ISSN 1004-9037,CODEN SCYCE4
Journal of Data Acquisition and Processing Vol. 39,No. 6,Nov. 2024,pp. 1479－1492
DOI:10. 16337/j. 1004-9037. 2024. 06. 016

http://sjcj. nuaa. edu. cn
E-mail:sjcj@nuaa. edu. cn
Tel/Fax:＋86-025-84892742

多目标跟踪中基于SOT和重匹配的防遗漏机制

张毅锋，张嘉成，李元浩

（东南大学信息科学与工程学院,南京 211102）

摘　要：数据关联是多目标跟踪（Multiple object tracking，MOT）中的重要步骤,一般需要根据特征相似性实现目标和检测物体之间的身份匹配。部分目标或检测物体可能在匹配结束后仍处于孤立状态，可能导致轨迹中断或身份错乱的遗漏现象。为改善MOT的精度和稳定性,抑制数据关联中的遗漏现象,提出了一种基于高性能单目标跟踪器（Single object tracker，SOT）和重匹配的防遗漏机制。该机制运用Transformer和扩散模型,设计了一款契合MOT需求的SOT用于追踪遗漏目标,并通过记忆目标信息对遗漏检测物体实施重匹配。通过消融实验验证了SOT和重匹配方法在防遗漏机制中的作用,并在标准数据集上测试了该机制对MOT算法跟踪性能的影响。结果表明,各算法加入该机制后性能获得全面改善,该机制可有效抑制MOT中的遗漏现象。

关键词：多目标跟踪;数据关联;遗漏现象;单目标跟踪器;重匹配

中图分类号：TP391　　**文献标志码：**A

Anti-missing Mechanism Based on SOT and Rematching in Multiple Object Tracking

ZHANG Yifeng，ZHANG Jiacheng，LI Yuanhao

（School of Information Science and Engineering，Southeast University，Nanjing 211102，China）

Abstract：Data association is an important step in multiple object tracking（MOT），which generally requires identity matching between objects and detections based on feature similarity. Some objects or detections may remain isolated after match is completed，which is the missing phenomenon that may lead to track interruption or identity confusion. Therefore，in order to improve the accuracy and stability of MOT and suppress the missing phenomenon in data association，this paper proposes an anti-missing mechanism based on high-performance single object tracker（SOT）and rematching. The mechanism uses Transformer and diffusion model to design a SOT that meets the requirements of MOT to track missing objects and rematch missing detections by remembering the object information. The effect of SOT and rematching methods in anti-missing mechanism is verified by ablation experiments，and the effect of this mechanism on the tracking performance of MOT algorithm is tested on standard datasets. The results show that the performance of all algorithms is improved comprehensively with the addition of this mechanism，which can effectively suppress the missing phenomenon in MOT.

Key words: multiple object tracking; data association; missing phenomenon; single object tracker; rematching

---

基金项目：国家自然科学基金(61673108);江苏省自然科学基金(BK20201267)。
收稿日期：2023-07-19;**修订日期：**2023-10-16

引　言

多目标跟踪(Multiple object tracking, MOT)[1]是计算机视觉(Computer vision, CV)中的重点研究领域之一,兼具实用价值和技术难度。对该领域的研究,旨在从包含前后逻辑关系的连续图像中,通过标记了身份(Identification, ID)信息的边界框反映特定目标的行进轨迹和大小变化。许多MOT算法依赖检测结果进行跟踪,对于ID和位置已知的目标,算法会将现有检测物体与这些目标进行特征匹配,明确检测物体的ID,这一过程通常被称为数据关联[2]。数据关联的结果对定位准确度和轨迹连续性有直接影响,且整个匹配过程一般需要具备理想的演算效率,因此对数据关联方法的研究成为MOT领域长期以来的重点工作之一。

最简易的数据关联方法以贪心思想[3],根据轨迹点之间的欧氏距离或交并比(Intersection over union, IoU)进行配对,这种方法匹配速度较快但高度依赖检测质量,且不擅处理多个IoU接近甚至相等的情况。更为常见的数据关联方法是基于构造代价矩阵的线性分配方法,如计算二分图最大匹配的匈牙利算法[4]和KM算法[5],前者更适合力图标识出更多目标但对精度和稳定性要求有限的情况,后者则更适合目标之间存在一定联系并需要细致甄别的情况。此外,两种算法均具有较低的时间复杂度,并支持动态跟踪中的在线处理,这些优势更提升了它们在MOT任务中的实用价值。还有一些基于图论且实现难度较大的关联方法,包括最小代价流[6]和图神经网络(Graph neural network, GNN)[7]等。最小代价流利用多条跟踪轨迹和检测物体的数据构造出网络图的节点和边,可综合考虑多个目标之间的相似度和运动交互等信息,但随之而来的是存储空间和计算开销的飞速增加,以及在线处理能力的丧失。基于GNN的方法在建图思路上与最小代价流类似,但其采用神经讯息传送的策略更新节点和边,能够在线处理视频而无需等待所有待关联数据的送达。

以上方法各自侧重于满足数据关联对准确度和速度等不同方面的需求,但均无法彻底避免遗漏现象[8]。当具体观察数据关联每一轮的结果时,通常会发现个别目标或检测物体没有找到足够相似的异类个体,这即是严重阻碍MOT性能提升的遗漏现象。这类现象有几率导致目标识别错乱甚至跟踪意外停止,从而制约了MOT算法向技术指标的理论上限逼近。基于图网络的多目标跟踪(Graph networks for MOT, GNMOT)[8]曾提出可利用单目标跟踪器(Single object tracker, SOT)处理一些跟踪难度较大的特殊目标。在此基础上,本文利用单目标跟踪和图像生成领域的部分研究,设计了一款更契合MOT任务特性、性能更佳的SOT,并辅以重匹配方法来限制遗漏个体的数量。

因此,为了抑制遗漏现象及其诱发的多种后果,提出了一种基于高性能SOT和重匹配的防遗漏机制(Anti-missing mechanism based on SOT and rematching, AMM-SOTR)。该机制使用一款经过精心训练且擅长于把握先后逻辑的SOT复现遭到遗漏的目标,并在更早的目标轨迹中尝试重匹配遭到遗漏的检测物体。为验证此机制对遗漏现象的抑制效果,将其安排到DAHRMOT[9]等MOT算法的数据关联步骤后,并在MOT的常用数据集上测试了多项重要指标。实验结果表明,AMM-SOTR能够为多种MOT算法带来精度、稳定性和ID准确性等方面的全面提升,同时不会引发跟踪速度的严重下降,从而有效减少了遗漏现象对MOT性能的损害。

1　相关工作

数据关联的本质是两类个体的特征和ID等信息之间的匹配,因此决定关联正确与否,主要是信息的可靠程度和匹配方式的合理性。如果目标或检测物体在匹配完成后有部分遗漏,一般是由检测质量较差、关联策略偏颇和目标属性突变等因素所导致。

在真实的日常视频案例中,目标属性的突变通常不会太过频繁,因此以往的MOT算法主要依靠升级检测手段和更新关联策略等方式来降低遗漏个体的比例。MOT中检测手段的演进历经多代,从早

期需要添加配套特征学习等模块的R-CNN[10]、YOLOv3[11]和SSD[12]等检测器,到集成在统一MOT框架内的CenterNet[13]等检测器,检测与跟踪两项任务的结合正愈发紧密,检测器的抗噪和区分能力也愈发强大。

在关联策略方面,概率数据关联(Probabilistic data association, PDA)[14]通过统计目标和检测物体的关联概率进行模糊匹配,这种方法能有效抵御杂波和噪声造成的检测质量下滑,但当大量目标发生重叠或外形相近时效果较差,且只能在目标数量固定,即没有新目标入场或旧目标离开的画面中使用。联合PDA(Joint PDA, JPDA)[15]将PDA中的关联概率改进为边缘关联概率,用于统计某一对关联个体在所有匹配假设中最终被确认的概率,再据此确定最优或次优的关联方案。该方法能在目标间有相互干扰的情况下减少遗漏个体,且可以用于目标数量持续变化的场景,但无法回避在遍历匹配假设时带来的计算量膨胀。基于树状结构的MHT[16]为所有目标生成多条假设轨迹,以递归思想在轨迹构成的树中经过扩展和剪枝等操作,求出每条轨迹的后验概率,再选出其中的最优假设作为最终的关联结果,这种方法不仅不受常规干扰的影响,还能合理应对目标发生分裂或合并等特殊情况,但对维护水平和计算资源的巨大需求限制了它的实际应用面。

以上研究从遗漏现象的产生原因出发,一般需要复杂的技术实现方案或抽象的数学理论推导,部分方法还极大程度地受制于硬件设备的性能上限。因此,在保证设备通用性,以及方法的可泛化性基础上,本文选择从既有的关联结果入手,使用SOT和重匹配方法分类处理遗漏个体。考虑到MOT任务的复杂性和特殊性,用于处理遗漏个体的SOT需要进行针对性的设计,进而能保证跟踪器具有更强的单个目标处理能力。

相比常规的视觉目标跟踪,MOT任务的难度主要来自目标自身运动中复杂的先后逻辑,以及多个目标之间在运动和外形上的彼此干扰。关于梳理先后逻辑方面,2017年面世的Transformer[17]得益于其独特的注意力机制,已成为包括TrSiam[18]和TransT[19]在内的擅长先后逻辑推理的SOT的内嵌模型;关于解决目标间干扰方面,近年诞生的扩散模型[19]可用于改进SOT的训练数据,最初的DDPM[20]通过向数据中引入噪声再逐步将其删除的方法,在生成的灵活度上取得了重要进展,随后出现的DDIM[21]、ADM-G[22]和SDG[23]等模型又在生成速度和控制手段等方面做出了改进。

本文通过SDG添加图像和语言双模态的语义引导,使得扩散模型能去噪生成不同姿态、布局以及风格的目标图片,极大地丰富了SOT模型训练所需要的数据,从而增强其对目标尺寸和位置的推断能力;通过在SOT模型中引入Transformer模块,能够增强其时间维度的梳理能力;在SOT模型中引入长、短时模块,丰富所提取的特征信息。进一步,将强化后的SOT模型与重匹配机制一起引入到多目标追踪中来作为防遗漏机制。其中,通过SOT定位被遗漏目标在新一帧中的位置,提升了其被继续跟踪的可能性;通过重匹配的记忆机制来存储未匹配成功的轨迹,当其在一定时间内与检测重新匹配上时,便能重新被激活。该防遗漏机制即插即用,能被应用到多种多目标追踪算法中。经实验验证,该机制能明显地提升多目标追踪器的性能。

2 本文方法

提出的防遗漏机制主要由两部分组成:一部分通过特制的SOT重新描绘被忽略目标的轨迹,另一部分通过重匹配寻找被忽略检测物体在更早的画面中存在的痕迹。如此形成的名为AMM-SOTR的防遗漏机制理论上能有效减少每一轮关联后出现的遗漏个体,且不涉及对MOT主干网络的更改,因此可以便捷地安插在大部分基于检测和相邻两帧关联的MOT算法中。为方便示例,本文将基于MOT中经典的相邻帧检测-跟踪结构分析该机制如何在一轮完整的MOT过程中处理遗漏个体。

2.1 经典MOT框架中的防遗漏机制

所提AMM-SOTR防遗漏机制的MOT框架如图1所示。所有检测工作完成后，算法将分别提取相邻两帧中目标和检测物体的特征，不同图案体现了特征之间的差异。由于现有目标的ID已经明确，因此使用标有数字序号的各色圆形加以区分；而新检测物体的ID尚待分配，使用无序号的同色方形表示。

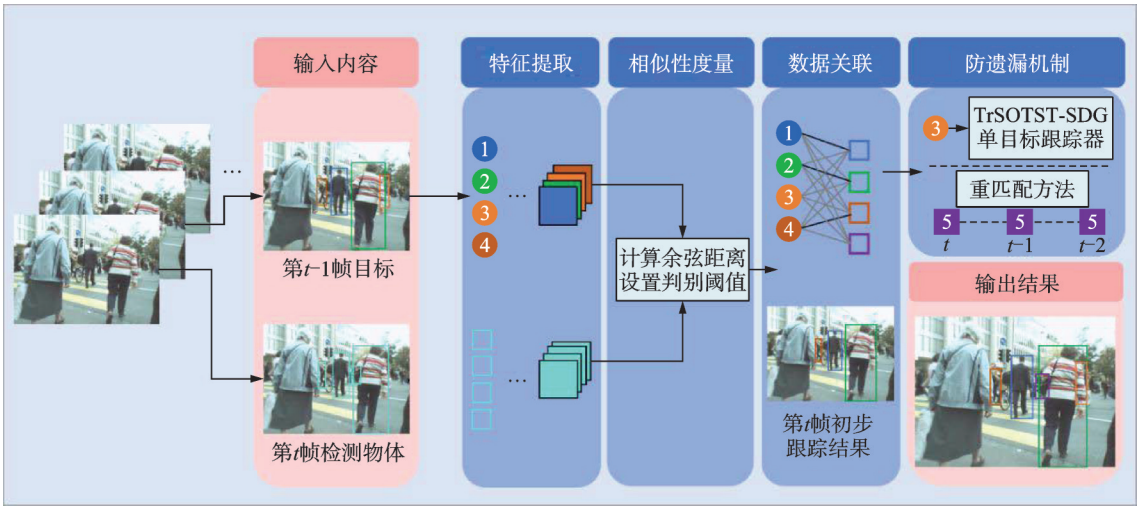

图1 具备防遗漏机制的MOT框架

Fig.1 Framework of MOT with anti-missing mechanism

特征提取后进行的是常规的计算特征相似性和数据关联的步骤，匹配过程中的黑色连线意为两者关系得到确认，此时检测物体将变为与对应目标相同的颜色。本次关联结束后分别有一个目标和一个检测对象没有达成任何匹配，即前文中所提及的遗漏个体，相应地，初步的跟踪结果中也不存在这两个遗漏个体。

此时，防遗漏机制一方面通过名为TrSOTST-SDG的SOT重新追踪目标，准确得出目标现处位置；一方面利用事先记忆的早期目标数据与检测物体展开重匹配，在更靠前的画面中发现了该物体对应的目标。得益于防遗漏机制的存在，输出结果的画面中所有应当被确认为目标的物体都已被准确标识。以上为一对目标和检测物体未成功匹配的情况，当有多对时，只需按序分别执行SOT追踪和重匹配即可。至此，防遗漏机制在MOT中的运作原理现已基本明晰，后续内容将探讨其中的SOT和重匹配这两大组成部分的具体实现方法。

2.2 防遗漏机制的SOT部分

如前文所述，构思适用于防遗漏机制的SOT的关键，在于对先后逻辑的把握和目标间干扰的抵御。因此，可考虑通过当下热门的扩散模型改进训练数据，增加SOT在训练时面对干扰物体的种类和数量，并依靠分段式结构[24]增加浅层特征对结果的影响，提升SOT对目标间干扰的抵御能力。同时，还可利用在梳理先后逻辑上具有先天优势的Transformer加工目标特征[18]，拓宽SOT在时间层面的视野。本文将提出的这种新型SOT命名为TrSOTST-SDG（Transformer-enhanced SOT based on sectional training with semantic diffusion guidance）。

2.2.1 TrSOTST-SDG训练数据的改进

在LaSOT[25]和GOT-10k[26]这些用于训练SOT的标准数据集中，画面中物体的个数一般远少于

MOT数据集,这对于实现防遗漏机制所需的SOT极为不利。因此,本文采用SDG[24]这一条件扩散模型,增加训练图像中的物体数量,在一定程度上模拟MOT场景。

传统的扩散模型缺乏控制手段,可能会忽略目标自身的某些重要信息,因此不适合在此使用。而作为时下自由度和可控性结合最佳的生成模型之一,SDG通过采用包括图像在内等多媒介信息调控的随机扩散策略,能够生成丰富度极高的高清图像,同时保持生成结果与调控图像的一定联系,利用SDG改进训练数据的方法如图2所示。

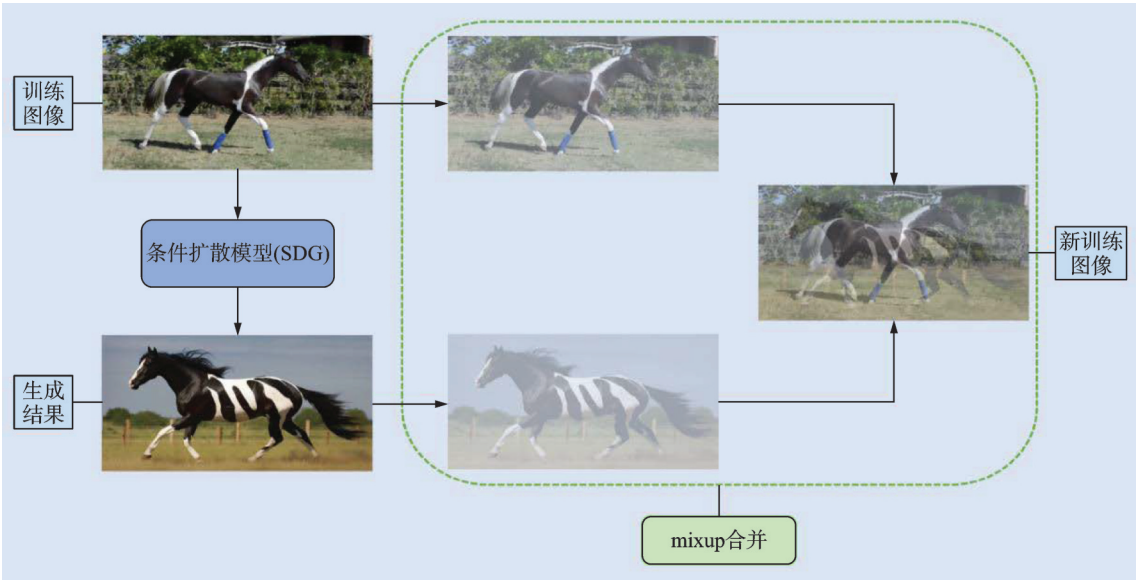

图2 利用SDG改进训练数据的方法

Fig.2 Method of using SDG to improve training data

在图2中,SDG以一幅散步中花马的训练图像充当调控信息,生成结果为类似环境中一匹奔跑中的花马,两者在内容上存在区别又有所联系,这样的结果满足本文对训练数据的改进要求。在普通的分类或分割等任务中,生成结果可以直接添加到数据集中,但目标跟踪的每张训练图像都是一段连续视频的组成部分,它们必须在按照指定顺序紧密排列,相邻的两张训练图像不仅不能随意颠倒位置,还不能接受任何其他图像被插入到它们中间,因为那将破坏视频的既定逻辑。为此,本文将训练图像和生成结果用mixup[27]合并为一张新训练图像,使其覆盖原来的训练图像。mixup是一种通过Beta分布来调和像素值的图像合并方法,设$\boldsymbol{x}_{ij}$和$\boldsymbol{y}_{ij}$分别表示训练图像和生成结果中坐标为(i,j)的像素值,合并后的新训练图像中对应点的像素值$\boldsymbol{z}_{ij}$为

$$\boldsymbol{z}_{ij}=\lambda\boldsymbol{x}_{ij}+(1-\lambda)\boldsymbol{y}_{ij}\quad \lambda\in[0.5,1] \tag{1}$$

式中:$\lambda\sim B(\beta_1,\beta_2)$为合并参数,它控制着$\boldsymbol{x}_{ij}$和$\boldsymbol{y}_{ij}$对$\boldsymbol{z}_{ij}$的影响程度。$\lambda$所属Beta分布的期望由$\beta_1$和$\beta_2$决定,常规的mixup会将期望设为0.5以实现均衡合并,但对SOT的训练数据而言,应当在合并后的图像中保留更多原目标的内容,即使得$\boldsymbol{x}_{ij}$的系数λ大于$\boldsymbol{y}_{ij}$的系数$1-\lambda$。因此,本文虽然仍使用0.5期望值的Beta分布获取λ,但会将其取值范围限定在$[0.5,1]$而非默认的$[0,1]$,这一操作可由简单的线性变换实现。当新训练图像逐一覆盖了原有对应图像后,训练数据的改进工作也宣告完成。

在通过SDG改进的训练数据中,不仅新增了与目标有关的物体,还可以达到遮挡和模糊等干扰效果。经过这些数据的滋养,TrSOTST-SDG可以更熟练地应对MOT中的目标间干扰。

2.2.2 TrSOTST-SDG的分段式结构

大多数SOT使用层数较深的主干网络,并直接在末端获得具有宏观类别含义的深层目标特征,这样做的好处是SOT不容易受颜色等表面因素的误导,但代价是对物体之间的细节差异不够敏感,因为决策时缺少了浅层特征的参与,这一缺陷对MOT任务来说相当致命。为避免这一问题,TrSOTST-SDG将网络划分为两个各有所长的模块,通过先后训练它们实现了图3中展现的分段式结构。在广为采用的孪生网络[28]中,前部的短时模块负责提取浅层特征,擅长区分有细微差异的同类物体;后部的长时模块负责提取深层特征,擅长区分有宏观界限的异类物体。为实现这种划分效果,需要先训练短时模块以固定基本参数,再同步训练两个模块。需要指出的是,模板图在跟踪中并非一成不变,且短时模块的模板变更比长时模块更频繁,因为前者需要时刻关注目标的最新动态。由初始目标框裁定的模板图和面积更大的搜索图输入孪生网络后,将经过多层网络提取出多种层次的特征。这些以矩阵形式存在的特征将被进一步送往特征加工模块,其中的Transformer会根据先后逻辑对特征进行一定程度的矫正,后文将详细介绍这个过程。

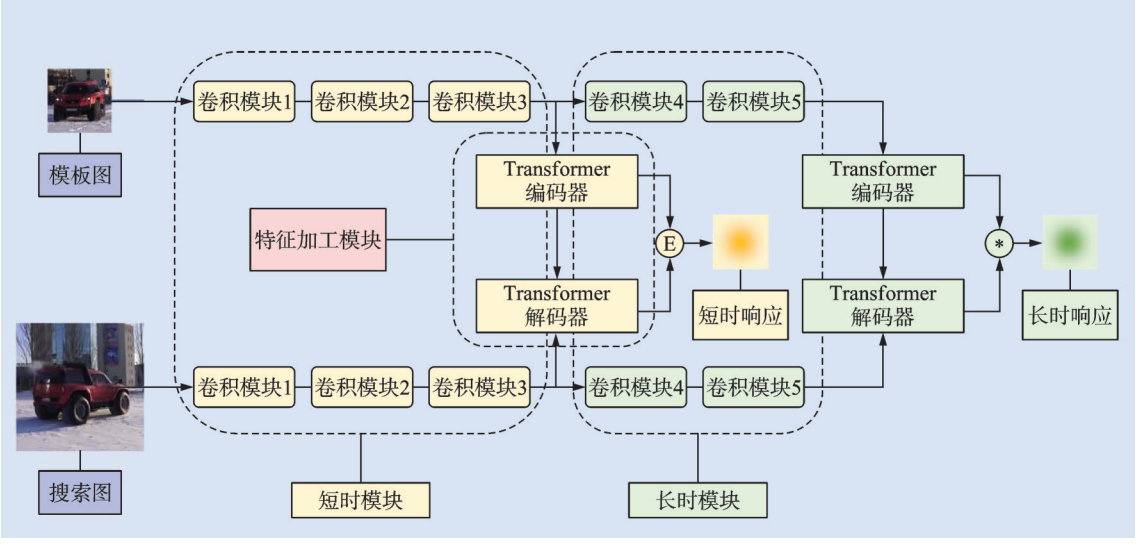

图3 TrSOTST-SDG的分段式结构

Fig.3 Sectional structure of TrSOTST-SDG

加工完毕后,算法进入相似度计算步骤。TrSOTST-SDG会对短时模块的两种特征计算欧氏距离,即图3中的“E”操作;而对长时模块的两种特征计算卷积,即图3中的“*”操作,这是由浅层特征和深层特征不同的分布密度决定的。对两个模块输出数据的相似度计算均结束后,算法即知悉了由二维矩阵构成的短时响应图和长时响应图。为综合两类特征的信息,可直接叠加两图并找出峰值位置,由此确定新的目标框。

2.2.3 特征加工模块

对于一些物体繁多的长视频而言,目标的当前属性可能不仅仅取决于最近几帧的画面,还可能与距离较远的某些画面有关,而常规的SOT内部网络在时间层面的视野较窄,这就需要动用Transformer的注意力机制加以解决,特征加工模块的内部结构如图4所示。

图4中的编码器以模板作为原料,实现视频中不同时期目标信息的交汇;解码器根据经编码后的模板特征,估计暗含目标位置的掩膜,再用其屏蔽模板中可能存在的各种干扰信息,将纯净又具备严谨先

后逻辑的模板特征关联至搜索图特征中,完成解码工作。

在编码器中,对于输入的模板特征组合矩阵 $\boldsymbol{T}$,可得其自相似度矩阵为

$$\boldsymbol{A}_{T\to T}=\mathrm{Atten}(\phi(\boldsymbol{T}'),\phi(\boldsymbol{T}')) \tag{2}$$

式中:Atten(·)为Transformer中必备的注意力函数,$\boldsymbol{T}'$是$\boldsymbol{T}$为了式(2)中计算的变形,$\phi(\cdot)$表示一个用于降低$\boldsymbol{T}'$通道数的简易函数。

各模板之间的先后逻辑即通过$\boldsymbol{A}_{T\to T}$体现,再进行实例归一化可得交汇后的模板特征为

$$\hat{\boldsymbol{T}}=\mathrm{Ins.Norm}(\boldsymbol{T}'+\boldsymbol{A}_{T\to T}\boldsymbol{T}') \tag{3}$$

式中:Ins.Norm(·)用函数形式表示实例归一化操作。

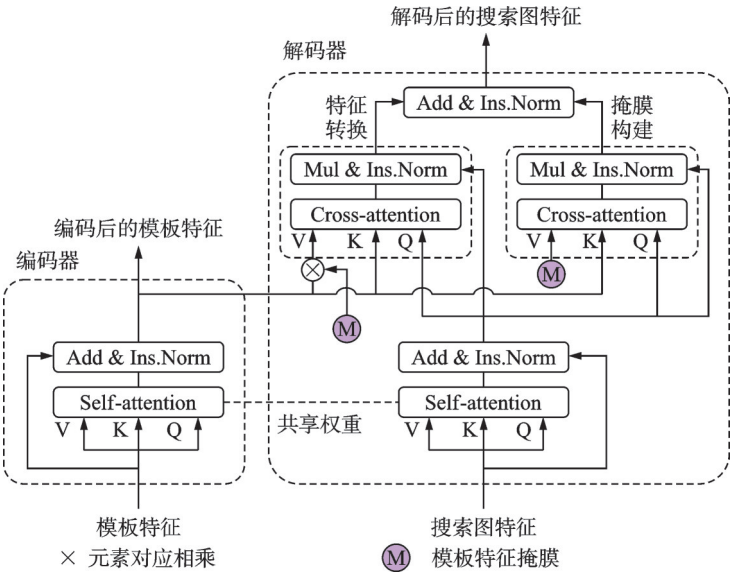

图4 特征加工模块的内部结构

Fig.4 Internal structure of feature processing module

在解码器中,与编码器的考量类似,输入的搜索图特征$\boldsymbol{S}$变为$\boldsymbol{S}'$,再仿照式(2)和式(3),则加工后的搜索图特征$\hat{\boldsymbol{S}}$为

$$\hat{\boldsymbol{S}}=\mathrm{Ins.Norm}(\boldsymbol{S}'+\boldsymbol{A}_{S\to S}\boldsymbol{S}') \tag{4}$$

式中:$\boldsymbol{A}_{S\to S}$为搜索图特征的自相似度矩阵,其计算方式参考式(2)。为了将交汇后模板中的信息传递至搜索图中,接下来通过交叉注意力机制,计算模板与特征图的互相似度矩阵

$$\boldsymbol{A}_{T\to S}=\mathrm{Atten}(\varphi(\hat{\boldsymbol{S}}),\varphi(\hat{\boldsymbol{T}})) \tag{5}$$

式中:$\varphi(\cdot)$的结构和功能与式(2)中$\phi(\cdot)$类似。如此一来,视频内在的先后逻辑便愈发明朗。但由于既往模板背景中存在噪声等干扰因素可能损害特征质量,还需根据模板中已知的目标位置,通过高斯分布估计出着重体现目标所在区域的掩膜$\boldsymbol{M}'$。经掩膜处理后的搜索图特征$\hat{\boldsymbol{S}}_{\mathrm{mask}}$可以强调目标的大致位置

$$\hat{\boldsymbol{S}}_{\mathrm{mask}}=\mathrm{Ins.Norm}(\boldsymbol{A}_{T\to S}\boldsymbol{M}'\otimes\hat{\boldsymbol{S}}) \tag{6}$$

式中:$\boldsymbol{A}_{T\to S}\boldsymbol{M}'\otimes\hat{\boldsymbol{S}}$描述的是两个矩阵元素对应相乘的过程。

接着,通过掩膜屏蔽$\hat{\boldsymbol{T}}$的无用背景,仿照式(3)可将纯净有效的模板特征关联至搜索图特征$\hat{\boldsymbol{S}}$,得到$\hat{\boldsymbol{S}}_{\mathrm{feat}}$

$$\hat{\boldsymbol{S}}_{\mathrm{feat}}=\mathrm{Ins.Norm}(\hat{\boldsymbol{S}}+\boldsymbol{A}_{T\to S}(\hat{\boldsymbol{T}}\otimes\boldsymbol{M}')) \tag{7}$$

最终,解码器输出的特征矩阵为$\hat{\boldsymbol{S}}_{\mathrm{final}}$,由$\hat{\boldsymbol{S}}_{\mathrm{feat}}$和$\hat{\boldsymbol{S}}_{\mathrm{mask}}$相加得到。值得注意的是,为了便于后续处理,需要适当调整解码器和编码器输出结果的尺寸,将$\hat{\boldsymbol{T}}$和$\hat{\boldsymbol{S}}_{\mathrm{final}}$变形成$\boldsymbol{T}_{\mathrm{encoded}}$和$\boldsymbol{S}_{\mathrm{decoded}}$。

得益于训练数据的改进,分段式结构的科学划分和特征加工模块对视频中先后逻辑的充分挖掘,TrSOTST-SDG这款SOT能够适应长期跟踪中目标的属性变化,对物体间干扰的抵御能力较强,因此在理论上非常适合在MOT防遗漏机制中追踪目标。

2.3 重匹配方法

随着检测技术的不断进步,现行检测器受外界干扰而虚报物体的情况已较为罕见,因此检测物体

的遗漏一般是由对应目标在近期画面中彻底不可见导致的。此时使用TrSOTST-SDG这样高性能的SOT也不可能得知目标的具体位置,这就需要MOT算法在运行时腾出部分空间,先行记忆尚存在于画面中的目标ID、特征、位置和相应时刻。

然而,受限于庞大的目标数量和视频帧数,MOT算法只能选择性地记忆上述目标信息。事实上,算法只需要记录那些轨迹即将停滞的目标,由于防遗漏机制中SOT的存在,这些目标指的是SOT无法成功追踪的被遗漏目标,如此可大大减少MOT算法的额外记忆空间和重匹配时间。当某个检测物体与所有目标均匹配失败,防遗漏机制将实施该物体与记忆目标的重匹配,假若两者特征相似性满足判决条件,则将目标ID传递至该物体,并将其位置作为目标轨迹的新起点。同样不能排除的一种可能性是,此检测物体在之前从未出现过,那么对其应用重匹配方法将不会有任何结果,MOT算法会将其标记为新目标。一般情况下,较久不可见的目标回到画面的可能性相对较小,这在重匹配时目标和检测物体间的代价函数中也有所体现

$$\tilde{F}(o,d)=F(o,d)\eta^{\Delta t-1} \tag{8}$$

式中:$F(o,d)$为常规匹配时两者的代价函数,Δt为两者所处帧数的差值,η为大于1的超参数。式(8)中的重匹配代价函数将随着时间的推移而增大,达成的匹配几率也相应降低。

到此为止,本文已明确了这种由高性能SOT和重匹配方法组成的AMM-SOTR防遗漏机制的基本思想和运作方法。但此机制对遗漏现象的真实抑制效果还需要通过实验加以证明,下节将对AMM-SOTR进行一些消融实验和对比实验,通过数据集指标的变动得出结论。

3 实验验证与分析

实验使用的主要设备型号和软件版本为:i9-10900K CPU、32GB RAM、GTX 3080Ti GPU×2、python 3.7和PyTorch1.7.1。

关于AMM-SOTR自身的实验需要一种MOT算法充当测试平台,本文选择了DAHRMOT[9]这一在各大数据集上表现优异的算法,防遗漏机制不会影响MOT主干网络的结构、参数和关键函数,因此关于其自身实验的结论准确性与平台算法无关。

所有测试工作主要基于MOT17数据集[29],这是MOT中的经典数据集,包含了拥挤行人、视角变化和相机移动等复杂场景。实验涉及了数据集上的6个指标,它们的含义如下:(1)MOTA代表跟踪准确率;(2)IDF1代表目标ID正确率;(3)IDSW代表目标ID变更数;(4)MT代表跟踪成功率;(5)ML代表跟踪失败率;(6)FPS代表跟踪速率。实验根据以下标准判断跟踪是否成功:假若目标的预测框和真值框的相交面积超过阈值,则判断跟踪成功,否则失败。其中,除表示次数和帧数的IDSW和FPS外,其他指标均为百分数。关于指标数值大小代表的性能优劣含义,除ML和IDSW在数值更低时对应性能更优外,其他指标均在数值更高时对应性能更优。

训练平台算法DAHRMOT所用的数据集与FairMOT[30]相同,全程共经历了30轮的训练,学习率将在Adam优化器的作用下从首轮的10^{-4}逐渐下降到20轮时的10^{-5}。对于防遗漏机制中的SOT(TrSOTST-SDG)的训练,本文先单独训练短时模块50轮,首轮学习率为10^{-2},同样以Adam优化器调控学习率,采用0.2的衰减参数,再以10^{-4}的学习率开始两个模块的同步训练,保持优化器设置不变。超参η设置为1.3,记忆帧数为50帧。相似性度量中,余弦距离和重叠率的阈值设置分别为0.4和0.5,若计算结果大于以上阈值,则可确定目标和检测对象的匹配关系。

3.1 SOT性能验证

为了验证本文所提出的SOT跟踪器的性能,将其在LaSOT数据集上与常见的单目标跟踪器进行比对,结果如图5所示。

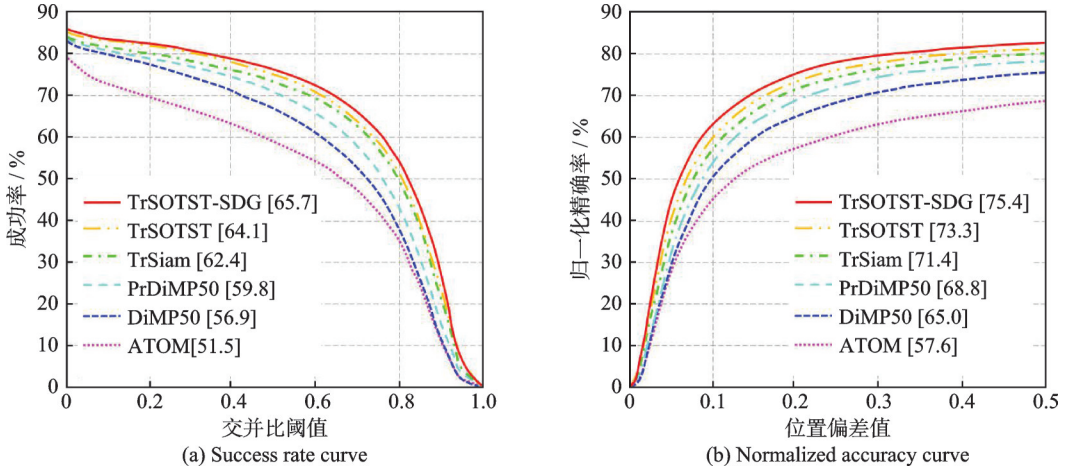

图5 TrSOTST-SDG与其他跟踪算法在LaSOT数据集上的性能对比

Fig.5 Performance comparison between TrSOTST-SDG and other algorithms on LaSOT dataset

在引入扩散模型生成更多的训练数据后,TrSOTST-SDG相比TrSOTST有了1.6%的成功率提升和2.1%的归一化精确率提升。与其余算法相比,TrSOTST-SDG的性能也有了明显提升。

在另一个数据集GOT-10k上进行算法性能比对,如表1所示。对于只引入Transformer的TrSOTST,其增强了时间维度的梳理能力,获得了性能上的提升;而在此基础上引入扩散模型的TrSOTST-SDG则进一步提升了追踪的性能,原因在于扩散模型能将与目标近似的物体引入到训练中,从而极大增强了模型的辨别能力。

表1 TrSOTST-SDG与其他跟踪算法在GOT-10k数据集上的性能对比

Table 1 Performance comparison between TrSOTST-SDG and other algorithms on GOT-10k dataset

跟踪算法	AO/%	SR0.5/%	SR0.75/%
TrSOTST-SDG	70.2	80.8	61.9
TrSOTST	69.0	78.5	60.5
TrSiam[18]	67.3	78.7	58.6
Siam R-CNN[31]	64.9	72.8	59.7
PrDiMP50[32]	63.4	73.8	54.3
DiMP50[33]	61.1	71.7	49.2
ATOM[34]	55.6	63.4	40.2
SiamFC[28]	37.4	40.4	14.4
ECO[35]	31.6	30.9	11.1

3.2 SOT跟踪起点的选择

AMM-SOTR中SOT的跟踪起点可以有两种选择,即目标当前轨迹的末端和首端。末端的距离优势可以节省一定时间,而从首端开始可获得更多的高价值信息,便于SOT细致梳理先后逻辑。两种选择孰优孰劣需要通过实验验证,因此本文使用两种SOT跟踪起点不同的AMM-SOTR,以DAH-RMOT为平台于MOT17验证集上进行了实验,结果如表2所示。表2中,SOT以轨迹首端开启跟踪的算法在前3项反映跟踪质量的指标上表现更加优异,仅在FPS一项指标上落后,但每秒18帧的跟踪速率已能够满足在线跟踪的需要。花费些许时间从首端开启跟踪,相当于能提供给SOT模型更多的目标信息,从而使得其能达成对目标更稳定的跟踪过程。因此,AMM-SOTR中的SOT应当从目标当前轨迹的首端开启跟踪。

表2 MOT17验证集上不同SOT跟踪起点的影响

Table 2 Influence of different trace starting points for SOT on MOT17 validation set

跟踪起点	MOTA/%	IDF1/%	IDSW	FPS
轨迹末端	87.4	85.7	390	20.8
轨迹首端	87.7	86.0	376	18.0

3.3 AMM-SOTR对MOT算法的作用验证

为验证AMM-SOTR能适配于多种MOT算法，减少其中的遗漏现象使性能改善，本文于MOT17数据集上比较了DAHRMOT[9]、FairMOT[30]、TransCenter[36]、GSDT[37]和PermaTrack[38]加入AMM-SOTR后4项指标的变化，如图6所示。

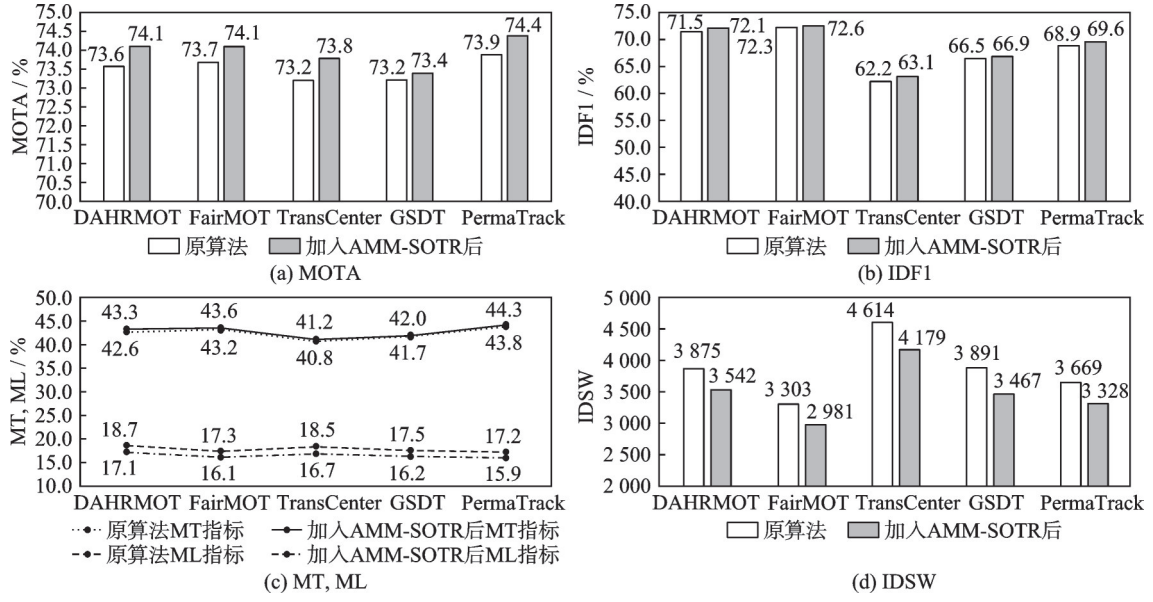

图6 多种MOT算法加入AMM-SOTR前后各项指标的变化

Fig.6 Changes of various metrics before and after adding AMM-SOTR to various MOT algorithms

图6中所有MOT算法在加入AMM-SOTR后，无论是关于跟踪精度的MOTA，还是关于目标ID准确性的IDF1和IDSW，抑或是关于跟踪成败几率的MT和ML，都有了全面的改善，这正是由于防遗漏机制及时处理了遗漏个体，避免了它们对跟踪性能的持续损害。因此可得出结论，基于TrSOTST-SDG和重匹配的防遗漏机制能够有效抑制遗漏现象，并可顺利安置在风格迥异的MOT算法中。

3.4 SOT和重匹配的作用验证

AMM-SOTR中的SOT和重匹配处理的遗漏个体不同，因此在理论上对遗漏现象的影响重点也有区别。本文改变了DAHRMOT中防遗漏机制的完整度，于MOT17验证集上分别对4种情况进行了实验，如表3所示。表3中，防遗漏机制在含有SOT时，MOT算法的MOTA数值较高；而在含有重匹配方法时，IDF1和IDSW的数值更理想。因此可得出结论，SOT的主要贡献在于跟踪精度的改善，而重匹配方法的主要贡献在于目标ID正确率的改善。另一方面，FPS数值在AMM-SOTR含有SOT时普遍较低，意味着SOT是执行该机制所花费时间的主要来源。

表3 MOT17验证集上不同完整度的AMM-SORT的影响

Table 3 Influence of AMM-SORT with different integrities on MOT17 validation set

SOT	重匹配	MOTA/%	IDF1/%	IDSW	FPS
√	√	87.7	86.0	376	18.0
√	×	87.5	85.4	419	19.1
×	√	87.3	85.7	398	20.9
×	×	87.2	85.2	435	21.5

3.5 AMM-SOTR在GNMOT追踪器中的性能验证

GNMOT采用DaSiamRPN[39]作为SOT追踪器，来应对目标无检测对象匹配的情况；而所提出的

AMM-SOTR与之相比,能发挥出更佳的性能。

表4所示为在MOT17测试集上,验证GNMOT有无AMM-SOTR的性能对比实验,加"*"表示其原有的SOT追踪器被替换成了本文中的AMM-SOTR。从实验结果可以看出,有了AMM-SOTR的辅助,并在保证回溯机制相同的基础上,GNMOT在多目标追踪的指标上获得了全方位的提升。其中,ML跟踪失败率下降了0.8%,提升最为明显。

表4 MOT17测试集上GNMOT有无AMM-SOTR的性能对比

Table 4 GNMOT performance comparison with and without AMM-SOTR on MOT17 test set

多目标跟踪器	MOTA/%	IDF1/%	IDSW	MT/%	ML/%
GNMOT[8]	50.2	47.0	5273	19.3	32.7
GNMOT*	50.7	47.3	5181	19.4	31.9

3.6 AMM-SOTR处理遗漏个体实例

除了以上的总体指标外,分析跟踪中一些具体的实例可更直观感受到防遗漏机制的作用。图7为AMM-SOTR处理遗漏个体的两个实例,所有图像均来自MOT17数据集。

图7 AMM-SOTR处理遗漏个体实例

Fig.7 Examples of AMM-SOTR handling missing entities

图7(a)实例中,在虚线上方,作为先前目标的行人此时未被检测到,因此该目标将成为遗漏个体,这触发了AMM-SOTR中的SOT。在虚线下方,左边的跟踪结果由SOT给出,而右边的结果来自MOT算法中内置的卡尔曼滤波器[40]。显而易见,SOT的结果更确切地描述了目标位置,AMM-SOTR对遗漏个体的处理更精细。这也进一步表明了,在扩散模型的反向去噪过程中加入语义引导的SDG模型能为SOT提供丰富的训练数据,加强其泛化能力。其中,图片引导包括内容和风格引导,内容引导能使得生成的目标具有不同的姿势、结构以及布局,风格引导能使得产生的图片具有未知的风格。而细粒度的语言引导能准确地生成所要求的图片。整合这两种引导方式的多模态引导能提供有效的互补信息,最终能生成多样化的高质量目标图片,进而能加强SOT对目标外貌的区分和追踪能力,使得遗漏目标能较好地复原。

图7(b)实例中,在虚线上方,2号目标逐渐变为不可见状态,在其基本消失以至于SOT都无法追踪

后，AMM-SOTR将记下该目标的信息。在虚线下方，该目标后续又回到画面中，此时的检测物体不可能与上一帧的目标完成匹配，因而该检测物体成为遗漏个体。左边的跟踪结果为：通过重匹配在历史目标中找到了记录，使目标ID得以延续，因此其跟踪ID为2。而右边结果未经过重匹配，跟踪器将会认为这是一个从未出现的物体，将会为其分配一个全新的ID即3号。由此可得，由于闭塞、遮挡导致的目标遗漏，通过一定时间内的遗漏轨迹记忆，当其在画面中重新出现时，便能通过与以往轨迹的匹配来重新激活。

以上结果表明，经过SOT机制和重匹配机制，某些遗漏个体能有效地重新被关联追踪。

4 结束语

提出了一种基于SOT和重匹配的防遗漏机制AMM-SOTR。一方面，TrSOTST-SDG单目标跟踪器通过扩散模型丰富训练数据，增强了目标区分能力；并采用Transformer来梳理时间维度信息，从而能精确定位被遗漏的目标。另一方面，AMM-SOTR机制利用提前记忆的目标信息重匹配孤立的检测物体，保持了目标ID的准确。实验表明，本文提出的防遗漏机制可有效抑制遗漏现象，减少遗漏个体，全面改善各种MOT算法的性能。

本文工作还存在一定不足，防遗漏机制可看作对部分遗漏结果的补救，但无法从根本上避免遗漏个体的出现，加入该机制后的MOT算法依然存在数量可观的目标ID错乱和跟踪失败案例。因此，在后续工作中可以尝试通过更严密的关联逻辑进一步减少遗漏现象的发生几率。

参考文献：

[1] LUO W, XING J, MILAN A, et al. Multiple object tracking: A literature review[J]. Artificial Intelligence, 2021, 293: 103448.

[2] EMAMI P, PARDALOS P M, ELEFTERIADOU L, et al. Machine learning methods for data association in multi-object tracking[J]. ACM Computing Surveys (CSUR), 2020, 53(4): 1-34.

[3] 张良，王运锋. 基于贪心策略的多目标跟踪数据关联算法[J]. 四川大学学报：自然科学版，2018(1): 56-60.
ZHANG Liang, WANG Yunfeng. Multi-target tracking data association algorithm based on greedy strategy[J]. Journal of Sichuan University (Natural Science Edition), 2018(1): 56-60.

[4] KUHN H W. The Hungarian method for the assignment problem[J]. Naval Research Logistics Quarterly, 1955, 2(1/2): 83-97.

[5] MUNKRES J. Algorithms for the assignment and transportation problems[J]. Journal of the Society for Industrial and Applied Mathematics, 1957, 5(1): 32-38.

[6] FORD L R, FULKERSON D R. Maximal flow through a network[J]. Canadian Journal of Mathematics, 1956, 8: 399-404.

[7] ZHOU J, CUI G, HU S, et al. Graph neural networks: A review of methods and applications[J]. AI Open, 2020, 1: 57-81.

[8] LI J, GAO X, JIANG T. Graph networks for multiple object tracking[C]//Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision. [S.l.]: IEEE, 2020: 719-728.

[9] 张毅锋，陈曦，张嘉成. 基于深度学习高分辨率网络的多目标跟踪[J]. 东南大学学报，2023, 53(1): 14-20.
ZHANG Yifeng, CHEN Xi, ZHANG Jiacheng. Multi-object tracking based on deep aggregation high-resolution network[J]. Journal of Southeast University (Natural Science Edition), 2023, 53(1): 14-20.

[10] GIRSHICK R, DONAHUE J, DARRELL T, et al. Rich feature hierarchies for accurate object detection and semantic segmentation[C]//Proceedings of 2014 IEEE Conference on Computer Vision and Pattern Recognition. Columbus: IEEE, 2014: 580-587.

[11] REDMON J, FARHADI A. YOLOv3: An incremental improvement[EB/OL]. (2018-04-08)[2023-05-08]. https://arxiv.org/abs/1804.02767.

[12] LIU W, ANGUELOV D, ERHAN D, et al. SSD: Single shot multibox detector[C]//Proceedings of Computer Vision-ECCV 2016. Amsterdam: Springer, 2016: 21-37.

[13] DUAN K, BAI S, XIE L, et al. CenterNet: Keypoint triplets for object detection[C]//Proceedings of 2019 IEEE/CVF International Conference on Computer Vision (ICCV). Seoul: IEEE, 2019: 6569-6578.

[14] BAR-SHALOM Y, TSE E. Tracking in a cluttered environment with probabilistic data association[J]. Automatica, 1975, 11(5): 451-460.

[15] FORTMANN T E, BAR-SHALOM Y, SCHEFFE M. Multi-target tracking using joint probabilistic data association[C]//Proceedings of 1980 19th IEEE Conference on Decision and Control including the Symposium on Adaptive Processes. [S.l.]: IEEE, 1980: 807-812.

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

[17] VASWANI A, SHAZEER N, PARMAR N, et al. Attention is all you need[C]//Proceedings of Neural Information Processing Systems 2017. Long Beach: [s.n.], 2017: 6000-6010.

[18] WANG N, ZHOU W, WANG J, et al. Transformer meets tracker: Exploiting temporal context for robust visual tracking[C]//Proceedings of 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Nashville: IEEE, 2021: 1571-1580.

[19] CHEN X, YAN B, ZHU J, et al. Transformer tracking[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. [S.l.]: IEEE, 2021: 8126-8135.

[20] HO J, JAIN A, ABBEEL P. Denoising diffusion probabilistic models[J]. Advances in Neural Information Processing Systems, 2020, 33: 6840-6851.

[21] SONG J, MENG C, ERMON S. Denoising diffusion implicit models[EB/OL]. (2020-10-06)[2023-05-08]. https://arxiv.org/abs/2010.02502.

[22] DHARIWAL P, NICHOL A. Diffusion models beat GANs on image synthesis[J]. Advances in Neural Information Processing Systems, 2021, 34: 8780-8794.

[23] LIU X, PARK D H, AZADI S, et al. More control for free! image synthesis with semantic diffusion guidance[C]//Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision. [S.l.]: IEEE, 2023: 289-299.

[24] 张卓翼 . 基于孪生网络的目标跟踪算法研究[D]. 南京: 东南大学, 2020.
ZHANG Zhuoyi .Research on object tracking based on siamese network[D]. Nanjing: Southeast University, 2020.

[25] FAN H, LING H, LIN L, et al. LaSOT: A high-quality benchmark for large-scale single object tracking[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Long Beach: IEEE, 2019: 5369-5378.

[26] HUANG L, ZHAO X, HUANG K. GOT-10k: A large high-diversity benchmark for generic object tracking in the wild[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2021, 43(5): 1562-1577.

[27] ZHANG H, CISSE M, DAUPHIN Y N, et al. Mixup: Beyond empirical risk minimization[EB/OL]. (2017-10-25)[2023-05-08]. https://arxiv.org/abs/1710.09412.

[28] BERTINETTO L, VALMADRE J, HENRIQUES J F, et al. Fully-convolutional siamese networks for object tracking[C]//Proceedings of European Conference on Computer Vision. Amsterdam: Springer, 2016: 850-865.

[29] MILAN A, LEAL-TAIXÉ L, REID I, et al. MOT16: A benchmark for multi-object tracking[EB/OL]. (2016-03-02)[2023-05-08]. https://arxiv.org/abs/1603.00831.

[30] ZHANG Y, WANG C, WANG X, et al. FairMOT: On the fairness of detection and re-identification in multiple object tracking[J]. International Journal of Computer Vision, 2021, 129(11): 3069-3087.

[31] VOIGTLAENDER P, LUITEN J, TORR PH, et al. Siam R-CNN: Visual tracking by re-detection[C]//Proceedings of 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Seattle: IEEE, 2020: 6578-6588.

[32] DANELLJAN M, GOOL LV, TIMOFTE R. Probabilistic regression for visual tracking[C]//Proceedings of 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Seattle: IEEE, 2020: 7183-7192.

[33] BHAT G, DANELLJAN M, VAN GOOL L, et al. Learning discriminative model prediction for tracking[C]//Proceedings of 2019 IEEE/CVF International Conference on Computer Vision (ICCV). Seoul: IEEE, 2019: 6181-6190.

[34] DANELLJAN M, BHAT G, KHAN F S, et al. ATOM: Accurate tracking by overlap maximization[C]//Proceedings of 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Long Beach: IEEE, 2019: 4655-4664.

[35] DANELLJAN M, BHAT G, SHAHBAZ-KHAN F, et al. ECO: Efficient convolution operators for tracking[C]//

Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Honolulu: IEEE, 2017: 6638-6646.

[36] XU Y, BAN Y, DELORME G, et al. TransCenter: Transformers with dense representations for multiple-object tracking[EB/OL]. (2021-03-28) [2023-05-08]. https://arxiv.org/abs/2103.15145.

[37] WANG Y, KITANI K, WENG X. Joint object detection and multi-object tracking with graph neural networks[C]// Proceedings of 2021 IEEE International Conference on Robotics and Automation (ICRA). Xi'an: IEEE, 2021: 13708-13715.

[38] TOKMAKOV P, LI J, BURGARD W, et al. Learning to track with object permanence[C]//Proceedings of 2021 IEEE/CVF International Conference on Computer Vision (ICCV). Montreal: IEEE, 2021: 10860-10869.

[39] PEI Y, BISWAS S, FUSSELL D S. An elementary introduction to Kalman filtering[J]. Communications of the ACM, 2019, 62(11): 122-133.

[40] ZHU Z, WANG Q, LI B, et al. Distractor-aware siamese networks for visual object tracking[C]//Proceedings of the European Conference on Computer Vision (ECCV). [S.l.]: ECCU, 2018: 101-117.

作者简介:

张毅锋(1963-),男,博士,副教授,研究方向:计算机视觉、目标跟踪和医学图像分析,E-mail:yfz@seu.edu.cn。

张嘉成(1997-),**通信作者**,男,硕士研究生,研究方向:目标跟踪,E-mail:1273228954@qq.com。

李元浩(1998-),男,硕士研究生,研究方向:多目标追踪。

(编辑:夏道家)