

# 基于通道注意力机制的高光谱目标跟踪方法

唐靖杰, 赵 林, 欧阳尔, 李 彬, 胡文静

(湖南理工学院 信息科学与工程学院, 湖南 岳阳 414006;  
三维重建与智能应用技术湖南省工程研究中心, 湖南 岳阳 414006)

**摘 要:** 传统目标跟踪算法根据彩色图像中目标的颜色、纹理等视觉信息进行目标跟踪。当目标发生形变或被遮挡时, 彩色图像中目标的视觉信息容易发生改变, 导致传统目标跟踪器失效。高光谱图像(HSI)中包含连续的波段光谱信息, 在上述场景下具有更强的稳定性, 有助于提高目标跟踪算法的鲁棒性。提出一种基于通道注意力机制的目标跟踪算法(CAM)。该方法将 HSI 不同通道的全局光谱信息和图像通道间的相关信息融合表征为权重向量; 然后, 通过加权重新校准 HSI 中的光谱响应值, 增强图像中目标波段光谱信息的有效性, 抑制冗余的背景光谱信息的干扰, 使得 HSI 中目标和背景之间的可分离性增强; 最后将加权后的图像输入到跟踪网络中得到预测结果。在高光谱视频数据集上的实验结果表明, 该算法具有良好的跟踪性能, 优于多数现有的目标跟踪算法。

**关键词:** 目标跟踪; 高光谱图像; 通道注意力机制

**中图分类号:** TQ46; Q51

**文献标识码:** A

**文章编号:** 1672-5298(2023)01-0033-06

## Channel Attention Mechanism Method for Hyperspectral Object Tracking

TANG Jingjie, ZHAO Lin, OUYANG Er, LI Bin, HU Wenjing

(School of Information Science and Engineering, Hunan Institute of Science and Technology, Yueyang 414006, China;  
Hunan Engineering Technology Research Center for 3D Reconstruction and Intelligent Application, Yueyang 414006, China)

**Abstract:** Traditional object tracking algorithms track targets based on the color and texture of the target in the color image. When the target is occlusion or deformed in complex scenarios, the visual information is easy to change leads to the traditional target tracker failed. However, spectral information in hyperspectral images (HSI) can improve the stability of object tracking. In this paper, an object tracking method was proposed based on channel attention mechanism (CAM), which can extract the global spectral information and correlation information of different channels to generate the weight vector of channels. Through a channel weighting operation, the spectral response value in HSI was recalibrated to enhance the valuable band spectral information and suppress the interference of useless background spectral information, and the separability between target and background was enhanced in HSI. Finally, the weighted image was input into the tracking network to obtain the prediction result. The experimental results on hyperspectral datasets show that the proposed tracker obtains satisfactory performance, and is superior to most of the current object tracker.

**Key words:** object tracking; hyperspectral image; channel attention mechanism

目标跟踪是计算机视觉领域一个基础却又充满挑战的课题, 在视频监控、人机交互、军事侦查等领域得到广泛应用。目标跟踪算法通过在第一帧图像中给定目标信息, 不断预测更新后续图像中目标的状态信息, 实现对目标的稳定跟踪。然而, 目标跟踪的实际场景十分复杂, 在跟踪过程中存在诸如光照变化、目标遮挡、目标形变以及运动模糊等不确定因素, 会导致跟踪性能下降。因此, 如何提高复杂场景下目标跟踪算法的精度及鲁棒性仍是视觉跟踪研究领域的难点之一。

目前, 视觉目标跟踪方法主要分为相关滤波方法和深度学习方法。在基于相关滤波的目标跟踪领域中, Bolme<sup>[1]</sup>等引入最小输出平方和误差(MOSSE)算法, 将图像特征初始化后生成相关滤波器进行目标跟踪。Danelljan<sup>[2]</sup>等提出空间正则化相关滤波器算法(SRDCF), 在更大范围的区域中寻找目标, 使跟踪精度得到提升。在基于深度学习的目标跟踪领域中, Nam<sup>[3]</sup>等提出一种多域学习的算法(MDNet), 利用

收稿日期: 2020-07-12

基金项目: 湖南省自然科学基金项目(2019JJ40110, 2019JJ40104); 湖南省教育厅科学研究项目(18B349, 19A201, 20A214); 湖南省研究生科研创新项目(CX20201141, CX20211160)

作者简介: 唐靖杰, 男, 硕士研究生。主要研究方向: 图像处理与计算机视觉, 目标跟踪

通信作者: 胡文静, 男, 博士, 教授。主要研究方向: 计算机视觉, 深度学习

卷积神经网络模型学习不同域中目标的通用特征,以区分目标和背景. Li<sup>[4]</sup>等提出孪生区域建议网络(SiamRPN),通过孪生网络提取目标特征,再利用区域建议网络来对目标进行分类和定位,让网络分别输出目标的分类和边框回归,缓解边界框精度低的问题.

针对彩色图像中目标在复杂场景下跟踪性能下降的问题,部分研究人员提出利用记录了连续光谱空间信息的高光谱图像(HSI, Hyperspectral Image)进行目标跟踪,光谱空间信息能够提高 HSI 的可判别能力<sup>[5]</sup>,有利于目标跟踪. Uzkent<sup>[6]</sup>等提出一种基于深度核相关滤波的图像处理方法(DeepHKCF),在图像传递到深度卷积神经网络之前将 HSI 转换为 3 通道的伪彩色图像,但是这样做无法有效利用 HSI 中大量的光谱信息. Xiong<sup>[7]</sup>等提出一种基于 HSI 材质信息的目标跟踪方法(MHT),通过 HSI 中目标材质的不同光谱响应来构建 HSI 中的 3D 局部光谱空间结构以及构成材质分布的丰度矩阵,将这两种类型的特征嵌入到相关滤波器中进行基于材质的目标跟踪,此方法获得不错的跟踪精度以及性能指标.

高光谱图像承载了大量可区分的目标与背景光谱信息<sup>[8]</sup>,与此同时,高光谱图像也存在数据维度高、光谱信息利用率低的问题,再加上目前高光谱视频数据集有限,这些都制约了深度学习网络模型在高光谱图像目标跟踪领域的应用. 本文提出一种基于通道注意力机制模块(Channel Attention Module, CAM) 的目标跟踪算法,训练过程中通过获取 HSI 的光谱信息和波段间相关性生成波段权重,对 HSI 图像进行波段加权,根据权值的不同对图像中的光谱信息进行增强或者抑制,再将不同域中加权后的 HSI 输入到卷积神经网络(CNN)模型中,根据图像中目标的光谱响应以及空间分布信息,捕捉到不同域中目标的材质信息,通过这些信息让网络获得对目标与背景更强的区分能力,提高网络在复杂场景下目标跟踪的鲁棒性.

## 1 算法设计

### 1.1 整体框架

本文所使用的神经网络结构如图 1 所示,主要由通道注意力模块、5 个卷积单元和两个全连接层组成. 其中 Conv1~Conv5 表示卷积单元, Conv1 和 Conv2 由二维卷积层、Relu 层、批量归一化层和最大池化层组成, Conv3~Conv5 仅由二维卷积层和 Relu 层组成. Fc6 和 Fc7 表示全连接层,  $\delta_+(x)$  表示网络预测候选样本为目标的分,  $\delta_-(x)$  表示网络预测候选样本为背景的分.

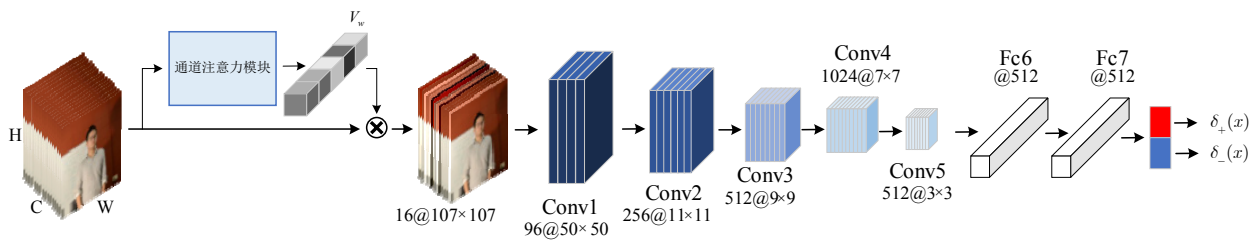


图 1 整体网络结构框架

### 1.2 通道注意力机制

为了能够从复杂的背景中将需要跟踪的目标准确区分出来,需要对 HSI 中目标特征进行聚焦,增加目标和背景的可判别性. 因此,引入通道注意力机制模块,提前赋予输入图像不同通道的重要性,增强重要的通道光谱信息,抑制无用的通道光谱信息.

通道注意力模块如图 2 所示,其包含三个部分:数据压缩、特征捕获和特征聚合. 首先,将输入 HSI 的全局光谱信息压缩为能够体现 HSI 整体光谱特征的通道信息向量. 然后,利用两层一维卷积来捕获信息向量的深层特征以及不同通道间的相关性特征. 最后,利用获得的通道特征信息生成通道权重向量,重新校准输入的 HSI 数据.

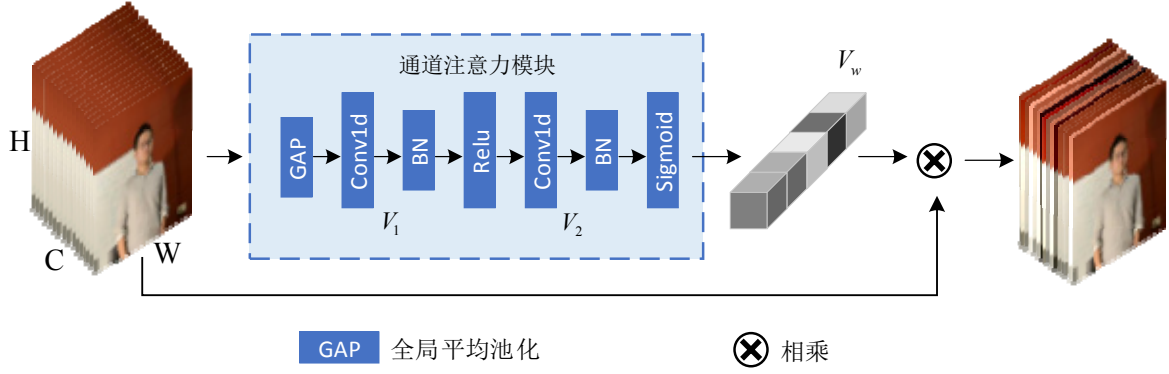


图 2 通道注意力机制

数据压缩部分. 根据数据标签从 HSI 中截取样本块  $I$  作为输入数据, 样本块  $I$  可表示为 2D 矩阵  $\{I_1, I_2, \dots, I_c\}$ . 其中  $I_i \in \mathbb{R}^{H \times W}$ ,  $c$  为通道数目,  $H$  和  $W$  分别为输入数据的高度和宽度. 利用一个全局平均池化层对 HSI 每个通道的全局光谱信息进行求和压缩, 获得 HSI 每个通道光谱特征值  $v_s^c$ . 该光谱特征值能够体现 HSI 对应通道的全局特征信息, 间接扩大了网络的感受野. 每个通道的光谱特征值  $v_s^c$  可描述为

$$v_s^c = \mu(I_c) = \frac{1}{H \times W} \sum_{i=1}^H \sum_{j=1}^W I_c(i, j), \quad c=1, 2, \dots, N,$$

其中  $\mu(\cdot)$  为全局平均池化函数,  $I_c(i, j)$  为对应于像素空间位置  $(i, j)$  的光谱强度. 输入样本数据  $I$  的光谱特征向量  $\mathbf{v}_s = \{v_s^1, v_s^2, \dots, v_s^c\}$ .

特征捕获部分. 通过两个一维卷积层去捕获每个通道的特征权重和通道之间的相关性, 得到光谱特征向量的权重参数  $V_1$  和  $V_2$ , 每个一维卷积层后接一个 BN 层, 对得到的权重参数进行批量规范化处理. 最后通过 Sigmoid 函数将权重参数进行归一化得到对应 HSI 每个通道的权重向量

$$\mathbf{V}_w = \sigma(C_2(C_1(\mathbf{v}_s))),$$

其中  $\sigma$  为 Sigmoid 激活函数,  $C_1$  为第一层一维卷积层,  $C_2$  为第二层一维卷积层.

特征聚合部分. 将权重向量  $\mathbf{V}_w$  与输入 HSI 对应通道光谱值进行加权聚合, 重新校准 HSI 光谱响应值, 增强贡献度高的通道光谱信息并抑制无用的背景通道信息, 从而提升后续 CNN 网络对 HSI 中目标和背景的判别能力. 加权过程可表示为

$$I_{re} = \mathbf{V}_w \otimes I,$$

其中  $\otimes$  表示向量元素对应相乘,  $I_{re}$  为重新校准后的 HSI 样本数据.

### 1.3 学习算法

本文算法的目的是训练一个能够消除目标和背景歧义的卷积神经网络(CNN)模型, 但在很多具有挑战性的场景下, 这是困难的. 例如光照改变、运动模糊、目标形变遮挡. 为了有效提取出不同场景下目标的通用特征, 将通道注意力模块嵌入到网络模型中, 针对不同的视频序列, 对输入的高光谱图像进行通道加权, 以训练出能够更好区分不同视频场景下目标信息和背景信息的通用网络模型. 我们的 CNN 模型采用随机梯度下降(SGD)算法来进行优化, 模型总共迭代  $L$  次, 每次迭代都对模型参数进行  $K$  次更新, 每次更新都是基于小批量样本数据集对模型进行训练. 小批量样本数据集由 32 个正样本块和 96 个负样本块组成, 这些样本块是从训练的视频序列中随机抽取的 8 帧图像, 然后从图像中截取得来. 最后将小批量训练样本数据集输入到网络中, 通过计算二值交叉熵损失对网络模型进行更新. 一直重复, 直到网络收敛或达到预定的迭代次数.

### 1.4 在线跟踪过程

在线跟踪过程中, 针对新的视频序列进行测试. 首先对带有目标数据标签的视频序列第一帧进行预处理, 生成 500 个正样本以及 5000 个负样本, 用这些正负样本数据对通道注意力模块以及 Fc6 和 Fc7 进

行在线微调, 总共经过 50 次迭代, 目的是让模型能够更好地拟合需要测试的视频序列, 增强预训练模型的泛化能力.

在上一帧的预测目标周围, 通过高斯分布获取  $N(N = 256)$  个样本区域  $x_i, i = 1, 2, \dots, N$ . 通过网络模型来评估每一个样本的得分  $\delta_+(x_i)$ , 得分最高的样本区域被视为当前帧的预测目标  $x_i^*$ , 即

$$x_i^* = \arg \max_{x_i} (\delta_+(x_i)).$$

## 2 实验结果与分析

### 2.1 高光谱跟踪数据集

高光谱数据集由训练集和测试集组成, 其中训练集包含 40 个视频序列, 测试集包含 35 个视频序列. 每个视频序列下包含高光谱图像序列 HSI、与 HSI 对应的伪彩色(False-Color)图像序列以及 RGB 彩色图像序列. 每个视频图像序列平均约 392 帧, 视频速度约 25 帧每秒. 每一段视频都会在 11 个属性中标注相关挑战性因素, 包括照明变化(IV)、比例变化(SV)、遮挡(OCC)、变形(DEF)、运动模糊(MB)、快动作(FM)、平面内旋转(IPR)、平面外旋转(OPR)、视线外(OV)、背景杂波(BC)和低分辨率(LR). 数据集包含多个目标类别、多种场景、丰富的活动和不同内容, 从而保证了数据集的通用性及复杂性.

### 2.2 评价指标

跟踪器性能可以通过精确度曲线图、成功率曲线图以及成功率曲线下面积(AUC)来描述.

精确度曲线图记录了每一帧图像目标的预测位置比较标签位置在指定的像素距离阈值内的分数值, 精确度的评估标准使用平均中心位置误差(CLE), CLE 是指算法预测结果的中心位置与目标标签位置之间的距离:

$$CLE = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2},$$

其中  $(x_1, y_1)$  表示目标预测框的中心坐标,  $(x_2, y_2)$  表示目标标签框的中心坐标.

精确度是计算 CLE 小于某一像素距离阈值(本文算法阈值为 20 个像素值, 记为 DP@20pixels)的视频帧数与该视频序列总帧数的百分比.

目标的预测框与数据标签框之间的重叠率(IOU)的计算公式为

$$IOU = \frac{|G \cap A|}{|G \cup A|},$$

其中  $G$  为目标标签框,  $A$  为目标预测框,  $|\cdot|$  为该区域内的像素点个数.

成功率曲线图展示了跟踪成功图像帧的百分比, 成功率是指 IOU 大于某个在 0 到 1 之间变化阈值(本文算法阈值为 0.5)的视频帧数占该视频序列总帧数的百分比.

AUC 为成功率曲线下方面积值, 可以通过 AUC 来评估跟踪器在所有测试视频序列中的整体性能.

### 2.3 实验结果与分析

将本文提出的 CAM 跟踪器与基于 HSI 序列设计的跟踪器 MHT<sup>[6]</sup>进行比较, 并且与其他 6 种基于 3 通道彩色图像序列跟踪器进行比较, 包括 KCF<sup>[9]</sup>、FDSST<sup>[10]</sup>、SRDCF<sup>[2]</sup>、SAMF<sup>[11]</sup>、BACF<sup>[12]</sup>和 MDNet<sup>[3]</sup>. 这 6 种跟踪器都是针对彩色视频图像而开发的, 所以在与高光谱视频序列对应的彩色图像和伪彩色图像上分别测试这几种跟踪器的性能, 测试结果如图 3、图 4 所示. 结果表明, 本文所提出的跟踪器与其他跟踪器相比, 无论是精度值还是 AUC 都展现出更优秀的性能.

表 1 和表 2 记录了所有在 RGB、False-color 和 HSI 视频序列上测试的跟踪器的性能指标数值. 不难看出, 本文 CAM 跟踪器的性能与其他跟踪器相比具有明显的优势, 这是因为 CAM 跟踪器在对 HSI 特征提取前通过通道注意力机制, 选择性增强有用、信息量大的通道, 并抑制无用、冗余的通道, 实现了高性能. MHT 跟踪器的性能也表现良好, 因为 MHT 能够有效利用 HSI 中的物质信息来区分目标和背景. BACF 跟踪器可以通过空间正则化来缓解边界效应从而获得了不错的性能. 此外, MDNet 跟踪器在 RGB 视频上

的跟踪性能要优于在 False-color 视频上的跟踪性能, 这可能是因为该跟踪器无法有效应对伪彩色视频数据中存在的冗余噪声。

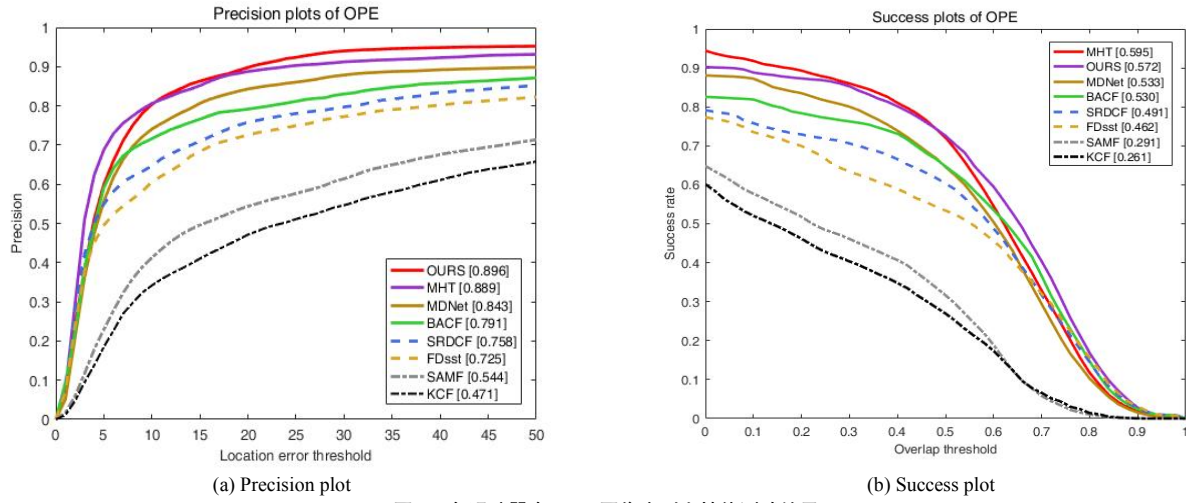


图 3 各跟踪器在 RGB 图像序列上性能测试结果

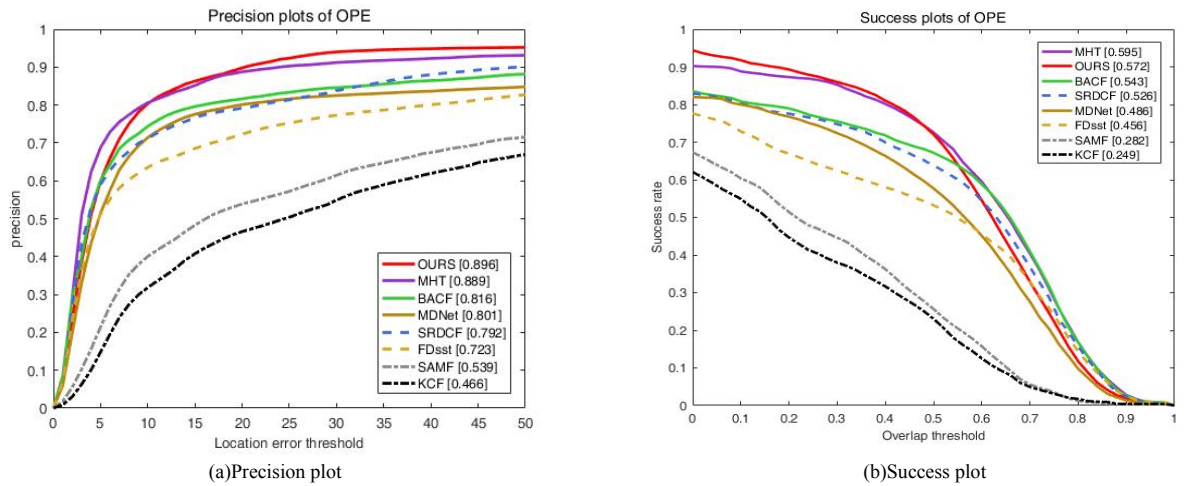


图 4 各跟踪器在 HSI-Falsecolor 图像序列上性能测试结果

表 1 各算法在 DP@20pixels 指标下的性能比较

Video	Ours	MHT	MDNet	BACF	SRDCF	FDSST	SAMF	KCF
Color	—	—	0.843	0.791	0.758	0.725	0.544	0.471
HSI/False-color	<b>0.896</b>	<b>0.889</b>	0.801	0.816	0.792	0.792	0.539	0.466

表 2 各算法在 AUC 指标下的性能比较

Video	Ours	MHT	MDNet	BACF	SRDCF	FDSST	SAMF	KCF
Color	—	—	0.533	0.530	0.491	0.462	0.291	0.261
HSI/False-color	<b>0.572</b>	<b>0.595</b>	0.486	0.543	0.526	0.456	0.282	0.249

图 5 展示了本文算法与其他 4 个跟踪算法在高光谱或伪彩色视频图像序列上定性评估的跟踪结果。从图 5 可以看出, 当目标的姿态和视角发生变化并且目标正处于快速移动状态(Kangaroo 序列, Car 序列), 当目标被遮挡(Campus 序列), 或与背景颜色环境相似(Paper 系列, Campus 序列)时, 这些挑战因素都会让目标在一定程度上融入背景干扰信息, 使得 BACF、MDnet 和 SRDCF 这 3 种基于视觉信息对目标进行跟踪的跟踪器对目标的判别能力不足, 导致跟踪出现飘移。而本文算法具有更强的鲁棒性, 其原因是: (1)本文设计的跟踪器能够利用 HSI 中更具鲁棒性的光谱信息对目标进行跟踪, 而光谱信息面对形变、遮挡这些因素不易发生改变; (2)算法结合通道注意力机制, 在 HSI 特征提取上能够有选择性地关注目标显著特征、抑制背景噪声。





图 5 多个序列中部分跟踪结果

### 3 结束语

本文从如何提升 HSI 目标波段光谱信息的有效性出发, 提出一种基于通道注意力机制的目标跟踪算法. 该算法通过探索 HSI 不同通道的深层特征和相关性获得通道权重, 并重新校准了 HSI 的光谱信息, 增强了 HSI 目标波段光谱信息的有效性, 抑制了冗杂无用的背景波段光谱信息, 提升了网络模型判别目标和背景的能力. 在高光谱数据集上进行的实验表明, 该算法在图像分辨率低、目标发生形变旋转或者被复杂的背景遮挡等情况下, 仍然能够获得优异的跟踪性能.

#### 参考文献:

- [1] Bolme D S, Beveridge J R, Draper B A, et al. Visual object tracking using adaptive correlation filters[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, IEEE, 2010: 2544~2550
- [2] Danelljan M, Hager G, Khan F S, et al. Learning spatially regularized correlation filters for visual tracking[C]//Proceedings of the IEEE International Conference on Computer Vision, IEEE, 2015: 4310~4318
- [3] Nam H, Han B. Learning multi-domain convolutional neural networks for visual tracking[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, IEEE, 2016: 4293~4302
- [4] Li B, Yan J, Wu W, et al. High performance visual tracking with siamese region proposal network[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, IEEE, 2018: 8971~8980
- [5] 易嘉闻, 李 希, 欧阳尔, 等. 基于自编码的高光谱图像波段加权分类网络研究[J]. 湖南理工学院学报(自然科学版), 2021, 34(1): 34~39
- [6] Uzkent B, Rangnekar A, Hoffman M J. Tracking in aerial hyperspectral videos using deep kernelized correlation filters[J]. IEEE Transactions on Geoscience and Remote Sensing, 2019, 57(1): 449~461
- [7] Xiong F C, Zhou J, Qian Y T. Material based object tracking in hyperspectral videos[J]. IEEE Transactions on Image Processing, 2020, 29: 3719~3733
- [8] 张国云, 欧阳慧婷, 涂 兵, 等. 空间一致核协同优化的高光谱异常检测方法[J]. 湖南理工学院学报(自然科学版), 2022, 35(3): 10~16+43
- [9] Henriques J F, Caseiro R, Martins P, et al. High-speed tracking with kernelized correlation filters[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2015, 37(3): 583~596
- [10] Danelljan M, Hager G, Khan F S, et al. Discriminative scale space tracking[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2017, 39(8): 1561~1575
- [11] Li Y, Zhu J. A scale adaptive kernel correlation filter tracker with feature integration[C]//Proceedings of the ECCV 2014 Workshops, Part II, LNCS 8926, Springer, 2015: 254~265
- [12] Galoogahi H K, Fagg A, Lucey S. Learning background-aware correlation filters for visual tracking[C]//Proceedings of the IEEE International Conference on Computer Vision, IEEE, 2017: 1144~1152