引文格式:赖普坚,张雪娇,程塬.脉冲驱动光谱-空间-时序建模的多光谱无人机单目标跟踪[J].武汉大学学报(信息科学版),2026,51(11):2456-2468.DOI:10.13203/j.whugis20260279

Citation: LAI Pujian, ZHANG Xuejiao, CHENG Gong. A Spike-Driven Framework with Spectral-Spatial-Temporal Modeling for Multispectral UAV Single Object Tracking[J]. Geomatics and Information Science of Wuhan University, 2026, 51(11): 2456-2468.DOI:10.13203/j.whugis20260279

脉冲驱动光谱-空间-时序建模的多光谱无人机单目标跟踪

赖普坚[1] 张雪娇[1] 程 塬[1]

1 西北工业大学自动化学院,陕西 西安,710072

摘 要:得益于不同谱段提供的丰富光谱线索,多光谱无人机单目标跟踪受到广泛关注。然而,现有方法大多依赖功耗较高的人工神经网络进行设计,难以满足无人机平台低能耗的视觉感知需求。为此,提出一种脉冲驱动光谱-空间-时序建模的多光谱无人机单目标跟踪算法。具体而言,所提算法主要包含记忆扰动增强的光谱-空间-时序检索模块和脉冲驱动的光谱门控融合模块。记忆扰动增强的光谱-空间-时序模块通过构建紧凑的历史帧目标记忆缓存,引导搜索特征与历史记忆进行递归交互,逐步挖掘目标相关线索。针对历史帧目标记忆缓存中噪声累积或者表征偏移问题,引入记忆扰动机制,通过模拟潜在的记忆偏差,增强模型对不可靠历史信息和背景干扰的鲁棒性。此外,通过时序聚合生成紧凑搜索特征,并沿光谱组维度进行脉冲自注意力建模,实现不同光谱组之间的信息交互。脉冲驱动的光谱门控融合模块可自适应评估不同光谱组特征的重要性,并进行加权融合,从而增强关键光谱组的特征响应,同时抑制光谱冗余信息。在多光谱无人机单目标跟踪数据集上进行实验,所提算法的精度与归一化精度分别达到83.5%和82.2%,能耗仅32.4 mJ。结果表明,所提算法在提升多光谱无人机单目标跟踪精度与复杂场景鲁棒性的同时,仍保持极低的能耗开销。

关键词:无人机;单目标跟踪;光谱-空间-时序建模;脉冲驱动;低能耗;记忆扰动增强

中图分类号:P237 **文献标识码**:A **收稿日期**:2026-07-22

DOI:10.13203/j.whugis20260279 **文章编号**:1671-8860(2026)11-2456-13

A Spike-Driven Framework with Spectral-Spatial-Temporal Modeling for Multispectral UAV Single Object Tracking

LAI Pujian[1] *ZHANG Xuejiao*[1] *CHENG Gong*[1]

1 School of Automation, Northwestern Polytechnical University, Xi'an 710072, China

Abstract: **Objectives**: Multispectral unmanned aerial vehicle (UAV) single-object tracking leverages complementary spectral cues from different bands to enhance target representation in complex scenes. However, most existing trackers rely on artificial neural networks with high energy consumption, making them unsuitable for low-energy visual perception on UAV platforms. This paper aims to achieve accurate and robust multispectral tracking while maintaining low energy consumption. **Methods**: A spike-driven framework with spectral-spatial-temporal modeling is developed for multispectral UAV single object tracking. The proposed method comprises a memory perturbation-enhanced spectral-spatial-temporal retrieval module and a spike-driven spectral gated fusion module. The former constructs a compact memory cache of target representations from historical frames and recursively guides the search features to interact with the stored memory, progressively retrieving target-relevant cues. To address noise accumulation and representation drift in the historical memory cache, a memory perturbation mechanism is introduced to simulate potential memory biases, thereby improving the tracker's robustness to unreliable historical information and

基金项目:国家自然科学基金(62376223)。

第一作者:赖普坚,博士生,研究方向为视频目标跟踪。laipujian@mail.nwpu.edu.cn

通信作者:程塬,博士,教授。gcheng@nwpu.edu.cn

background interference. The retrieved cues are then integrated through temporal aggregation to generate compact search representations. Spike-driven self-attention is subsequently applied along the spectral-group dimension to facilitate information exchange among different spectral groups. The spike-driven spectral gated fusion module further estimates the importance of individual spectral-group features and performs weighted aggregation, enhancing informative spectral responses while suppressing redundant responses. **Results:** On the multispectral UAV single object tracking dataset, the proposed method achieves a precision of 83.5% and a normalized precision of 82.2%, with an energy consumption of only 32.4 mJ. **Conclusions:** The proposed method improves tracking accuracy and robustness in complex scenes while maintaining extremely low energy consumption, demonstrating its effectiveness for energy-efficient multispectral UAV single-object tracking.

Key words: unmanned aerial vehicle (UAV); single object tracking; spectral-spatial-temporal modeling; spike-driven framework; low-energy visual perception; memory perturbation-enhanced

无人机(unmanned aerial vehicle, UAV)凭借机动性强、部署灵活和观测范围广等优势,已广泛应用于军事侦察、安全监控、灾害救援和交通巡检等领域[1-4]。作为无人机视觉感知系统的核心任务之一,目标跟踪旨在根据初始帧中给定的目标状态,在后续视频序列中持续估计目标的位置与尺度。然而,无人机平台通常面临剧烈视角变化和频繁相机运动等问题,同时目标还易受到遮挡、低照度、背景杂波等因素的影响,这些因素显著增加了稳定目标表征的提取难度,并严重制约主要依赖可见光外观信息的跟踪器性能[5-7]。近年来,多光谱图像通过联合利用不同谱段对目标材质、结构及辐射特性的差异化响应,为复杂场景下的无人机目标跟踪提供了更加丰富的判别信息[8-9]。文献[10]提出了大规模多光谱无人机单目标跟踪(multispectral UAV single object tracking, MUST)数据集,并设计了统一光谱-空间-时序跟踪框架(unified spectral-spatial-temporal tracker, UNTrack)。文献[11]提出了一种面向无人机多光谱目标跟踪的上下文感知记忆框架,通过光谱-空间调制模块挖掘跨谱互补信息,并利用状态空间记忆机制建模长期时序依赖,从而提升复杂场景下的跟踪性能。文献[12]提出了一种正-负提示挖掘网络,通过将目标特征和背景干扰分别建模为正、负提示,并结合光谱-空间-时序注意力挖掘与提示调制机制增强目标与背景之间的判别性。然而,现有的多光谱无人机单目标跟踪方法大多基于人工神经网络(artificial neural network, ANN)或者视觉变换器(vision Transformer, ViT)进行设计。尽管这些方法取得了较好的跟踪性能,但其较高的能耗开销难以满足无人机平台的低能耗的端侧部署需求。

脉冲神经网络(spiking neural network, SNN)凭借其神经元动力学特性和事件驱动的计算机制[13-15],受到了广泛关注。SNN通过刻画生物神经元的脉冲发放过程,将连续信息转换为离散脉冲序列进行时空特征编码。相比于ANN或者ViT中高密度的矩阵乘法操作,SNN依赖稀疏加法计算完成信息传递与更新,在保证表达能力的同时显著降低计算复杂度和能耗。目前,已有研究尝试将SNN引入视觉目标跟踪任务,并取得了一定进展。文献[16]提出一种脉冲驱动的目标跟踪框架(spike-driven framework for efficient visual tracking, SpikeTrack),通过非对称时间步扩展、单向信息流以及记忆检索机制,在提升跟踪精度的同时显著降低计算能耗。文献[17]提出首个脉冲Transformer驱动的事件跟踪框架,通过全局轨迹提示增强事件时空表示,并采用全脉冲骨干网络实现端到端事件跟踪。文献[18]提出了一种纯脉冲事件目标跟踪框架,通过多搜索单模板训练策略增强时间依赖建模能力,并利用泄漏-积分-发放神经元(leaky integrate-and-fire neuron, LIF)自适应调节脉冲激活,在保持计算稀疏性的同时实现高效跟踪。文献[19]提出全脉冲帧-事件融合跟踪框架,通过结合卷积局部建模与Transformer全局建模实现跨模态信息融合,并利用随机图像块重组及时空正则化策略提升跟踪鲁棒性与能量效率。然而,现有脉冲目标跟踪方法主要面向可见光或事件单目标跟踪任务,对于多光谱无人机单目标跟踪中的脉冲光谱-空间-时序联合建模仍缺乏探索。

为此,本文提出了一种脉冲驱动光谱-空间-时序建模的多光谱无人机单目标跟踪算法,算法沿用SpikeTrack[16]中的基础框架,但SpikeTrack主要面向可见光目标跟踪,直接将其输入扩展至多光谱数据,难以有效处理历史模板退化、跨谱

互补信息利用不充分以及等权融合产生冗余响应等问题。针对上述局限,本文面向多光谱输入设计了记忆扰动增强的光谱-空间-时序检索(memory perturbation-enhanced spectral-spatial-temporal retrieval,MPSSTR)模块和脉冲驱动的光谱门控融合(spike-driven spectral gated fusion,SDSG)模块。前者通过历史记忆建模与扰动增强,缓解遮挡和背景杂波引起的历史模板表征漂移,并促进跨谱互补信息交互;后者依据不同谱段特征的响应自适应分配融合权重,抑制等权融合产生的冗余响应。具体而言,MPSSTR模块通过构建紧凑的历史帧记忆缓存,并引导搜索特征与历史记忆缓存进行递归交互,逐步挖掘目标相关线索。为了缓解历史帧目标记忆缓存中噪声累积或者表征偏移导致的记忆退化问题,引入记忆扰动机制,通过模拟潜在的记忆偏差增强模型对不可靠历史信息和复杂背景干扰的鲁棒性。之后,通过时序聚合生成紧凑搜索特征,并沿光谱组维度进行脉冲自注意力建模,实现不同光谱组之间的信息交互。SDSG模块自适应估计不同光谱特征的重要性,在增强目标判别性光谱信息的同时抑制冗余信息与背景响应。

1 本文方法

1.1 前期准备:脉冲神经元模型

本文的脉冲神经元模型采用归一化整数泄漏-积分-发放神经元(normalized integer leaky integrate-and-fire,NI-LIF)[20],该神经元以LIF[13]为基础,在训练阶段引入归一化整数机制,并在推理阶段将整数映射为0/1脉冲表示。NI-LIF的动态更新过程可表示为:

$$\boldsymbol{U}(t)=\beta_t\boldsymbol{H}(t-1)+\boldsymbol{Y}(t) \tag{1}$$

$$\boldsymbol{S}(t)=f_{\text{clip}}([\boldsymbol{U}(t)],0,D)/D \tag{2}$$

$$\boldsymbol{H}(t)=\boldsymbol{U}(t)-\boldsymbol{S}(t)\times D \tag{3}$$

式中,t表示时间步;$\boldsymbol{U}(t)$表示融合时间信息后的膜电位;$\boldsymbol{Y}(t)$表示当前时间步的空间输入;$\boldsymbol{H}(t)$表示当前时间步t的膜电位;$\beta_t\boldsymbol{H}(t-1)$表示由上一时间步膜电位衰减得到的时间输入;$\beta_t$为漏电因子,本文遵循文献[16]将$\beta_t$设计为可学习的参数;$\boldsymbol{S}(t)$表示当前时间步$t$的输出脉冲;$[\cdot]$为取整操作;$f_{\text{clip}}(\cdot,\min,\max)$表示将输入裁剪至区间$[\min,\max]$;$D$为控制最大脉冲输出数值的超参数。

1.2 方法框架

本文所提算法总体框架如图1所示。该算法以脉冲骨干网络为基础,围绕历史目标信息利用与多光谱特征融合,设计了MPSSTR模块和SDSG模块,提升多光谱场景下的目标表征与定位能力。具体地,采用双分支孪生脉冲骨干网络分别提取多帧模板和搜索图像的特征。MPSSTR将多帧模板信息压缩为紧凑的历史目标记忆,并通过记忆扰动训练增强模型对噪声和表征偏移的鲁棒性,进而引导搜索特征从光谱、空间和时序维度检索稳定的目标线索。在此基础上,对时

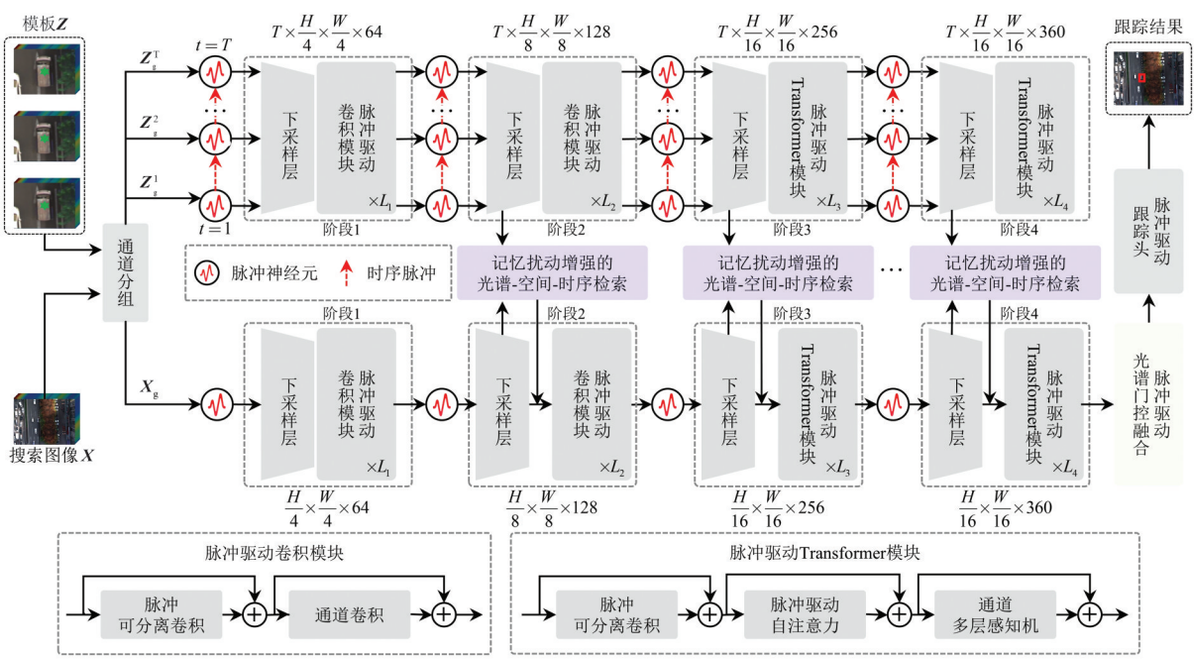


图1 基于脉冲驱动光谱-空间-时序建模的多光谱无人机单目标跟踪算法框架图

Fig. 1 Spike-Driven Framework with Spectral-Spatial-Temporal Modeling for Multispectral UAV Single Object Tracking

序聚合后的搜索特征沿光谱组维度进行脉冲自注意力计算,以建模不同光谱组之间的互补关系。随后,SDSG根据各光谱组的响应自适应生成门控权重,对其进行加权融合,增强有效谱段信息并抑制低质量谱段及冗余响应。

1.2.1　孪生脉冲骨干网络

本文方法采用脉冲驱动变换器v3[21]作为骨干网络,如图1所示,骨干网络共包含4个阶段,其中,阶段1和阶段2由下采样层和$L_j(j\in\{1,2\})$个脉冲驱动卷积模块构成,阶段3和阶段4则由下采样层和$L_j(j\in\{3,4\})$个脉冲驱动Transformer模块构成。每个阶段输出特征的通道数分别为64、128、256和360。给定模板输入$\boldsymbol{Z}\in\mathbb{R}^{T\times B\times H\times W}$和搜索图像$\boldsymbol{X}\in\mathbb{R}^{B\times H\times W}$,其中,$T$表示模板帧数,$B$表示光谱通道数,$H$表示图像的高度,$W$表示图像的宽度。为兼顾局部光谱连续性、跨谱组交互效率与预训练骨干的输入兼容性,本文对多光谱数据的波段通道进行分组。具体地,将每3个连续光谱通道划分为一组,共得到$S=\lceil B/3\rceil$组,其中$\lceil\cdot\rceil$表示向上取整。若最后一组不足3个通道,则采用零填充方式补齐。基于此,经过光谱分组后,模板$\boldsymbol{Z}$和搜索图像$\boldsymbol{X}$变换后为$\boldsymbol{Z}_g^t\in\mathbb{R}^{3S\times H\times W}$($t\in\{1,2,\cdots,T\}$)和$\boldsymbol{X}_g\in\mathbb{R}^{3S\times H\times W}$,并分别输入模板和搜索图像分支进行特征提取。

为了方便讨论,本文将脉冲驱动卷积模块或者脉冲驱动Transformer模块的输入特征记为$\boldsymbol{U}$。脉冲驱动卷积模块的处理流程可表示为:

$$\begin{cases}\bar{\boldsymbol{U}}=\boldsymbol{U}+\Phi_{\text{sep}}(\boldsymbol{U})\\ \tilde{\boldsymbol{U}}=\bar{\boldsymbol{U}}+\Phi_{\text{conv}}(\bar{\boldsymbol{U}})\end{cases}\tag{4}$$

$$\Phi_{\text{sep}}(\boldsymbol{U})=\text{Conv}_{\text{p}}(\text{SN}(\text{Conv}_{\text{d}}(\text{SN}(\text{Conv}_{\text{p}}(\text{SN}(\boldsymbol{U})))))))\tag{5}$$

$$\Phi_{\text{conv}}(\bar{\boldsymbol{U}})=\text{Conv}(\text{SN}(\text{Conv}(\text{SN}(\bar{\boldsymbol{U}}))))\tag{6}$$

式中,$\bar{\boldsymbol{U}}$、$\tilde{\boldsymbol{U}}$分别表示输入特征经第一个、第二个残差分支处理后得到的中间特征;$\Phi_{\text{sep}}(\cdot)$表示脉冲可分离卷积;$\Phi_{\text{conv}}(\cdot)$表示通道卷积;SN$(\cdot)$表示脉冲神经元层(spike neuron, SN);Conv$(\cdot)$、Conv$_{\text{p}}(\cdot)$、Conv$_{\text{d}}(\cdot)$分别表示常规卷积、逐点卷积、深度卷积。需要注意的是,为了公式的简洁性,本文忽略批归一化层(batch normalization,BN)。

脉冲驱动Transformer模块的处理流程可表示为:

$$\begin{cases}\bar{\boldsymbol{U}}=\boldsymbol{U}+\Phi_{\text{sep}}(\boldsymbol{U})\\ \tilde{\boldsymbol{U}}=\bar{\boldsymbol{U}}+\Phi_{\text{self_att}}(\bar{\boldsymbol{U}})\\ \hat{\boldsymbol{U}}=\tilde{\boldsymbol{U}}+\Phi_{\text{mlp}}(\tilde{\boldsymbol{U}})\end{cases}\tag{7}$$

$$\Phi_{\text{mlp}}(\tilde{\boldsymbol{U}})=\text{Linear}(\text{SN}(\text{Linear}(\text{SN}(\tilde{\boldsymbol{U}}))))\tag{8}$$

$$\Phi_{\text{self_att}}(\bar{\boldsymbol{U}})=\text{Linear}_{\text{O}}(\text{SN}((\boldsymbol{Q}_s\boldsymbol{K}_s^{\text{T}})\boldsymbol{V}_s\times k))\tag{9}$$

$$\begin{cases}\boldsymbol{Q}_s=\text{SN}(\text{Linear}_{\boldsymbol{Q}}(\bar{\boldsymbol{U}}))\\ \boldsymbol{K}_s=\text{SN}(\text{Linear}_{\boldsymbol{K}}(\bar{\boldsymbol{U}}))\\ \boldsymbol{V}_s=\text{SN}(\text{Linear}_{\boldsymbol{V}}(\bar{\boldsymbol{U}}))\end{cases}\tag{10}$$

式中,$\hat{\boldsymbol{U}}$表示脉冲驱动Transformer模块的输出特征;$\Phi_{\text{mlp}}(\cdot)$表示通道多层感知机(multilayer perceptron, MLP);$\Phi_{\text{self_att}}(\cdot)$表示脉冲自注意力;Linear$(\cdot)$表示线性层;$k$表示缩放系数;$\boldsymbol{Q}$、$\boldsymbol{K}$、$\boldsymbol{V}$分别表示查询、键、值,下标$s$表示张量的脉冲形式。关于脉冲骨干网络的更多细节,参见文献[21]。

1.2.2　记忆扰动增强的光谱-空间-时序检索

为充分利用模板中的目标线索增强搜索图像特征对目标的感知能力,本文设计了MPSSTR模块,如图2所示。该模块首先利用模板特征构建紧凑、可复用的历史帧目标记忆缓存,随后通过搜索分支特征与历史帧记忆缓存的递归交互逐步增强搜索分支特征的目标感知能力。在此基础上,给历史帧记忆缓存注入扰动,从而提高检索过程对目标外观变化的适应能力,并通过时序特征聚合和光谱组特征交互实现多维度特征融合。需要说明的是,为降低历史检索的推理开销,MPSSTR采用紧凑记忆缓存和单次递归交互($N=1$),避免对全部历史帧重复执行高成本匹配,记忆扰动仅在训练阶段注入,推理时不增加额外参数或分支,从而在保持遮挡和外观变化约束的同时控制递归计算量。

给定某一个MPSSTR模块的模板分支输入特征$\boldsymbol{U}_Z'\in\mathbb{R}^{T\times S\times C\times h\times w}$和搜索分支输入特征$\boldsymbol{U}_X'\in\mathbb{R}^{S\times C\times h\times w}$,其中$h$和$w$分别表示某一阶段特征的高度和宽度,首先经过线性层和脉冲神经元层获取键$\boldsymbol{K}_s$和值$\boldsymbol{V}_s$:

$$\begin{cases}\boldsymbol{K}_s=\text{SN}(\text{Linear}_{\boldsymbol{K}}(\text{SN}(\boldsymbol{U}_Z')))\\ \boldsymbol{V}_s=\text{SN}(\text{Linear}_{\boldsymbol{V}}(\text{SN}(\boldsymbol{U}_Z')))\end{cases}\tag{11}$$

而查询特征的获取方式为:

$$\begin{cases}\boldsymbol{Q}^0=\text{Linear}_{\boldsymbol{Q}}(\text{SN}(\boldsymbol{U}_X'))\\ \boldsymbol{Q}_s^0=\text{SN}(\text{Expand}(\boldsymbol{Q}^0))\end{cases}\tag{12}$$

式中,Expand$(\cdot)$表示沿时间维度进行扩展,即$\mathbb{R}^{S\times C\times h\times w}\rightarrow\mathbb{R}^{T\times S\times C\times h\times w}$;$\boldsymbol{Q}_s^0$表示递归交互过程中初始查询输入。随后构建紧凑、可复用的历史帧记忆缓存$\boldsymbol{M}=(\boldsymbol{K}_s)^{\text{T}}\otimes\boldsymbol{V}_s$,其中,$\otimes$表示矩阵乘法,$\boldsymbol{M}\in\mathbb{R}^{T\times S\times C\times C}$。考虑到跟踪过程中动态更新的历史帧模板记忆缓存可能受到遮挡、光照变化以及定位误差等因素的影响,从而引发噪声累积和表征偏移,导致历史记忆退化。为了缓解

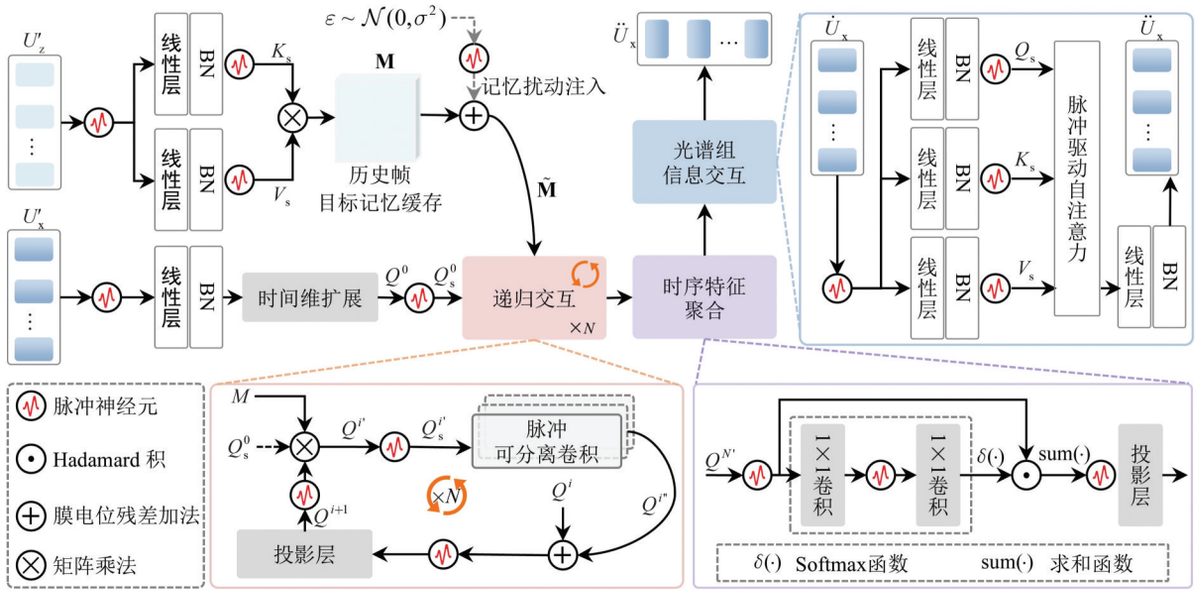


图2 记忆扰动增强的光谱-空间-时序检索模块

Fig. 2 Perturbation-Enhanced Spectral-Spatial-Temporal Retrieval Module

这一问题，本文引入记忆扰动机制，在训练阶段向记忆缓存 $\boldsymbol{M}$ 中注入扰动：

$$\tilde{\boldsymbol{M}}=\boldsymbol{M}+\mathrm{SN}(\gamma\varepsilon) \tag{13}$$

式中，ε 表示高斯噪声扰动，$\varepsilon\sim\mathrm{N}(0,\sigma^2)$；$\gamma$ 表示扰动系数，用来控制扰动强度。对于递归交互，首先利用记忆矩阵 $\tilde{\boldsymbol{M}}$ 对当前查询特征 $\boldsymbol{Q}_s^i$ 进行记忆引导，获取目标响应特征 $\boldsymbol{Q}^{i'}$。然后，考虑到不同时间步的特征响应存在差异性，采用时间步独立的脉冲可分离卷积进行特征增强和细化，得到增强特征 $\boldsymbol{Q}^{i''}$。最后，将增强特征 $\boldsymbol{Q}^{i''}$ 与当前未经过脉冲化的查询特征 $\boldsymbol{Q}^i$ 相加，获得更新后的查询特征 $\boldsymbol{Q}^{i+1}$，并将其用于下一轮交互。处理流程可表示为：

$$\begin{cases}\boldsymbol{Q}_s^{i'}=\mathrm{SN}(\boldsymbol{Q}^i)\\ \boldsymbol{Q}^{i'}=\boldsymbol{Q}_s^i\otimes\tilde{\boldsymbol{M}}\times k\end{cases} \tag{14}$$

$$\boldsymbol{Q}^{i''}=\Phi_{\mathrm{sep}}^{t}(\boldsymbol{Q}_s^{i'}(t)) \tag{15}$$

$$\begin{cases}\boldsymbol{Q}^{i+1}=\mathrm{Proj}(\mathrm{SN}(\boldsymbol{Q}^i+\boldsymbol{Q}^{i''}))\\ \boldsymbol{Q}_s^{i+1}=\mathrm{SN}(\boldsymbol{Q}^{i+1})\end{cases} \tag{16}$$

式中，i 表示第 i 次迭代，$i\in\{1,2,\cdots,N\}$；$\mathrm{Proj}(\cdot)$ 表示投影层。上述过程迭代 N 次，使模板目标先验能够逐步校正搜索特征中的目标表征。递归交互迭代完成后，通过时序信息聚合模块对不同时刻 t 的目标特征进行融合，充分整合历史时序线索，并获得最终的目标表征。给定递归交互的最终输出 $\boldsymbol{Q}^{N'}\in\mathbb{R}^{T\times S\times C\times hw}$，则时序聚合处理过程为：

$$\boldsymbol{W}=\delta(\mathrm{Conv}(\mathrm{SN}(\mathrm{Conv}(\mathrm{SN}(\boldsymbol{Q}^{N'}))))) \tag{17}$$

$$\begin{cases}\dot{\boldsymbol{U}}_X=\mathrm{Proj}(\mathrm{SN}(\mathrm{sum}(\boldsymbol{W}(t)\odot\boldsymbol{Q}^{N'}(t))))\\ \ddot{\boldsymbol{U}}_X=\mathrm{Linear}(\mathrm{SN}(\Phi_{\mathrm{self_att}}(\mathrm{SN}(\dot{\boldsymbol{U}}_X))))\end{cases} \tag{18}$$

式中，$\boldsymbol{W}(t)$ 表示 t 时刻的权重，$\boldsymbol{W}\in\mathbb{R}^{T\times S\times C\times 1}$；$\delta(\cdot)$ 表示 Softmax 函数；$\mathrm{sum}(\cdot)$ 表示求和函数；$\odot$ 表示逐元素相乘；$\dot{\boldsymbol{U}}_X\in\mathbb{R}^{S\times C\times h\times w}$ 表示经过时序聚合后的特征；$\ddot{\boldsymbol{U}}_X$ 表示增强的光谱特征。

本文通过脉冲驱动自注意力沿光谱组维度进行交互建模，充分挖掘不同光谱组之间的互补信息，获得 $\ddot{\boldsymbol{U}}_X$。最后，将 $\ddot{\boldsymbol{U}}_X$ 与搜索分支特征 $\boldsymbol{U}_X$ 相加得 $\breve{\boldsymbol{U}}_X$，并将其输入到后续网络模块进行进一步的特征编码。

1.2.3 脉冲驱动的光谱门控融合

为实现不同光谱组信息的自适应聚合，本文设计SDSG模块，如图3所示。该模块以搜索分支第4阶段的输出特征 $\breve{\boldsymbol{U}}_X^{\mathrm{stage_4}}\in\mathbb{R}^{S\times C\times h\times w}$ 作为输入，通过SDSG模块，自适应地确定各光谱组的融合权重。

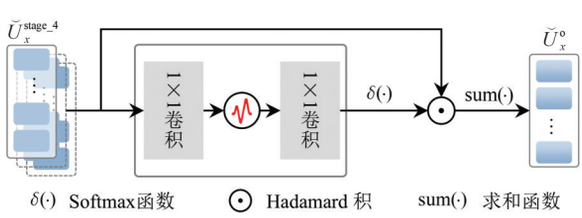


图3 脉冲驱动光谱门控融合模块

Fig. 3 Spike-Driven Spectral Gated Fusion Module

首先通过两个1×1卷积和脉冲神经元层进行特征变换，然后沿空间维度执行Softmax归一化，得到每个光谱组在不同空间位置上的响应权重 $\bar{\boldsymbol{W}}$：

$$\bar{\boldsymbol{W}}=\delta(\mathrm{Conv}_{1\times1}(\mathrm{SN}(\mathrm{Conv}_{1\times1}(\breve{\boldsymbol{U}}_X^{\mathrm{stage_4}})))) \tag{19}$$

最后，利用权重 $\bar{\boldsymbol{W}}$ 对不同光谱组特征进行加

权聚合得到特征 $\boldsymbol{U}_X^o \in \mathbf{R}^{C\times h\times w}$：

$$\boldsymbol{U}_X^o = \mathrm{sum}\left(\bar{\boldsymbol{W}} \odot \breve{\boldsymbol{U}}_X^{\mathrm{stage_4}}\right) \tag{20}$$

通过上述门控融合过程，不同通道能够自适应选择更具判别力的光谱响应，从而避免等权融合对有效光谱信息的稀释，并降低噪声光谱组对目标表征的干扰。最终，将融合特征 $\boldsymbol{U}_X^o$ 输入脉冲驱动跟踪头进行目标状态预测。

1.2.4　跟踪头和损失函数

本文采用与 SpikeTrack[16] 相同的跟踪头，并未对其结构进行额外修改。将经过光谱门控融合后的搜索特征输入 3 个并行预测分支，每个分支均由若干卷积层、BN 层和 NI-LIF 脉冲神经元堆叠构成，其中最后一层仅保留卷积操作，不引入 BN 和 NI-LIF 神经元。3 个预测分支分别用于预测目标中心置信图、由特征下采样引起的中心位置偏移量和归一化的目标边界框宽高。

模型的训练损失由分类损失和边界框回归损失组成，其中分类损失 L_{class} 采用加权焦点损失[22]，边界框回归损失则由 L_1 损失和广义交并比（generalized intersection over union，GIoU）损失[23] L_{GIoU} 联合构成。因此，模型的总体训练目标表示为：

$$L = L_{\mathrm{class}} + \lambda_1 L_1 + \lambda_{\mathrm{GIoU}} L_{\mathrm{GIoU}} \tag{21}$$

遵循 OSTrack[5] 和 SpikeTrack[16] 的设置，令系数 $\lambda_1 = 5.0$ 和 $\lambda_{\mathrm{GIoU}} = 2.0$。

2　实验与结果分析

2.1　实验数据集和评价指标

MUST[10] 是一个面向多光谱无人机单目标跟踪的大规模基准数据集，由搭载多光谱成像系统的无人机平台采集。该数据集包含 250 段视频序列，其中 160 段用于训练，90 段用于测试。每帧图像提供 8 个空间配准的光谱波段，覆盖 390～950 nm 的波长范围，图像分辨率为 1 200×900 像素。数据集涵盖完全遮挡、部分遮挡、背景杂波、低分辨率、相似目标、尺度变化、运动模糊、快速运动、颜色相近、目标出视野、光照变化以及相机运动等 12 种典型挑战。为全面评价跟踪器性能，本文采用成功率（success rate，SR）、精度（precision，P）、归一化精度（normalized precision，NP）以及成功率曲线下面积（area under the curve，AUC）作为评价指标。其中 $E_{\mathrm{SR}} = \frac{1}{T}\sum_{t=1}^{T} I(\mathrm{IoU}_t > \tau_{\mathrm{SR}})$，$I(\cdot)$ 为指示函数，表示预测边界框与真实边界框之间交并比（intersection over union，IoU）超过阈值 τ_{SR} 时的帧比例，本文选取 $\mathrm{SR}_{0.5}$（$\tau_{\mathrm{SR}} = 0.5$）和 $\mathrm{SR}_{0.75}$（$\tau_{\mathrm{SR}} = 0.75$）两个指标；$E_{\mathrm{P}} = \frac{1}{T}\sum_{t=1}^{T} I(\| \boldsymbol{c}_t - \boldsymbol{c}_t^{\mathrm{gt}} \|_2 < \tau_p)$ 表示预测目标中心 $\boldsymbol{c}_t$ 与真值框中心 $\boldsymbol{c}_t^{\mathrm{gt}}$ 误差小于阈值 τ_p 时的帧比例；$E_{NP} = \frac{1}{T}\sum_{t=1}^{T} I(\| \boldsymbol{W}_t(\boldsymbol{c}_t - \boldsymbol{c}_t^{\mathrm{gt}}) \|_2 < \tau_n)$，$\mathbf{W}_t = \mathrm{diag}(1/w_t^{\mathrm{gt}}, 1/h_t^{\mathrm{gt}})$，本文中 $\tau_p = 20$，$\tau_n = 0.5$。

2.2　实验设置

本文算法使用 PyTorch 2.12.0 实现，在 8 张 NVIDIA RTX 5090 GPU 上进行训练，批大小设置为 32。在训练阶段，采用 SpikeTrack[16] 在可见光跟踪数据集上的预训练权重初始化模型，并使用 AdamW[24] 优化器进行参数优化，权重衰减系数设为 1×10^{-4}。历史模板采用三模板缓存机制，缓存上限设为 3 个模板。模型在初始帧建立目标模板，随后每处理一帧搜索图像，便通过模板分支与搜索分支的递归交互更新历史模板表征，模板更新间隔为 1 帧，并将更新结果用于下一帧跟踪。训练与推理阶段采用相同的模板更新方式，不采用每隔若干帧重新采样模板的离散更新策略，也不额外引入显式的队列刷新机制。模型总共训练 30 个轮次，每个轮次包含 8 000 个样本对，并在第 20 个轮次后将学习率衰减至初始值的 0.1 倍。模板图像和搜索图像尺寸均设置为 384×384。

2.3　对比实验

表 1 比较了本文方法与多种代表性跟踪器的性能。由表 1 可知，本文方法在 AUC 和 $\mathrm{SR}_{0.75}$ 指标上取得最优结果，在 $\mathrm{SR}_{0.5}$、P 和 NP 指标上也表现出较强竞争力。与 UNTrack[10] 相比，本文方法在 AUC、$\mathrm{SR}_{0.5}$、$\mathrm{SR}_{0.75}$、P 和 NP 上分别提高 6.9、5.9、17.7、4.3 和 7.4 个百分点，其中 $\mathrm{SR}_{0.75}$ 的增益最为明显，说明该方法能够在较高 IoU 阈值下保持更准确的目标定位和边界框预测。与脉冲驱动跟踪器 SpikeTrack[16] 相比，上述指标分别提高 4.2、4.7、5.6、4.2 和 5.0 个百分点，表明所设计的多模态信息建模机制能够有效改善跟踪性能。相较于 HEATTrack[34]，所提方法的 AUC 和 $\mathrm{SR}_{0.75}$ 分别提高 0.6 和 5.4 个百分点，但 $\mathrm{SR}_{0.5}$、P 和 NP 分别降低 1.0、1.9 和 0.1 个百分点，这表明本文方法虽然在较高 IoU 阈值下具有更好的定位精度，但目标中心定位能力仍有进一步提升空间。与 SpecTrack[35] 相比，本文方法在 AUC、$\mathrm{SR}_{0.5}$、$\mathrm{SR}_{0.75}$、P 和 NP 上分别提高 1.4、1.9、2.9、0.9 和 2.7 个百分点，其中 $\mathrm{SR}_{0.75}$ 和 NP 的增益尤为突出。

表1 不同方法在MUST数据集上的跟踪性能对比/%

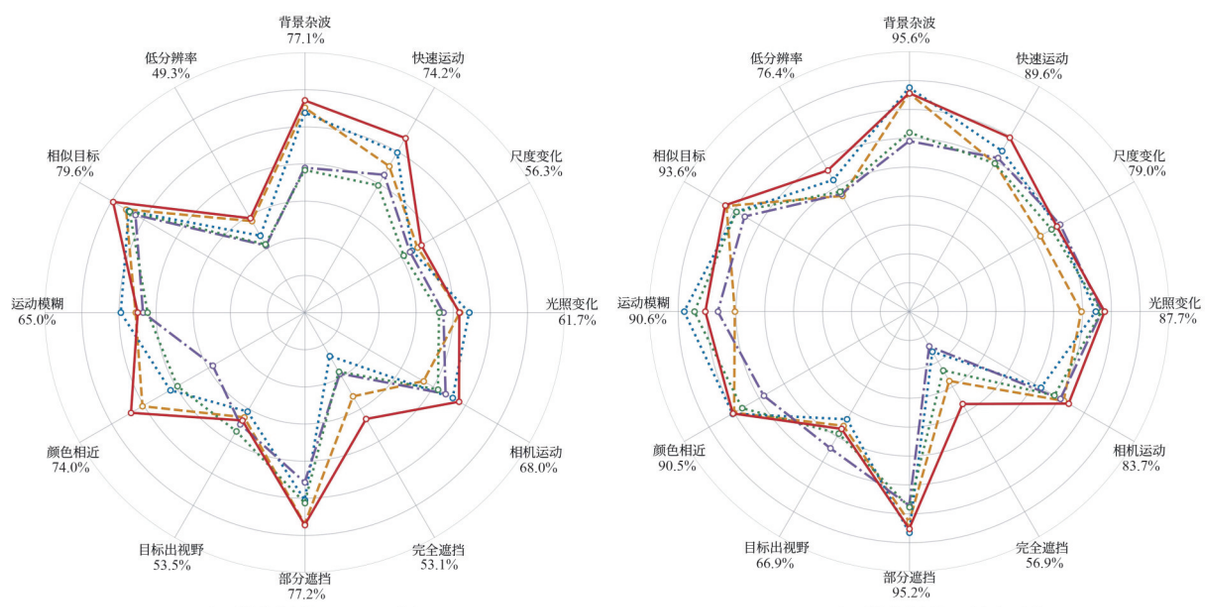
Table 1 Accuracy Comparison of Different Methods on the MUST Dataset/%

方法	AUC	$SR_{0.5}$	$SR_{0.75}$	P	NP
SiamRPN[25]	38.9	49.5	22.3	56.7	50.1
SiamMask[26]	39.1	51.0	25.3	56.1	52.8
DiMP[6]	47.3	60.7	31.6	69.8	57.1
TaMOs[27]	20.4	24.5	12.0	31.0	23.8
ToMP[28]	30.3	39.1	14.5	46.1	40.5
SGDViT[29]	38.1	50.2	22.8	54.4	52.1
$OSTrack_{256}$[5]	39.9	50.5	27.4	55.9	51.2
$OSTrack_{384}$[5]	44.5	56.7	27.1	63.9	56.8
ODTrack[30]	46.3	58.4	28.1	67.6	60.4
ZoomTrack[31]	44.0	56.2	25.2	63.5	55.9
$OSTrack_{256}$*[5]	52.3	64.8	36.3	67.5	65.2
$OSTrack_{384}$*[5]	55.1	69.6	44.1	73.3	68.8
ODTrack*[30]	52.7	67.3	38.8	71.3	66.2
ZoomTrack*[31]	53.3	67.7	40.4	72.3	66.7
UNTrack[10]	59.7	75.8	46.2	79.2	74.8
HEATTrack[34]	66.0	82.7	58.5	85.4	82.3
SpecTrack[35]	65.2	79.8	61.0	82.6	79.5
SpikeTrack[16]	62.4	77.0	58.3	79.3	77.2
本文方法	**66.6**	**81.7**	**63.9**	**83.5**	**82.2**

注：*表示使用插值重建的参数进行训练，更多细节请参考UNTrack[10]。

图4给出了本文方法与SpikeTrack[16]、UNTrack[10]、OSTrack[5]和ZoomTrack[31]在不同挑战属性下的AUC和P对比。总体而言，本文方法在大多数属性上均取得较优或具有竞争力的表现，在快速运动、完全遮挡、部分遮挡、相似目标和颜色相近等场景下优势尤为明显。上述结果表明，本文方法能够有效挖掘多光谱信息的互补性，并充分利用历史目标线索，从而提升复杂场景下的目标定位精度与跟踪鲁棒性。

进一步地，表2比较了不同跟踪器的参数量与推理能耗，其中推理能耗统计的更多细节可参考SpikeTrack[16]。本文方法的参数量为48.6 MB，能耗为32.4 mJ。与此同时，AUC和P分别由62.4%和79.3%提升至66.6%和83.5%，说明新增的历史检索、跨谱交互与门控融合在可控开销下带来了稳定的精度收益。与SpikeTrack[16]相比，本文方法由于引入MPSSTR和SDSG模块，参数量增加了11.8 MB，能耗增加了5.1 mJ，但仍保持了较低的推理能耗。与CSWinTT[32]、$OSTrack_{256}$[5]、STARK[7]和TransT[33]相比，本文方法的能耗分别降低了约57.0%、67.2%、35.3%和56.9%。此外，本文方法参数量较$OSTrack_{256}$[5]减少43.5 MB。特别地，与多光谱无人机单目标跟踪算法UNTrack[10]相比，本文在参数量和能耗分别降低了64.0 MB和556.5 mJ。结合表1的性能结果可知，本文方法以较低的能耗增量换取了显著的精度提升。表2可见，完整模型相较SpikeTrack[16]增加11.8 MB参数和5.1 mJ单次推理能耗，但AUC和P均提升了4.2个百分点。与其余方法相比，推理能耗仍处于较低水平。结合表3中的推理速度为18.51帧/s，本文方法在精度、能耗和实时性之间取得了较好的平衡。



(a) 不同挑战属性下AUC对比　　(b) 不同挑战属性下精度对比

图4 不同挑战属性下的性能比较

Fig. 4 AUC and Precision Under Different Challenge Attributes

表 2　不同方法的参数量和能耗

Table 2　Parameters and Power of Different Methods

对比项	SpikeTrack[16]	CSWinTT[32]	$OSTrack_{256}$[5]	STARK[7]	TransT[33]	UNTrack[10]	本文方法
能耗/mJ	27.3	75.4	98.9	50.1	75.2	588.9	32.4
参数量/MB	36.8	25.1	92.1	23.5	17.9	112.6	48.6

2.4　消融实验

2.4.1　模块有效性验证

表 3 展示了不同组件组合在 MUST 数据集上的消融实验结果,对数据进行分析可知:(1)对比变体 1 与变体 7,加入记忆扰动策略后,AUC 和 P 分别提升 2.1 和 2.2 个百分点。该策略通过模拟模板动态更新过程中可能出现的噪声累积和表征偏移,使模型能够从不可靠历史记忆中筛选稳定的目标线索,从而增强对背景干扰和历史信息退化的适应能力。(2)对比变体 2 与变体 7,加入 SDSG 模块后,AUC 和 P 分别提升 3.0 和 3.5 个百分点。SDSG 通过自适应门控聚合不同光谱组特征,增强具有判别性的光谱响应,并抑制冗余信息和背景干扰。(3)对比变体 3 与变体 7,引入光谱组信息交互后,AUC 和 P 分别提升 1.3 和 1.6 个百分点,表明显式建模不同光谱组之间的信息传递有助于利用其互补线索,增强目标多光谱特征表征。(4)MPSSTR 仅引入记忆扰动增强模块的变体 4,AUC 和 P 较基线模型分别提高 1.3 和 0.6 个百分点。该结果表明,记忆扰动增强能够改善历史信息筛选过程,并提高模型对复杂背景干扰的适应能力。(5)MPSSTR 仅引入光谱组信息交互的变体 5,AUC 和 P 较基线模型分别提高 3.3 和 3.2 个百分点。这说明跨光谱组交互能够促进不同光谱特征之间的互补信息传递,缓解单一光谱信息不足引起的定位偏差。(6)仅引入 SDSG 模块的变体 6,AUC 和 P 较基线模型分别提高 2.7 和 2.4 个百分点,说明自适应门控融合本身即可改善目标表征质量。与其他组件联合使用时,该模块能够进一步发挥协同作用。(7)完整模型,即变体 7,同时采用记忆扰动增强、光谱组信息交互和 SDSG 模块,AUC 和 P 分别达到 66.6% 和 83.5%,较基线模型均提高 4.2 个百分点。各模块单独引入时均可带来稳定增益,联合使用后取得最优结果,表明三者在历史记忆建模、跨光谱组交互和自适应融合方面具有互补作用。完整模型的推理速度为 18.51 帧/s,较基线模型下降 1.88 帧/s。尽管参数量和能耗有所增加,但性能提升较为明显,额外推理开销处于可接受范围。

表 3　本文提出的不同组件在 MUST 数据集上的消融实验结果

Table 3　Ablation Results of the Proposed Components on the MUST Dataset

变体	MPSSTR		SDSG	AUC/%	P/%	参数量/MB	能耗/mJ	推理速度/(帧·s^{-1})
	记忆扰动增强	光谱组信息交互						
基线模型				62.4	79.3	36.8	27.3	20.39
1		√	√	64.5	81.3	48.6	32.4	18.77
2	√	√		63.6	80.0	48.4	32.4	18.56
3	√		√	65.3	81.9	46.9	31.6	18.67
4	√			63.7	79.9	36.8	27.3	20.20
5		√		65.7	82.5	48.4	32.4	19.17
6			√	65.1	81.7	46.9	31.6	19.73
7	√	√	√	66.6	83.5	48.6	32.4	18.51

2.4.2　不同融合策略的影响

表 4 比较了不同光谱组融合策略在 MUST 数据集上的性能。求和融合与取均值融合均采用等权方式聚合各光谱组特征,两种策略的 AUC/P 分别达到 65.0%/81.9% 和 65.0%/81.7%。由于两种策略未区分不同光谱组的有效性,低质量或冗余光谱响应也会以相同权重参与融合,从而干扰目标表征。相比之下,SDSG 根据各光谱组的响应自适应生成融合权重,使 AUC 和 P 分别提升至 66.6% 和 83.5%。与性能次优的求和融合相比,AUC 和 P 均提高 1.6 个百分点,表明性能增益主要来自光谱组权重的自适应分配,而非简单增加特征响应。SDSG 能够强化目标相关的关键光谱信息,同时抑制低质量和冗余响应,从而形成更加稳定、判别性更强的目标表征。

此外,SDSG 的参数仅增加约 0.2 MB,能耗反而略低于等权融合策略,说明脉冲驱动的稀疏事件计算能够抵消部分门控操作带来的额外开销。

综合性能与计算代价可以看出,SDSG在提升融合质量的同时,保持了较好的参数和能耗效率。

表4 不同光谱融合策略的结果对比

Table 4 Comparison Results of Different Spectral Fusion Strategies

融合策略	AUC/%	P/%	参数量/MB	能耗/mJ
求和	65.0	81.9	48.4	32.44
取均值	65.0	81.7	48.4	32.44
SDSG	66.6	83.5	48.6	32.41

2.4.3 扰动系数的影响

表5分析了不同扰动系数γ对模型性能的影响。当扰动系数$\gamma=0$时,历史记忆缓存未注入扰动。随着扰动系数由0增大至0.001,模型性能逐步提升,并在$\gamma=0.001$时达到最优,AUC和P分别为66.6%和83.5%,较未注入扰动时提高2.1和2.2个百分点。扰动系数仅控制训练阶段注入历史记忆缓存的噪声强度,从而在训练阶段模拟多光谱无人机单目标跟踪中历史模板可能出现的不可靠状态。在实际跟踪过程中,短时遮挡、相似目标干扰、低照度、谱段响应波动以及前序帧定位误差,都可能使历史模板混入背景噪声或产生表征偏移。若训练阶段始终依赖未经扰动的历史记忆,模型难以学习识别和抑制模板退化带来的影响。当搜索帧受到背景干扰或不同谱段的成像质量存在差异时,历史记忆中的潜在偏差可能通过检索过程传递至当前目标表征,并进一步干扰光谱组交互与特征融合。将扰动系数设置为0.001后,训练过程中引入轻微的历史记忆偏移,使检索模块能够在接近实际退化的条件下学习筛选稳定的目标线索。经过筛选的历史信息与光谱组交互及SDSG门控融合协同作用,从而减少历史噪声传播,增强模型对不可靠历史信息和目标外观变化的适应能力。

表5 不同扰动系数对模型性能的影响

Table 5 Comparison Results of Different Perturbation Coefficients

γ	AUC/%	P/%	参数量/MB	能耗/mJ
0	64.5	81.3	48.6	32.41
0.1	65.0	81.5	48.6	32.42
0.01	65.9	82.6	48.6	32.40
0.001	66.6	83.5	48.6	32.41
0.000 1	65.5	82.4	48.6	32.21

当扰动系数增大至0.01或0.1时,过强扰动会破坏历史记忆中的目标结构和光谱响应,使检索模块难以区分有效目标线索与背景噪声,导致跟踪性能下降。因此,扰动强度需要在记忆可靠性与鲁棒性之间取得平衡,$\gamma=0.001$能够在增强退化适应性的同时,较好地保留历史记忆中的目标信息。不同扰动系数下模型参数量均为48.6 MB,能耗保持不变,表明该策略无需增加模型容量即可改善跟踪性能。

2.4.4 递归次数N的影响

表6分析了不同递归次数N对模型性能的影响。当$N=1$时,模型取得最佳性能。递归次数决定了当前搜索特征与历史记忆表征之间进行交互和更新的轮数。当递归次数为0时,搜索特征无法充分利用历史模板中积累的目标先验,模型主要依赖当前帧的局部外观和光谱响应进行定位,在遮挡、相似目标或低分辨率目标场景下,当前帧线索容易变得模糊,因而AUC和P明显低于引入递归检索后的结果。当递归次数为1时,历史记忆能够作为目标提示对搜索特征进行一次有效校正,使模型在当前帧中重新强调与历史目标一致的空间位置、外观结构和光谱响应,同时避免过多重复交互带来的噪声累积,因此取得最佳性能。当递归次数继续增加至2或3时,模型虽然能够反复访问历史记忆,但这些交互并不会引入新的观测信息,反而可能持续放大历史模板中由遮挡、目标漂移或尺度变化引起的偏差,使搜索特征逐渐向不可靠记忆靠拢。同时,多轮递归会重复执行记忆引导、脉冲可分离卷积和投影更新等操作,增加推理能耗,并可能造成目标判别特征被过度平滑或冗余背景响应被累积。因此,递归次数并非越多越好,单次递归在历史先验利用、噪声抑制和计算开销之间取得了较合理的平衡,本文最终将递归次数设置为1。

表6 不同递归次数对模型性能的影响

Table 6 Comparison Results of Different Recursive Iterations

N	AUC/%	P/%	参数量/MB	能耗/mJ
0	59.1	75.4	48.6	27.93
1	66.6	83.5	48.6	32.41
2	63.4	80.8	48.6	34.47
3	60.3	77.1	48.6	34.29

2.5 可视化分析

2.5.1 注意力热图分析

图5展示了本文方法与基线模型在多个跟踪序列中的注意力响应结果。从图5可以看出,基线模型在跟踪过程中容易在背景区域或邻近干扰物上产生错误的响应,导致目标表征受损。

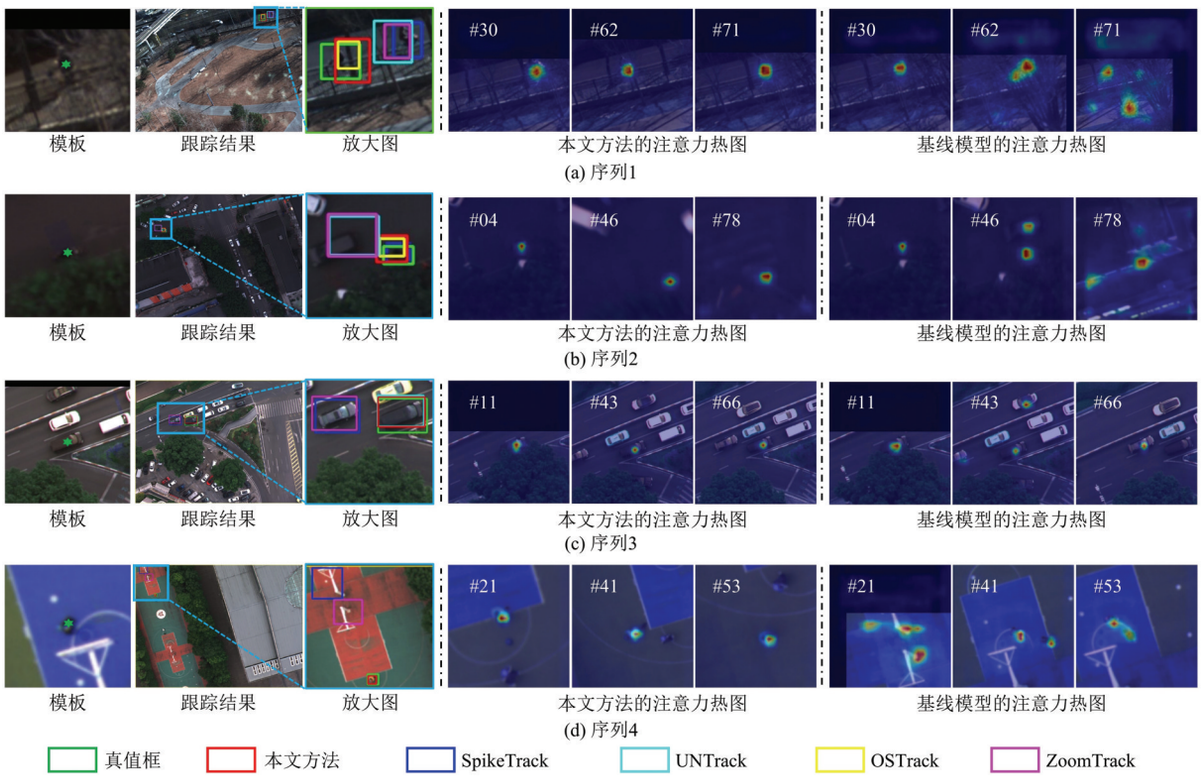


图 5　注意力热图可视化

Fig. 5　Visualization Results of Attention Heat Maps

相比之下,本文方法在不同帧中均能够将高响应区域稳定地聚焦于目标位置,有效抑制无关区域的干扰,从而实现更加鲁棒的目标跟踪。

2.5.2　定性结果分析

图 6 展示了本文方法与 SpikeTrack[16]、UNTrack[10]、OSTrack[5]和 ZoomTrack[31]等方法在完全遮挡、尺寸变化、低分辨率、运动模糊、相似目标和颜色相近等挑战性场景下的定性结果对比。从图 6 可以看出,在完全遮挡和低分辨率场景中,SpikeTrack[16]和 UNTrack[10]因目标外观线索缺失或退化而出现不同程度的定位偏差。在相似目标和颜色相近的场景中,SpikeTrack[16]、UNTrack[10]、OSTrack[5]和 ZoomTrack[31]容易受到干扰目标及相似背景的影响,导致预测框发生漂移。相比之下,本文方法在上述各类挑战场景中均能获得与真值框更为一致的预测结果,表现出更强的定位精度和跟踪鲁棒性。

2.5.3　失败案例与局限性分析

图 7 给出了本文方法在 MUST 数据集上的 4 类典型跟踪失败案例,其中红色框表示预测结果,绿色框表示真实目标位置。图 7(a)为颜色相近场景,目标与周围背景具有相似的颜色和纹理,判别线索较弱,预测框逐渐向相似区域偏移,表明仅依靠光谱信息难以有效区分目标与背景;图 7(b)为背景杂波场景,道路标线、阴影及复杂纹理产生较强响应,与目标形成竞争,导致预测框向背景区域漂移;图 7(c)为相似物体场景,邻近物体与目标具有相近的外观和运动状态,二者响应难以区分,预测框最终转移至干扰物体,发生目标切换;图 7(d)为相机运动场景,无人机运动造成目标位置快速变化,并伴随运动模糊,使历史目标信息与当前帧之间产生较大偏差,预测框出现明显滞后。

上述结果表明,本文方法在目标判别线索减弱或快速变化时仍存在一定局限。历史模板可能受到相似背景或干扰物的错误响应影响,而现有方法对目标运动及相机运动的显式建模仍显不足,后续可引入运动感知时序建模和相机运动补偿机制,以进一步提高复杂条件下的跟踪稳定性。

图 8 给出了本文方法在 MUST 数据集两类典型挑战场景下的中心位置误差(center location error, CLE)曲线。其中,红色曲线表示各帧 CLE,绿色虚线表示 20 像素的有效跟踪阈值。整体来看,CLE 在大多数帧中保持较低水平,表明本文方法具有较稳定的目标定位能力。图 8(a)序列 1 为相似物体干扰场景。前期目标能够稳定跟踪,CLE 基本低于阈值。当外观相近的干扰物出现后,曲线迅速升高并超过阈值,表明预测位

置发生明显偏移。结合放大区域可见,相似物体削弱了目标与背景的判别差异,使历史模板中的目标先验被错误匹配至干扰区域。随着干扰减弱,CLE重新回落,说明模型仍具备一定的重新定位能力。图8(b)序列2为低分辨率目标场景。由于目标占据像素较少,边缘、纹理和形状等细粒度信息不足,CLE整体波动更为明显。在目标尺寸进一步减小的帧段,曲线出现超过阈值的尖峰,表明目标表征退化使定位更易受到背景干扰。当目标线索恢复后,CLE随之回落。

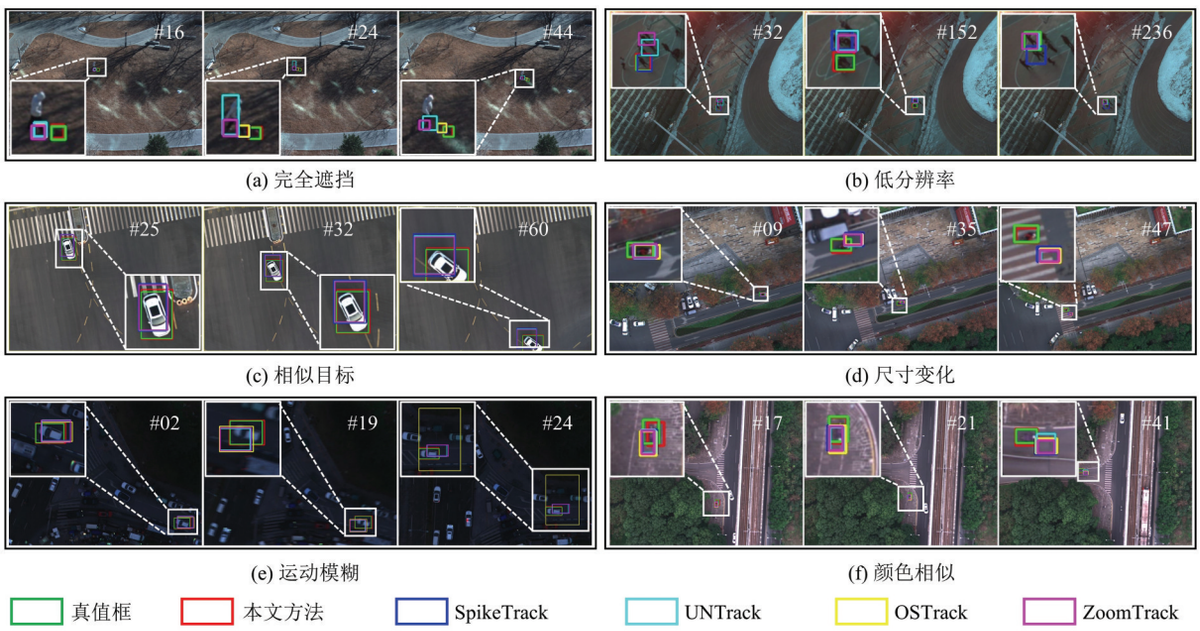


图6 MUST数据集不同挑战属性下定性结果比较

Fig. 6 Qualitative Comparisons Under Representative Challenging Attributes on the MUST Dataset

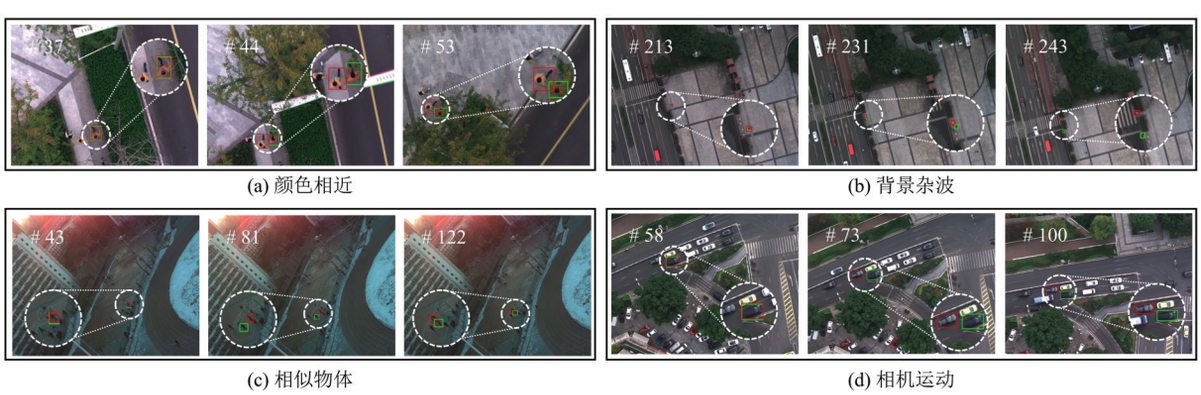


图7 MUST数据集不同挑战属性下的代表性跟踪失败案例

Fig. 7 Representative Tracking Failure Cases Under Different Challenging Attributes on the MUST Dataset

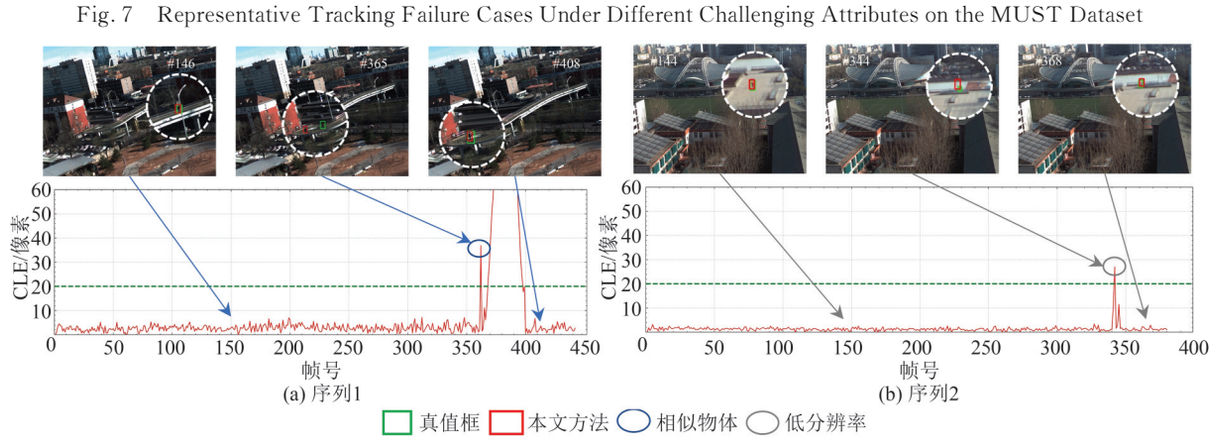


图8 MUST数据集不同挑战场景下的中心位置误差曲线

Fig. 8 Center Location Error Curves Under Different Challenging Scenarios on the MUST Dataset

总体而言，本文方法在大多数帧中能够保持稳定定位，但相似物体和低分辨率条件仍可能引起短时定位误差。上述现象表明，现有方法对相似目标的区分和弱目标表征仍存在不足，后续可结合运动约束、状态预测及小目标增强机制进一步提高跟踪稳定性。

3　结　语

本文提出了一种脉冲驱动光谱-空间-时序建模的多光谱无人机单目标跟踪算法，旨在保证跟踪性能的同时实现低功耗视觉感知。所提算法主要包含 MPSSTR 模块和 SDSG 模块。其中，MPSSTR 模块通过构建紧凑的历史目标记忆缓存，引导搜索特征与历史记忆进行递归交互，从光谱、空间和时序 3 个维度充分挖掘目标相关线索，并通过记忆扰动机制增强模型对历史记忆退化的鲁棒性；SDSG 模块则对不同光谱组特征的重要性进行自适应评估与加权融合，从而增强目标相关的判别性光谱信息并抑制冗余干扰。

在 MUST 数据集上将本文方法与现有方法进行了较全面的比较。实验结果表明，相较于基线模型，本文方法在 AUC、$SR_{0.5}$、$SR_{0.75}$、P 和 NP 上分别提高了 4.2、4.7、5.6、4.2 和 5.0 个百分点。此外，大量消融实验和可视化分析进一步验证了所提各组件与策略的有效性。

参　考　文　献

[1] WU X, LI W, HONG D F, et al. Deep Learning for Unmanned Aerial Vehicle-Based Object Detection and Tracking: A Survey[J]. *IEEE Geoscience and Remote Sensing Magazine*, 2022, 10(1): 91-124.

[2] SUN N Y, ZHAO J, SHI Q, et al. Moving Target Tracking by Unmanned Aerial Vehicle: A Survey and Taxonomy[J]. *IEEE Transactions on Industrial Informatics*, 2024, 20(5): 7056-7068.

[3] 肖云，曹丹，李成龙，等. 基于高空无人机平台的多模态跟踪数据集[J]. 中国图象图形学报，2025，30(2)：361-374.
XIAO Yun, CAO Dan, LI Chenglong, et al. A Benchmark Dataset for High-Altitude UAV Multimodal Tracking[J]. *Journal of Image and Graphics*, 2025, 30(2): 361-374.

[4] 许国伟，吴亮，翟晗，等. 耦合置换注意力和可形变卷积的输电线金具目标检测[J]. 武汉大学学报(信息科学版)，2026，51(3)：544-556.
XU Guowei, WU Liang, ZHAI Han, et al. A Model Coupled Shuffle Attention and Deformable Convolution for Overhead Transmission Line Hardware Object Detection[J]. *Geomatics and Information Science of Wuhan University*, 2026, 51(3): 544-556.

[5] YE B T, CHANG H, MA B P, et al. Joint Feature Learning and Relation Modeling for Tracking: A One-Stream Framework[C]//European Conference on Computer Vision, Tel Aviv, Israel, 2022.

[6] BHAT G, DANELLJAN M, VAN GOOL L, et al. Learning Discriminative Model Prediction for Tracking[C]//IEEE/CVF International Conference on Computer Vision (ICCV), Seoul, South Korea, 2019.

[7] YAN B, PENG H W, FU J L, et al. Learning Spatiotemporal Transformer for Visual Tracking[C]//IEEE/CVF International Conference on Computer Vision (ICCV), Montreal, Canada, 2021.

[8] BIAN L H, WANG Z, ZHANG Y Z, et al. A Broadband Hyperspectral Image Sensor with High Spatiotemporal Resolution[J]. *Nature*, 2024, 635(8037): 73-81.

[9] QIN H L, XU T F, LIU P F, et al. DMSSN: Distilled Mixed Spectral-Spatial Network for Hyperspectral Salient Object Detection[J]. *IEEE Transactions on Geoscience and Remote Sensing*, 2024, 62: 1-18.

[10] QIN H L, XU T F, LI T H, et al. MUST: The First Dataset and Unified Framework for Multispectral UAV Single Object Tracking[C]//IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Nashville, USA, 2025.

[11] CHEN Y Z, YUAN Q Q, ZHONG B N, et al. CASS: A Context-Aware Memory Framework with Spectral-Spatial Modulation for UAV Multispectral Video Tracking[J]. *IEEE Transactions on Circuits and Systems for Video Technology*, 2026, DOI: 10.1109/TCSVT.2026.3708187.

[12] TENG X, ZHAO D, XIANG P, et al. UAV-Based Multispectral Object Tracking with Positive-Negative Prompt Mining Network[J]. *IEEE Transactions on Geoscience and Remote Sensing*, 2026, 64: 5515418.

[13] MAASS W. Networks of Spiking Neurons: The Third Generation of Neural Network Models[J]. *Neural Networks*, 1997, 10(9): 1659-1671.

[14] ROY K, JAISWAL A, PANDA P. Towards Spike-Based Machine Intelligence with Neuromorphic Computing[J]. *Nature*, 2019, 575(7784): 607-617.

[15] GERSTNER W, RITZ R, VAN HEMMEN J L.

Why Spikes? Hebbian Learning and Retrieval of Time-Resolved Excitation Patterns[J]. *Biological Cybernetics*, 1993, 69(5): 503-515.

[16] ZHANG Q Y, CHENG J J, MAO Q C, et al. SpikeTrack: A Spike-Driven Framework for Efficient Visual Tracking[C]//IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Denver, USA, 2026.

[17] SHAN Y M, REN Z B, WU H D, et al. SDTrack: A Baseline for Event-Based Tracking via Spiking Neural Networks[C]//IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Denver, USA, 2026.

[18] WANG Y, ZHANG J Q, SUN C Y, et al. SpikeTrack: High-Performance and Energy-Efficient Event-Based Object Tracking with Spiking Neural Network[C]//IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Denver, USA, 2026.

[19] YANG J J, FAN L W, ZHANG J P, et al. Fully Spiking Neural Networks for Unified Frame-Event Object Tracking[C]//Advances in Neural Information Processing Systems, Sydney, Australia, 2026.

[20] LEI Z X, YAO M, HU J K, et al. Spike2Former: Efficient Spiking Transformer for High-Performance Image Segmentation[C]//The 39th Annual AAAI Conference on Artificial Intelligence, Philadelphia, USA, 2025.

[21] YAO M, QIU X R, HU T X, et al. Scaling Spike-Driven Transformer with Efficient Spike Firing Approximation Training[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2025, 47(4): 2973-2990.

[22] LAW H, DENG J. CornerNet: Detecting Objects as Paired Keypoints[C]//European Conference on Computer Vision, Munich, Germany, 2018.

[23] REZATOFIGHI H, TSOI N, GWAK J, et al. Generalized Intersection Over Union: A Metric and a Loss for Bounding Box Regression[C]//IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Long Beach, USA, 2019.

[24] LOSHCHILOV I, HUTTER F. Decoupled Weight Decay Regularization[C]//International Conference on Learning Representations, Vancouver, Canada, 2018.

[25] LI B, YAN J J, WU W, et al. High Performance Visual Tracking with Siamese Region Proposal Network[C]//IEEE/CVF Conference on Computer Vision and Pattern Recognition, Salt Lake City, USA, 2018.

[26] WANG Q, ZHANG L, BERTINETTO L, et al. Fast Online Object Tracking and Segmentation: A Unifying Approach[C]//IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Long Beach, USA, 2019.

[27] MAYER C, DANELLJAN M, YANG M H, et al. Beyond SOT: Tracking Multiple Generic Objects at Once[C]//IEEE/CVF Winter Conference on Applications of Computer Vision (WACV), Waikoloa, USA, 2024.

[28] MAYER C, DANELLJAN M, BHAT G, et al. Transforming Model Prediction for Tracking[C]//IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), New Orleans, USA, 2022.

[29] YAO L L, FU C H, LI S H, et al. SGDViT: Saliency-Guided Dynamic Vision Transformer for UAV Tracking[C]//IEEE International Conference on Robotics and Automation (ICRA), London, UK, 2023.

[30] ZHENG Y Z, ZHONG B N, LIANG Q H, et al. ODTrack: Online Dense Temporal Token Learning for Visual Tracking[C]//AAAI Conference on Artificial Intelligence, Vancouver, Canada, 2024.

[31] KOU Y T, GAO J, LI B, et al. ZoomTrack: Target-Aware Non-Uniform Resizing for Efficient Visual Tracking[C]//Advances in Neural Information Processing Systems, New Orleans, USA, 2023.

[32] SONG Z K, YU J Q, CHEN Y P, et al. Transformer Tracking with Cyclic Shifting Window Attention[C]//IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), New Orleans, USA, 2022.

[33] CHEN X, YAN B, ZHU J W, et al. Transformer Tracking[C]//IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Nashville, USA, 2021.

[34] YAN L, DONG X Y, YANG J, et al. HEATTrack: History Evidence Injection and Target-Aware Prompting for Aerial Multispectral Single-Object Tracking[J]. *IEEE Transactions on Geoscience and Remote Sensing*, 2026, 64: 5406719.

[35] TAN X Y, QIN Y R, HU M J. SpecTrack: Spectral Prompt Guided Adaptive Experts for Multispectral Object Tracking[PP/OL]. [2026-07-07]. https://doi.org/10.48550/arXiv.2607.05988.