第 60 卷 第 10 期/2023 年 5 月/激光与光电子学进展 研究论文

激光与光电子学进展 先进成像

基于时空轨迹融合的遮挡视频行人重识别

云霄*，宋凯莉，张晓光，袁新超

中国矿业大学信息与控制工程学院，江苏 徐州 221008

摘要 针对视频行人重识别中的目标行人大范围遮挡的问题，将具有时间关联性、不受遮挡影响的行人轨迹预测与行人重识别相结合，提出一种基于时空轨迹融合的遮挡视频行人重识别算法。首先，从时间与空间域出发，实现符合社会属性的精确行人轨迹坐标预测；其次，构建时空轨迹融合特征，将视频序列中的表观视觉特征与行人轨迹中的坐标数据有效结合，有效缓解查询集中遮挡问题对重识别性能造成的影响；最后，构建适用于所提算法的轨迹融合数据集MARS_traj，并通过实验证明所提算法对遮挡视频重识别性能的有效提升。

关键词 图像处理；机器视觉；视频行人重识别；目标遮挡；行人轨迹预测

中图分类号 TP391.41 **文献标志码** A **DOI**：10.3788/LOP220812

Occluded Video-Based Person Re-Identification Based on Spatial-Temporal Trajectory Fusion

Yun Xiao*, Song Kaili, Zhang Xiaoguang, Yuan Xinchao

School of Information and Control Engineering, China University of Mining and Technology, Xuzhou 221008, *Jiangsu, China*

Abstract Aiming at the problem of wide-range occlusion of target pedestrians in video pedestrian re-identification, a pedestrian re-identification algorithm based on spatio-temporal trajectory fusion is proposed by combining pedestrian trajectory prediction with pedestrian re-identification, which is time-related and not affected by occlusion. First, from the time and space domains, accurate pedestrian trajectory coordinate prediction in line with social attributes is realized. Second, the spatiotemporal trajectory fusion feature is constructed to effectively combine the apparent visual features in the video sequence with the coordinate data in the pedestrian trajectory, which effectively alleviates the impact of centralized occlusion on the re-identification performance. Finally, a trajectory fusion dataset MARS_traj suitable for the proposed algorithm is constructed, and experiments show that the proposed algorithm can effectively improve the performance of the occlusion video re-identification.

Key words image processing; machine vision; video-based person re-identification; target occlusion; pedestrian trajectory prediction

1 引 言

行人重识别指在跨摄像头环境下拍摄的行人图像中检索出具有相同身份的行人[1]，具有广阔的应用前景，如刑事侦查、智能安防[2]、查人寻踪[3]等方面。根据输入数据的不同，行人重识别又可分为图像行人重识别和视频行人重识别[4]。与基于图像的行人重识别不同，基于视频的行人重识别的输入信息是视频序列，即根据时间顺序排序的多个图像集合。视频序列中不仅包含行人的外观信息，还包含行人的时间信息等[5]。随着视频监控设备的发展[6]，利用时间信息线索的视频行人重识别受到了更多关注[7]，同时也面临光照变化、拍摄场景多样化等更多挑战，其中，目标遮挡问题是最大的难点之一。

常用的解决遮挡问题的视频行人重识别方法有注意力机制[8]和生成对抗网络[9]两种。注意力机制利用注意力模型从视频序列中选择有判别力的帧生成信息丰富的视频表征，但会丢弃部分遮挡的图像，如Chen

收稿日期：2022-02-25；**修回日期**：2022-03-29；**录用日期**：2022-04-19；**网络首发日期**：2022-04-29

基金项目：国家自然科学基金(61902404，51804304)

通信作者：*xyun@cumt.edu.cn

等[10]提出的协同注意力嵌入策略，Li等[11]提出的多样性正则化时空注意力模型等。因此有学者提出利用生成对抗网络来复现被遮挡部分外观表征，如Hou等[12]提出的时空补全网络。然而生成对抗网络只能恢复被小半部分遮挡的图像外观，难以恢复大范围遮挡视频序列中的行人图像外观。所述大范围遮挡视频序列是指单帧图像大范围遮挡（行人基本被完全遮挡）或多帧长时间遮挡（连续多帧图像行人受到大范围目标遮挡）两种情况。

行人轨迹预测通过观察行人的历史轨迹信息对其未来轨迹进行预测。行人轨迹描述的行人位置坐标随时间而变化，具有较强的时间关联性[13]。轨迹数据[14-15]由按时间顺序排列的一系列轨迹数据点组成，不受外界遮挡、目标表观变化等影响。经典的行人轨迹预测算法有Alahi等[16]提出的社会长短时记忆网络模型（Social LSTM）以及Gupta等[17]提出的社会力生成对抗网络（Social-GAN）等。本文提出一种基于时空轨迹融合（STTF）的大范围遮挡视频行人重识别算法，将轨迹预测与视频行人重识别结合。首先，从时间域与空间域角度验证行人轨迹预测的准确性；其次，构建STTF特征，将视频序列中的表观视觉特征与行人轨迹中的坐标数据有效结合，依据时序融合与空间融合挑选出新的查询视频序列，用不受遮挡影响的新的查询行人视频序列替换原本的大范围遮挡查询视频序列并引入重识别网络中，解决遮挡造成的表观视觉特征提取错误问题，有效缓解查询集中大范围物体遮挡问题对重识别性能造成的影响；最后，以MARS数据集为基础构建轨迹融合MARS_traj数据集，并为行人视频序列添加时间帧数和空间坐标信息，适用于基于轨迹预测的遮挡视频行人重识别问题。所提算法基于测试阶段进行改进，可以将STTF模型与任意视频行人重识别模型进行结合，提高重识别精度。

2 算法网络框架

2.1 整体框架

基于STTF模型的遮挡视频行人重识别框架如图1所示。首先，采用相似度判别方法[12]对原视频序列是否含有遮挡行人图像进行判断，将含有遮挡的视频序列输入STTF模型，从时间与空间域出发实现行人未来轨迹坐标计算；其次，构建STTF特征，将预测得到的轨迹序列集query_pred与gallery候选集进行时间域和空间域上的融合，获取新的视频序列查询集query_TP，用新的query_TP查询集代替原本含有大范围目标遮挡的查询集，与gallery候选集分别进行融合特征提取，有效缓解查询集中大范围遮挡问题对重识别性能造成的影响；最后，根据视频重识别模型提取的含有表观视觉信息与运动轨迹信息的STTF特征，利用特征相似度输出候选视频的top-N排序结果。

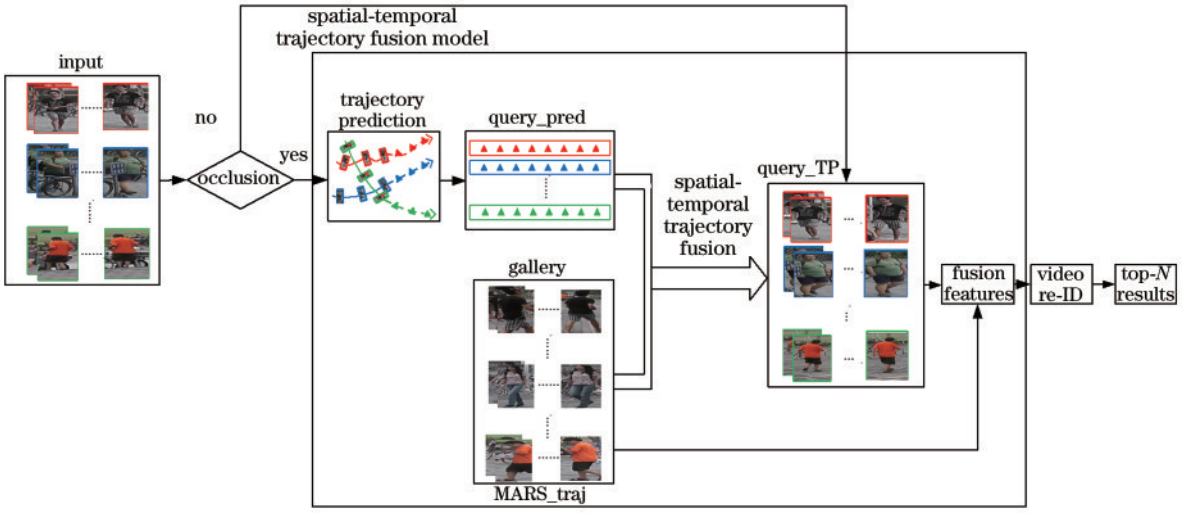

图1 基于STTF的遮挡视频行人重识别框架

Fig. 1 Video-based person re-identification framework based on spatial-temporal trajectory fusion

2.2 STTF模型

2.2.1 Social-GAN轨迹坐标计算

为了后续行人时空轨迹预测，采用Social-GAN[17]计算行人轨迹初步预测坐标。Social-GAN由生成器、鉴别器和池化模块组成，其中，生成器基于编码器-解码器框架构建，并通过池化模块链接编码器和解码器的隐藏状态，从合理性、多样性和预测速度等方面提升模型性能。

轨迹预测整体框架如图2所示。首先，从数据集图像序列中提取最近8帧行人轨迹坐标嵌入编码器的多层感知器中，获取固定长度的坐标向量$\boldsymbol{e}_i^t=\varphi\left(x_i^t, y_i^t, W_{\mathrm{ee}}\right)$，其中$x_i^t$和$y_i^t$分别表示行人$i$在$t$时刻位置的横坐标和纵坐标，$W_{\mathrm{ee}}$表示嵌入权重，$\varphi(\cdot)$表示ReLU非线性单元的嵌入函数。

然后，将坐标向量$\boldsymbol{e}_i^t$输入t时刻编码器的LSTM单元，对行人已知轨迹信息进行循环递归处理，从而获

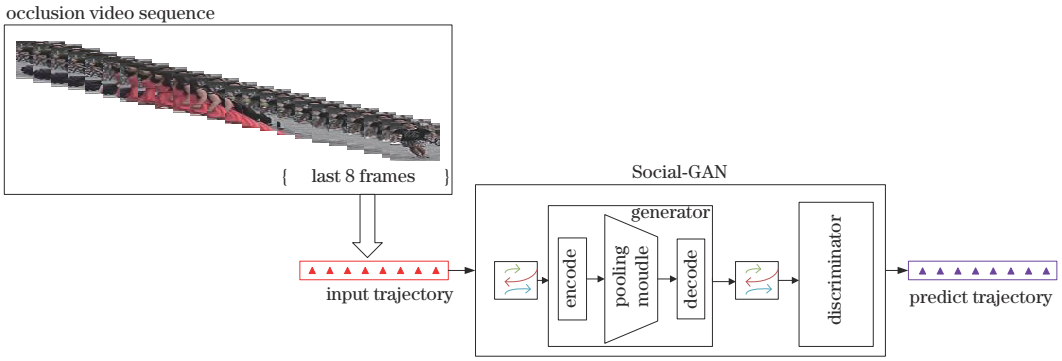

图 2　轨迹预测框架

Fig. 2　Structural framework of trajectory prediction

得编码器隐藏状态 h_{ei}^{t}：

$$h_{ei}^{t}=\mathrm{LSTM}\left(h_{ei}^{t-1},\boldsymbol{e}_{i}^{t};\ W_{\mathrm{encoder}}\right),\tag{1}$$

式中：W_{encoder} 表示一个场景中所有人共享的 LSTM 权重。利用池化模块对行人交互表现出的社会性进行建模，得到每个行人轨迹的池化向量 $\boldsymbol{P}_i$，并根据固定场景特点，将解码器的隐藏状态初始化为 $h_{di}^{t}=\left[\gamma\left(\boldsymbol{P}_i,h_{ei}^{t};\ W_{\mathrm{c}}\right),\boldsymbol{z}\right]$，以设定生成输出轨迹的限制条件。其中，$\gamma(\cdot)$ 表示具有 ReLU 非线性的多层感知器，W_{c} 表示嵌入权重，$\boldsymbol{z}$ 表示从标准正态分布中取样得到的噪声向量。

最后，在完成解码器的隐藏状态初始化后进行后续行人坐标的预测，获得预测的行人轨迹坐标向量 $\left(\hat{x}_i^t,\hat{y}_i^t\right)$：

$$\begin{cases}\boldsymbol{e}_i^t=\varphi\left(x_i^{t-1},y_i^{t-1};\ W_{\mathrm{ed}}\right)\\ \boldsymbol{P}_i=\mathrm{PM}\left(h_{d1}^{t-1},\cdots,h_{dn}^{t}\right)\\ h_{di}^{t}=\mathrm{LSTM}\left[\gamma\left(\boldsymbol{P}_i,h_{di}^{t-1}\right),\boldsymbol{e}_i^t;\ W_{\mathrm{decoder}}\right]\\ \left(\hat{x}_i^t,\hat{y}_i^t\right)=\gamma\left(h_{di}^{t}\right)\end{cases},\tag{2}$$

式中：$\boldsymbol{P}_i$ 表示池化张量；W_{ed} 表示嵌入权重；W_{decoder} 表示解码器共享 LSTM 权重；$\gamma(\cdot)$ 表示具有 ReLU 非线性的多层感知器。得到生成的预测坐标后，将其输入鉴别器（discriminator），根据社会互动规则判别预测轨迹的“真”或“假”，并将判别为“真”的轨迹序列作为真实预测轨迹进行输出，输出轨迹序列长度为 8 帧。

2.2.2　STTF

所提 STTF 模型，从时间域与空间域的角度提取预测轨迹信息，实现行人预测轨迹与行人表观视觉特征的有效结合。

2.2.2.1　时间轨迹融合

考虑预测轨迹与已知历史轨迹在时间上的连续性问题，需要在时间域上计算时间融合损失，如下：

$$l_j^{\mathrm{tem}}=\max\left[\varphi(\Delta t-T),0\right],\tag{3}$$

式中：Δt 为 query 查询集中视频序列最终帧与 gallery 候选集中视频序列第 1 帧的帧数差值；T 为帧数常量值；φ 为较大常量值。通过实验对比帧数常量 T 的取值，选取 $T=4$ 为最优解。由式（3）可知：如果帧数差 Δt 小于帧数常量阈值 T，则时间融合损失 l_j^{tem} 取值为 0；反之，时间融合损失 l_j^{tem} 取值为较大常量值。图 3 表示 $T=4$ 时对 gallery 中视频序列的选取。gallery 中各个视频序列后绿色的√表示时序相符，红色的×表示时序不符。时序相符表示 gallery 候选集中该视频序列的时序信息符合该 query 查询视频的后续时序范围。

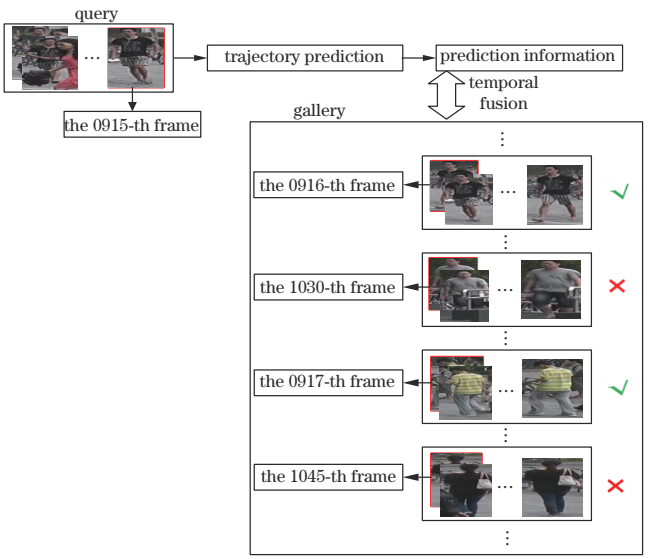

图 3　时间融合选取示例

Fig. 3　Example of temporal trajectory fusion model

2.2.2.2　空间轨迹融合

在实际场景中，存在相邻视频序列间的帧数不连续等问题，造成预测轨迹序列与 gallery 候选集中视频序列的帧数出现错位问题。因此，考虑可能出现的帧数错误情况，计算空间融合损失 l_j^{spa}：

$$l_j^{\mathrm{spa}}=\min\left(l_t\right),\forall t\in 1,2,\cdots,N,N=2,3,\cdots,7,\tag{4}$$

式中：$l_t=\dfrac{\sum_{i=1}^{n}p_i}{n}$，$n=9-t$，$p_i$ 表示预测轨迹序列与 gallery 候选视频序列对应坐标间的欧氏距离；N 表示

允许预测轨迹序列与候选序列帧数的偏离范围。通过实验比较得出$N=4$时为最优解。当$N=4$时，l_t的计算方式如图4所示。

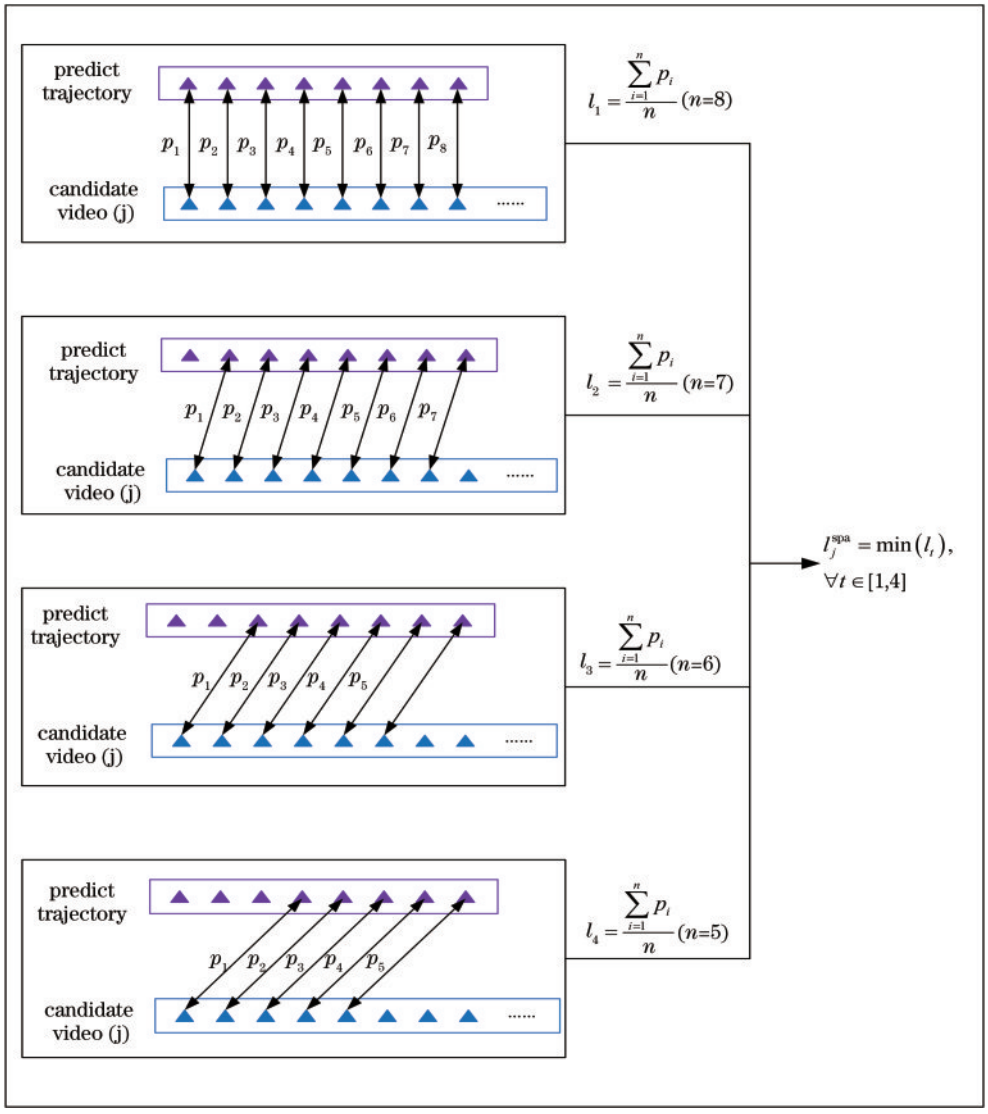

图4　空间融合损失计算示例

Fig. 4　Example of spatial fusion loss calculation

根据式(3)、(4)得到时间融合损失和空间融合损失后，利用式(5)计算query中第i个视频序列与gallery中第j个视频序列间的时空融合损失l_{i-j}：

$$l_{i-j}=\min\left(l_j^{tem}+l_j^{sap}\right),\forall j\in 1,2,\cdots,N_2, \tag{5}$$

式中：N_2为gallery中视频序列总数量。根据式(5)求出使l_{i-j}最小的j值，从而将gallery中第j个视频序列送入query_TP查询集，进行后续的STTF特征提取。

2.3　视频行人重识别模型

视频行人重识别主要有基于循环神经网络、非局部运算、3D卷积、时间池化等时序建模方法。循环卷积网络和3D卷积网络擅长处理局部时间关系和编码相对位置；非本地视频注意力网络虽不善于编码相对位置，但能够较好地模拟长时间依赖关系；时间池化简单粗暴，具有很高的效率，应用广泛，同时可以进一步充分利用具有注意力机制的图像层次特征。

与单独使用视频行人重识别模型相比，结合所提STTF模型能够克服重识别的遮挡问题。将STTF模型分别与基于循环神经网络的COSAM模型[18]、基于非局部运算的STE-NVAN模型[19]、基于3D卷积的AP3D模型[20]和基于时间池化的TCLNet模型[21]这4种常用视频行人重识别模型相结合。通过这4种方法的实验可知，基于时间池化的视频重识别方法TCLNet模型与STTF模型的结合在实验中表现性能最优。

2.4　整体算法流程

算法1为所提算法流程，其中，N_1和N_2分别表示MARS_traj数据集中query和gallery中视频序列的数量。

Algorithm 1: video-based person re-identification based on spatial-temporal trajectory fusion

Input: MARS_traj dataset; trajectory prediction model Social-GAN; video-based person re-identification model

Output: mAP and Rank-k

1) spatial coordinates and temporal information of person ID from query dataset of video sequences are input into Social-GAN model;
2) possible predicted trajectories are generated by generator in Social-GAN based on spatial coordinates and temporal information;
3) discriminator in Social-GAN discriminates generated predicted trajectory, and obtains matching predicted trajectory dataset query_pred.
4) For $i=1:N_1$ do
5) For $j=1:N_2$ do
6) temporal fusion loss l_j^{tem} and spatial fusion loss l_j^{sap} between j-th video sequence in gallery and i-th video prediction trajectory in query_pred are computed by equation (3) and (4), respectively;
7) end
8) $l_{i-j}=\min\left(l_j^{\mathrm{tem}}+l_j^{\mathrm{sap}}\right);\ \forall j\in[1,N_2]$;
9) value of j corresponding to l_{i-j} is obtained and assigned to i_j;
10) sent i_j-th video sequence of gallery into query_TP ;
11) end
12) fusion feature of query_TP and gallery are extracted, respectively;
13) feature distance metrics based on query_TP and gallery are calculated, and feature vectors of all gallery video sequences are ranked according to distance metrics;
14) probability of correct match within ranked gallery is calculated according to query;
15) return mAP and Rank-k.

3　轨迹融合数据集 MARS_traj

MARS 数据集[22]是目前用于视频行人重识别任务中最大、最常用的数据集，由 1261 个不同身份的行人组成，包含 20478 个视频序列，使用 6 个不重叠摄像头进行数据采集。无论是行人身份还是视频序列的数量，MARS 数据集都远远大于 iLIDS-VID 数据集[23]和 PRID2011 数据集[24]，它的边框和标签也是自动生成的，这样的数量与复杂度也使得 MARS 数据集更加具有挑战性。但是，MARS 数据集不具有轨迹信息，无法进行轨迹预测。

在 MARS 数据集基础上，构建适用于基于轨迹预测的遮挡视频行人重识别的轨迹融合数据集 MARS_traj。从 MARS 数据集中手动选取部分行人视频序列放入 MARS_traj 数据集中，为了测试模型对大范围目标遮挡问题的处理能力，MARS_traj 数据集选取的 query 查询集中 90% 以上的行人视频序列含有大范围目标遮挡情况。大范围遮挡视频序列如图 5 所示：第 1 行视频序列中查询行人为黄色短袖男子，但连续多帧基本被完全遮挡；第 2 行视频序列中查询行人为短袖长发女子，同样连续多帧基本被完全遮挡。

为挑选出的 MARS_traj 上的每个行人添加时间帧数和空间坐标信息，并对行人标签进行修改，如图 6 所示。坐标的时空信息由真实轨迹预测 ETH-UCY 数据集提供。ETH 和 UCY 是在行人轨迹预测中广泛使用的公开数据集，共包含有 5 个不同场景，其中，2 个来自 ETH、3 个来自 UCY。数据集总共包含 1600 条行人轨迹，并且每 0.4 s 记录一次行人位置[25]。

所提 MARS_traj 数据集共有行人身份 744 个，视频序列 9659 个，主要信息如表 1 所示。MARS_traj 测试数据集由查询集 query 和候选集 gallery 组成，如图 7 所示，每一行代表同一身份行人的不同视频序列片段和标签，gallery 中右上角打勾的视频序列片段代表与 query 中对应行人的真实后续图像序列片段。采用 MARS_traj 数据集进行实验，验证所提 STTF 模型对遮挡视频的处理能力。

图 5　大范围遮挡视频序列示例

Fig. 5　Example of large-scale occlusion video sequences

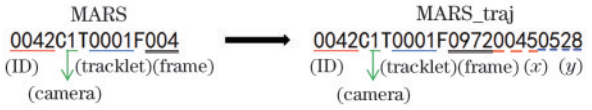

图 6　MARS_traj 数据集中序列标签修改示例

Fig. 6　Example of sequence label modification in MARS_traj dataset

表 1　MARS_traj 数据集

Table 1　MARS_traj dataset

Subset	MARS_traj	
	ID	Tracklets
query	90	90
gallery	119	1271
total	119	1361

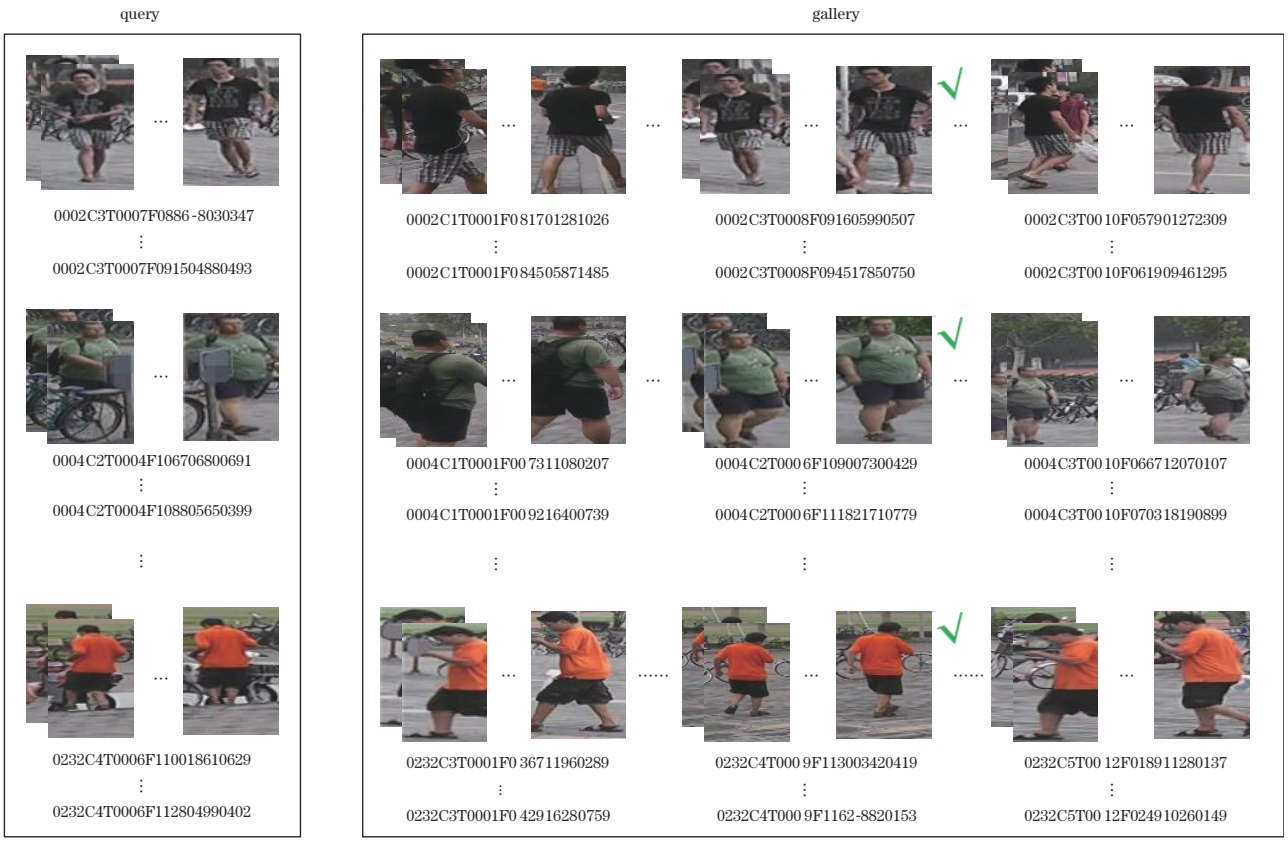

图 7　MARS_traj 数据集组成

Fig. 7　Composition of MARS_traj dataset

4　实验结果与分析

4.1　评价指标与实验配置

实验均在 MARS 数据集和所提 MARS_traj 数据集上进行。行人重识别实验采用常用的 2 个通用评价指标：累计匹配特征（CMC）和平均精度均值（mAP）进行评价。CMC 曲线表示在已排序 gallery 中前 k 个视频匹配正确的可能性，反映出算法的检索精度。mAP 从准确率和召回率上综合评价方法的性能，主要反映算法精度的稳定性。

所提算法是在配备 NVIDIA GeForce GTX 1080 Ti GPU 的机器上使用 PyTorch 1.1.0 版本深度学习框架实现的，并获取了最终重识别准确率及 gallery 中视频序列排序结果。训练过程中，输入视频帧大小都被调整为 256 pixel×128 pixel，并使用 Adam 优化网络。为了更直观地比较所提算法的重识别性能，利用 MATLAB2021b 版本，根据参数分析实验获取的重识别准确率、画出折线对比图，同时利用 Visio 2019 版本，根据 gallery 中部分视频序列排序结果画出直观的可视化比较展示图。

4.2　实验结果分析

所提 STTF 模型可用于任何常用视频行人重识别网络，因此，将经典视频行人重识别算法 STE-NVAN 模型、COSAM 模型、TCLNet 模型以及 AP3D 模型分别与所提 STTF 框架相结合，并测试得到最终结果，如表 2 所示。

从表 2 可以看出，4 种重识别模型在 MARS_traj 数据集上的 mAP 和 Rank-1 值均比在 MARS 数据集上有不同程度的下降，如 COSAM 模型的 mAP 和 Rank-1 分别降低了 8.6 个百分点和 11.8 个百分点，STE-NAVN 模型的 mAP 和 Rank-1 分别降低了 11.65 个百分点和 14.46 个百分点，AP3D 模型的 mAP 和 Rank-1 分别降低了 21.2 个百分点和 26.7 个百分点，TCLNet 模型的 mAP 和 Rank-1 分别降低了 15.65 个百分点和

表 2　MARS 和 MARS_traj 数据集上的性能评估

Table 2　Performance evaluation on MARS and MARS_traj datasets　　unit: %

Method	MARS		MARS_traj	
	mAP	Rank-1	mAP	Rank-1
COSAM	80.50	81.20	71.90	69.40
COSAM+STTF			**93.00**	**92.90**
STE-NAVE	77.80	85.05	66.15	70.59
STE-NAVE+STTF			**92.88**	**96.47**
AP3D	84.10	89.10	62.90	62.40
AP3D+STTF			**90.10**	**91.70**
TCLNet	85.10	89.80	69.45	72.94
TCLNet+STTF			**94.82**	**96.47**

16.86 个百分点。综上，没有添加 STTF 模型的视频行人重识别模型在处理大范围遮挡问题上性能不佳。

为验证所提 STTF 模型对视频序列中大范围遮挡问题的处理能力，将单独视频重识别模型和与 STTF 模型结合后的视频重识别模型在 MARS_traj 数据集上进行对比实验。表 2 中两行为一组，表示一个视频行人重识别模型与所提 STTF 模型结合与否在 MARS_traj 数据集上的性能比较。从表 2 可以看出，将单独视频重识别模型与 STTF 模型结合后在 MARS_traj 数据集上得到的 mAP 和 Rank-1 分别都有不同程度的提升，如 COSAM+STTF 模型的 mAP 和 Rank-1 分别提升了 21.1 个百分点和 23.5 个百分点，STE-NAVN+STTF 模型的 mAP 和 Rank-1 分别提升了 26.73 个百分点和 25.88 个百分点，AP3D+STTF 模型的 mAP 和 Rank-1 分别提升了 27.2 个百分点和 29.3 个百分点，TCLNet+STTF 模型的 mAP 和 Rank-1 分别提升了 25.18 个百分点和 23.53 个百分点。综上，将视频行人重识别模型与所提 STTF 模型结合可以有效缓解大范围遮挡问题对重识别性能的影响。

通过实验发现，COSAM 模型专注于人像主体的特征提取，能够缓解一定程度的背景干扰，因此受遮挡问题影响最小，与 STTF 模型相结合所提升的性能幅度也最小。AP3D 模型对外观表征变化较敏感，遮挡对特征提取造成较大干扰，因此受遮挡影响最大，与 STTF 模型相结合所提升的性能幅度也最大。由于非本地视频注意力网络在视频时空特征提取上的优势，STE-NAVN 模型受遮挡影响没有太大。

4.3　参数分析

4.3.1　时间融合损失常量

参数时间融合损失常量 T 影响时间轨迹融合中对视频序列的融合选择，为了分析不同 T 值对模型性能的影响，选取 2、3、4、5、6、7 等 6 个不同的数值以 TCLNet 模型为例将 STTF 与 TCLNet 模型结合，在 MARS_traj 数据集上进行实验分析，实验结果如图 8(a)所示。

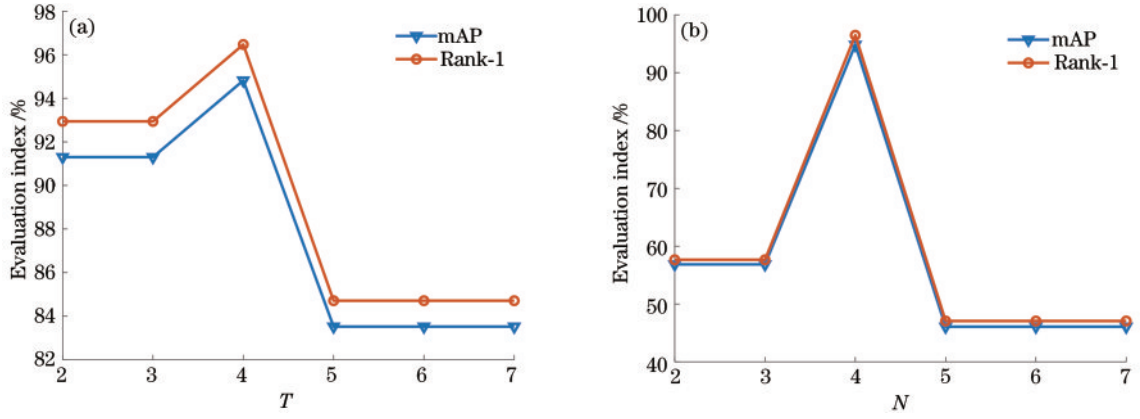

图 8　参数分析实验结果。(a)时间融合损失常量 T 对 Rank-1 和 mAP 的影响；(b)空间融合损失常量 N 对 Rank-1 和 mAP 的影响

Fig. 8　Parameter analysis experimental results. (a) Influence of temporal fusion loss value T on Rank-1 and mAP; (b) influence of spatial fusion loss value N on Rank-1 and mAP

在图 8(a)中，保持空间融合损失常量 N 取值不变，改变式(1)中 T 的取值，研究 T 的取值变化对 Rank-1 和 mAP 的影响。通过实验发现：$T=4$ 时，效果最好；T 值过大，会使时序融合匹配失误的可能性增大；T 值过小，会减少时序融合匹配的选择范围，这两种情况都会导致 Rank-1 和 mAP 下降。

4.3.2　空间融合损失常量

参数空间融合损失常量 N 影响空间轨迹融合中对视频序列的融合选择，为了分析不同 N 值对模型性能的影响，选取 2、3、4、5、6、7 等 6 个不同的数值以 TCLNet 模型为例将 STTF 与 TCLNet 模型结合，在 MARS_traj 数据集上进行实验分析，实验结果如图 8(b)所示。

在图 8(b)中，保持时间融合损失常量 T 取值不变，改变式(2)中 N 的取值，研究 N 的取值变化对 Rank-1 和 mAP 的影响。通过实验发现：$N=4$ 时，效果最好；N 值过大，会增大空间融合匹配失误的概率；N 值过小，会降低空间融合匹配的灵活性，这两种情况都会导致 Rank-1 和 mAP 下降。

4.4　可视化比较

为了更直观地观察所提算法对视频行人重识别效果的提升，分别对 TCLNet、TCLNet＋STTF、AP3D、AP3D＋STTF 在 MARS_traj 数据集上的测试结果进行了可视化实验。可视化实验能够更直观地观察重识别排序结果中的视频序列是否为正确匹配结果，常作为对比算法和参数分析等定量分析实验的定性补充。实验选取其中任意 2 个身份行人的可视化结果进行对比展示，如图 9 所示。图 9 中上下 2 幅图分别表示 2 个不同身份行人各自的排序可视化结果。以图 9 上半部分图像为例，左侧为检索视频序列，右侧表示候选库中 Rank-1～Rank-5（从左到右）的检索结果，视频序列的绿色边框和红色边框分别表示正确和错误的匹配结果，黄色边框表示干扰项，在计算精确度时会被忽略。

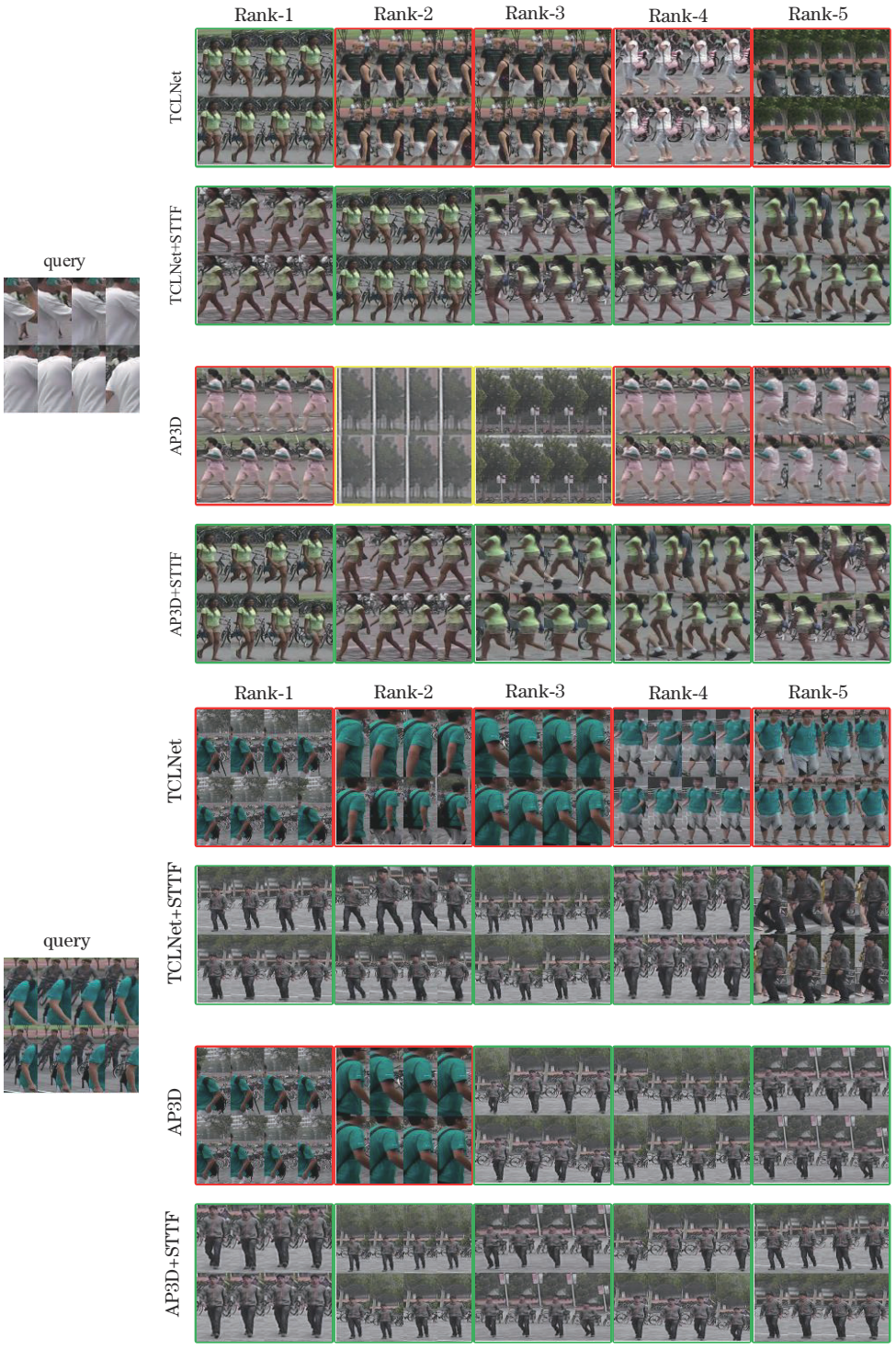

图 9　视频行人重识别结果可视化

Fig. 9　Visualization of video-based person re-identification results

从图 9 可以明显看出，与单独的视频行人重识别模型 TCLNet 或 AP3D 相比，TCLNet+STTF 或 AP3D+STTF 在候选库中 Rank-1～Rank-5 的正确匹配情况得到大幅改善(绿色边框视频序列增加)。由于行人重识别的准确率与正确匹配结果的排序密切相关，Rank-k 表示重识别排序结果中前 k 个视频序列匹配正确的概率，因此 STTF 模型对行人重识别的准确率(尤其是 Rank-1 的准确率)有大幅度提升作用。

5 结　论

提出基于 STTF 模型的遮挡视频行人重识别方法，将行人轨迹预测和视频行人重识别结合，有效缓解查询集中大范围遮挡造成的重识别性能下降问题。以 MARS 数据集为基础构建 MARS_traj 数据集，并添加真实轨迹坐标，使其更好地适用于基于行人轨迹的重识别课题研究。在 MARS_traj 数据集上的大量实验表明，视频行人重识别与 STTF 模型的结合可以明显改善物体大范围遮挡问题对重识别性能的影响，对识别精度有明显提升。随着定位技术的发展，将视频序列与其坐标相结合更容易，使得轨迹预测和视频行人重识别的结合更具有实际应用意义。

参 考 文 献

[1] 张正一，丁建伟，魏慧雯，等. 基于注意力机制的多级特征级联行人重识别[J]. 激光与光电子学进展，2021，58(22)：2215003.
Zhang Z Y，Ding J W，Wei H W，et al. Multi-level features cascade for person re-identification based on attention mechanism[J]. Laser & Optoelectronics Progress，2021，58(22)：2215003.

[2] 张德祥，袁培成，王俊. 基于多尺度批量特征丢弃网络的行人重识别研究[J]. 激光与光电子学进展，2022，59(12)：1215009.
Zhang D X，Yuan P C，Wang J. Person reidentification based on multiscale batch feature-discarding network[J]. Laser & Optoelectronics Progress，2022，59(12)：1215009.

[3] 王凤随，刘芙蓉，陈金刚，等. 融合注意力机制的多损失联合跨模态行人重识别方法[J]. 激光与光电子学进展，2022，59(8)：0810010.
Wang F S，Liu F R，Chen J G，et al. Multi-loss joint cross-modality person re-identification method integrating attention mechanism[J]. Laser & Optoelectronics Progress，2022，59(8)：0810010.

[4] 李梦静，吉根林. 视频行人重识别研究进展[J]. 南京师大学报(自然科学版)，2020，43(2)：120-130.
Li M J，Ji G L. Research progress on video-based person re-identification[J]. Journal of Nanjing Normal University (Natural Science Edition)，2020，43(2)：120-130.

[5] Jiang M，Leng B，Song G L，et al. Weighted triple-sequence loss for video-based person re-identification[J]. Neurocomputing，2020，381：314-321.

[6] 武加文，王世勇. 基于统计的灰度视频自适应背景建模算法[J]. 中国激光，2021，48(3)：0309001.
Wu J W，Wang S Y. Statistical-based adaptive background modeling algorithm for grayscale video[J]. Chinese Journal of Lasers，2021，48(3)：0309001.

[7] Li P K，Pan P B，Liu P，et al. Hierarchical temporal modeling with mutual distance matching for video based person re-identification[J]. IEEE Transactions on Circuits and Systems for Video Technology，2021，31(2)：503-511.

[8] Xu K，Ba J，Kiros R，et al. Show，attend and tell：neural image caption generation with visual attention[EB/OL]. (2015-02-10)[2022-02-05]. https://arxiv.org/abs/1502.03044.

[9] Goodfellow I J，Pouget-Abadie J，Mirza M，et al. Generative adversarial networks[EB/OL]. (2014-06-10)[2022-02-05]. https://arxiv.org/abs/1406.2661.

[10] Chen D P，Li H S，Xiao T，et al. Video person re-identification with competitive snippet-similarity aggregation and Co-attentive snippet embedding[C]//2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition，June 18-23，2018，Salt Lake City，UT，USA. New York：IEEE Press，2018：1169-1178.

[11] Li S，Bak S，Carr P，et al. Diversity regularized spatiotemporal attention for video-based person re-identification[C]//2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition，June 18-23，2018，Salt Lake City，UT，USA. New York：IEEE Press，2018：369-378.

[12] Hou R B，Ma B P，Chang H，et al. VRSTC：occlusion-free video person re-identification[C]//2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)，June 15-20，2019，Long Beach，CA，USA. New York：IEEE Press，2019：7176-7185.

[13] 张志远，倪国新，徐艳国. 轨迹预测技术的现状及发展综述[J]. 电子测量技术，2020，43(13)：111-116.
Zhang Z Y，Ni G X，Xu Y G. Review of the status and development of trajectory prediction technology[J]. Electronic Measurement Technology，2020，43(13)：111-116.

[14] 曾氢菲，刘雪梅，冯焱，等. 双光束激光焊接机器人轨迹优化[J]. 中国激光，2021，48(18)：1802020.
Zeng Q F，Liu X M，Feng Y，et al. Trajectory optimization of dual beam laser welding robot[J]. Chinese Journal of Lasers，2021，48(18)：1802020.

[15] 王宵宵，周骛，王芳婷，等. 基于离焦成像的粒子轨迹测速[J]. 光学学报，2021，41(19)：1912004.
Wang X X，Zhou W，Wang F T，et al. Particle streak velocimetry based on defocused imaging[J]. Acta Optica Sinica，2021，41(19)：1912004.

[16] Alahi A，Goel K，Ramanathan V，et al. Social LSTM：human trajectory prediction in crowded spaces[C]//2016 IEEE Conference on Computer Vision and Pattern Recognition，June 27-30，2016，Las Vegas，NV，USA. New York：IEEE Press，2016：961-971.

[17] Gupta A，Johnson J，Li F F，et al. Social GAN：socially acceptable trajectories with generative adversarial networks[C]//2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition，June 18-23，2018，Salt Lake City，UT，USA. New York：IEEE Press，2018：

2255-2264.

[18] Subramaniam A, Nambiar A, Mittal A. Co-segmentation inspired attention networks for video-based person re-identification[C]//2019 IEEE/CVF International Conference on Computer Vision (ICCV), October 27-November 2, 2019, Seoul, Korea (South). New York: IEEE Press, 2019: 562-572.

[19] Liu C T, Wu C W, Wang Y C F, et al. Spatially and temporally efficient non-local attention network for video-based person re-identification[EB/OL]. (2019-08-05)[2021-05-03]. https://arxiv.org/abs/1908.01683.

[20] Gu X Q, Chang H, Ma B P, et al. Appearance-preserving 3D convolution for video-based person re-identification[M]//Vedaldi A, Bischof H, Brox T, et al. Computer vision-ECCV 2020. Lecture notes in computer science. Cham: Springer, 2020, 12347: 228-243.

[21] Hou R B, Chang H, Ma B P, et al. Temporal complementary learning for video person re-identification [EB/OL]. (2020-07-18)[2021-05-03]. https://arxiv.org/abs/2007.09357.

[22] Zheng L, Bie Z, Sun Y F, et al. MARS: a video benchmark for large-scale person re-identification[M]// Leibe B, Matas J, Sebe N, et al. Computer vision-ECCV 2016. Lecture notes in computer science. Cham: Springer, 2016, 9910: 868-884.

[23] Hirzer M, Beleznai C, Roth P M, et al. Person re-identification by descriptive and discriminative classification [C]//SCIA'11: Proceedings of the 17th Scandinavian conference on Image analysis, May, 2011, Ystad, Sweden. New York: ACM Press, 2011: 91-102.

[24] Wang T Q, Gong S G, Zhu X T, et al. Person re-identification by video ranking[M]//Fleet D, Pajdla T, Schiele B, et al. Computer vision-ECCV 2014. Lecture notes in computer science. Cham: Springer, 2014, 8692: 688-703.

[25] Zamboni S, Kefato Z T, Girdzijauskas S, et al. Pedestrian trajectory prediction with convolutional neural networks[J]. Pattern Recognition, 2022, 121: 108252.