E-mail: jig@aircas.ac.cn
Website: www.cjig.cn
Tel: 010-58887035
中国图象图形学报
JOURNAL OF IMAGE AND GRAPHICS

中图法分类号:TP391　　**文献标识码**:A　　**文章编号**:1006-8961(2024)07-1970-14

论文引用格式:Guo W, Liu Q G and Ding X M. 2024. Multi-object tracking using adaptive-IoU loss and hierarchical association. Journal of Image and Graphics, 29(07):1970-1983(郭文, 刘其贵, 丁昕苗. 2024. 自适应IoU损失和层级关联的多目标跟踪. 中国图象图形学报, 29(07):1970-1983)[DOI:10.11834/jig.230390]

自适应IoU损失和层级关联的多目标跟踪

郭文,刘其贵,丁昕苗*

山东工商学院信息与电子工程学院,烟台　264005

摘　要:**目的**　针对模糊行人特征造成身份切换的问题和复杂场景下目标之间遮挡造成跟踪精度降低的问题,提出AIoU-Tracker多目标跟踪算法。**方法**　首先根据骨干网络检测头设计了一个特殊的AIoU(adaptive intersection over union)回归损失函数,从重叠面积、中心点距离和纵横比3个方面去衡量,缓解了由于模糊行人特征判别性不足造成的身份切换现象;其次提出了一种简单有效的层级(hierarchical)关联策略,在高分检测框和低分检测框分别关联之后,充分利用关联失败检测框周围的嵌入信息再次进行关联,提高了在遮挡条件下多目标跟踪的关联精度。**结果**　通过一系列的对比实验,提出的AIoU-Tracker跟踪方法相比于FairMOT跟踪方法在MOT16数据集上,HOTA(higher order tracking accuracy)值由58.3%提高至59.8%,IDF1(ID F1 score)值由72.6%提高至73.1%,MOTA(multi-object tracking accuracy)值由69.3%提高至74.4%;在MOT17数据集上,HOTA值由59.3%提高至59.9%,IDF1值由72.3%提高至72.9%。**结论**　本文提出的特征平衡性跟踪方法,使边界框大小特征、热图特征和中心点偏移量特征在训练测试中达到了更好的平衡,使多目标跟踪结果更加准确。

关键词:多目标跟踪(MOT);数据关联;回归损失;特征平衡性;级联匹配方法

Multi-object tracking using adaptive-IoU loss and hierarchical association

Guo Wen, Liu Qigui, Ding Xinmiao*

School of Information and Electronic Engineering, Shandong Technology and Business University, Yantai　264005, China

Abstract: **Objective** Multiple object tracking (MOT) is a mainstream task in computer vision, which aims mainly to estimate the tracklets of multiple objects in videos and has important applications in the fields of autonomous driving, human-computer interaction, and human activity recognition. A large number of methods focus on improving the tracking performance based on the given detection results. Re-ID based trackers can be divided into two categories: separate detection and embedding (SDE) tracking models and joint detection and embedding (JDE) tracking models. The SDE tracking model tunes the detection model and the Re-ID model separately to optimize the model, but this leads to the disadvantage of the SDE tracking model being unable to perform real-time detection. The JDE tracking model performs object detection while outputting the object location and appearance embedding information for the next step of object association, thus improving the algorithm's operational speed. However, the JDE tracking method suffers from the problem of identity switching due to ambiguous pedestrian features and the degradation of tracking accuracy due to occlusion between objects in complex scenes. An adaptive intersection-over-union (AIoU)-tracker multi-object tracking algorithm is proposed to

收稿日期:2023-06-20;**修回日期**:2023-10-17;**预印本日期**:2023-10-23

*通信作者:丁昕苗　dingxinmiao@126.com

基金项目:国家自然科学基金项目(62072286, 61876100, 61572296);山东省研究生教育创新计划项目(SDYAL21211)

Supported by: National Natural Science Foundation of China (62072286, 61876100, 61572296); Shandong Provincial Graduate Education Innovation Plan Project (SDYAL21211)

address these issues. **Method** First, we utilize the backbone network detection head to design a special AIoU regression loss function that measures the overlap area, center point distance, and aspect ratio. This approach helps alleviate the problem caused by identity switching due to ambiguous pedestrian features. Second, we propose a simple and effective hierarchical association method to leverage the embedding information around association failure detection frames for Re-ID. The high-score detection frames and low-score detection frames are associated separately, improving the association accuracy of multi-object tracking under occlusion conditions. We utilize a variant of the DLA-34 network architecture as the backbone network. The model parameters are trained on the common objects in context (COCO) dataset and used to initialize the model. The experiments are conducted on a system running Ubuntu 16.04 with 64 GB of memory and a GTX2080Ti GPU. The software configuration includes CUDA 10.2. We train the model using the Adam optimizer for 30 epochs, with an initial learning rate of 10^{-4}. The learning rate is decayed to 10^{-5} after 20 epochs, and the batch size is set to 16. We apply standard data augmentation techniques, including rotation, scaling, and color jittering. The input image size is adjusted to 1 088 × 608 pixels, and the feature map resolution is set to 272 × 152 pixels. We evaluate our approach on the MOT Challenge benchmark, specifically the MOT16 and the MOT17 datasets. The experiments utilize various datasets, including CrowdHuman, MIX dataset (ETH, CityPerson, CUHKSYSU, Caltech, and PRW). The ETH and CityPerson datasets only provide bounding box annotations, so we only train the detection branch on these datasets. The Caltech, MOT17, CUHKSYSU, and PRW datasets provide both bounding box positions and ID annotations, allowing for training of both branches. To ensure a fair comparison, we remove the overlapping videos between the ETH dataset and the MOT17 test dataset. The CrowdHuman dataset only contains bounding box annotations, so we perform self-supervised training on it. To evaluate the tracking performance, we use several well-defined metrics, including higher-order tracking accuracy (HOTA), multi-object tracking accuracy (MOTA), ID F1 score (IDF1), false positive, false negative, and number of identity switches (IDs). MOTA primarily assesses the performance of the detection branch, IDF1 evaluates identity preservation, focusing on the association performance, and HOTA provides a comprehensive evaluation of both the detection branch and the data association performance. **Result** The performance of our method is compared with that of existing methods on two datasets. The comparison results are as follows: 1) our HOTA value is 59.8% on the MOT16 dataset, which is increased by 1.5% compared with the FairMOT. Our MOTA value is 74.4% on the MOT16 dataset, which is increased by 5.1% compared with the FairMOT. Our IDF1 value is 73.1% on the MOT16 dataset, which is increased by 0.5% compared with the FairMOT. 2) The HOTA value is 59.9% on the MOT17 dataset, which is increased by 0.6% compared with the FairMOT. The IDF1 value is 72.9% on the MOT17 dataset, which is increased by 1.6% compared with the FairMOT. In addition, we conduct ablation studies on the MOT17 dataset to verify the effectiveness of different components in our method, which demonstrates that the proposed method significantly outperforms the competition in multiple object tracking. In the ablation studies, we observe a decrease in the number of identity switches through the added AIoU regression loss function. We also visualize the predicted Re-ID feature extraction positions, bounding box size feature, heat map feature, and center point offset feature. The visualization results show that our method is more robust than FairMOT. Moreover, our hierarchical association method makes the association more robust. For example, even after two frames, obscured IDs can still be associated. **Conclusion** The proposed feature balancing tracking method achieves better balance among the bounding box size feature, heat map feature, and center point offset feature during training and testing, resulting in more accurate multi-object tracking results. In this study, we propose two improvement measures for the FairMOT framework. First, we design an AIoU regression loss module to optimize the detection branch, enabling it to optimize targets based on the current optimal distance and extract more accurate appearance features. Second, we optimize the Re-ID branch through a hierarchical association strategy module, utilizing three-level matching to enhance the tracking system's association performance. Experimental results demonstrate significant improvements on the MOT17 dataset, with HOTA increasing to 59.9%, IDF1 increasing to 72.9%, and MOTA increasing to 70.8%. However, a competition issue exists between the detection and Re-ID branches in the JDE tracking model, which can lead to a decrease in MOTA. Future research will focus on investigating this competition in the JDE tracking model.

Key words: multi-object tracking (MOT); data association; regression loss; feature balance; hierarchical association method

0 引 言

多目标跟踪(multiple object tracking, MOT)的目的是在视频中估计多个对象的轨迹,在自动驾驶、人机交互和人类活动识别等领域具有重要意义,是计算机视觉的主流任务之一(Park 等,2021;Luo 等,2021;Wang 等,2021)。多目标跟踪可以根据检测模块和特征提取模块的结合状态分为两类:1)分离检测和嵌入(separate detection and embedding, SDE)跟踪模型;2)联合检测和嵌入(joint detection and embedding, JDE)跟踪模型。SDE 跟踪模型先输出目标的检测特征,然后再执行特征提取和数据关联操作,造成了 SDE 跟踪模型不能实时检测的后果。JDE 跟踪方法同时输出目标的检测特征和外观嵌入特征,然后再进行数据关联操作,极大地提高了算法的整体运行速度。在 MOT17 数据集上不同跟踪器的 HOTA(higher order tracking accuracy)-IDF1(ID F1 score)-MOTA(multi-object tracking accuracy)比较如图 1 所示,其中,横轴为 IDF1,纵轴为 HOTA,圆圈半径为 MOTA。

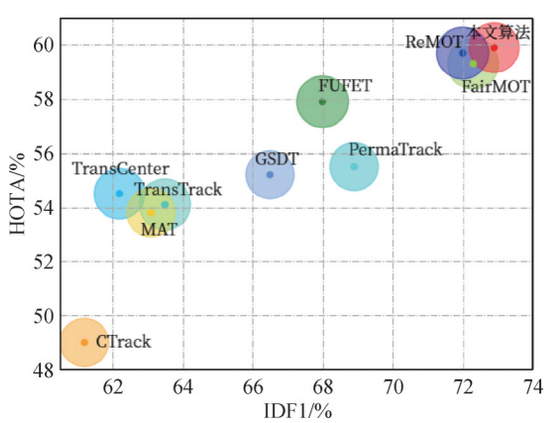


图 1 在 MOT17 数据集上不同跟踪器的 HOTA-IDF1-MOTA 比较

Fig. 1 Comparison of HOTA-IDF1-MOTA for different trackers on MOT17 dataset

SDE 跟踪模型将 MOT 建模为两个独立的任务:1)目标检测任务;2)特征提取与数据关联任务。首先,SDE 跟踪模型检测当前帧的对象,然后通过特征提取与数据关联模块将检测到的物体进行关联,从而生成目标的轨迹信息。由于 SDE 跟踪模型具有通俗易懂的特点和出色的跟踪精度,一度成为主流的多目标跟踪模型。SORT(simple online and realtime tracking)(Bewley 等,2016)通过卡尔曼滤波器来预测下一帧中的候选框,并使用匈牙利算法通过测量边界框的重叠面积进行关联。IoU-Tracker(Bochinski 等,2017)使用交并比(intersection over union, IoU)匹配直接将最后一帧边界框与当前帧候选框关联。DeepSORT(Wojke 等,2017)通过深度学习来提取外观特征,极大程度上减少了 ID(identity document)切换的次数;另外,还设计了一个级联匹配策略,提高了模型的关联能力。

为了平衡跟踪精度和跟踪速度之间的关系,MOTS(Voigtlaender 等,2019)提出通过单个卷积网络 Mask R-CNN(region convolutional neural network)(He 等,2017)解决检测和跟踪问题,通过多帧并行输入操作和 3D 卷积操作来增强模型提取多帧之间时域信息的能力,又添加了一个全连接层来提取数据关联的 ID 嵌入信息。Xiao 等人(2017)利用单个卷积网络(Faster R-CNN)(Ren 等,2017)处理行人检测任务和行人重新识别(re-identification, Re-ID)任务,又在 Faster R-CNN 网络后添加了一个全连接层来提取更鲁棒的 ID 嵌入信息。而 JDE(joint detection and embedding)(Wang 等,2020)重新设计了骨干网络的预测头,使用特征金字塔网络(feature pyramid network, FPN)(Lin 等,2017)来处理 MOT 任务,但其并没有处理好目标检测分支和 Re-ID 分支之间的竞争现象,它通常将 Re-ID 分支作为次要任务,使跟踪精度很大程度上受检测分支的影响。CTracker(Peng 等,2020)将目标检测模块、特征提取模块和数据关联模块融合到一个网络中,提出了一个链式跟踪框架,将相同目标相邻两帧信息作为骨干网络的输入,并设计了一个配对注意力模块来帮助模型更精确地回归,提高了模型的关联能力。

CenterTrack(Zhou 等,2020)将 CenterNet(Duan 等,2019)网络作为检测模块,使用检测轨迹中心点的空间距离作为数据关联的度量。FairMOT(Zhang 等,2021)将深度层级聚合网络(deep layer aggregation, DLA-34)作为骨干网络,在检测分支使用基于 anchor-free 的 CenterNet 检测网络,将 Re-ID 分支和检测分支集成到一个联合空间关联的框架中,减少了 anchor 带来的 ID 切换次数。为了更好地平衡检测分支和 Re-ID 分支之间的竞争,CSTrack(Liang 等,2022)提出了一种具有自我关系和交叉关系的新型互惠网络(reciprocal network, REN),帮助两个分支

学习更鲁棒的特征，为了平衡检测分支和 Re-ID 分支之间的竞争，又提出了一个可防止语义级别错位的尺度感知注意力网络（scale-aware attention network，SAAN）来提取更鲁棒的 ID 嵌入信息。

边界框回归是目标检测的关键步骤，但是在现有工作中，通常是采用 Ln 级范数去进行边界框回归，并使用 IoU 为评判标准，然而大部分工作并未针对 IoU 度量标准进行专门设计。有的工作虽然对 IoU 进行了设计，但是大多直接使用 IoU 作为回归损失，当检测框和目标框没有重叠时，IoU 回归损失表现出了极差的性能。Rezatofighi 等人（2019）提出 GIoU（generalized IoU）回归损失，它解决了 IoU 在检测框和目标框没有重叠时不能优化的缺点，但是 GIoU 在收敛速度和回归精度上仍有很大的提升空间。

Zheng 等人（2020）提出 DIoU（distance IoU）回归损失和 CIoU（complete IoU）回归损失。DIoU 回归损失结合了预测框和目标框之间的归一化距离，在训练中能够快速地收敛。CIoU 回归损失从边界框回归的中心点距离和纵横比两个几何因素去设计，使得收敛速度更快、性能更好。

SORT 使用分配矩阵计算所有预测框和目标框之间的 IoU 距离，较好地处理了目标遮挡的问题。JDE 针对前景和背景的分类使用了交叉熵回归损失，对于边界框回归使用了 L1 回归损失。FairMOT 针对检测分支使用了 L1 回归损失，针对 Re-ID 分支使用了交叉熵回归损失。

在数据关联上 SORT 直接使用了匈牙利算法，利用检测框的位置和大小进行数据关联。DeepSORT 在 SORT 基础上添加了深度关联度量模块和外观信息提取模块，深度关联度量模块类似于 Re-ID 网络，通过比较两个向量之间的距离来判断是否属于同一个目标，外观信息提取模块实现了长时间遮挡的目标跟踪。它还提出了一个级联匹配策略，首先将检测框和轨迹匹配，然后再与丢失的轨迹进行二次匹配，提高了模型数据关联的鲁棒性。

乐应英等人（2023）针对多目标跟踪场景中全局遮挡的问题，提出了一种自适应抗遮挡特征；又采用了一种级联筛查机制，减少了由于遮挡而产生信息误报的现象；并提出了一种自适应干扰模板更新机制，提高了模型的抗遮挡能力。ByteTrack（Zhang 等，2022）并未直接去掉低分检测框，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，从低分检测框中挖掘真正的目标（例如由运动模糊、遮挡问题导致的低分检测框），从而提高了数据关联的准确度。

上述方法部分解决了多目标跟踪中存在的问题，但是仍存在以下问题限制了其性能的提升：

1）SDE 跟踪模型存在运行速度慢、不能进行实时检测的缺点。而 JDE 跟踪模型，尤其是在 FairMOT 中，仅仅使用了简单的 L1 回归损失和交叉熵回归损失进行边界框回归，这在评判中必然会使得跟踪性能有所降低。尽管 IoU 回归损失、GIoU（Rezatofighi 等，2019）回归损失和 CIoU（Zheng 等，2020）回归损失都有利于 IoU 度量，但是它们却存在收敛缓慢和回归不准确的缺点。

2）在多目标跟踪进行关联操作时，并未充分利用检测框周围的外观嵌入信息，这使得检测分数低的物体（例如未被完全遮挡的物体）被直接删除，造成轨迹碎片化增多和关联精度下降的现象。

为了缓解上述问题，本文提出了 ***AE***（AIoU and embedding）矩阵模块和层级关联模块。首先，使用骨干网络对当前帧图像进行特征提取，然后将特征输出到检测分支和 Re-ID 分支；在检测分支输出边界框大小、热图和中心点偏移量特征后，使用 AIoU（adaptive IoU）回归损失矩阵进行融合，Re-ID 分支输出 ID 嵌入分支后，使用嵌入回归矩阵进行融合；再将 AIoU 回归损失矩阵和嵌入回归矩阵融合为 ***AE*** 回归损失矩阵，最后通过匈牙利算法和层级关联策略模块进行数据关联，具体操作流程如图 2 所示。关于模块的解释可分为两部分：

1）***AE*** 矩阵模块。在 FairMOT 中，利用骨干网络提取图像特征后，通过两个同构分支，分别提取检测特征和 Re-ID 特征，在检测分支输出边界框大小、热图和中心点偏移量。本文根据 3 个检测头输出的重叠面积、中心点距离和纵横比信息，专门设计了 AIoU 回归损失，帮助跟踪器更快、更精确地回归，还使用 Re-ID 分支输出的外观嵌入矩阵对目标进行远程跟踪。

2）层级关联策略模块。该模块在 ByteTrack 关联策略基础上，首先使用 ***AE*** 矩阵代替相似度矩阵进行匹配，在第 2 次匹配结束后，计算低分检测框周围 3 × 3 范围内嵌入向量与存储在未匹配跟踪对象嵌入向量之间的最小余弦距离，如果最小余弦距离小于阈值，则进行再次匹配。

本文的主要贡献如下：1）设计了一个 ***AE*** 矩阵模

块，不仅考虑了骨干网络预测头输出的信息，还使用IoU回归损失，解决了由于回归损失设计不当而造成的收敛速度慢和回归不准确的问题。2）提出一种更有效的层级关联模块，在层级关联匹配中使用***AE***矩阵代替相似度矩阵，经过二次匹配之后，挖掘低分检测框和未匹配成功检测框周围的外观嵌入信息，进行再次匹配，提高了跟踪器对外观嵌入信息的利用能力。3）得益于以上两个模块，设计了一个跟踪器AIoU-Tracker，跟踪器在MOT16和MOT17数据集上取得了较好的性能。

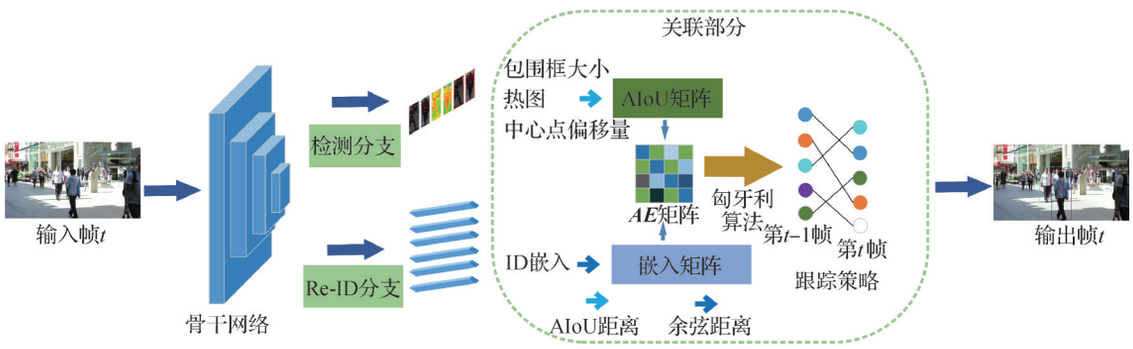


图2　跟踪流程

Fig. 2　Overview of the tracking process

1　AIoU-Tracker跟踪器

1.1　骨干网络

本文使用DLA-34网络改良版作为骨干网络，如图3所示。深层聚合网络（deep layer aggregation，DLA）具有更多底层特征和高层特征之间的跳跃连接，它巧妙地将稠密卷积（dense convolutional network，DenseNet）（Huang等，2018）和空间特征金字塔网络（FPN）进行了结合。稠密卷积可以聚合语义信息，而空间特征金字塔可以聚合空间信息。将所有卷积层替换为可变形卷积，可以更好地适应物体尺度和姿态的变化。

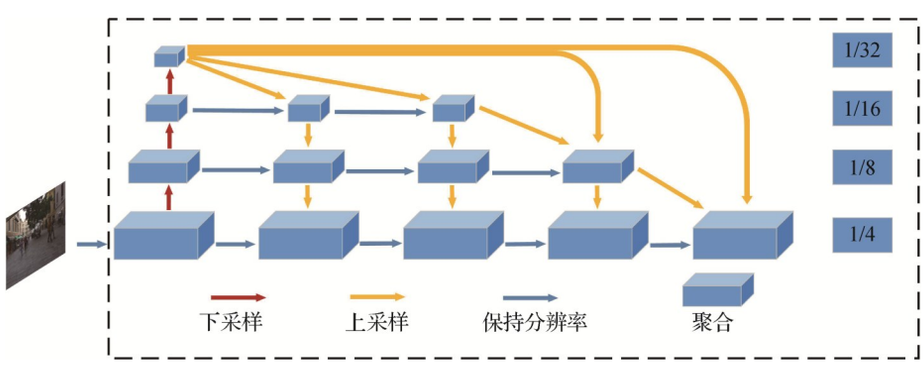


图3　骨干网络框架图

Fig. 3　Backbone network architecture diagram

1.2　*AE*矩阵模块

通过研究骨干网络预测头输出的边界框大小、热图和中心点偏移量后，对本文回归损失进行了针对性的设计。AIoU损失在IoU损失上添加了3个偏置项，分别为衡量跟踪对象边界框与检测对象边界框之间的中心点距离、纵横比（aspect ratio，AR）和重叠面积。AIoU损失考虑了3个几何属性，即重叠区域、中心点距离和纵横比，具有更快的收敛速度和更好的性能。

相似度矩阵通常由位置信息、运行信息和外观信息构成，假设$\boldsymbol{A}$，$\boldsymbol{M}$，$\boldsymbol{E}$分别为距离矩阵、运动矩阵和外观信息矩阵，将$\boldsymbol{A}$和$\boldsymbol{E}$融合为相似度量矩阵，简称为***AE***矩阵。L_{AIoU}可以表示为

$$L_{\mathrm{AIoU}}=1-\frac{|\boldsymbol{P}\cap\boldsymbol{Q}|}{|\boldsymbol{P}\cup\boldsymbol{Q}|}+L_{\mathrm{heat}}+L_{\mathrm{AR}}+L_{\mathrm{S}} \tag{1}$$

式中，$\boldsymbol{P}$和$\boldsymbol{Q}$分别表示跟踪对象的边界框和检测对

象的边界框，L_{heat}为衡量跟踪对象边界框与检测对象边界框之间的中心点距离，L_{heat}可以表示为

$$L_{heat}=\frac{\rho^2\left(b,b^{gt}\right)}{C^2} \tag{2}$$

式中，$\rho^2(b,b^{gt})$表示跟踪对象边界框中心点与检测对象边界框中心点之间的欧几里得距离，C为覆盖跟踪对象边界框和检测对象边界框最小封闭边界框的最大对角线距离。

L_{AR}为衡量跟踪对象边界框与检测对象边界框之间的纵横比，L_{AR}可以表示为

$$L_{AR}=\frac{\rho^2\left(w,w^{gt}\right)}{\left(w^c\right)^2}+\frac{\rho^2\left(h,h^{gt}\right)}{\left(h^c\right)^2} \tag{3}$$

式中，$\rho^2(w,w^{gt})$和$\rho^2(h,h^{gt})$分别表示跟踪对象边界框和检测对象边界框宽度和长度之间的欧几里得距离，w^c和h^c分别表示覆盖跟踪对象边界框和检测对象边界框最小封闭边界框的宽度和高度。

L_S为衡量跟踪对象边界框与检测对象边界框之间的重叠面积，L_S可以表示为

$$L_S=\frac{\left|S-\boldsymbol{P}\cup\boldsymbol{Q}\right|}{\left|S\right|} \tag{4}$$

式中，S为覆盖跟踪对象边界框和检测对象边界框最小封闭边界框的面积。

L_E可以表示为

$$L_E=\frac{\boldsymbol{O}_e^1\times\boldsymbol{O}_e^2}{\left\|\boldsymbol{O}_e^1\right\|\left\|\boldsymbol{O}_e^2\right\|} \tag{5}$$

式中，$\boldsymbol{O}_e^1$和$\boldsymbol{O}_e^2$分别表示不同的外观嵌入向量信息。

式(5)中的矩阵定义了余弦距离矩阵，可以与式(1)融合表示为$\boldsymbol{AE}$矩阵，L_{AE}可以表示为

$$L_{AE}=\lambda_1 L_{A\ IoU}+\lambda_2 L_E \tag{6}$$

式中，$\lambda_1=1.0$，$\lambda_2=0.5$。

1.3 层级关联策略模块

本文设计了一种层级匹配思想，即在原有二次匹配的基础上，充分利用未匹配检测框周围的嵌入信息完成第3次匹配，在级联匹配中使用$\boldsymbol{AE}$矩阵代替相似度矩阵。在二次匹配之后，根据检测框与未匹配轨迹周围嵌入信息之间的余弦距离来进行再次匹配，提高了匹配的成功率。

在算法的检索过程中，使用了卡尔曼滤波器预测未匹配跟踪目标的中心点位置，并且为了保证卡尔曼滤波器准确运行，又提取预测目标周围的外观特征。然后，计算这些向量与未匹配跟踪对象嵌入向量之间的最小余弦距离。如果最小余弦距离小于阈值，则启动跟踪检测机制，其中被遮挡的检测框可以通过卡尔曼滤波器来进行恢复。最后，将未匹配检测框和目标框进行第3次匹配。

假设AIoU-Tracker输入视频序列为$\boldsymbol{V}$，对象检测器为Det，检测分数阈值为$\boldsymbol{T}_{high}$，跟踪得分阈值为ε，跟踪检索阈值为ε_r。AIoU-Tracker的输出为视频轨迹$\boldsymbol{T}$，当前轨迹包含当前帧对象的边界框和ID标识。

在视频的当前帧中，使用检测器Det预测检测框和置信度分数。根据检测阈值分数$\boldsymbol{T}_{high}$将检测框分为$\boldsymbol{D}_{high}$和$\boldsymbol{D}_{low}$两部分。对于高于$\boldsymbol{T}_{high}$的检测框，将其放到高分检测框$\boldsymbol{D}_{high}$中，对于低于$\boldsymbol{T}_{high}$的检测框，将其放到低分检测框$\boldsymbol{D}_{low}$中，然后使用卡尔曼滤波器预测$\boldsymbol{T}$中所有轨迹在当前帧的新位置。

首先，计算高分检测框$\boldsymbol{D}_{high}$和所有轨迹$\boldsymbol{T}$之间的$\boldsymbol{AE}$相似度矩阵距离，然后使用匈牙利算法完成第1次相似度匹配。将未匹配的检测框保留在$\boldsymbol{D}_{low}$中，将未匹配的轨迹框保留在$\boldsymbol{T}_{remain}$中。然后，计算轨迹$\boldsymbol{T}_{remain}$与低得分检测框$\boldsymbol{D}_{low}$之间的$\boldsymbol{AE}$相似度矩阵距离，使用匈牙利算法完成第2次相似度匹配。将未匹配的检测框保留在$\boldsymbol{D}_{last}$中，将未匹配的轨迹框保留在$\boldsymbol{T}_{re\text{-}remain}$中，对$\boldsymbol{T}_{re\text{-}remain}$进行遍历操作，计算未匹配轨迹框$t_u$周围3×3的外观向量嵌入信息$\boldsymbol{E}_u$和低分检测框周围3×3的嵌入向量信息$\boldsymbol{E}_d$，如果$\boldsymbol{E}_u$和$\boldsymbol{E}_d$的最小余弦绝对值小于跟踪检索阈值$\varepsilon_r$，则将$t_u$暂存到$\boldsymbol{T}_{reback}$。计算检测框$\boldsymbol{D}_{last}$与轨迹$\boldsymbol{T}_{reback}$之间的$\boldsymbol{AE}$相似度矩阵距离，使用匈牙利算法完成第3次相似度匹配。

在进行3次关联之后将不匹配的轨迹从$\boldsymbol{T}$和$\boldsymbol{T}_{re\text{-}remain}$中删除。为了简单起见，在算法中并未列出轨迹复现的过程。在实际情况中，进行第3次关联之后，会将复现的轨迹存储到$\boldsymbol{T}_{lost}$中，只有当它们存在的时间超过特定帧数(例如30帧)后，才会将它们从轨迹$\boldsymbol{T}$中删除。在第3次关联之后，从未匹配的检测框$\boldsymbol{D}_{remain}$中初始化新的轨迹。

2 实验结果与分析

2.1 实验细节

本文使用DLA-34网络变体结构作为骨干网络，将在COCO(common objects in context)数据集上预先训练的模型参数用于初始化模型，本文实验的运行环境为Ubuntu 16.04系统，内存64 GB，GPU为

GTX2080Ti。软件配置为CUDA10.2，使用Adam优化器训练30个epoch，初始学习率为10^{-4}，学习速率在20个epoch后衰减到10^{-5}，批处理大小设置为16。本文使用标准的数据增强技术，包括旋转、缩放和颜色抖动。输入图像大小调整为1 088×608像素，特征图分辨率为272×152像素。

在跟踪阶段，默认高检测分数阈值$\boldsymbol{T}_{\text{high}}$设置为0.3，将跟踪分数阈值$\epsilon$设置为0.6，跟踪检索阈值$\epsilon_r$设置为0.1。在线性分配步骤中，对于高置信度检测，将分配阈值设置为0.8，对于低置信度检测，将分配阈值设置为0.4。

2.2 数据集

本文在MOT Challenge基准上进行评估，特别是MOT16和MOT17数据集（Milan等，2016）。实验使用了CrowdHuman数据集（Shao等，2018）、MIX数据集（ETH（Ess等，2008）、CityPerson（Zhang等，2017）、CUHKSYSU（Li等，2014）、Caltech（Dollár等，2009）和PRW（Zheng等，2017）数据集）、MOT16和MOT17数据集。由于ETH数据集和CityPerson数据集只提供框注释，所以只是在这两个数据集上训练了检测分支。Caltech数据集、MOT17、CUHKSYSU数据集和PRW数据集提供了边界框位置和ID注释，所以同时训练了两个分支。由于ETH数据集和MOT17测试数据集的一些视频相同，所以将相同的视频进行了删除。由于CrowdHuman数据集只包含边界框注释，所以只在上面进行了自监督训练，以上实验操作都与FairMOT操作保持一致。

为了评估跟踪性能指标，使用清晰的度量标准，包括HOTA、MOTA、IDF1、FP（false positive）、FN（false negative）和IDs（number of identity switches）等。MOTA比较关注检测分支性能，IDF1评估身份保持能力，更关注关联性能，HOTA可以综合地评估检测分支和数据关联的性能。

2.3 实验结果与分析

本文算法的实验在多目标跟踪数据集MOT16和MOT17上进行，并与现有方法进行实验对比。

2.3.1 实验结果与分析

为了验证所提出的损失函数对跟踪效果的影响，本文将框架中的回归损失函数分别替换为IoU回归损失、CIoU回归损失和AIoU回归损失，实验结果如表1所示。AIoU回归损失对HOTA、MOTA和IDF1都有所提升，由于AIoU回归损失是根据本文骨干网络预测头专门设计，不仅将IoU回归损失考虑了进去，还从重叠面积、中心点距离和纵横比3个方面去衡量，提高了模型收敛速度和跟踪器的鲁棒性。而FP和IDs上升的原因是由于AIoU损失模块使得本文跟踪器更加灵敏，从而使得误检的数量和目标切换的次数增多。但是，从FN可以看出，虽然误检的数量和目标切换的次数增多，但是漏检的数量却减少了，而且，HOTA、MOTA和IDF1都有不同程度的提升，所以适当地增多FP和IDs是有必要的。

表1 不同回归损失函数在MOT17数据集上对跟踪效果的影响

Table 1 The effect of different regression loss functions on tracking results on MOT17 dataset

损失	HOTA /%↑	MOTA /%↑	IDF1 /%↑	FP↓	FN↓	IDs↓
IoU	56.3	69.3	68.2	**14 466**	157 146	**1 710**
GIoU	58.0	72.0	70.4	33 681	122 619	2 115
CIoU	58.2	72.2	71.0	32 265	122 536	2 103
AIoU	**58.7**	**72.4**	**71.3**	36 057	**117 762**	2 082

注：加粗字体表示各列最优结果，“↑”表示值越高越好，“↓”表示值越低越好。

图4展示了GIoU损失、CIoU损失和AIoU损失对比。由表1和图4可以看出，GIoU损失、CIoU损失和AIoU损失的差异性。图4中右上角代表检测框，中心点代表目标框。GIoU损失仅仅扩大两个框的重叠区域，这导致了GIoU损失更倾向于增加预测框的大小，从而使得预测框向目标移动的速度非常缓慢；特别是对于水平和垂直的边界框，GIoU损往往需要更多的迭代才可以收敛。CIoU损失仅仅从归一化中心点距离和纵横比两个方面进行了设计，而AIoU损失从中心点距离、重叠面积和纵横比3方面考虑，使得检测框回归得更快、更精确。AIoU损失不仅在精度上比CIoU损失和GIoU损失更好，而且回归速度也比CIoU损失和GIoU损失更快。

为了验证本文所提出模块对跟踪效果的影响，对本文框架中的模块分别做了消融实验，结果如表2所示。从MOTA指标来看，AIoU回归损失函数明显地提高了跟踪器的跟踪准确度；从IDF1指标来看，层级匹配模块明显提高了关联精度。但是当AIoU回归损失函数和层级关联策略模块结合时，跟踪器

的MOTA反而不如两个单独模块时高，这主要是由于JDE跟踪模型中的竞争，从而产生许多错检。由于JDE跟踪模型将检测分支和Re-ID分支集成到一个网络中，而检测分支是为了将背景和前景区别开，Re-ID分支却是为了将不同的前景（例如行人）区别开，这两个任务必然会存在竞争关系。所以，当IDF1提升的同时，必然会造成MOTA有所下降。但是从HOTA来看，本文跟踪器较好地平衡了检测分支和Re-ID分支的竞争，使得HOTA提升了2.4%。

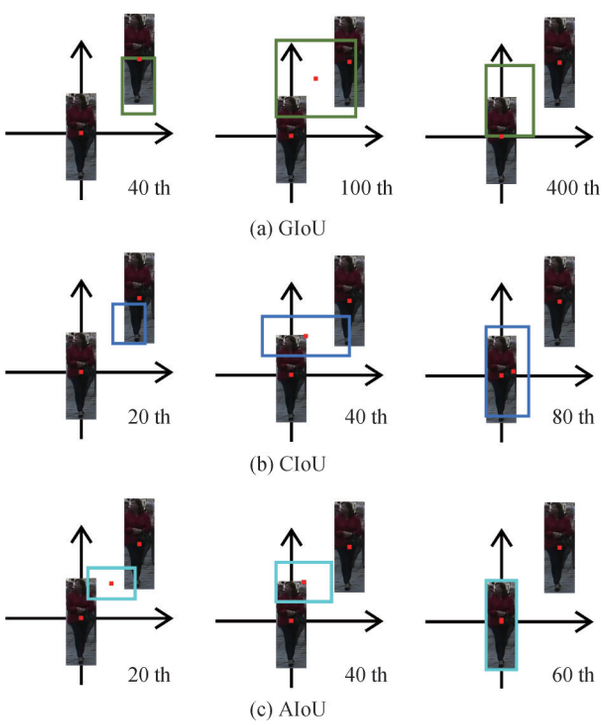


图4　GIoU损失、CIoU损失和AIoU损失对比

Fig. 4　Comparison of GIoU loss, CIoU loss, and AIoU loss ((a)GIoU;(b)CIoU;(c)AIoU)

表2　在MOT17数据集上不同模块对跟踪效果的影响

Table 2　The effect of different components on tracking results on MOT17 dataset

模块	HOTA /% ↑	MOTA /% ↑	IDF1 /% ↑	FP ↓	FN ↓	IDs ↓
Baseline	57.5	71.3	70.1	**23 376**	136 098	2 349
AIoU	59.0	**72.6**	71.4	31 233	121 338	**1 983**
Hierarchical	57.8	72.0	71.9	26 235	127 416	2 037
AIoU+ Hierarchical	**59.9**	70.8	**72.9**	57 480	**104 805**	2 694

注：加粗字体表示各列最优结果，“↑”表示值越高越好，“↓”表示值越低越好。

同时，为了验证***AE***矩阵模块中不同几何属性损失的作用，对本文框架中的***AE***模块中不同几何属性损失在MOT验证集上做了消融对比实验，结果如表3所示。

表3　基于*AE*矩阵模块中不同几何属性损失的消融研究

Table 3　The ablation effect of different geometric attribute losses in the *AE* matrix module

模块					HOTA /% ↑	IDF1 /% ↑	MOTA /% ↑	IDs ↓
IoU	L_{heat}	L_{AR}	L_S	L_E				
√	–	–	–	–	60.6	79.1	82.8	409
√	√	–	–	–	61.1	80.2	81.6	359
√	–	√	–	–	61.8	80.1	83.6	393
√	–	–	√	–	62.3	81	83.2	378
√	√	√	–	–	62.6	81.6	82.4	376
√	–	√	√	–	63.2	82.2	83.6	388
√	√	–	√	–	63.6	82.6	81.9	379
√	√	√	√	–	64.8	84.6	84.1	323
–	–	–	–	√	59.8	76.6	78.6	416
√	–	–	–	√	61.2	81.4	80.6	392
√	√	√	√	√	**65.9**	**82.4**	**83.1**	**293**

注：加粗字体表示各列最优结果，“√”表示采用；“–”表示未采用。

2.3.2　与现有方法对比

为了验证本文提出的多目标跟踪性能，与现有方法在MOT16和MOT17数据集上进行了对比，结果如表4和表5所示。可以看出，本文方法在HOTA指标和IDF1指标超过大部分现有方法。虽然本文MOTA指标可能不太高，但是从HOTA指标可以看出，本文方法较好地平衡了边界框大小特征、热图特征和中心点偏移量特征之间的平衡性。

2.3.3　可视化结果

图5为本文跟踪器与FairMOT跟踪器在MOT17数据集上的一部分可视化效果。可以看出，目标间存在比较严重的遮挡，但是借助本文提出的AIoU回归损失使得目标仍然能够进行精确的识别。尤其是在放大的地方，经过不同程度的遮挡后，FairMOT已经不能进行正常检测，但是本文的跟踪器仍然可以进行正常的检测。

图6—图8为本文跟踪器在MOT17数据集上的一部分可视化效果。图6为AIoU-Tracker在MOT17

表4　本文方法与其他方法在MOT16数据集上的效果对比

Table 4　The tracking performance comparison between our method and other methods on MOT16 dataset

方法	HOTA/%↑	IDF1/%↑	MOTA/%↑	IDs↓	FP↓	FN↓	帧速率(帧/s)↑
FairMOT(Zhang等,2021)	58.3	72.6	69.3	815	13 501	41 653	25.9
JDE(Wang等,2020)	–	55.8	64.4	1544	–	–	22.2
CTracker(Peng等,2020)	48.8	57.2	67.6	1897	**8 934**	48 350	6.8
MeMOT(Cai等,2022)	57.4	69.7	72.6	845	14 595	34 595	–
本文	**59.8**	**73.1**	**74.4**	**638**	12 561	**33 468**	**26.4**

注:加粗字体表示各列最优结果,"–"表示未公布结果。

表5　本文方法与其他方法在MOT17数据集上的效果对比

Table 5　The tracking performance comparison between our method and other methods on MOT17 dataset

方法	HOTA/%↑	IDF1/%↑	MOTA/%↑	IDs↓	FP↓	FN↓	帧速率(帧/s)↑
CenterTrack(Zhou等,2020)	48.2	59.6	61.5	2 583	14 076	200 672	17.0
TransCenter(Xu等,2023)	54.5	62.2	73.2	4 614	23 112	123 738	1.0
SOTMOT(Zheng等,2021)	–	71.9	71.0	5 184	39 537	118 983	16.0
TransTrack(Sun等,2021a)	54.1	63.5	75.2	3 603	50 157	**86 442**	10.0
MOTR(Zeng等,2022)	57.8	68.6	73.4	2 439	20 268	133 440	4.5
MAT(Han等,2022)	56.0	69.2	67.1	**1 279**	22 756	161 547	11.5
GSDT(Wang等,2021)	55.5	68.7	66.2	3 318	43 368	144 261	4.9
TPAGT(Shan等,2020)	57.9	68.0	76.2	3 237	32 796	98 475	6.8
PermaTrack(Tokmakov等,2021)	55.5	68.9	73.8	3 699	28 998	115 104	11.9
MeMOT(Cai等,2022)	56.9	69.0	72.5	2 724	37 221	115 248	–
YOLOTracker(Chan等,2022)	53.5	65.1	67.1	4 983	37 701	142 914	24.9
FairMOT(Zhang等,2021)	59.3	72.3	73.7	3 303	27 507	117 477	18.9
CSTrack(Liang等,2022)	59.3	72.6	74.9	3 567	23 847	114 303	15.8
CTracker(Peng等,2020)	49.0	61.2	66.6	5 529	**22 284**	160 491	**34.4**
ReMOT(Yang等,2021)	59.7	72.0	**77.0**	2 853	33 204	93 612	1.8
本文	**59.9**	**72.9**	70.8	2 694	57 480	104 805	20.3

注:加粗字体表示各列最优结果,"–"表示未公布结果。

测试数据集上的实例跟踪轨迹结果。由图7可发现,得益于本文的层级关联策略,在行人经过连续两帧消失后,仍然能够正确地进行关联,提高了本文跟踪器的鲁棒性。图8为AIoU-Tracker与FairMOT在MOT17数据集上Re-ID特征判别能力的可视化对比结果。

为了方便读者了解本文所提出方法的局限性,在图9展示了一些移动相机下挑战性和结果不太好的例子,推测可能是因为移动相机的优化仍然具有局限性,也有可能是本文的检测网络DLA-34基于ResNet-34,导致检测器的检测能力较差。

3　结　论

本文针对模糊的行人特征造成的身份切换问题和复杂场景下目标之间的遮挡跟踪精度降低的问题,提出AIoU-Tracker多目标跟踪算法。首先,根据本文的骨干网络设计自适应损失函数,从重叠面积、中心点距离和纵横比3个方面去考虑,增强了模型对模糊行人特征的抗干扰能力,降低了身份切换次

(a) FairMOT　　(b) AIoU-Tracker

图5　AIoU-Tracker与FairMOT在MOT17数据集上可视化对比结果

Fig. 5　Visualization comparison results of AIoU-Tracker and FairMOT on MOT17 dataset

((a)FairMOT;(b)AIoU-Tracker)

图6　AIoU-Tracker在MOT17测试集上的实例跟踪轨迹结果

Fig. 6　Instance tracking trajectory results of AIoU-Tracker on the MOT17 test set

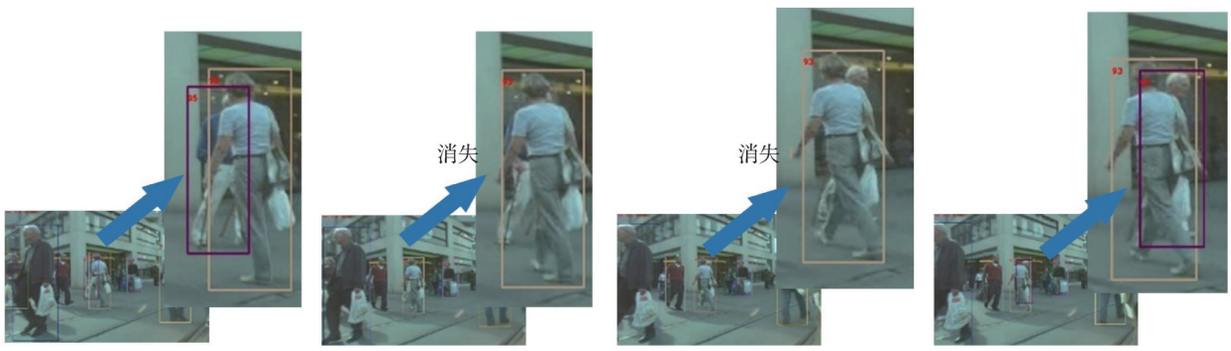


图7　经过两帧之后被遮挡的ID仍然可以被关联

Fig. 7　IDs that are occluded after two frames can still be associated

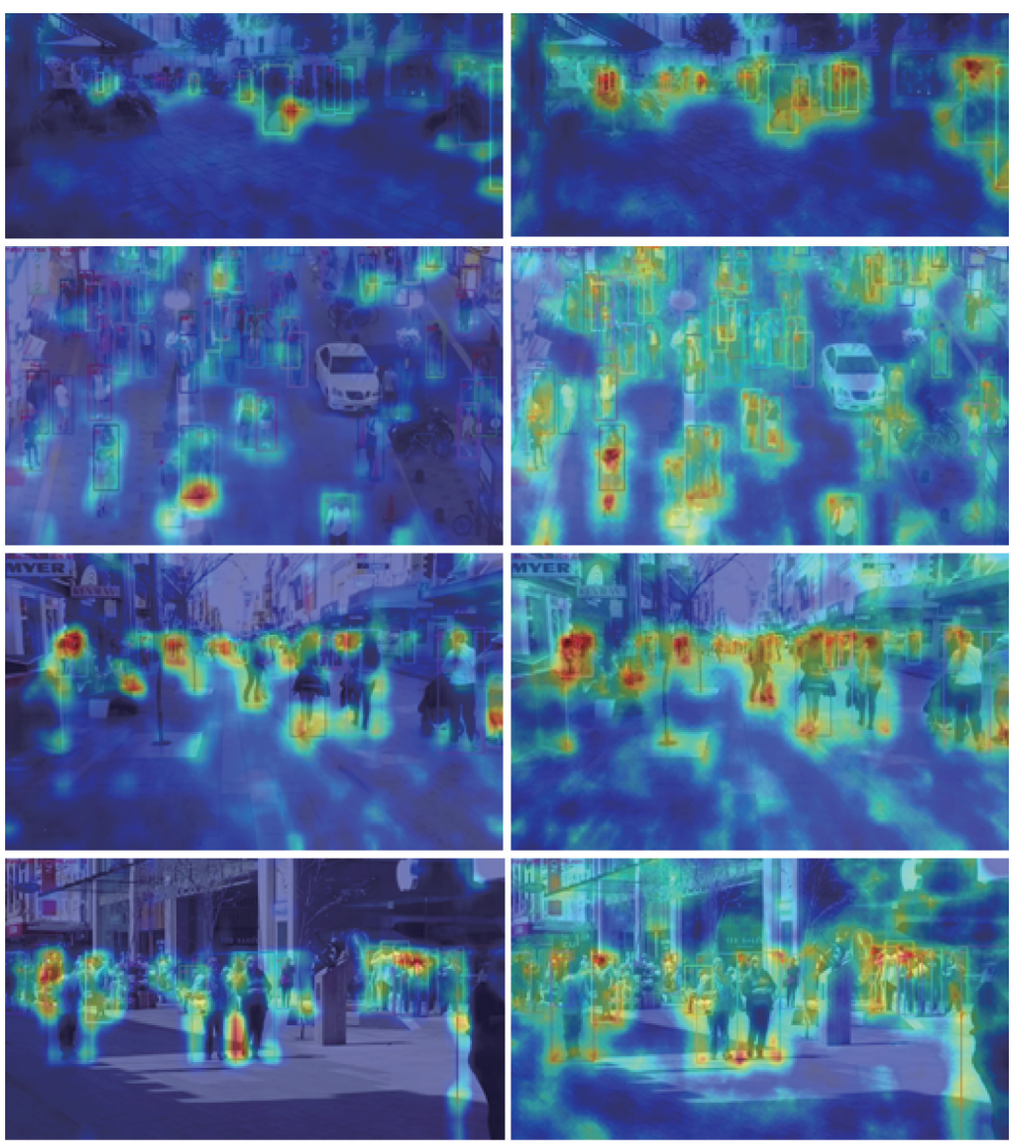
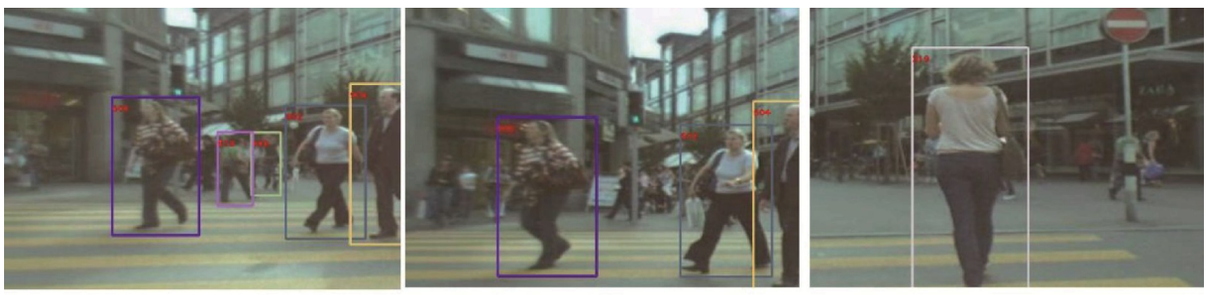


图8　AIoU-Tracker与FairMOT在MOT17数据集上Re-ID特征判别能力的可视化对比结果

Fig. 8　Visualization of comparative results of Re-ID feature discriminability between AIoU-Tracker and FairMOT on MOT17 dataset



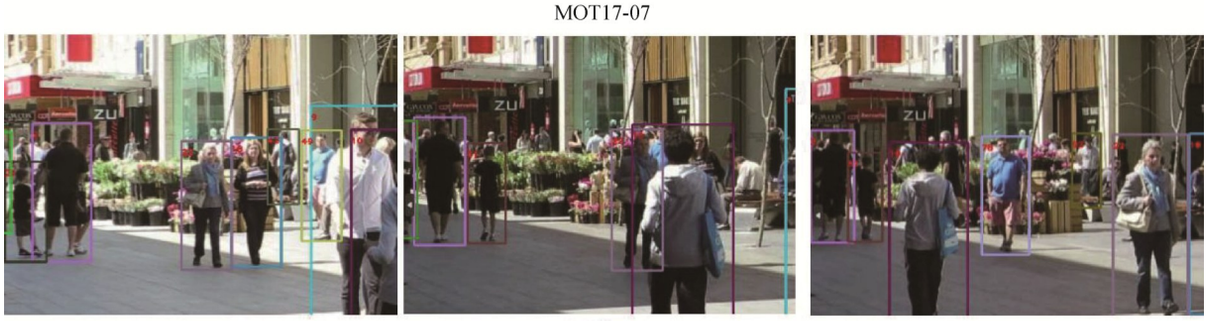


图9　AIoU-Tracker在移动相机下的可视化结果

Fig. 9　Visualization results of AIoU-Tracker under moving camera conditions

数。其次，采用层级关联策略，在高分检测框和低分检测框分别关联之后，利用关联失败检测框周围的嵌入信息再次进行关联，进一步提高了模型在遮挡情况下的适应能力。实验结果表明，本文提出的算法在模糊的行人特征和目标遮挡的场景下，跟踪性能得到提升，大多数评价指标优于FairMOT、JDE、

CTracker、CSTrack和YOLOTracker等典型算法，对模糊的行人特征和目标遮挡具有一定的抵御作用。

但是，本文算法仍然存在一些瓶颈。1）对于在移动相机下拍摄的数据集，跟踪性能仍然存在一定的挑战性。当相机移动时，由于目标特征发生形变，使提取到的特征不够鲁棒，从而产生许多错检。2）本文算法是基于回归损失和层级匹配策略的跟踪算法，并未对骨干网络结构进行优化，使得骨干网络提取的目标特征不够鲁棒。因此，后续可以进一步提高跟踪器在移动相机下拍摄的数据集的跟踪性能；设计特征解耦模块或使用更深层的网络结构，提高骨干网络的目标检测能力。

参考文献（References）

Bewley A, Ge Z Y, Ott L, Ramos F and Upcroft B. 2016. Simple online and realtime tracking//Proceedings of 2016 IEEE International Conference on Image Processing (ICIP). Phoenix, USA: IEEE: 3464-3468 [DOI: 10.1109/ICIP.2016.7533003]

Bochinski E, Eiselein V and Sikora T. 2017. High-speed tracking-by-detection without using image information//Proceedings of the 14th IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS). Lecce, Italy: IEEE: 1-6 [DOI: 10.1109/AVSS.2017.8078516]

Cai J R, Xu M Z, Li W, Xiong Y J, Xia W, Tu Z W and Soatto S. 2022. MeMOT: multi-object tracking with memory//Proceedings of 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). New Orleans, USA: IEEE: 8080-8090 [DOI: 10.1109/CVPR52688.2022.00792]

Chan S X, Jia Y W, Zhou X L, Bai C, Chen S Y and Zhang X Q. 2022. Online multiple object tracking using joint detection and embedding network. Pattern Recognition, 130: #108793 [DOI: 10.1016/j.patcog.2022.108793]

Dollár P, Wojek C, Schiele B and Perona P. 2009. Pedestrian detection: a benchmark//Proceedings of 2009 IEEE Conference on Computer Vision and Pattern Recognition. Miami, USA: IEEE: 304-311 [DOI: 10.1109/CVPR.2009.5206631]

Duan K W, Bai S, Xie L X, Qi H G, Huang Q M and Tian Q. 2019. CenterNet: keypoint triplets for object detection//Proceedings of 2019 IEEE/CVF International Conference on Computer Vision (ICCV). Seoul, Korea (South): IEEE: 6568-6577 [DOI: 10.1109/ICCV.2019.00667]

Ess A, Leibe B, Schindler K and van Gool L. 2008. A mobile vision system for robust multi-person tracking//Proceedings of 2008 IEEE Conference on Computer Vision and Pattern Recognition. Anchorage, USA: IEEE: 1-8 [DOI: 10.1109/CVPR.2008.4587581]

Han S, Huang P, Wang H, Yu E, Liu D and Pan X.2022. Mat: motion-aware multi-object tracking. Neurocomputing, 476: 75-86 [DOI: 10.1016/j.neucom.2021.12.104]

He K M, Gkioxari G, Dollar P and Girshick R. 2017. Mask R-CNN//Proceedings of 2017 IEEE International Conference on Computer Vision (ICCV). Venice, Italy: IEEE: 2980-2988 [DOI: 10.1109/ICCV.2017.322]

Huang G, Liu S C, van der Maaten L and Weinberger K Q. 2018. CondenseNet: an efficient DenseNet using learned group convolutions//Proceedings of 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City, USA: IEEE: 2752-2761 [DOI: 10.1109/CVPR.2018.00291]

Li W, Zhao R, Xiao T and Wang X G. 2014. DeepReID: deep filter pairing neural network for person re-identification//Proceedings of 2014 IEEE Conference on Computer Vision and Pattern Recognition. Columbus, USA: IEEE: 152-159 [DOI: 10.1109/CVPR.2014.27]

Liang C, Zhang Z P, Zhou X, Li B, Zhu S Y and Hu W M. 2022. Rethinking the competition between detection and ReID in multiobject tracking. IEEE Transactions on Image Processing, 31: 3182-3196 [DOI: 10.1109/ TIP.2022.3165376]

Lin T Y, Dollár P, Girshick R, He K M, Hariharan B and Belongie S. 2017. Feature pyramid networks for object detection//Proceedings of 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Honolulu, USA: IEEE: 936-944 [DOI: 10.1109/CVPR.2017.106]

Luo W H, Xing J L, Milan A, Zhang X Q, Liu W and Kim T K. 2021. Multiple object tracking: a literature review. Artificial Intelligence, 293: #103448 [DOI: 10.1016 /j.artint.2020.103448]

Milan A, Leal-Taixe L, Reid I, Roth S and Schindler K. 2016. MOT16: a benchmark for multi-object tracking [EB/OL]. [2023-11-01]. https://arxiv.org/pdf/1603.00831.pdf

Park Y, Dang L M, Lee S, Han D and Moon H. 2021. Multiple object tracking in deep learning approaches: a survey. Electronics, 10(19): #2406 [DOI: 10.3390/ electronics10192406]

Peng J L, Wang C G, Wan F B, Wu Y, Wang Y B, Tai Y, Wang C J, Li J L, Huang F Y and Fu Y W. 2020. Chained-tracker: chaining paired attentive regression results for end-to-end joint multiple-object detection and tracking//Proceedings of the 16th European Conference on Computer Vision. Glasgow, UK: Springer: 145-161 [DOI: 10.1007/978-3-030-58548-8_ 9]

Ren S Q, He K M, Girshick R and Sun J. 2017. Faster R-CNN: towards real-time object detection with region proposal networks. IEEE Transactions on Pattern Analysis and Machine Intelligence, 39(6): 1137-1149 [DOI: 10.1109/TPAMI.2016.2577031]

Rezatofighi H, Tsoi N, Gwak J, Sadeghian A, Reid I and Savarese S. 2019. Generalized intersection over union: a metric and a loss for bounding box regression//Proceedings of 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Long Beach, USA: IEEE: 658-666 [DOI: 10.1109/CVPR.2019.00075]

Shan C B, Wei C B, Deng B, Huang J Q, Hua X S, Cheng X L and Liang K W. 2020. Tracklets predicting based adaptive graph tracking [EB/OL]. [2023-11-01]. https:// arxiv.org/pdf/2010.09015.pdf

Shao S, Zhao Z J, Li B X, Xiao T T, Yu G, Zhang X Y and Sun J. 2018. CrowdHuman: a benchmark for detecting human in a crowd [EB/OL]. [2023-11-01]. https://arxiv.org/pdf/1805.00123.pdf

Sun P Z, Cao J K, Jiang Y, Zhang R F, Xie E Z, Yuan Z H, Wang C H and Luo P. 2021a. TransTrack: multiple object tracking with Transformer [EB/OL]. [2023-11-01]. https://arxiv.org /pdf/2012.15460.pdf

Tokmakov P, Li J, Burgard W and Gaidon A. 2021. Learning to track with object permanence//Proceedings of 2021 IEEE/CVF International Conference on Computer Vision (ICCV). Montreal, Canada: IEEE: 10840-10849 [DOI: 10.1109/ICCV48922.2021.01068]

Voigtlaender P, Krause M, Osep A, Luiten J, Sekar B B G, Geiger A and Leibe B. 2019. MOTS: multi-object tracking and segmentation//Proceedings of 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Long Beach, USA: IEEE: 7934-7943 [DOI: 10.1109/CVPR.2019.00813]

Wang Y X, Kitani K and Weng X S. 2021. Joint object detection and multi-object tracking with graph neural networks//Proceedings of 2021 IEEE International Conference on Robotics and Automation (ICRA). Xi'an, China: IEEE: 13708-13715 [DOI: 10.1109/ICRA4850 6.2021.9561110]

Wang Z D, Zheng L, Liu Y X, Li Y L and Wang S J. 2020. Towards real-time multi-object tracking//Proceedings of the 16th European Conference on Computer Vision. Glasgow, UK: Springer: 107-122 [DOI: 10.1007/978-3- 030-58621-8_7]

Wojke N, Bewley A and Paulus D. 2017. Simple online and realtime tracking with a deep association metric//Proceedings of 2017 IEEE International Conference on Image Processing (ICIP). Beijing, China: IEEE: 3645-3649 [DOI: 10.1109/ICIP.2017.8296962]

Xiao T, Li S, Wang B C, Lin L and Wang X G. 2017. Joint detection and identification feature learning for person search//Proceedings of 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Honolulu, USA: IEEE: 3376-3385 [DOI: 10.1109/CVPR.2017.360]

Xu Y H, Ban Y T, Delorme G, Gan C, Rus D and Alameda-Pineda X. 2023. TransCenter: Transformers with dense representations for multiple-object tracking. IEEE Transactions on Pattern Analysis and Machine Intelligence, 45(6): 7820-7835 [DOI: 10.1109/TPAM I.2022.3225078]

Yang F, Chang X, Sakti S, Wu Y and Nakamura S. 2021. ReMOT: a model-agnostic refinement for multiple object tracking. Image and Vision Computing, 106: #104091 [DOI: 10.1016/j. imavis. 2020. 104091]

Yue Y Y, Xu D, He K J and Zhang H. 2023. An adaptive occlusion-aware multiple targets tracking algorithm for low viewpoint. Journal of Image and Graphics, 28(2): 441-457 (乐应英, 徐丹, 贺康建, 张浩. 2023. 低视点下遮挡自适应感知的多目标跟踪算法. 中国图象图形学报, 28(2): 441-457) [DOI: 10.11834/jig.210853]

Zeng F G, Dong B, Zhang Y A, Wang T C, Zhang X Y and Wei Y C. 2022. MOTR: end-to-end multiple-object tracking with Transformer//Proceedings of the 17th European Conference Computer Vision. Tel Aviv, Israel: Springer: 659-675 [DOI: 10.1007/978-3-031-19812-0_ 38]

Zhang S S, Benenson R and Schiele B. 2017. CityPersons: a diverse dataset for pedestrian detection//Proceedings of 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Honolulu, USA: IEEE: 4457-4465 [DOI: 10.1109/CVPR.2017.474]

Zhang Y F, Sun P Z, Jiang Y, Yu D D, Weng F C, Yuan Z H, Luo P, Liu W Y and Wang X G. 2022. ByteTrack: multi-object tracking by associating every detection box//Proceedings of the 17th European Conference on Computer Vision. Tel Aviv, Israel: Springer: 1-21 [DOI: 10.1007/978-3-031-20047-2_1]

Zhang Y F, Wang C Y, Wang X G, Zeng W J and Liu W Y. 2021. FairMOT: on the fairness of detection and re-identification in multiple object tracking. International Journal of Computer Vision, 129(11): 3069-3087 [DOI: 10.1007/s11263-021-01513-4]

Zheng L, Zhang H H, Sun S Y, Chandraker M, Yang Y and Tian Q. 2017. Person re-identification in the wild//Proceedings of 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Honolulu, USA: IEEE: 3346-3355 [DOI: 10.1109/CVP R.2017.357]

Zheng L Y, Tang M, Chen Y Y, Zhu G B, Wang J Q and Lu H Q. 2021. Improving multiple object tracking with single object tracking//Proceedings of 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Nashville, USA: IEEE: 2453-24 62 [DOI: 10.1109/CVPR46437.2021.00248]

Zheng Z H, Wang P, Liu W, Li J Z, Ye R G and Ren D W. 2020. Distance-IoU Loss: faster and better learning for bounding box regression//Proceedings of the 34th AAAI Conference on Artificial Intelligence. New York, USA: AAAI: 12993-13000 [DOI: 10.1609/aaai.v34i07.6999]

Zhou X Y, Koltun V and Krähenbühl P. 2020. Tracking objects as points//Proceedings of the 16th European Conference on Computer Vision. Glasgow, UK: Springer: 474-490 [DOI: 10.1007/978-3-030-58548-8_28]

作者简介

郭文，男，教授，主要研究方向为计算机视觉和多媒体计算。E-mail: wguo@sdtbu.edu.cn

丁昕苗，通信作者，女，教授，主要研究方向为计算机视觉和视频理解。E-mail: dingxinmiao@126.com

刘其贵，男，硕士研究生，主要研究方向为计算机视觉。E-mail: 2021420049@sdtbu.edu.cn