

DOI:10.11918/201910048

一种基于视频预测的红外行人目标跟踪方法

柳恩涵¹, 张 锐¹, 赵 硕², 王 茹¹

(1. 哈尔滨理工大学 自动化学院, 哈尔滨 150080; 2. 哈尔滨工程大学 计算机科学与技术学院, 哈尔滨 150001)

摘 要: 红外视频与普通彩色视频相比易受周围环境的影响, 在红外行人跟踪中行人目标外观轮廓和灰度分布常有较大幅度变化, 导致跟踪困难. 为解决此问题, 本文提出了一种 VPSiamRPN (Video Prediction with Siamese Region Proposal Network) 红外行人目标跟踪系统. 在 SiamRPN (Siamese Region Proposal Network) 网络的基础上, 针对目标形变、目标遮挡和背景杂波等严重影响红外跟踪效果的因素进行研究, 设计将 PredNet (Deep Predictive Coding Networks for Video Prediction and Unsupervised) 网络的图像预测功能结合应用到 SiamRPN 网络上, 以提高跟踪模板与被检测目标的相似度, 增强目标跟踪中的模型匹配能力, 从而提高对红外行人目标的跟踪能力. 通过改变网络层数、预测所用的目标图像及图像帧数、网络的跟踪策略, 对网络进行优化, 设计了 9 组对比试验, 在 PTB-TIR 数据集上, 与 SiamRPN 网络客观定量对比, 通过跟踪的成功率和重叠率在 10 种属性上对网络进行评估. 实验结果表明: 本文网络对红外目标的识别在热交叉、强度变化、遮挡和尺寸变化等多种属性上的跟踪成功率和重叠率均较 SiamRPN 网络有较大提高, 显示出对红外行人跟踪的良好性能, 在这一领域将有广阔的应用前景.

关键词: 目标跟踪; 视频预测; 孪生网络; 红外; PredNet

中图分类号: TP391 **文献标志码:** A **文章编号:** 0367-6234(2020)10-0192-09

Infrared pedestrian target tracking method based on video prediction

LIU Enhao¹, ZHANG Rui¹, ZHAO Shuo², WANG Ru¹

(1. College of Automation, Harbin University of Science and Technology, Harbin 150080, China; 2. College of Computer Science and Technology, Harbin Engineering University, Harbin 150001, China)

Abstract: Compared with the common color video, infrared video is easily affected by the surrounding environment. In infrared pedestrian tracking, the appearance contour and gray distribution of the pedestrian target often have great changes, which lead to the difficulty of tracking. To solve this problem, this paper proposes a VPSiamRPN (Video prediction with Siamese Region Proposal Network) infrared pedestrian target tracking system. Aiming at the factors that seriously affect the performance of infrared pedestrian tracking (such as target deformation, target occlusion, and background clutter), the image prediction function of PredNet (Deep Predictive Coding Networks for Video Prediction and Unsupervised) was designed and applied to SiamRPN (Siamese Region Proposal Network) to improve the similarity between the tracking template and the detected target, so as to improve the tracking ability to the infrared pedestrian target. Nine comparative experiments were carried out by changing the number of layers of the network, the number of target images and frames used for prediction, and the tracking strategy of the network. On PTB-TIR dataset, experimental results show that the success plots and precision of theinfrared target recognition in thermal crossover, intensity change, occlusion, scale variation, and other attributes were much higher than those of SiamRPN, indicating good performance of infrared pedestrian tracking, which will have broad application prospects in this field.

Keywords: target tracking; video prediction; Siamese network; IR image; PredNet

红外视频与普通彩色视频相比易受周围环境的影响, 目标的外观常有较大的变化幅度, 主要表现为轮廓变化和灰度分布变化, 导致跟踪困难^[1-3]. 目前, 红外行人跟踪广泛应用的是三维中值滤波、图像差分法、卡尔曼滤波、粒子滤波和 mean-shift 等方法^[4]. 这些方法往往只能针对图像的某些特征对目标进行识别, 由于特征种类的限制, 跟踪的效果往往

仅在几种属性上有较强效果但不够全面.

深度学习方法具有强大的特征建模能力, 近年它在完成目标跟踪任务的准确度上表现出了突出优势. 目前主要有两类方法. 第一类是将卷积特征与相关滤波器结合. 例如, Danelljan 等提出了 C-COT^[5] 算法, 通过在连续的分辨率序列中学习, 创建时域连续的的相关滤波器, 可以将不同分辨率的特征图作为滤波器的输入, 使得传统特征和深度特征能够深度结合. 在下一年 Danelljan 等又提出了 C-COT 的改进算法 ECO^[6], 通过构造一组更小的滤波器, 有

收稿日期: 2019-10-10

作者简介: 柳恩涵(1994—), 男, 硕士研究生

通信作者: 张 锐, zr_gh@sina.com

效地利用矩阵分解操作, 降低了模型的大小, 防止过高的维度导致的效率低下和过拟合问题. 这一类方法的缺点是跟踪速度较慢, 在跟踪数据集上训练易造成过拟合. 第二类是使用孪生网络的方法. 例如, 文献[7]提出了 SINT 算法, SINT 可以说是第一个将孪生网络应用到跟踪领域的通过相似学习解决跟踪问题. Bertinetto 等又通过改进 SINT 算法提出了一种新的、经过端到端训练的、全卷积的孪生网络的跟踪算法 SiamFC^[8], 它也是通过使用相似度学习的方法来解决追踪任意目标的问题. SiamRPN^[9]是由 Li 等提出的基于 SiamFC 改进来的算法, 引入了 Faster RCNN^[10]中的 RPN 模块, 让网络可以回归位置、形状, 省掉多尺度测试, 提高算法性能. 第二类方法极大地增强了跟踪速度, 使待搜索图像大小不受限制, 并且避免了在跟踪数据集上训练造成的过拟合. 但此类方法的跟踪模版始终为上一帧的被跟踪目标, 对被跟踪帧的目标特征没有预判, 在处理目标

形状快速变化的跟踪任务时仍有欠缺.

本文设计了 VPSiamRPN 目标跟踪网络. 将 SiamRPN 网络应用到红外目标跟踪任务中, 根据红外成像的特点, 将视频预测加入到网络中, 增强网络对目标快速形变的适应能力, 采用 Pred Net^[11]网络对当前帧目标外观进行预测, 用预测出的目标作为模版在被检测帧中进行搜索, 完成目标跟踪任务.

1 相关原理介绍

1.1 SiamRPN^[8] 目标跟踪原理

SiamRPN 将跟踪问题抽象成单样本检测问题. SiamRPN 结合了跟踪中的孪生网络和检测中的区域推荐网络: 孪生网络实现对跟踪目标的适应, 让算法可以利用被跟踪目标的信息, 完成检测器的初始化; 区域推荐网络让算法可以对目标位置进行更精准地预测. 经过两者的结合, SiamRPN 可以进行端到端的训练. SiamRPN 结构如图 1 所示.

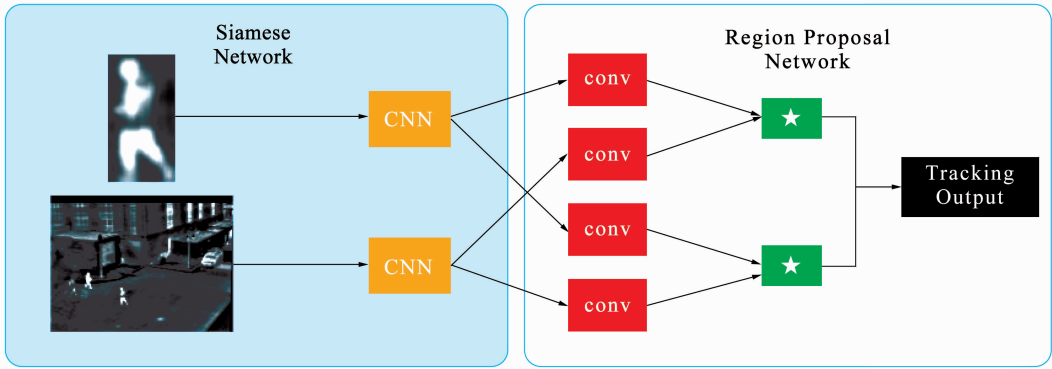


图 1 SiamRPN 网络结构^[9]
Fig. 1 Diagram of SiamRPN network structure

孪生神经网络可计算两个输入的相似度, 同时将跟踪模版图像和被检测帧图像由上下两个支路输入, 两个支路由一个卷积网络实现, 它们共享权值, 此网络必须为全卷积网络, 以满足平移不变性, 该网络的作用是分别提取模板帧和检测帧的图像特征. RPN 网络由两部分组成, 一是分类分支, 用于区分目标和背景, 另一个是回归分支, 它将候选区域进行微调. 网络的卷积基部分本文采用 resnet50^[12]结构.

网络训练采用 COCO、YOUTUBE11、DET、VID 数据集进行训练.

- 1) 损失函数: cross-entropy loss 用于分类, smooth L1 用于回归;
- 2) 训练方法: 随机梯度下降;
- 3) 数据增强: 包括 affine transformation (仿射变换) 在内的数据增强;

- 4) 正负样本选择策略: 设置高低两个阈值, 大于 0.6 的为正样本, 小于 0.3 的为负样本, 一个 pair 中有 64 个样本, 其中正样本最多 16 个.

1.2 PredNet^[11] 视频帧预测原理

该网络采用 CNN + LSTM^[13] 思路做视频预测, 通过提取前 n 帧图像特征, 预测目标图像. 网络包含一系列重复堆叠的模块, 每个模块作为一个概念层, 每层包含 4 个基本部分: 输入卷积层 A_1 , 递归表示层 R_1 , 预测层 $\hat{A}_1$, 误差表示层 E_1 .

- 1) E_1 : 误差表示层, $f(A_1 - \hat{A}_1)$.
- 2) A_1 : 输入卷积层, 对于第一层, 为目标图像; 对于更高层, 为前一层预测误差 E 的卷积 + Relu.
- 3) R_1 : 卷积 LSTM 层.
- 4) $\hat{A}_1$: 预测层, 对 R 单元卷积 + Relu 得到.

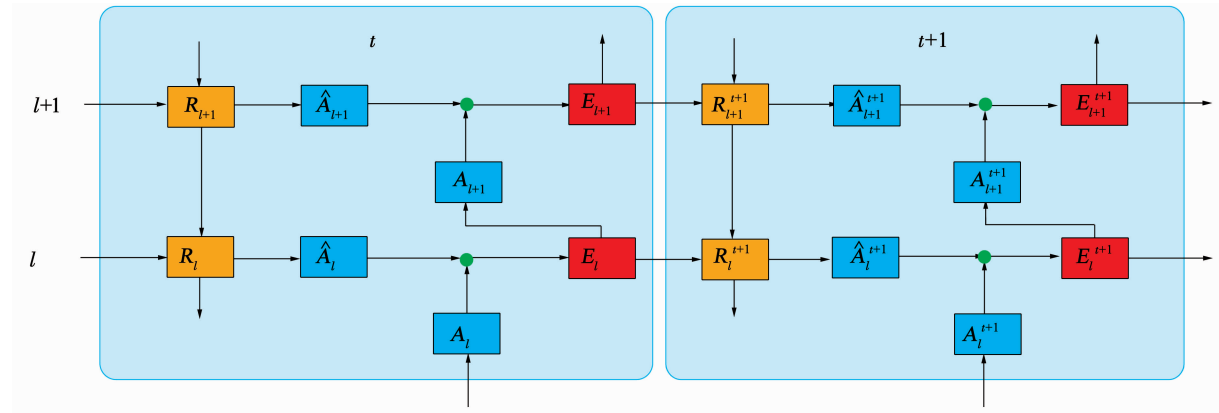


图 2 PredNet 网络结构^[10]

Fig. 2 Diagram of PredNet network structure^[10]

图 2 为 PredNet 网络结构,激活函数采用 Relu, A_l 与 $\hat{A}_l$ 之差小于零的部分置零,通过 A_l 与 $\hat{A}_l$ 相交互作差,拼接,再输入 Relu 层. E_l 传给 A_{l+1} ,作为下一层的输入. R_l^t 单元接受的输入是前一刻本层误差 E_l^t ,本层状态 R_l^{t-1} ,本时刻高层预测特征 R_{l+1}^t ,根据这三者进行特征级的预测. 预测的特征在 $\hat{A}_l$ 单元卷积,得到特征图像,与 A_l 相比较. 总体 loss 是各层、各时刻预测误差的加权和. 各层误差权重 λ_l ,各时刻误差权重 λ_t 由实验确定. 网络状态更新存在水平(时间)方向和竖直(各层)方向两方向的更新. 竖直方向先更新,先自下而上前向传播计算得各层误差 E_l ,再自上而下计算 RNN 单元的状态 R_l . t 时刻网络更新好后,进行 $t+1$ 时刻的更新. 因此对于各 t 的网络,输入是前一刻 RNN 状态 R^{t-1} 和本时刻目标输出图像 A_0 .

2 本文提出的基于视频预测的红外行人跟踪网络 VPSiamRPN

本文设计了一种基于视频预测的红外行人跟踪网络(Video Prediction with Siamese Region Proposal Network, VPSiamRPN),将用于视频预测的 PredNet 网络添加到 SiamRPN 网络的模版和与模版图像所对应的卷积网络之间. PredNet 采用四层网络,通过 5 帧目标图像预测下一帧目标外观,为使被预测包含部分视频之初的目标特征,设计一个可以存储 5 帧图像的序列 X ,前 3 帧固定为被跟踪视频的第 1~3 帧图像中的目标,后两帧将跟踪得到的目标图像做为模版依次存入序列中,序列始终按照时间顺序存储最新的两帧目标图像. 跟踪过程最初的前 5 帧应用 SiamRPN 网络进行跟踪,从第 6 帧开始预测,将 X 中的数据输入 PredNet 网络,预测和生成下一帧的目标,将生成的目标图像作为 SiamRPN 网络的模板,在被检测帧中,根据模版搜索目标并输出. 由

于 SiamRPN 网络在跟踪时将上一帧目标图像做为模版,这就造成一旦某一帧的跟踪发生偏差,下一帧的模版就会随之出现偏差. 如果偏差过大,下一帧目标就会丢失,并再也无法找回目标. 为防止由预测误差造成的目标丢失,10 帧中的后两帧将第 1 帧的目标做为模版通过 SiamRPN 网络进行跟踪,不经过预测网络.

由于孪生网络计算相似度要求两支路所表达的特征必须源自一种网络结构,所以 PredNet 网络依据前几帧的目标信息预测出待测目标信息,并以输出的目标图像为媒介,将网络所预测的特征信息转换为 resnet50 网络所表达的特征信息,将得到的特征输入 RPN 网络进行跟踪. 图 3 为本文描述的网络和 SiamRPN 网络的跟踪效果截图,其中绿色框为被跟踪目标的 ground truth,红色框为 SiamRPN 网络的跟踪效果,蓝色框为本文设计网络的跟踪效果. 图 4 为 VPSiamRPN 网络的结构.

非预测的网络跟踪过程为:将被检测帧重塑为 $255 \times 255 \times 3$ 并输入卷积网络作为被检测帧. 由于 Prednet 网络只能输出长宽为偶数的图像,所以首先将第 1 帧目标图像重塑为 $128 \times 128 \times 3$,将重塑后的图像存放在序列 X 中,再将图像重塑为 $127 \times 127 \times 3$ 作为模版帧输入卷积网络,将得到的两组特征输入 RPN 网络完成跟踪. 根据目标位置信息对被检测帧进行裁剪,将裁剪后得到的目标图像重塑为 $128 \times 128 \times 3$ 存入序列 X 中替换掉第 1 帧并将序列 X 按图像存入的顺序重新排序.

预测的网络跟踪过程为:PredNet 网络通过存储在序列 X 中的 8 帧图像预测出当前帧的目标图像,输出的图像经重塑后大小为 $127 \times 127 \times 3$ 作为模版帧. 模板帧和被检测帧经过相同的卷积结构分别得到 $6 \times 6 \times 256, 22 \times 22 \times 256$ 的特征,模版帧所产生的特征通过 3×3 的卷积核分别产生了 $4 \times 4 \times (2k \times$

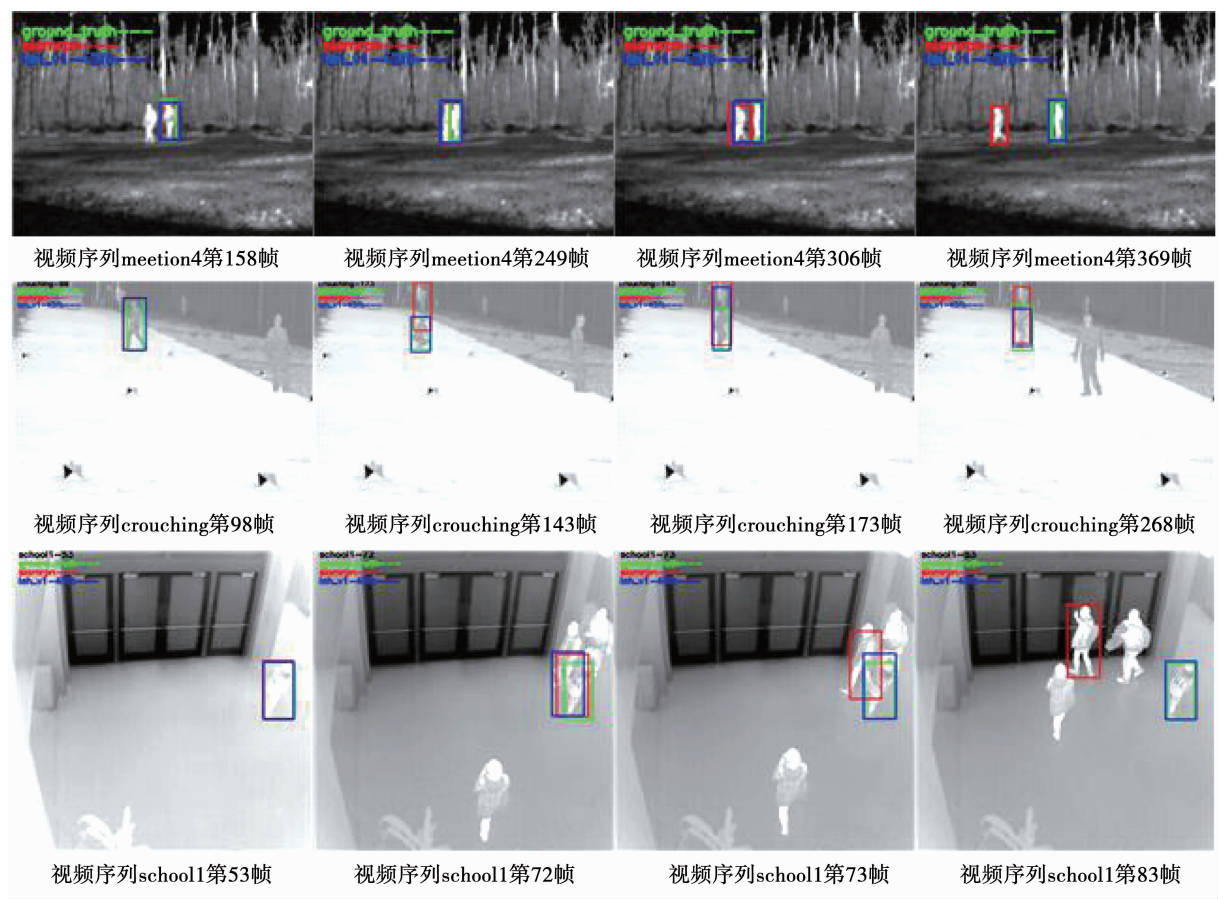


图 3 本文描述的网络和 SiamRPN 网络的跟踪效果

Fig. 3 Screenshots of tracking effects of the proposed network and SiamRPN network

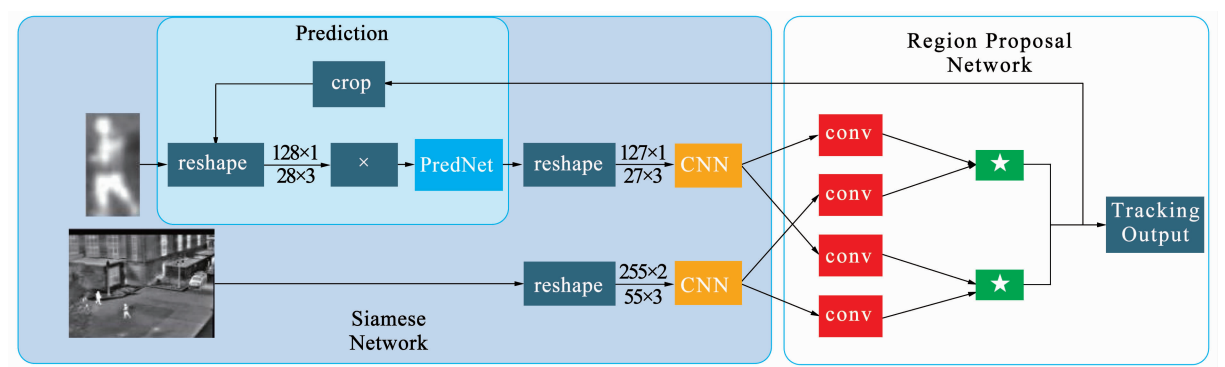


图 4 VPSiamRPN 网络结构

Fig. 4 Diagram of VPSiamRPN network structure

256) 和 $4 \times 4 \times (4k \times 256)$ 的特征, k 值为 RPN 网络所需的矩形框数. 通过得到的两组特征输入 RPN 网络得到目标位置信息. 根据目标位置信息对被检测帧进行裁剪, 将裁剪后得到的目标图像重塑为 $128 \times 128 \times 3$ 存入序列 X 中替换掉第 1 帧并将序列 X 按图像存入的顺序重新排序.

3 结果分析

本实验软件运行环境为 Ubuntu18. 04, 平台为 Python3. 7 + PyTorch0. 4. 1 开源机器学习框架, 硬件配置为: Intel i7-8700k 3. 7 GHz CPU, 64 GB RAM,

single GPU NVIDIA GTX 2 080 Ti.

本研究设计的网络采用非端到端的训练方法, 将 PredNet 网络和 SiamRPN 网络分开训练, PredNet 网络: 从 OBT50、OBT100、VOT2013、VOT2014、VOT2015、VOT2016 和 VOT2018 数据集中选取 129 个红外行人视频, 将每个被标注的视频序列中的每帧图像依据标注进行裁剪得到一个只有目标没有背景的图像序列做为预测网络的训练集. SiamRPN 网络: 采用经商汤开源的 SiamRPN 网络卷积基采用 resnet50 版本的权重文件.

通过改变预测网络层数、预测过程所需的被选

目标图像及图像帧数、预测网络参与跟踪的频率,设计九组对比试验,以找出性能最佳的网络结构,具体设置如表 1 所示.

表 1 9 种网络具体结构
Tab. 1 Nine network structures

实验编号	预测网络层数	预测序列所需帧数/f	预测序列更新情况	预测网络不参与跟踪的频率
45fb	4	5	更新后两帧	前 5 帧不预测 每 10 帧中后两帧不预测
45fq	4	5	全部更新	前 5 帧不预测 每 10 帧中后两帧不预测
45gb	4	5	更新后两帧	前 5 帧不预测
48fq	4	8	全部更新	前 8 帧不预测 每 10 帧中后两帧不预测
48fb	4	8	更新后两帧	前 8 帧不预测 每 10 帧中后两帧不预测
35fb	3	5	更新后两帧	前 5 帧不预测 每 10 帧中后两帧不预测
35fq	3	5	全部更新	前 5 帧不预测 每 10 帧中后两帧不预测
38fb	3	8	更新后两帧	前 8 帧不预测 每 10 帧中后两帧不预测
38fq	3	8	全部更新	前 8 帧不预测 每 10 帧中后两帧不预测

采用 PTB-TIR 数据集^[14]对本研究的 9 种网络与 SiamRPN 网络进行对比评估,测试网络性能. 该数据集包含 60 组红外行人视频序列,根据每个序列所包含的特性组合出了 9 种属性(热交叉、强度变化、遮挡、尺寸变化、背景杂波、低分辨率、快速运动、运动模糊和脱离视野)的数据集,评估通过 OTB 数据集^[15]的评价标准计算本研究设计的网络和 SiamRPN 网络的跟踪成功率和跟踪精度,对网络性能进行评价.

1)整体数据集. 在整体数据集上,45fb 网络(PredNet 中采用 4 层网络,通过 5 帧目标图像预测当前帧目标,其中前 3 帧固定为视频的第 1~3 帧中的目标,每 8 帧的下 2 帧不预测)在跟踪成功率和跟踪精度上都表现出了较大优势,得分为 0. 426 和 0. 571,SiamRPN 网络的得分为 0. 419 和 0. 557.

2)热交叉. 热交叉属性指视频中具有相似强度的两个目标相互交叉. 在此属性中 45fb 网络的跟踪成功率的得分为 0. 407,跟踪精度的得分为 0. 610;SiamRPN 网络跟踪成功率的得分为 0. 398,跟踪精度的得分为 0. 581. 在这种属性上相对于 SiamRPN 网络表现出了较明显的优势. 48fb 网络在此属性跟踪成功率略高于 SiamRPN 网络,得分为 0. 403. 此项结果表明本文网络对具有热交叉属性的红外视频在跟踪性能上有提高.

3)强度变化. 强度变化属性指由于目标的温度变化,目标区域的强度发生了变化. 在此属性中 45fb 网络的成功率的得分为 0. 374,SiamRPN 网络的得分为 0. 350;45fb 网络跟踪精度的得分为 0. 328,SiamRPN 网络的得分为 0. 334. 网络在此属性中跟踪成功率略高于 SiamRPN 网络. 在此属性中 38fq 网络在成功率和跟踪精度上都表现出了较大优势,得

分为 0. 551 和 0. 509. 此项结果表明本文网络对具有强度变化属性的红外视频在跟踪成功率上有所提高.

4)遮挡. 遮挡属性指目标被部分或完全遮挡. 在此属性中 45fb 网络在跟踪成功率和跟踪精度上都表现出了较大优势,得分为 0. 380 和 0. 496,SiamRPN 网络的得分为 0. 370 和 0. 486. 此项结果表明本文网络对具有遮挡属性的红外视频在跟踪成功率上有所提高.

5)尺寸变化. 尺寸变化属性指第 1 帧和当前帧中,目标大小的比率超出范围 $[1/2,2]$. 在此属性中 45fb 网络在跟踪成功率上高于 SiamRPN 网络,45fb 网络得分为 0. 408,SiamRPN 网络的得分为 0. 404;在跟踪精度上略低,45fb 网络得分为 0. 508,SiamRPN 网络的得分为 0. 516. 此项结果表明本文网络对具有尺寸变化属性的红外视频在跟踪成功率上有所提高.

6)背景杂波. 背景杂波属性指目标附近的背景具有相似的纹理或强度. 在此属性中 45fb 网络的成功率得分为 0. 411,48fb 网络的成功率得分为 0. 409;45fb 网络的跟踪精度得分为 0. 548,48fb 网络的跟踪精度得分为 0. 551;SiamRPN 网络的得分分别为 0. 398 和 0. 534,本文网络在成功率和跟踪精度上对 SiamRPN 网络都表现出了较大优势. 结果表明本文网络对具有热交叉属性的红外视频在跟踪性能上有提高.

7)低分辨率. 低分辨率属性指目标小于 600 像素. 在此属性中 45fb 网络在跟踪成功率上的得分比 SiamRPN 网络低 0. 016,但在跟踪精度上高 0. 033.

8)快速运动. 快速运动属性指目标在相邻帧中的移动距离大于 20 像素. 在此属性中 45fb 网络在

跟踪成功率上的得分比 SiamRPN 网络低了 0. 014, 跟踪精度上高 0. 049.

9) 运动模糊. 运动模糊属性指目标区域由于目标或摄像机运动而变得模糊. 在此属性中 45fb 网络在跟踪成功率上和 SiamRPN 网络取得了相同的得分 0. 456,跟踪精度上低了 0. 001;48fb 网络在跟踪精度上的得分比 SiamRPN 网络高了 0. 008.

10)脱离视野. 脱离视野属性指目标部分不在图像区域内. 在此属性中 45fb 网络在跟踪成功率和

跟踪精度上都强于 SiamRPN 网络. 在此属性中 45fb 网络的跟踪成功率的得分为 0. 456, SiamRPN 网络的得分为 0. 453; 跟踪精度的得分为 0. 497, SiamRPN 网络的得分为 0. 466. 在两种属性上相对于其他网络(包括 SiamRPN 网络)都表现出了较明显的优势. 此项结果表明本文网络对具有脱离视野属性的红外视频在跟踪性能上有提高.

表 2 和表 3 分别列出了 10 种网络在各个属性上的成功率和精度.

表 2 10 种网络在各个属性上的跟踪成功率
Tab. 2 Success plots of 10 networks in each attribute

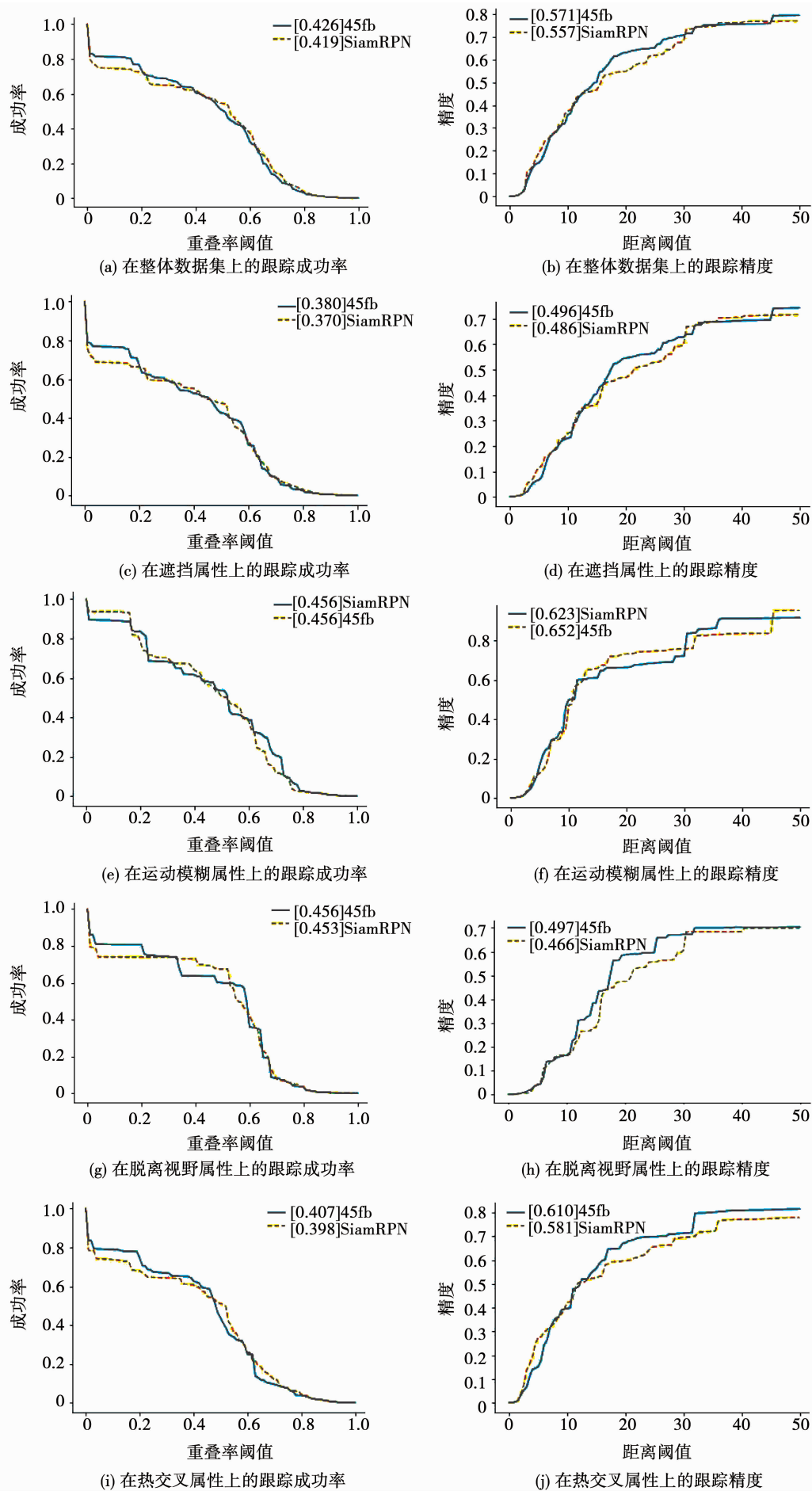
实验编号	总体	热交叉	强度变化	遮挡	尺寸变化	背景杂波	低分辨率	快速运动	运动模糊	脱离视野
45fb	0. 426	0. 407	0. 374	0. 380	0. 408	0. 411	0. 493	0. 393	0. 456	0. 456
45fq	0. 402	0. 419	0. 539	0. 352	0. 384	0. 396	0. 467	0. 303	0. 413	0. 368
45gb	0. 400	0. 394	0. 526	0. 343	0. 367	0. 392	0. 444	0. 297	0. 413	0. 379
48fq	0. 406	0. 417	0. 542	0. 349	0. 389	0. 398	0. 470	0. 300	0. 422	0. 366
48fb	0. 419	0. 403	0. 374	0. 369	0. 405	0. 409	0. 491	0. 295	0. 456	0. 415
35fb	0. 414	0. 402	0. 375	0. 364	0. 400	0. 400	0. 496	0. 372	0. 451	0. 422
35fq	0. 401	0. 415	0. 544	0. 347	0. 388	0. 395	0. 460	0. 305	0. 411	0. 365
38fb	0. 414	0. 404	0. 371	0. 363	0. 405	0. 404	0. 493	0. 323	0. 456	0. 411
38fq	0. 399	0. 400	0. 551	0. 341	0. 383	0. 393	0. 442	0. 303	0. 406	0. 362
SiamRPN	0. 419	0. 398	0. 350	0. 370	0. 404	0. 398	0. 509	0. 407	0. 456	0. 453

表 3 10 种网络在各个属性上的跟踪精度
Tab. 3 Precision plots of 10 networks in each attribute

实验编号	总体	热交叉	强度变化	遮挡	尺寸变化	背景杂波	低分辨率	快速运动	运动模糊	脱离视野
45fb	0. 571	0. 610	0. 328	0. 496	0. 508	0. 548	0. 811	0. 466	0. 652	0. 497
45fq	0. 548	0. 618	0. 483	0. 480	0. 499	0. 544	0. 748	0. 356	0. 603	0. 424
45gb	0. 548	0. 585	0. 478	0. 462	0. 473	0. 536	0. 745	0. 350	0. 621	0. 435
48fq	0. 559	0. 616	0. 481	0. 473	0. 508	0. 547	0. 772	0. 355	0. 623	0. 424
48fb	0. 569	0. 607	0. 325	0. 494	0. 515	0. 551	0. 805	0. 406	0. 661	0. 468
35fb	0. 553	0. 607	0. 325	0. 471	0. 504	0. 531	0. 805	0. 434	0. 639	0. 436
35fq	0. 551	0. 621	0. 489	0. 474	0. 507	0. 545	0. 741	0. 358	0. 605	0. 420
38fb	0. 553	0. 612	0. 317	0. 471	0. 510	0. 534	0. 809	0. 405	0. 645	0. 425
38fq	0. 544	0. 592	0. 509	0. 465	0. 497	0. 539	0. 698	0. 357	0. 597	0. 420
SiamRPN	0. 557	0. 581	0. 334	0. 486	0. 516	0. 534	0. 778	0. 417	0. 623	0. 466

根据跟踪结果可以看出 45fb 网络在热交叉、强度变化、遮挡、尺寸变化、背景杂波、脱离视野和总体数据集上相对于 SiamRPN 网络在跟踪成功率上有很大提高;在热交叉、遮挡、背景杂波、低分辨率、脱离视野和总体数据集上相对于 SiamRPN 网络在重叠率上表现出了明显提升,综合性能具有突出优势,

表明该方法有效、可行,图 5 为 45fb 网络和 SiamRPN 在 PTB-TIR 数据集上的跟踪成功率和跟踪精度的对比. 此外,38fb 网络在热交叉属性上的成功率和重叠率都表现出了极大优势,表明此网络非常适合跟踪带有热交叉属性的视频.



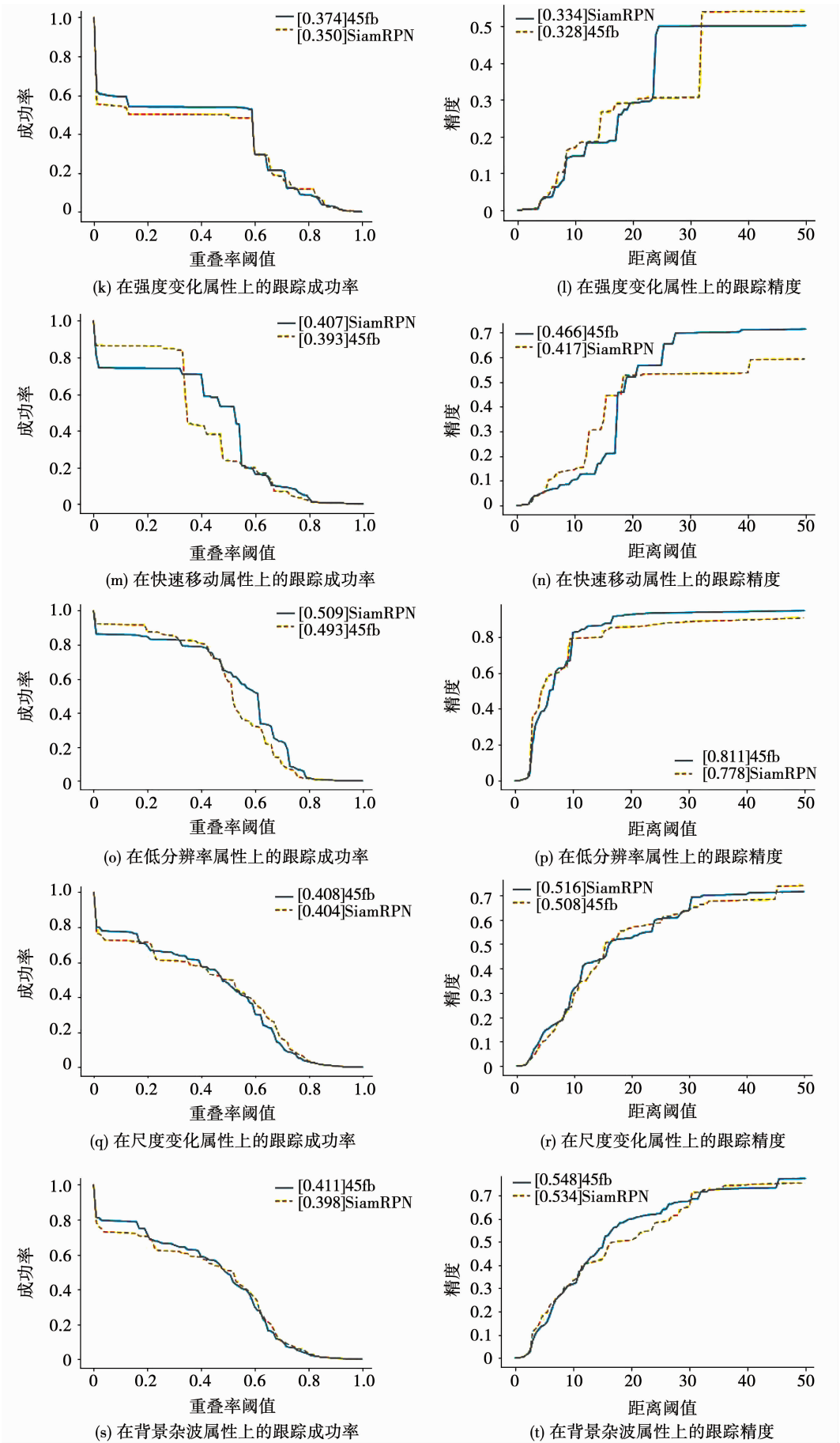


图 5 45fb 网络和 SiamRPN 在 PTB-TIR 数据集上的跟踪成功率和跟踪精度

Fig. 5 Success plots and precision tracking of 45fb network and SiamRPN on PTB-TIR dataset

4 总 结

1) 针对红外行人目标的跟踪问题,提出了使用视频预测网络对跟踪模板进行更新,该方法能够提高跟踪模板与被跟踪目标的相似度,增强网络对红外行人目标外观轮廓和灰度分布变化所导致跟踪困难的问题的处理能力,有利于目标跟踪中的模型匹配。

2) 通过改变预测网络层数、预测过程所需的被选目标图像及图像帧数、预测网络参与跟踪的频率,设计了 9 组对比试验。在 PTB-TIR 数据集上,与 SiamRPN 网络客观定量对比和评估,实验结果表明,45fb 网络对红外目标的识别在热交叉、强度变化、遮挡和尺寸变化等多种属性上的跟踪成功率和重叠率均较 SiamRPN 网络有较大提高,显示出对红外行人跟踪的良好性能,在这一领域将有广阔的应用前景。

参考文献

- [1] RAJKUMAR S, CHANDRA M. Pedestrian detection in Infrared Images using local thresholding [C]//2015 2nd International Conference on Electronics and Communication Systems (ICECS). Cairo: IEEE, 2015: 259. DOI: 10.1109/ECS.2015.7124904
- [2] 高小明. 影响红外热像仪测量精度的因素分析[J]. 华电技术, 2008, 30(11): 4
GAO Xiaoming. Analysis of factors affecting the measurement accuracy of infrared thermal imager[J]. Huadian Technology, 2008, 30(11): 4. DOI: 10.3969/j.issn.1674-1951.2008.11.002
- [3] 郭永彩, 胡瑞光, 高潮. 红外图像中的行人检测[J]. 重庆大学学报, 2009, 32(9): 1070
GUO Yongcai, HU Ruiguang, GAO Chao. Pedestrian detection in infrared image[J]. Journal of Chongqing University, 2009, 32(9): 1070. DOI: 10.11835/j.issn.1000-582x.2009.09.017
- [4] 孟磊, 杨旭. 目标跟踪算法综述[J]. 自动化学报, 2019, 45(7): 1244
MENG Lu, YANG Xu. A survey of target tracking algorithms[J]. Acta Automatica Sinica, 2019, 45(7): 1244. DOI: 10.16383/j.aas.c180277
- [5] DANELLJAN M, ROBINSON A, FELSBERG M, et al. Beyond correlation filters: learning continuous convolution operators for visual tracking [C]//ECCV 2016: Computer Vision-ECCV 2016

- Workshops. Amsterdam: IEEE, 2016: 472. DOI: 10.1007/978-3-319-46454-1_29
- [6] DANELLJAN M, BHAT G, FELSBERG M, et al. ECO: efficient convolution operators for tracking [C]//The IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Honolulu: IEEE, 2017: 6638. DOI: 10.1109/CVPR.2017.733
- [7] TAO R, GAVVES E. Siamese instance search for tracking [C]//The IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Las Vegas: IEEE, 2016: 1420. DOI: 10.1109/CVPR.2016.158
- [8] BERTINETTO L, VALMADRE J. Fully-convolutional Siamese networks for object tracking [C]//ECCV 2016: Computer Vision-ECCV 2016 Workshops. Amsterdam: IEEE, 2016: 850
- [9] LI Bo, YAN Jiejun, WU Wei, et al. High performance visual tracking with Siamese region proposal network [C]//The IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Salt Lake City: IEEE, 2018: 8971. DOI: 10.1109/CVPR.2018.00935
- [10] REN Qingshao, HE Minghai, GIRSHICK R, et al. Faster R-CNN: towards real-time object detection with region proposal networks [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2017, 39(6): 2805
- [11] LOTTERW, KREIMAN G, COX D. Deep predictive coding networks for video prediction and unsupervised learning [Z]. arXiv:1605.08104, 2016
- [12] HE Minghai, ZHANG Yuxiang, REN Qingshao, et al. Deep residual learning for image recognition [C]//The IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Las Vegas: IEEE, 2016: 770. DOI: 10.1109/CVPR.2016.90
- [13] SHI Jianxing, CHEN Rongzhou, WANG Hao, et al. Convolutional LSTM network: a machine learning approach for precipitation nowcasting [C]//NIPS'15 Proceedings of the 28th International Conference on Neural Information Processing Systems-Volume 1. Montreal: MIT Press, 2015: 802. DOI: 10.1007/978-3-319-21233-3_6
- [14] LIU Qiao, HE Yuzhen, LI Xin, et al. PTB-TIR: A thermal infrared pedestrian tracking benchmark [Z]. arXiv:1801.05944, 2018
- [15] WU Y, LIM J, YANG M. Online object tracking: a benchmark [C]//Proceedings of CVPR. Portland: IEEE, 2013: 2411. DOI: 10.1109/CVPR.2013.312

(编辑 苗秀芝)