

# 基于弱监督表示学习的热红外目标跟踪

袁 笛

(西安电子科技大学 广州研究院, 广东 广州 510555)

**摘 要:** 由于热红外成像技术具有更强的穿透雾、霾、雨、雪的能力, 在恶劣天气条件下的成像效果几乎不受影响, 使得基于热红外图像的目标跟踪任务越来越被研究者重视。针对基于卷积神经网络的热红外目标跟踪算法在模型训练过程中需要的带有标签的数据不足的问题, 提出了一种基于弱监督表示学习的方法, 利用少量的标签数据及大量的无标签数据进行模型训练, 从而用于热红外目标跟踪任务。首先, 利用主动学习的指导在大量无标签的数据中挑选最具有代表性的训练样本; 然后, 给定每个样本序列的首帧目标的真实标签, 利用基础跟踪器生成该序列中其他图像帧中目标的伪标签; 之后, 利用带有真实标签和伪标签的训练数据进行模型训练; 最后, 利用训练好的模型在热红外目标跟踪算法测试数据集上进行模型测试。实验结果表明: 该方法可以在减少模型训练对标签数据需求的同时保证跟踪器的准确性。

**关键词:** 弱监督表示学习; 主动学习; 训练样本挑选; 伪标签生成; 热红外目标跟踪

中图分类号: TP393. 0

文献标识码: A

文章编号: 1673-629X(2024)04-0035-07

doi: 10. 20165/j. cnki. ISSN1673-629X. 2024. 0006

## Weakly Supervised Based Representation Learning for Thermal Infrared Target Tracking

YUAN Di

(Guangzhou Institute of Technology, Xidian University, Guangzhou 510555, China)

**Abstract:** Since thermal infrared imaging technology has a stronger ability to penetrate fog, haze, rain and snow, the imaging effect is almost unaffected in bad weather conditions, which makes the target tracking task based on thermal infrared images has been paid more and more attention by researchers. Aiming at the problem of insufficient labeled data in the model training of the thermal infrared target tracking algorithm based on convolutional neural network, a method based on Weakly Supervised Representation Learning (WSRL) is proposed, which uses a small amount of labeled data and a mass of unlabeled data for model training, so as to be used in thermal infrared target tracking tasks. Firstly, the guidance of active learning is used to select the most representative training samples from a large amount of unlabeled data. Then, given the ground-truth label of the target in the first frame of each sample sequence, the basic tracker is used to generate pseudo-labels for other frames in the same sequence. Then, the training data with ground-truth labels and pseudo-labels is used for model training. Finally, the trained model is used to test the algorithm on the thermal infrared target tracking algorithm test data set. The experimental results show that the proposed method can ensure the accuracy of the tracker while reducing the demand for label data for model training.

**Key words:** weakly supervised representation learning; active learning; training sample selection; pseudo-label generation; thermal infrared target tracking

### 0 引言

目标跟踪技术可以给视频监控、智能交通、人机交互等领域提供技术支撑<sup>[1-5]</sup>。热红外目标跟踪由于其在夜间、雨天和雾天等极端天气下的良好性能, 得到了广泛研究<sup>[6-9]</sup>。近年来随着深度学习的发展, 热红外目标跟踪取得了较好的跟踪效果。但是基于深度学习

的热红外目标跟踪算法需要大量带有标签的数据用于卷积神经网络模型的训练。由于带有标签的训练数据十分有限, 无标签的数据虽然可以获取到, 数据标注又十分费时费力, 因此极大地限制了基于深度学习的热红外目标跟踪的发展。

一种有效的解决无标签数据的方法是在大量的无

收稿日期: 2023-06-14

修回日期: 2023-10-18

基金项目: 国家自然科学基金(62202362); 中国博士后科学基金(2021TQ00247); 中央高校基本科研业务费专项资金(XJS222503)

作者简介: 袁 笛(1991-), 男, 博士, 讲师, 硕导, 通讯作者, CCF 会员(J8304M), 研究方向为计算机视觉、图像处理、目标跟踪。

标签数据中随机选取一个小规模的子集进行标注,然后用带有标签的这部分样本来训练卷积神经网络模型。而在随机选择的前提下,只有选择足够大的子集,才能保证所挑选的训练样本集能够训练出性能较好的卷积神经网络模型。与随机选择不同的是,该文使用主动学习的方式来选择样本,这样可以提高训练数据样本空间的多样性。在选择相同规模训练数据集的情况下,基于主动学习方法选择的样本集上训练的模型比在随机选择的样本集上训练的模型具有更强的表示能力<sup>[10-11]</sup>。基于此,在给定样本数量的情况下可以得到具有多样性的训练样本数据。

在大规模无标签的数据集中挑选具有代表性的训练样本子集可以降低所需标注数据的规模。但是标注数据仍然需要很高的成本。因此,该文尝试使用一种基于前后向跟踪一致性<sup>[12-14]</sup>的方法对无标注的训练数据生成伪标签,然后采用这些伪标签训练模型。具体来说,首先,对挑选得到的无标签训练数据序列的第一帧中的目标进行手动标注。然后,利用基础跟踪器对序列后续帧进行自动标注生成对应的伪标签,采用前后向跟踪一致性的方法对伪标签进行优化,从而得到相对准确的伪标签。最后,利用带有首帧真实标签和其他帧伪标签的数据进行卷积神经网络模型的训练,从而在保证模型性能的基础上减少对标签数据的需求。

## 1 相关工作

基于深度学习的目标跟踪算法以其良好的表示能力受到了越来越多研究人员的青睐<sup>[15-17]</sup>。文献[17]提出通过基于判别相关滤波的跟踪框架结合多级卷积特征来实现由粗到细的跟踪模型。文献[18]提出一种基于目标感知的 TADT 跟踪器,该跟踪器使用两个辅助任务在线选择目标感知特征,以实现模板和候选目标之间更精确的匹配,通过寻找具有最大相似度的候选目标定位目标位置。在文献[19]中,作者提出了一种采用深度卷积神经网络提取特征的模型来实现热红外目标跟踪。Liu 等人<sup>[7]</sup>构建了一种集成的热红外目标跟踪方法,使用预训练的卷积神经网络提取目标的多个卷积特征,通过多特征融合的方式获得相对完整的目标表示。大多数基于深度卷积神经网络模型提取特征的热红外目标跟踪方法使用在普通的可见光训练样本上训练的现成特征提取网络来提取热红外目标的特征。为了更好地对热红外目标进行表示,文献[20]给热红外目标跟踪任务提出了一个多任务框架。尽管这些跟踪器已经取得了不错的跟踪效果,但与真实场景需求相比,热红外目标跟踪器的跟踪性能仍有很大的提升空间。尤其是带有标签的训练样本不足,

使得此类跟踪器的跟踪精度受到了很大限制。

主动学习可以在大规模样本中指导挑选出最具有代表性的样本子集。在主动学习方法中存在以下几种查询策略:基于信息量的查询方法<sup>[21]</sup>、基于表示性的查询策略<sup>[10]</sup>、混合型查询策略<sup>[22]</sup>以及基于性能的查询策略<sup>[23]</sup>等。基于上述不同的查询策略,主动学习已经被成功应用于文本分类、图像分类、目标检测等任务中。在文献[21]中,BALD 方法将贝叶斯深度学习结合到高维数据的主动学习框架中。在文献[24]中,作者提出了一种任务不可知的主动学习方法,可以从单个损失预测块训练网络。其他的方法试图减少样本标注的成本,关注于稀疏子集近似或核心集选择等方面。该文尝试采用主动学习的方式对未标注的视频序列样本进行选择。

目标在运动过程中,在时间序列上的循环一致性已经在许多文献中得到了广泛的研究。Wang 等人<sup>[25]</sup>提出使用周期一致性来学习视觉表示,主要集中在单个视频中统一光流和跟踪,以使用自监督学习方法实现更好的嵌入表示。Dwibedi 等人<sup>[26]</sup>尝试使用可微分的时间周期一致性损失来训练网络,以便能够在多个视频中寻找对应的关系。Li 等人<sup>[27]</sup>提出跟踪大图像块并在连续的视频帧之间建立起彼此的关联。作为时间序列中周期一致性的代表,前后向一致性已被广泛用于跟踪任务。TLD<sup>[13]</sup>跟踪器通过计算前后向跟踪误差来估计跟踪轨迹的可靠性。通过反向轨迹和前向轨迹的比较来纠正跟踪器的跟踪结果。MTA<sup>[12]</sup>跟踪器通过预测多个弱跟踪器的前后向一致性并通过最大鲁棒性评分确定最佳跟踪器来执行前向跟踪任务。UDT<sup>[14]</sup>跟踪器回顾了前后向跟踪框架,提出一种基于无监督学习的方式对跟踪器的特征提取网络进行训练。在基于主动学习挑选得到无标注训练序列子集的基础上,给定序列初始帧中目标的真实标签,然后利用前后一致性约束基础跟踪器生成序列中剩余帧的伪标签。利用带有真实标签的初始帧和带有伪标签的其他帧的训练数据训练热红外目标跟踪器的卷积神经网络模型,可以在保证模型跟踪性能的基础上,尽可能地减少模型训练对标签数据的需求。

## 2 基于伪标签训练的弱监督跟踪算法

基于伪标签进行模型训练的弱监督热红外目标跟踪算法(WSRL)的总体框架如图 1 所示。首先,采用主动学习指导从大量无标注数据集中挑选得到具有代表性的样本子集。然后,采用基于前后项一致性约束进行样本的伪标签生成;利用带有伪标签的数据进行特征提取网络参数的训练。最后,利用基于判别式跟踪框架来完成目标跟踪任务。

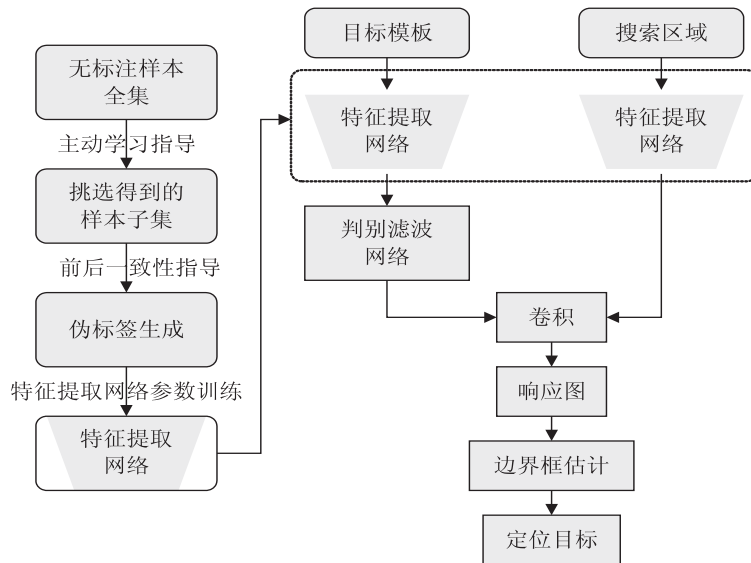


图1 基于伪标签训练的弱监督跟踪算法框架

## 2.1 基于主动学习的样本挑选方法

基于有限的带有标签的样本训练的卷积神经网络模型往往对目标缺乏较强的表示能力,从而使得基于这些模型的跟踪器的跟踪性能不佳。海量的无标签数据虽然可以很容易地从互联网上获取到,但是标注海量的数据非常耗时且成本很高。一种折衷的方法是随机选取部分样本进行标注,但这种随机性往往会使得训练得到的模型表现很不稳定。为此,该文提出了一种基于主动学习的样本选择方法。大多数主动学习的方法都是根据图像样本之间的相似度来选择训练样本的。受启发于 PointNet++<sup>[28]</sup> 采用最远点采样(Farthest Point Sampling, FPS)算法选择能够覆盖整个集合的子集,使用 FPS 方法对样本进行初始选择可以得到一个有代表性的样本子集。FPS 算法总是从所选子集的所有样本中选取距离最远的样本并将其加入到子集中,以保证所选子集中样本的多样性。

由于未标注的样本数据是从互联网上直接得到的,数据的质量不能得到保证。采用最远点采样的方式可以选择出样本集中最具有多样性的样本子集,但是也会引入一些噪声数据。为了保证所选择的样本具有代表性,换句话说就是为了保证所选取的样本不是孤立数据,该文采用一种基于  $k$  近邻的方式在最远点采样的基础上进行训练数据的挑选。

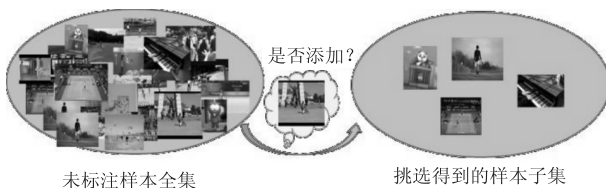


图2 基于主动学习的样本挑选

图2给出了一个采用基于  $k$  近邻的最远点采样选择样本子集的示例。给定未标注的数据序列样本集合

$A(A = \{s_1, s_2, \dots, s_n\})$ , 使用基于  $k$  近邻的最远点采样(FPS)方法来选择序列样本子集  $\{s_{i_1}, s_{i_2}, \dots, s_{i_k}\}$ 。在已经选择了  $j-1 (j \leq k)$  个样本的情况下,挑选样本  $s_{i_j}$  的方法在  $A$  剩余样本  $(A - \{s_{i_1}, s_{i_2}, \dots, s_{i_{j-1}}\})$  中挑选离已经选择的子集  $\{s_{i_1}, s_{i_2}, \dots, s_{i_{j-1}}\}$  最远并且其  $k$  近邻的样本未被选取的样本,这样可以保证所选样本子集的多样性和代表性。该文的  $k$  值选为 3。

## 2.2 基于前后一致性的伪标签生成

在得到小规模未标注的训练样本序列之后,首先对每个样本序列的首帧中的目标进行标注,然后利用基础跟踪器对样本序列的后续帧生成对应的伪标签。为了使生成的伪标签更加准确,采用基于前后一致性约束对生成的伪标签进行优化。根据视频数据本身的性质,在同一视频数据内目标在时间上是存在着运动一致性。如图3所示,给定首帧视频图像帧中目标的真实标签(实线边界框),使用前向跟踪到下一帧图像帧得到伪标签(灰色虚线边界框),然后下一帧反向跟踪到第一帧得到黑色虚线边界框,理论上灰色虚线边界框和黑色边界框应该是一致的,因此最小化两者之间的差异就可以生成相对准确的伪标签。通过这种前后项一致性的方式就可以生成后续图像帧中目标的伪标签。采用带有真实标签和伪标签的训练样本进行卷积神经网络模型的训练,在减少对带有标签的数据需求的同时保证模型对目标的表示能力。

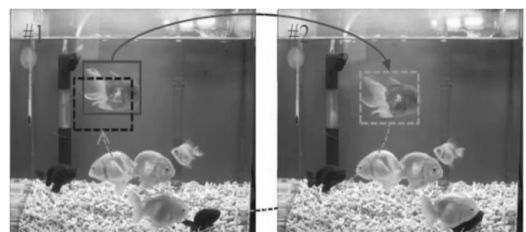


图3 基于前后项一致性的伪标签生成



### 2.3 基于判别式跟踪框架

得到带有真实标签和伪标签的训练数据之后,该文采用一种判别式跟踪框架进行热红外目标跟踪<sup>[29]</sup>。该判别式跟踪框架包括两个部分:目标分类和边界框估计。目标分类包含一个卷积模块,用于提取图像块的特征。在给定带有标签的训练数据的情况下,分类模型可以生成目标分类器的权重。将得到的权重应用于新的测试图像块中提取的特征上,从而可以得到目标置信度得分。目标分类损失的定义如下:

$$L(f) = \|r(x * f, c)\|^2 + \lambda \|f\|^2 \quad (1)$$

其中,  $x$  表示图像块,  $f$  表示滤波权重,  $*$  表示卷积操作,  $x * f$  表示目标置信度得分,  $c$  表示目标的中心位置,  $r(x * f, c)$  表示残差部分,  $\lambda$  表示正则化参数。通过引入判别学习损失,公式 1 可以重写为:

$$L_{cl} = \|l(x * f', z_c)\|^2 \quad (2)$$

其中,  $z_c$  表示设置为以  $c$  为中心的高斯函数回归标签,  $f'$  表示为特征提取网络的权值参数。

为了使模型能够更加关注目标部分,目标分类部分的损失函数也可以采用 Focal 损失的形式:

$$L(f) = \|f * \varphi(x) - y\|^2 - \xi \|f * \varphi(x)\|^2 + \lambda \|f\|^2 \quad (3)$$

其中,  $\xi$  控制不同部分的重要性,  $\|f * \varphi(x)\|^2$  表示目标关注部分,用于增加目标部分在回归模型中的作用。由于采用训练过的外观模型来估计当前帧的候选对象,响应图中的最大值总是低于期望值。背景样本的期望响应接近于零。因此,公式 3 致力于增加目标与背景响应的相对差异,并使目标样本的响应最大化,使得模型更加关注于目标部分。

边界框估计部分采用了 ATOM<sup>[30]</sup> 跟踪器中提出的重叠最大化方法。根据在目标分类部分中获得的目标置信度得分,给定目标的候选边界框,训练边界框估计模型,找出这些估计边界框 ( $B$ ) 与目标真实边界框 ( $B^{gt}$ ) 之间的最大的 IoU 分数。IoU 损失函数的定义为:

$$L_{ibb} = 1 - \text{IoU} \quad (4)$$

其中,  $L_{ibb}$  表示目标边界框估计的 IoU 损失,  $\text{IoU} = \frac{B \cap B^{gt}}{B \cup B^{gt}}$ 。结合边界框估计损失和目标分类损失,判别跟踪模型可以通过最小化损失函数  $L_{tot} = \eta L_{cl} + L_{ibb}$  来训练。

## 3 实验分析

主要采用 PTB-TIR<sup>[31]</sup> 和 LSOTB-TIR<sup>[32]</sup> 两个数据集来验证提出的基于弱监督表示学习的热红外目标跟踪算法的有效性。

### 3.1 实验环境及配置

主要参考 DiMP<sup>[29]</sup> 跟踪器采用带有 0.9 动量的随机梯度下降来训练深度卷积神经网络模型。采用 ResNet-50 作为跟踪器的骨干网络。权重衰减系数为  $5e-4$ , 学习率为  $1e-5$ , 目标分类权重  $\eta$  设置为  $1e-2$ , 网络训练 50 轮。采用基于主动学习在 GOT10k<sup>[33]</sup> 训练数据集上挑选的前 2 500 个训练样本序列来训练网络。实验基于 Python3.7, 在单个 3080Ti 的 GPU 上进行。

### 3.2 参数分析

主要目的是在保证跟踪性能的前提下降低模型训练对标注数据的需求,因此仅对挑选得到的训练数据的个数对模型的影响进行讨论。跟踪器中其他参数均保持和原有跟踪器一致。表 1 给出了文中方法在训练数据为 500, 1 000, 1 500, 2 000 以及 2 500 时在 PTB-TIR 数据集上的跟踪结果。从该结果可以看出,随着训练数据的增多,跟踪器的性能越来越好;同时在训练数据从 2 000 增加到 2 500 时,性能提升已经不太显著,这也说明只需要足够的具有代表性的训练数据就可以得到不错的跟踪结果。

表 1 在不同训练数据下的对比结果

数据量	精度/%	成功率/%
500	68.6	56.6
1 000	72.9	59.8
1 500	75.1	61.7
2 000	76.8	62.9
2 500	78.1	63.8

### 3.3 算法对比分析

为了验证文中方法的有效性,主要通过定量分析和定性分析的方式和其他一些具有代表性的热红外目标跟踪算法进行对比。在定量分析中主要通过跟踪精度和成功率对不同的跟踪算法进行对比。精度显示了目标估计的位置中心和真实位置的中心在距给定阈值下的准确率。成功率同时考虑了目标的尺度大小和位置信息,通过重叠率曲线来计算。在定性分析中,主要通过不同算法在一些具有挑战性的跟踪序列上的跟踪结果的可视化来直观地展示跟踪结果。

#### 3.3.1 定量对比

在 PTB-TIR 数据集上的对比结果:这是一个包含 60 个测试序列的热红外行人跟踪基准。首先展示提出的 WSRL 跟踪器和 DiMP<sup>[29]</sup>、MMNet<sup>[20]</sup>、STAMT<sup>[6]</sup>、ASTMT<sup>[34]</sup>、MLSSNet<sup>[35]</sup>、CREST<sup>[36]</sup>、UDT<sup>[14]</sup>、MCFTS<sup>[7]</sup>、HSSNet<sup>[8]</sup> 以及 SiamTir<sup>[37]</sup> 跟踪器的实验对比结果,如表 2 所示。从表 2 可以看出,提出的 WSRL 跟踪器在成功率指标上获得了最好的分数。

相比于 DiMP 有 2 个百分点的提升,相比于其他跟踪器也有明显的优势。尽管 WSRL 跟踪器在精度图上的得分低于 MMNet 跟踪器的得分,但其在成功率得分上远高于 MMNet 跟踪器。该实验结果表明,提出的弱监督表示学习可以对目标进行良好的表示,并能够使跟踪器准确地定位目标位置。

表 2 在 PTB-TIR 数据集上的对比结果

跟踪器	精度/%	成功率/%
WSRL	78.1	63.8
DiMP	74.9	61.8
MMNet	78.3	55.7
STAMT	78.1	57.3
ASTMT	77.7	62.1
MLSSNet	73.1	51.6
CREST	71.1	52.4
UDT	69.9	52.9
MCFTS	69.0	49.2
HSSNet	68.9	46.8
SiamTir	60.8	45.9

在 LSOTB-TIR 数据集上的对比结果:作为常用的热红外目标跟踪算法测试数据集,LSOTB-TIR 数据集包含 120 个测试视频序列。表 3 给出了提出的 WSRL 跟踪器和 DiMP<sup>[29]</sup>、MMNet<sup>[20]</sup>、STAMT<sup>[6]</sup>、ASTMT<sup>[34]</sup>、MLSSNet<sup>[35]</sup>、CREST<sup>[36]</sup>、UDT<sup>[14]</sup>、MCFTS<sup>[7]</sup>、HSSNet<sup>[8]</sup>以及 SiamTir<sup>[37]</sup>跟踪器的实验对比结果。从表 3 可以看出,提出的 WSRL 跟踪器在精

度和成功率指标上都获得了最高分。与其他跟踪器相比,WSRL 跟踪器在这两个性能指标上都有显著提高。上述实验结果证明了所提出的弱监督表示学习的 WSRL 跟踪方法的有效性。

表 3 在 LSOTB-TIR 数据集上的对比结果

跟踪器	精度/%	成功率/%
WSRL	81.6	68.6
DiMP	78.7	66.2
MMNet	56.3	49.9
STAMT	71.2	57.3
ASTMT	78.7	62.1
MLSSNet	59.6	45.9
CREST	59.7	50.4
UDT	62.9	52.3
MCFTS	63.5	47.9
HSSNet	51.5	40.9
SiamTir	70.0	55.4

### 3.3.2 定性对比

为了更加直观地展示不同跟踪器的跟踪结果,给出了提出的 WSRL 跟踪器和一些具有代表性的跟踪器在几个最具有挑战性的跟踪测试序列上的对比结果。图 3 展示了提出的 WSRL 跟踪器和 DiMP、HSSNet、ASTMT 跟踪器在一些具有挑战性的热红外目标测试序列上的对比结果。为了更清楚地显示不同跟踪器的跟踪结果,图 4 同时给出了目标的真实标签



图 4 不同跟踪器跟踪结果可视化对比

(灰色边界框)。从图中可以看出,DiMP 跟踪器很容易受到相似物干扰。HSSNet 跟踪器不能很好地处理相似物干扰以及目标尺度变化等问题。ASTMT 跟踪器在相似物干扰以及遮挡问题处理方面表现不佳。与对比的算法不同的是,提出的 WSRL 跟踪器在这些测试序列上取得了不错的跟踪结果,这说明所提出的基于弱监督表示学习的跟踪方法在热红外目标跟踪任务是可行的。

#### 4 结束语

针对基于深度卷积神经网络模型的热红外目标跟踪器需要大量带有标签的数据进行模型训练,带有标签的数据又十分有限的问题,提出了采用弱监督学习的方式进行模型训练来降低跟踪模型对标签数据的需求。采用主动学习的思想来挑选代表性样本序列,然后给定样本序列的首帧标签,采用基础跟踪器生成序列的其他伪标签。利用带有真实标签和伪标签的样本序列进行训练得到的模型可以对目标进行良好的表示,从而在降低模型训练对标签数据需求的同时获得不错的跟踪性能。热红外目标跟踪算法模型训练采用的是可见光的训练数据,热红外视频训练数据较少,这也显著影响了跟踪器的精度。

#### 参考文献:

- [1] 刘 洋,杨小军. 基于孪生网络特征融合与阈值更新的跟踪算法[J]. 计算机技术与发展,2022,32(3):65-70.
- [2] 张 玲,田 鹏,王 溢,等. 基于多空间显著特征的目标跟踪算法[J]. 计算机技术与发展,2019,29(11):71-76.
- [3] 李 玺,查宇飞,张天柱,等. 深度学习的目标跟踪算法综述[J]. 中国图象图形学报,2019,24(12):2057-2080.
- [4] 陈志浩,杨小军. 基于孪生网络多级注意特征融合目标跟踪算法[J]. 计算机技术与发展,2021,31(11):58-63.
- [5] 张汝佳,杨小军,王 海. 多特征融合相关粒子滤波器视频目标跟踪算法[J]. 计算机技术与发展,2021,31(6):29-34.
- [6] YUAN D, SHU X, LIU Q, et al. Structural target-aware model for thermal infrared tracking [J]. Neurocomputing, 2022,491:44-56.
- [7] LIU Q, LU X, HE Z, et al. Deep convolutional neural networks for thermal infrared object tracking [J]. Knowledge-Based Systems,2017,134:189-198.
- [8] LI X, LIU Q, FAN N, et al. Hierarchical spatial-aware siamese network for thermal infrared object tracking [J]. Knowledge-Based Systems,2019,166:71-81.
- [9] 刘 乔. 基于深度表征学习的热红外目标跟踪方法研究[D]. 哈尔滨:哈尔滨工业大学,2021.
- [10] SENER O, SAVARESE S. Active learning for convolutional neural networks: a core-set approach [C]//Proceedings of the international conference on learning representations. Vancouver:STAT,2018:1-13.
- [11] YUAN D, CHANG X, LIU Q, et al. Active learning for deep visual tracking [J]. IEEE Transactions on Neural Networks and Learning Systems,2023,35(1):1-13.
- [12] LEE D Y, SIM J Y, KIM C S. Multihypothesis trajectory analysis for robust visual tracking [C]//Proceedings of the IEEE conference on computer vision and pattern recognition. Boston:IEEE,2015:5088-5096.
- [13] KALAL Z, MIKOLAJCZYK K, MATAS J. Tracking learning detection [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence,2011,34(7):1409-1422.
- [14] WANG N, SONG Y, MA C, et al. Unsupervised deep tracking [C]//Proceedings of the IEEE conference on computer vision and pattern recognition. Long Beach:IEEE,2019:1308-1317.
- [15] 王红涛,邓森磊,赵文君,等. 基于深度学习的单目标跟踪算法综述[J]. 计算机系统应用,2022,31(5):40-51.
- [16] 王 林,郑有玲. 结合孪生网络重检的长期目标跟踪算法[J]. 计算机系统应用,2022,31(4):188-195.
- [17] MA C, HUANG J B, YANG X, et al. Hierarchical convolutional features for visual tracking [C]//Proceedings of the IEEE international conference on computer vision. Santiago:IEEE,2015:3074-3082.
- [18] LI X, MA C, WU B, et al. Target-aware deep tracking [C]//Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. Long Beach:IEEE,2019:1369-1378.
- [19] GUNDOGDU E, KOC A, SOLMAZ B, et al. Evaluation of feature channels for correlation-filter-based visual object tracking in infrared spectrum [C]//Proceedings of the IEEE conference on computer vision and pattern recognition workshops. Las Vegas:IEEE,2016:24-32.
- [20] LIU Q, LI X, HE Z, et al. Multi-task driven feature models for thermal infrared tracking [C]//Proceedings of the AAAI conference on artificial intelligence. New York:AAAI,2020:11604-11611.
- [21] GAL Y, ISLAM R, GHAHRAMANI Z. Deep bayesian active learning with image data [C]//Proceedings of the international conference on machine learning. Sydney:PMLR, 2017:1183-1192.
- [22] HUANG S J, JIN R, ZHOU Z H. Active learning by querying informative and representative examples [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence,2014,36(10):1936-1949.
- [23] FU W, WANG M, HAO S, et al. Scalable active learning by approximated error reduction [C]//Proceedings of the 24th ACM SIGKDD international conference on knowledge discovery & data mining. London:ACM,2018:1396-1405.
- [24] YOO D, KWEON I S. Learning loss for active learning [C]//Proceedings of the IEEE conference on computer vi-

- sion and pattern recognition. Long Beach; IEEE, 2019: 93–102.
- [25] WANG X, JABRI A, EFROS A A. Learning correspondence from the cycle-consistency of time[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. Long Beach; IEEE, 2019: 2566–2576.
- [26] DWIBEDI D, AYTAZ Y, TOMPSO J, et al. Temporal cycle-consistency learning[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. Long Beach; IEEE, 2019: 1801–1810.
- [27] LI X, LIU S, MELLO S D, et al. Joint-task self-supervised learning for temporal correspondence[C]//Proceedings of the advances in neural information processing systems. Vancouver; MIT Press, 2019: 317–327.
- [28] QI C R, YI L, SU H, et al. PointNet++: deep hierarchical feature learning on point sets in a metric space[C]//Proceedings of the advances in neural information processing systems. Long Beach; MIT Press, 2017: 5099–5108.
- [29] BHAT G, DANELLJAN M, GOOL L V, et al. Learning discriminative model prediction for tracking[C]//Proceedings of the IEEE international conference on computer vision. Seoul; IEEE, 2019: 6182–6191.
- [30] DANELLJAN M, BHAT G, KHAN F S, et al. ATOM: accurate tracking by overlap maximization[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. Long Beach; IEEE, 2019: 4660–4669.
- [31] LIU Q, HE Z, LI X, et al. PTB-TIR: a thermal infrared pedestrian tracking benchmark[J]. IEEE Transactions on Multimedia, 2019, 22(3): 666–675.
- [32] LIU Q, LI X, HE Z, et al. LSOTB-TIR: a large-scale high-diversity thermal infrared object tracking benchmark[C]//Proceedings of the 28th ACM international conference on multimedia. Seattle; ACM, 2020: 3847–3856.
- [33] 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.
- [34] YUAN D, SHU X, LIU Q, et al. Aligned spatial-temporal memory network for thermal infrared target tracking[J]. IEEE Transactions on Circuits and Systems II: Express Briefs, 2023, 70(3): 1224–1228.
- [35] LIU Q, LI X, HE Z, et al. Learning deep multi-level similarity for thermal infrared object tracking[J]. IEEE Transactions on Multimedia, 2020, 23: 2114–2126.
- [36] SONG Y, MA C, GONG L, et al. Crest: convolutional residual learning for visual tracking[C]//Proceedings of the IEEE international conference on computer vision. Venice; IEEE, 2017: 2555–2564.
- [37] DONG X, SHEN J. Triplet loss in Siamese network for object tracking[C]//Proceedings of the European conference on computer vision. Munich; Springer, 2018: 459–474.