of ByteTrack and other algorithms. The qualitative experimental results show that the proposed algorithm can effectively detect small objects and maintain object IDs in sea-surface scenarios. The algorithm proposed in this paper has the characteristics of high tracking accuracy and low error detection rate, and is suitable for ship multi-object tracking in sea-surface scenarios.

Yuan Z A, Gu Y, Ma G. Improved CTrack algorithm for multi-class ship multi-object tracking[J]. *Opto-Electron Eng*, 2023, 50(12): 230218; DOI: [10.12086/oe.2023.230218](https://doi.org/10.12086/oe.2023.230218)

Foundation item: Project supported by Natural Science Foundation of Zhejiang Province (LY21F030010, LZ23F030002)

School of Automation, Hangzhou Dianzi University, Hangzhou, Zhejiang 310018, China

* E-mail: guyu@hdu.edu.cn