智能系统学报
CAAI TRANSACTIONS ON INTELLIGENT SYSTEMS

基于模板特征缓存与自适应注意力的无人机目标跟踪

陈泷, 丁锰, 石磊, 黎智辉, 许晓宇, 潘亦伦

引用本文:
陈泷, 丁锰, 石磊, 等. 基于模板特征缓存与自适应注意力的无人机目标跟踪[J]. 智能系统学报, 2026, 21(3): 688-700.
CHEN Long, DING Meng, SHI Lei, et al. UAV object tracking based on template feature buffer and adaptive attention[J]. *CAAI Transactions on Intelligent Systems*, 2026, 21(3): 688-700.

在线阅读 View online: https://dx.doi.org/10.11992/tis.202507009

DOI: 10.11992/tis.202507009
网络出版地址: https://link.cnki.net/urlid/23.1538.tp.20260204.1049.006

基于模板特征缓存与自适应注意力的无人机目标跟踪

陈泷[1], 丁锰[1,2], 石磊[3], 黎智辉[4], 许晓宇[5], 潘亦伦[1]
（1. 中国人民公安大学 侦查学院, 北京 100038; 2. 中国人民公安大学 公共安全行为科学实验室, 北京 100038; 3. 中国传媒大学 媒体融合与传播国家重点实验室, 北京 100024; 4. 公安部鉴定中心, 北京 100038; 5. 广东省证据材料司法鉴定工程技术研究中心, 广东 深圳 518033）

摘　要: 现有无人机目标跟踪方法通常仅保留最近帧的模板信息, 容易造成重要历史外观信息的丢失。为解决无人机目标跟踪中的目标信息丢失和历史信息有效利用问题, 本文提出一种基于模板特征缓存和自适应注意力的无人机目标跟踪方法。1) 设计模板特征缓存模块, 通过特征缓存区系统性保存多样化的历史目标外观信息, 有效解决传统方法中历史信息丢失的问题。2) 提出自适应注意力机制, 采用通道级注意力动态评估存储特征的重要性, 实现对历史模板信息的自适应加权利用。3) 采用即插即用架构, 可无缝集成到现有主流跟踪器中, 增强了算法的实用性和通用性, 并设计对称序列评估方法验证历史目标信息的有效保持和利用。实验结果表明, 所提方法在 5 种主流跟踪算法上均取得显著性能提升, 在 UAV123 数据集上 AUC 平均提升 2.34 百分点, 在 UAV20L 数据集上提升 7.24 百分点, 在扩展数据集 UAV123-L 和 UAV20L-L 上分别提升 4.11 和 9.14 百分点, 验证了方法的有效性和适用性。

关键词: 目标跟踪; 无人机; 模板特征缓存; 自适应注意力; 即插即用; 历史信息; 计算机视觉; 深度学习
中图分类号: TP391.4　　**文献标志码:** A　　**文章编号:** 1673−4785(2026)03−0688−13

中文引用格式: 陈泷, 丁锰, 石磊, 等. 基于模板特征缓存与自适应注意力的无人机目标跟踪 [J]. 智能系统学报, 2026, 21(3): 688–700.
英文引用格式: CHEN Long, DING Meng, SHI Lei, et al. UAV object tracking based on template feature buffer and adaptive attention [J]. CAAI transactions on intelligent systems, 2026, 21(3): 688–700.

UAV object tracking based on template feature buffer and adaptive attention

CHEN Long[1], DING Meng[1,2], SHI Lei[3], LI Zhihui[4], XU Xiaoyu[5], PAN Yilun[1]
(1. College of Investigation, People's Public Security University of China, Beijing 100038, China; 2. Public Security Behavioral Science Lab, People's Public Security University of China, Beijing 100038, China; 3. State Key Laboratory of Media Convergence and Communication, Communication University of China, Beijing 100024, China; 4. Institute of Forensic Science of China, Beijing 100038, China; 5. Guangdong Provincial Forensic Science of Evidence Materials Engineering Technology Research Center, Shenzhen 518033, China)

Abstract: Existing UAV object tracking methods typically retain only the template information from recent frames, leading to the loss of critical historical appearance information. To address the challenges of information loss and effective utilization of historical data in UAV tracking, this paper proposes a novel tracking method based on template feature buffer and adaptive attention mechanisms. 1)we design a template feature buffer module that maintains a comprehensive repository of historical target appearances through a sliding window mechanism, effectively addressing the information loss problem inherent in traditional methods. 2) we introduce an adaptive attention mechanism that employs channel-level attention to dynamically evaluate the relevance of stored features, enabling intelligent weighting of historical template information. 3) we adopt a plug-and-play architecture that integrates seamlessly with existing mainstream trackers, enhancing the algorithm's practicality and versatility, and design a symmetric sequence evaluation method to validate the effective retention and utilization of historical target information. Experimental results demonstrate significant performance improvements across five mainstream tracking algorithms, with average AUC improvements of 2.34 percentage points on the UAV123 dataset, 7.24 percentage points on the UAV20L dataset, 4.11 percentage points on UAV123-L, and 9.14 percentage points on UAV20L-L, validating the effectiveness and broad applicability of our method.

Keywords: object tracking; UAV; template feature buffer; adaptive attention; plug-and-play; historical information; computer vision; deep learning

无人机凭借其机动性强、部署灵活等优势, 已被广泛应用于军事侦察、安全监控、搜救救援等多个领域[1-2]。目标跟踪[3]作为无人机视觉系统

收稿日期: 2025−07−07.　**网络出版日期:** 2026−02−04.
基金项目: 广东省证据材料司法鉴定 (南天) 工程技术研究中心开放课题 (2024-NT-03).
通信作者: 丁锰. E-mail: dingmeng@ppsuc.edu.cn.

的核心技术, 对于提升无人机的跟踪能力和任务执行效率具有重要意义。然而, 现有无人机目标跟踪方法通常仅保留最近帧的模板信息, 在跟踪过程中容易遗忘原来的目标外观, 造成重要历史外观信息的丢失[4-5]。这种历史信息利用不足的问题导致跟踪器无法有效记忆和利用目标的历史外观表示, 影响了跟踪的鲁棒性和准确性。因此, 研究能够有效保存和利用历史目标外观信息的跟踪方法, 对于提升无人机目标跟踪性能具有重要的意义。

无人机目标跟踪面临的挑战之一在于如何在复杂动态环境中有效保存和利用历史目标信息。一方面, 现有跟踪方法通常仅保留最近帧的模板信息, 在目标出现外观变化、遮挡和重现等情况时, 容易造成重要历史外观信息的丢失[6]; 另一方面, 缺乏对历史信息有效性的评估机制, 无法自适应地从历史模板中筛选和利用最相关的特征信息。为了应对上述挑战, 现有方法主要从模板更新和记忆增强两个方向进行探索。Zhou 等[7] 提出了 RFGM(reading relevant feature from global representation memory) 方法, 通过全局表示记忆和相关性注意力机制从历史模板特征中自适应选择相关信息, 但是该方法主要关注特征选择而非系统性的历史信息保存。Chen 等[8] 提出的 MemVLT (memory-based vision-language tracker) 通过记忆存储模块保存历史目标信息, 但该方法主要面向视觉−语言跟踪任务, 与传统视觉目标跟踪的应用场景存在差异。Yuan 等[9] 提出 MT-Track(multi-step temporal modeling for UAV tracking) 通过多步时序建模框架利用历史帧的时序上下文增强无人机跟踪, 但该方法的时序建模主要集中在相关图层面, 缺乏在模板特征层面的精细化历史信息管理。Zhang 等[10] 提出基于动态模板更新的跟踪方法 DTU(dynamic template updating), 采用平均峰值能量方法判断是否进行模板更新, 并通过像素级融合策略生成新的动态模板, 但该方法在复杂场景下容易受到噪声干扰影响更新质量。Zhang 等[11] 提出 SiamDEPU(dynamic template pool updating siamese tracker), 利用多个动态模板构建模板池来最大化时间信息利用, 并设计在线动态模板池更新器选择具有不同外观的帧作为新模板, 但其整合额外的专用模块影响了跟踪的实时性以及轻量化。Zheng 等[12] 设计了时空记忆网络来动态检索存储在外部记忆中的历史模板特征, 以适应目标外观变化。然而, 该方法需要维护大规模记忆矩阵, 计算开销过大, 难以适应无人机平台的资源限制。Zhang 等[13] 设计了多记忆集成网络来多维度表示目标状态, 并提出门控细化策略来过滤干扰信息, 但该方法的循环神经网络结构增加了在线更新训练的复杂度, 参数量大, 对实时性能造成较大影响。Cai 等[14] 提出的 HIPTrack(visual tracking with historical prompts) 通过历史提示编码器将历史目标特征编码为提示信息, 证明了仅通过提供高质量的历史模板信息就能显著提升跟踪精度, 但该方法主要关注提示学习机制, 对于如何有效管理和优化历史信息的存储策略考虑不足。Wang 等[15] 提出了一种基于时空信息的动态模板更新方法, 通过跟踪置信度网络来判断更新时机, 并利用时空信息融合多个历史模板, 但该方法增加了训练复杂度, 且在模板融合时缺乏对不同历史特征通道级重要性的细粒度建模。Wu 等[16] 提出了基于遮挡鲁棒表示学习的视觉 Transformer (vision Transformer, ViT) 跟踪方法 ORTrack(occlusion-robust tracking), 通过空间 Cox 过程模拟的随机掩码操作, 使目标特征表示对遮挡保持不变性, 不仅提高了单流 ViT 模型在面对遮挡时的鲁棒性, 还设计了自适应特征知识蒸馏策略以满足无人机平台的实时性需求, 但该方法主要针对遮挡问题设计, 对于其他挑战考虑不足。Xue 等[17] 提出了一种基于相似性引导的层自适应视觉 Transformer 方法, 通过动态禁用冗余层优化了无人机跟踪的计算效率, 该方法主要专注于网络结构优化, 而非目标历史表示的记忆与利用。

综上所述, 针对无人机目标跟踪中历史信息丢失和有效利用困难的问题, 本文提出基于模板特征缓存与自适应注意力的无人机目标跟踪方法。该方法通过构建模板特征缓存模块系统性保存历史目标外观表示, 利用自适应注意力机制智能评估和利用存储的历史特征, 并采用即插即用架构可无缝集成到现有主流跟踪器中。

本文主要贡献如下:

1) 设计模板特征缓存模块, 系统性保存多样化的历史目标外观表示, 有效解决传统跟踪方法中目标历史表示记忆不足的问题。

2) 提出自适应注意力机制, 采用通道级注意力动态评估存储特征的重要性, 实现对历史模板信息的自适应加权利用, 提升历史信息的利用效率和跟踪精度。

3) 采用即插即用的架构设计, 所提方法能够无缝集成到现有主流跟踪器中, 在保持计算效率的同时显著提升跟踪性能, 具有良好的实用性和通用性。通过在 5 个主流跟踪器上的实验验证,

证明了方法的适用性和有效性。

1　无人机目标跟踪算法设计

无人机目标跟踪中的目标表示记忆不足与历史特征有效利用不足问题严重影响了跟踪性能的稳定性。针对该关键问题，本文提出一种基于模板特征缓存与自适应注意力的跟踪方法，整体框架如图 1 所示。

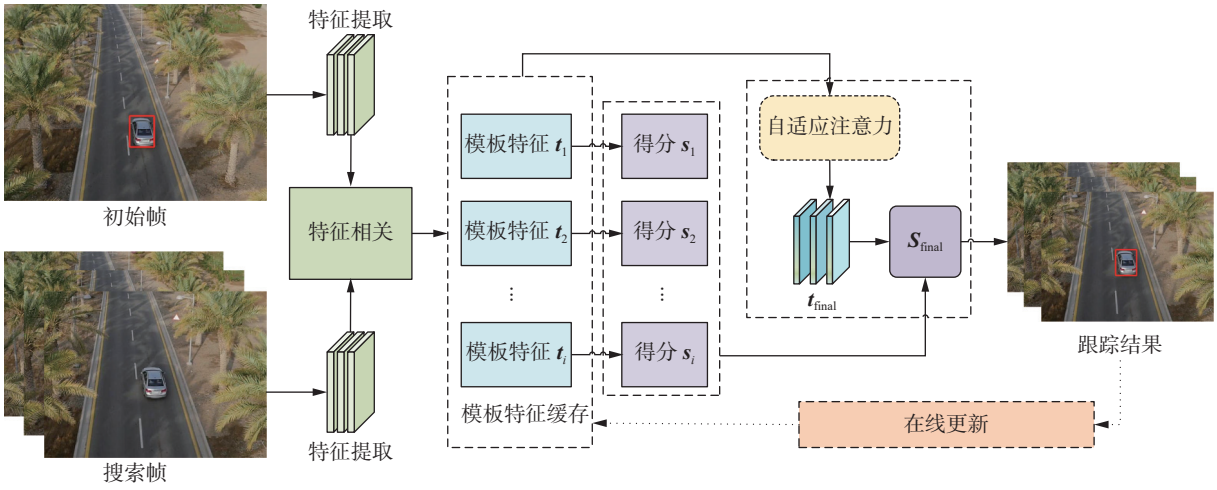

图 1　基于模板特征缓存与自适应注意力的无人机目标跟踪框架

Fig. 1　UAV object tracking framework based on template feature buffer and adaptive attention

所提方法通过构建模板特征缓存模块和自适应注意力机制，系统性地解决历史信息保存和利用问题。在跟踪过程中，首先从初始帧和当前帧中提取特征，模板特征缓存模块将历史模板特征存储在缓存区中，形成多样化的目标外观信息库。对于当前搜索帧，提取的搜索特征与缓存区中的每个模板特征进行相关性计算，生成对应的得分图。随后，自适应注意力机制通过通道注意力权重动态评估各历史特征的重要性，实现对多个得分图的智能融合，最终生成综合得分图用于目标定位，形成了从特征提取、缓存存储、自适应加权到结果融合的完整流程，有效增强跟踪器对历史信息的保存和利用能力。

1.1　模板特征缓存

模板特征缓存模块旨在保持对历史目标外观的记忆，其核心思想是存储和利用历史模板特征。现有无人机目标跟踪方法通常仅保留最近帧的模板信息，这种做法容易导致重要历史外观信息的丢失，从而影响跟踪的连续性和准确性。为了系统性地解决历史信息保存问题，本文设计模板特征缓存模块来存储和管理多样化的历史目标外观表示。如图 2 所示，该模块的在线跟踪框架中包含两个缓存区。其中，短期缓存区存储来自某些历史时刻的图像信息，用于更新模板特征；长期缓存区存储在先前时刻生成的模板特征，用于目标跟踪。这种双缓存区设计能够在保持计算效率的同时，有效平衡短期适应性和长期记忆能力。短期缓存区主要关注目标在连续帧间的渐进变化，捕获目标外观的动态演化过程。长期缓存区则专注于保存具有代表性的关键外观特征，为应对目标重现、视角变化等长期跟踪挑战提供支撑。

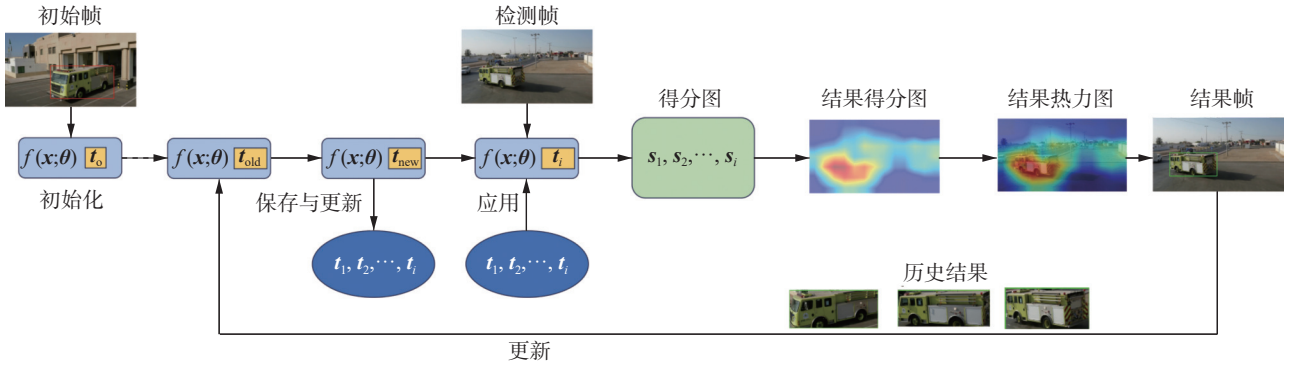

图 2　模板特征缓存的在线跟踪流程

Fig. 2　Illustration of the template feature buffer in online tracking

基于该缓存区结构, 本文建立了长度为i的模板特征缓存模块来存储更新的模板特征, 获取模板特征序列$\boldsymbol{t}_1,\boldsymbol{t}_2,\cdots,\boldsymbol{t}_i$。模板特征序列表示为

$$\mathcal{T}=\{\boldsymbol{t}_1,\boldsymbol{t}_2,\cdots,\boldsymbol{t}_i\},\boldsymbol{t}_j\in\mathbf{R}^{c_s\times h\times w}$$

式中: c_s表示通道数, h和w分别表示特征图的高度和宽度。该缓存区采用滑动窗口机制[18], 存储最新的i个模板特征。当新的模板特征到达时, 按照先进先出原则更新缓存区内容, 确保缓存区始终保持最相关的历史信息, 通过这种固定容量的设计在记忆能力和计算效率之间取得平衡。

在线跟踪过程中, i个模板特征用于导出i个得分图, 实现多模板协同跟踪。对于每个模板特征$\boldsymbol{t}_j$, 其对应的得分图$\boldsymbol{s}_j$通过与搜索区域特征的相关运算得出:

$$\boldsymbol{s}_j=\mathcal{F}(\boldsymbol{t}_j,\boldsymbol{x})$$

式中: $\mathcal{F}$表示相关运算函数, $\boldsymbol{x}$表示搜索区域的特征。这种多模板相关计算策略能够从不同历史时刻的角度评估当前搜索区域中的目标似然度, 提供更全面和鲁棒的目标定位信息。各个得分图最终融合生成最终得分图$\boldsymbol{s}_{\text{final}}$:

$$\boldsymbol{s}_{\text{final}}=\sum_{j=1}^{i}w_j\boldsymbol{s}_j$$

式中: w_i表示对应于得分图$\boldsymbol{s}_j$的权重, $w_j\geqslant 0,\sum_{j=1}^{i}w_j=1$。为突出最新模板信息的重要性, 本文采用线性递增权重分配策略。对于缓存区中的i个模板特征, 权重分配函数定义为

$$w_j=\frac{j}{\sum_{k=1}^{i}k}=\frac{2j}{i(i+1)},\ j=1,2,\cdots,i$$

式中: j表示模板在缓存区中的时间索引, $j=1$对应最早加入缓存区的模板, $j=i$对应最新的模板。该权重函数具有以下性质。1) 单调递增性: $w_1<w_2<\cdots<w_i$, 满足$w_j-w_{j-1}=\dfrac{2}{i(i+1)}$, 确保最新模板获得最高权重; 2) 归一化约束: $\sum_{j=1}^{i}w_j=\dfrac{1}{i(i+1)}\sum_{j=1}^{i}2j=\dfrac{2}{i(i+1)}\cdot\dfrac{i(i+1)}{2}=1$; 3) 权重范围: 最小权重$w_1=\dfrac{2}{i(i+1)}$, 最大权重$w_i=\dfrac{2i}{i(i+1)}=\dfrac{2}{i+1}$。

当缓存区大小设为$i=20$时, 最新模板的权重$w_{20}\approx 0.095$, 最早模板的权重$w_1\approx 0.004\,8$, 两者权重比为 20∶1。这种线性递增的权重分配策略在保持计算简洁性的同时, 有效平衡了最新信息的优先利用与历史外观的长期记忆。跟踪器利用最终得分图$\boldsymbol{s}_{\text{final}}$来定位目标, 相比仅利用单一最新特征进行目标定位, 引入模板特征缓存模块后的跟踪器能够同时利用当前和历史的目标外观信息, 有效避免仅依赖最新特征可能导致的信息丢失问题, 从而增强跟踪能力。

然而, 这种人工设定的权重策略仅考虑了时间维度的先验知识, 无法根据当前跟踪场景的具体特点对不同特征通道进行动态调整。为此, 本文在 1.2 节进一步提出自适应注意力机制, 通过数据驱动的方式在特征层面实现更精细的权重学习。

1.2　自适应注意力

模板特征缓存模块中采用人工设定的权重来衡量模板特征的重要性, 但这种静态权重分配策略存在明显局限性: 无法准确反映不同历史特征在当前跟踪场景下的实际重要性, 难以适应目标外观的动态变化和复杂的跟踪环境。静态权重往往基于时间远近进行简单分配, 忽略了不同历史特征所包含信息的差异性和互补性, 从而影响跟踪效果。对此, 本文设计自适应注意力机制, 动态学习每个通道特征的权重, 实现对历史模板信息的智能化利用。

自适应注意力机制的核心思想是通过数据驱动的方式自动发现和强调最有利于当前跟踪任务的历史特征[19]。与传统的静态权重分配不同, 该机制能够根据当前搜索区域的内容和历史模板特征的特点, 动态调整各个特征的贡献度, 从而实现更精准的特征融合。这种自适应能力使得跟踪器能够在面对不同跟踪场景时自动选择最相关的历史信息, 提升跟踪的鲁棒性和准确性。

该设计在结构上采用了以下技术决策: 1) 空间池化采用求和操作而非平均池化, 以保留特征的激活强度; 2) 权重生成仅使用 Softmax 作为非线性激活函数, 确保权重归一化; 3) 通道加权融合不引入额外的激活函数, 保持特征原始分布特性; 4) 整个结构不包含可学习参数, 完全依靠数据驱动的自适应计算。这些设计决策既确保了结构的轻量化, 又提供了足够的表达能力。

图 3 给出了自适应注意力机制对单个模板特征的处理过程: 首先将缓存区中的模板特征按通道重组, 然后对每个通道组进行空间求和池化, 接着通过 Softmax 操作生成注意力权重, 最后进行通道特征的加权融合并拼接得到该模板的增强特征。此过程对缓存区中的每个模板特征$(j=1,2,\cdots,i)$独立执行, 输入为i个原始模板特征$\{\boldsymbol{t}_1,\boldsymbol{t}_2,\cdots,\boldsymbol{t}_i\}$, 输出为$i$个增强模板特征$\{\boldsymbol{t}'_1,\boldsymbol{t}'_2,\cdots,\boldsymbol{t}'_i\}$, 为后续的时间加权融合提供高质量的特征表示。

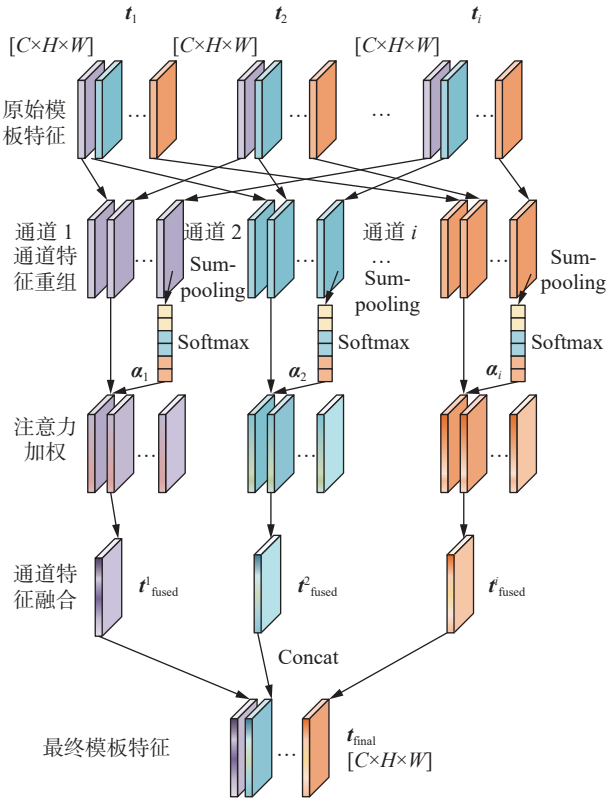

图 3　自适应注意力机制对单个模板特征的处理流程

Fig. 3　Processing flow of adaptive attention mechanism for a single template feature

对于模板特征序列$\mathcal{T}=\{\boldsymbol{t}_1,\boldsymbol{t}_2,\cdots,\boldsymbol{t}_i\}$，每个模板特征具有相同的通道维度$c_s$，可按通道索引重新组织为$c_s$个通道组。自适应注意力机制对每个特征通道分别进行处理，学习不同模板特征的重要性。这种分通道处理方式的优势在于能够捕获不同语义层面的特征：低层通道关注边缘、纹理等基础视觉特征，而高层通道则更多地编码语义信息。通过对每个通道独立建模，注意力机制能够在细粒度层面优化特征选择，相比于将模板作为整体处理的粗粒度方法，这种通道分析方法能够实现更精细的特征选择和权重分配，有效捕获跟踪过程中最相关的历史信息。

首先，将缓存区中的模板特征按通道进行重组。对于第c个通道$(c=1,2,\cdots,c_s)$，提取所有模板在该通道上的特征，形成通道组$\{\boldsymbol{t}_1^c,\boldsymbol{t}_2^c,\cdots,\boldsymbol{t}_i^c\}$，其中$\boldsymbol{t}_j^c\in\mathbf{R}^{h\times w}$表示第$j$个模板在第$c$个通道上的特征图。

然后，在每个通道组内，对每个模板特征在空间维度$h\times w$上应用求和池化，得到通道特征的紧凑表示：

$$f^c(\boldsymbol{t}_j)=\sum_{h,w}\boldsymbol{t}_j^c(h,w)$$

式中：$f^c(\boldsymbol{t}_j)$表示第j个模板特征在通道c上的池化表示。这种空间池化操作能够将二维特征图压缩为标量值，既降低了计算复杂度，又保留了该通道的关键信息。求和池化相比于平均池化能够更好地保持特征的激活强度，有利于后续的注意力权重计算。

对池化表示进行 Softmax 操作，得到每个通道内各模板特征的归一化注意力权重：

$$\boldsymbol{\alpha}^c=\text{Softmax}(f^c(\boldsymbol{t}_1),f^c(\boldsymbol{t}_2),\cdots,f^c(\boldsymbol{t}_i))$$

式中：$\boldsymbol{\alpha}^c=[\boldsymbol{\alpha}_1^c,\boldsymbol{\alpha}_2^c,\cdots,\boldsymbol{\alpha}_i^c]$表示通道$c$的注意力权重向量，$\sum_{j=1}^{i}\boldsymbol{\alpha}_j^c=1$。所得权重表示模板特征序列中各模板在该通道上的重要性分布，权重值越大表明对应的历史特征在该通道上包含更多有价值的信息。Softmax 函数确保了权重的归一化特性，使得所有模板特征在同一通道上的权重和为 1，避免了数值不稳定问题。

然后将注意力权重应用于对应通道的特征，进行加权融合，以获得该通道融合特征：

$$\boldsymbol{t}_{\text{fused}}^c=\sum_{j=1}^{i}\boldsymbol{\alpha}_j^c\boldsymbol{t}_j^c$$

式中：$\boldsymbol{t}_{\text{fused}}^c\in\mathbf{R}^{h\times w}$表示通道$c$的融合特征。这一步骤实现了基于注意力权重的特征加权融合，使得重要的历史特征能够获得更大的影响力，而相对不重要的特征则被适当抑制。

最后，将各通道融合特征连接，形成增强的模板特征：

$$\boldsymbol{t}'_j=\text{concat}(\boldsymbol{t}_{\text{fused}}^{1,(j)},\boldsymbol{t}_{\text{fused}}^{2,(j)},\cdots,\boldsymbol{t}_{\text{fused}}^{c_s,(j)})$$

式中：$\boldsymbol{t}'_j\in\mathbf{R}^{c_s\times h\times w}$表示第$j$个模板经过通道级自适应注意力处理后的增强特征，$\boldsymbol{t}_{\text{fused}}^{c,(j)}$表示第$j$个模板在第$c$个通道上的融合特征。concat操作沿通道维度连接来自所有c_s个通道的融合特征。对于缓存区中的每个模板$\boldsymbol{t}_j(j=1,2,\cdots,i)$，都独立应用通道级注意力处理，得到对应的增强特征$\boldsymbol{t}'_j$。通过这种通道级的独立处理和最终的特征重组，自适应注意力机制能够在保持特征完整性的同时实现精细化的权重分配。该过程对缓存区中的所有模板特征独立进行，输入为i个原始模板特征$\{\boldsymbol{t}_1,\boldsymbol{t}_2,\cdots,\boldsymbol{t}_i\}$，输出为$i$个增强模板特征$\{\boldsymbol{t}'_1,\boldsymbol{t}'_2,\cdots,\boldsymbol{t}'_i\}$，保持了模板的数量和独立性，为后续的时间加权融合提供了高质量的特征表示。

本文方法中，模板特征缓存 (1.1 节) 与自适应注意力 (1.2 节) 采用两层权重机制串联工作。首先通过通道级自适应注意力将缓存区中的每个原始模板特征$\boldsymbol{t}_j$增强为$\boldsymbol{t}'_j$，然后对增强后的特征计算得分图，最后使用模板级时间权重进行加权融合：

$$\boldsymbol{s}_{\text{final}}=\sum_{j=1}^{i}w_j\cdot\mathcal{F}(\boldsymbol{t}'_j,\boldsymbol{x})$$

式中：w_j为模板级时间权重，$\mathcal{F}(\cdot,\cdot)$表示相关运算

函数, $\boldsymbol{x}$表示搜索区域特征, $\boldsymbol{t}'_j$为第j个模板经过通道级自适应注意力增强后的特征。这种双层权重串联机制使得通道级自适应注意力在特征层面实现精细化的历史信息筛选, 模板级时间权重在得分图层面确保最新信息的优先性, 两者协同工作, 既保证了基于内容的细粒度通道选择, 又实现了基于时间先验的粗粒度加权。

通过这种数据驱动的权重学习方式, 自适应注意力机制作为智能记忆控制器, 能够根据当前跟踪场景动态选择相关历史信息并排除冗余特征, 显著提升模板特征缓存模块的有效性, 为无人机长期目标跟踪提供更加智能和鲁棒的解决方案。

在模型的训练与优化方面, 本文提出的自适应注意力机制采用端到端训练方式, 无需额外的监督信号。该模块与基线跟踪器共享统一的训练目标, 损失函数包含分类损失和边界框回归损失:

$$\mathcal{L}_{\text{total}} = \mathcal{L}_{\text{cls}} + \lambda \mathcal{L}_{\text{reg}}$$

式中: $\mathcal{L}_{\text{cls}}$为分类损失, 用于判断搜索区域是否包含目标; $\mathcal{L}_{\text{reg}}$为回归损失, 用于优化边界框位置; λ为平衡系数。

自适应注意力模块中的通道级权重$\boldsymbol{\alpha}^c$通过反向传播自动学习。梯度传播路径为

$$\frac{\partial \mathcal{L}_{\text{total}}}{\partial \boldsymbol{\alpha}_j^c} = \frac{\partial \mathcal{L}_{\text{total}}}{\partial \boldsymbol{s}_{\text{final}}} \cdot \frac{\partial \boldsymbol{s}_{\text{final}}}{\partial \boldsymbol{t'}_j} \cdot \frac{\partial \boldsymbol{t}'_j}{\partial \boldsymbol{\alpha}_j^c}$$

梯度从最终得分图$\boldsymbol{s}_{\text{final}}$依次反向传播到增强特征$\boldsymbol{t}'_j$、融合特征$\boldsymbol{t}^c_{\text{fused}}$和注意力权重$\boldsymbol{\alpha}^c_j$。由于 Softmax 函数和求和池化操作均可微, 且 Softmax 具有良好的数值稳定性 (输出范围 [0,1], 导数为$\boldsymbol{\alpha}^c_j(1-\boldsymbol{\alpha}^c_j)$), 注意力权重能够稳定优化。

这种端到端训练方式的优势在于: 1) 注意力权重自动适应跟踪任务需求, 无需人工标注; 2) 避免多阶段训练的复杂性; 3) 由于仅引入轻量级操作 (池化和 Softmax), 继承了基线跟踪器的收敛特性。

2　实验验证

2.1　数据集与评价指标

本文采用已建立的无人机目标跟踪基准数据集 UAV123(a dataset of 123 video sequences for unmanned aerial vehicle target tracking)[20] 和 UAV20L[20] 进行实验评估。UAV123 提供了 123 个完全标注的无人机航拍视频, 而 UAV20L 包含 20 个专门为评估长期跟踪而设计的长时序列。

为了进一步测试跟踪器的记忆能力, 本文设计了一种创新的对称序列评估方法。具体而言, 将每个原始序列与其时间反向序列进行连接, 创建扩展数据集, 命名为 UAV123-L 和 UAV20L-L。对于包含 n 帧的原始视频序列, 该方法生成包含 $2n-1$ 帧的对称序列, 其中第 i 帧与第 $2n-i$ 帧在内容上完全相同。这种对称设计的核心思想是让跟踪器在序列的后半段重新遇到早期出现过的目标外观, 从而直接检验其对历史信息的保存和利用能力。相比传统的单向序列评估, 这种对称结构显著增加了跟踪难度, 能够更加清晰地评估跟踪器在跟踪过程中的记忆容量和信息保持能力。理想情况下, 具有良好记忆能力的跟踪器在重新遇到先前见过的目标外观时, 应能够利用存储的历史信息保持稳定的定位精度, 而非从零开始重新学习目标特征。表 1 给出了标准无人机目标跟踪数据集与扩展数据集的统计特性对比。

表 1　标准 UAV 跟踪数据集与扩展数据集的数据统计
Table 1　Statistics of standard and extended UAV tracking datasets

数据集	序列数	最小帧数	最大帧数	平均帧数	总帧数
UAV123	123	109	3 085	915	112 578
UAV20L	20	1 717	5 527	2 934	58 670
UAV123-L	123	217	6 169	1 829	225 033
UAV20L-L	20	3 433	11 053	5 867	117 320

为了全面评估跟踪算法的性能, 本文采用了目标跟踪领域广泛使用的标准评价指标[21-22], 包括曲线下面积 (area under curve, AUC)、精确率 (Precision)、50% 重叠精度 (overlap precision at 50%, OP50) 和 75% 重叠精度 (overlap precision at 75%, OP75)。其中, AUC 通过测量成功率曲线下的面积评估整体跟踪性能; 精度指标用于评估中心点定位的准确性; OP50 和 OP75 分别测量预测边界框与真实边界框重叠比超过 50% 和 75% 时的成功率。这些指标从不同角度衡量跟踪器的定位精度和成功率, 能够客观反映算法在各种跟踪场景下的综合表现。

2.2　参数设置

实验在配备 NVIDIA GeForce RTX 4090 24GB GPU 的工作站上进行。所提出的模板特征缓存模块和自适应注意力机制基于 PyTorch 框架实现。经过性能与计算效率的权衡考虑, 将模板特征缓存区大小设置为 20。自适应注意力的通道维度与各基线跟踪器的原始特征维度保持一致。

为确保训练稳定性和性能提升, 本文方法沿用了各基线跟踪器的原始训练配置, 包括优化器类型、学习率策略和损失函数设计。具体而言, 基线跟踪器参数采用预训练权重初始化, 保持原

有特征表示能力；训练过程采用两阶段策略：首先冻结基线跟踪器参数进行初步训练，使自适应注意力机制适应现有特征分布；随后以较小学习率联合微调所有参数，实现整体性能优化。为保证实验公平性，所有基线跟踪器均采用其原始论文及官方 PyTracking 实现中的推荐参数设置。

2.3　实验结果与分析

本节给出了所提方法在标准无人机目标跟踪数据集 (UAV123 和 UAV20L) 及扩展数据集 (UAV123-L 和 UAV20L-L) 上的综合性能评估结果。

为验证所提方法的通用性，本文将模块集成到 5 个代表性的主流跟踪器中：ATOM(accurate tracking by overlap maximization)[23]、DiMP(discriminative model prediction)[24]、PrDiMP(probabilistic discriminative model prediction)[25]、ToMP(transformer-based model prediction)[26] 和 STARK(spatio-temporal advanced werk for tracking)[27]。这些跟踪器代表不同的跟踪范式和架构设计，为全面评估提供了坚实基础。

为便于区分基线方法与集成本文所提方法的改进版本，在后续实验结果中，集成了模板特征缓存模块和自适应注意力的跟踪器统一标记为“-BA”的形式。例如，ATOM-BA 表示集成了本文所提方法的 ATOM 跟踪器，DiMP50-BA 表示集成了本文所提方法的 DiMP50 跟踪器，以此类推。表 2 的结果分析表明，集成了模板特征缓存模块和自适应注意力的方法 (-BA) 在不同架构的跟踪器上均能带来持续性能提升，但提升幅度存在显著差异。这种差异主要源于各跟踪架构在时序建模和历史信息利用方面的能力不同。

表 2　在标准数据集和扩展数据集上的跟踪性能比较

Table 2　Comparison of tracking performance on standard datasets and extended datasets　%

方法	UAV123		UAV20L		UAV123-L		UAV20L-L	
	AUC	Precision	AUC	Precision	AUC	Precision	AUC	Precision
ATOM	64.23	80.59	53.43	69.55	58.68	76.12	45.17	59.70
ATOM-BA(本文方法)	**66.71**	**83.59**	**62.51**	**80.15**	**62.87**	**80.48**	**56.68**	**74.40**
DiMP50	65.32	84.42	59.83	78.28	60.38	78.60	51.63	70.26
DiMP50-BA(本文方法)	**67.13**	**88.24**	**65.01**	**84.42**	**64.43**	**81.54**	**60.63**	**78.97**
PrDiMP50	68.01	85.45	61.29	79.78	62.94	79.96	53.24	70.83
PrDiMP50-BA(本文方法)	**70.09**	**89.03**	**67.60**	**85.66**	**67.64**	**83.82**	**61.53**	**79.22**
ToMP	69.02	85.85	61.53	80.04	64.25	81.12	54.22	70.74
ToMP-BA(本文方法)	**71.86**	**87.99**	**68.89**	**85.17**	**67.93**	**85.40**	**61.99**	**79.92**
STARK	71.53	85.84	62.24	80.85	67.92	82.38	56.63	72.85
STARK-BA(本文方法)	**74.03**	**89.65**	**70.53**	**86.98**	**71.83**	**85.65**	**65.76**	**81.00**
平均提升	+2.34	+3.27	+7.24	+6.78	+4.11	+3.74	+9.14	+9.83

注：加粗表示效果最好。

1) 对于基于判别式学习的 ATOM 和 DiMP 系列跟踪器，它们主要依赖单帧目标表示和在线判别模型更新，缺乏系统性的历史信息保持机制。因此，集成本文方法后获得了较大提升，在 UAV20L 上 AUC 分别提升 9.08 百分点和 5.18 百分点。这表明模板特征缓存模块有效弥补了这类架构在长期记忆方面的不足。

2) 对于引入概率回归的 PrDiMP 和 ToMP 跟踪器，它们虽在目标状态估计上具有不确定性建模能力，但在目标表示记忆方面仍有局限。集成本文方法后，在 UAV20L 数据集上分别获得了 6.31 百分点和 7.36 百分点的提升，证明了历史特征缓存与状态估计的良好互补性。

3) 对于基于 Transformer 的 STARK 跟踪器，其架构已包含一定的时间建模能力，能够捕捉帧间依赖关系，在标准数据集上的提升相对较小 (UAV123 上提升 2.50 百分点)。然而在长期跟踪场景 (UAV20L) 中，仍获得了 8.29 百分点的显著提升，表明本文的模板特征缓存机制与 Transformer 的自注意力机制存在有效互补，尤其在处理长时序跟踪时更为明显。

在扩展数据集上，所提方法的优势更加突出。在 UAV123-L 上实现了平均 4.11 百分点的 AUC 增益，在 UAV20L-L 上达到平均 9.14 百分点的显著提升。这一现象深层次地反映了所提方法的核心优势：随着序列长度增加，传统跟踪器容易出

现目标遗忘和记忆衰减, 而模板特征缓存模块通过系统性保存历史外观信息, 防止关键特征丢失; 自适应注意力机制则根据当前场景动态选择相关历史信息, 从根本上提升了长期跟踪的稳定性。

这种不同架构的差异化提升证明了所提方法具有良好的适配性和通用性。对于原本缺乏历史信息保持能力的架构, 模板特征缓存提供了关键的长期记忆支持; 对于已具备一定时序建模能力的架构, 自适应注意力机制则通过精细化的特征选择进一步增强了其性能, 为无人机视觉跟踪系统提供了有效的性能增强方案。

2.4　计算效率与资源占用分析

为全面评估所提方法的计算效率, 本节从理论复杂度、实际推理速度以及嵌入式平台部署 3 个方面进行分析。

2.4.1　理论复杂度分析

所提方法的额外计算开销主要来源于两个部分: 1) 模板特征缓存模块, 采用固定大小滑动窗口机制, 单次更新时间复杂度为$O(1)$, 空间复杂度为$O(i\cdot c_s\cdot h\cdot w)$, 其中$i$为缓存区大小, c_s为特征通道数, h和w为特征图尺寸; 2) 自适应注意力机制, 包括空间池化 ($O(i\cdot c_s\cdot h\cdot w)$)、Softmax 归一化 ($O(i\cdot c_s)$) 和加权融合 ($O(i\cdot c_s\cdot h\cdot w)$), 总体复杂度为$O(i\cdot c_s\cdot h\cdot w)$。相比基线跟踪器的主干网络特征提取 ($O(c_s\cdot h\cdot w\cdot k^2)$, k为卷积核大小) 和特征相关计算, 所提模块的额外开销相对有限, 且操作均为轻量级, 实际计算开销可控。

2.4.2　推理速度与显存占用

本文在 NVIDIA GeForce RTX 4090 GPU 上测试了所提方法的推理速度。采用 STARK 作为基线时, 输入模板分辨率 128×128, 搜索区域 256×256。基线 STARK 的推理速度为 39.2 帧/s, 当缓存区大小为 20 时, 推理速度为 34.9 帧/s, 下降约 11.0%, 但仍保持实时性。GPU 显存占用增加约 15%, 在现代 GPU 的显存容量下可接受。

对于其他基线跟踪器, 集成本文方法后的速度下降幅度小于 STARK, 下降幅度约为 7%~9%, 这是因为 Transformer 架构本身计算复杂度较高, 引入额外模块后相对下降幅度更大, 而卷积神经网络 (convolutional neural network, CNN) 架构的基线计算量较小, 所提模块的额外开销占比相对较低。结果表明, 所提方法的计算开销在不同架构上均可控, 体现了良好的通用性。

2.4.3　嵌入式平台部署验证

为验证所提方法在资源受限平台上的实用性, 本文在典型的嵌入式计算平台 Jetson Xavier NX(配备 8 GB 内存) 上进行了部署测试。实验选用 STARK-BA 作为代表性实现, 输入分辨率与桌面平台保持一致 (模板 128×128, 搜索区域 256×256)。

如表 3 所示, 尽管嵌入式平台上的整体帧率较桌面平台有所下降, 但所提方法在资源受限环境中仍保持了可接受的性能。当缓存区大小为 10 时, 跟踪速度达 12.5 帧/s; 当缓存区大小增至 20 时, 帧率降至 10.3 帧/s, 仍能保持实时性。

表 3　在嵌入式平台 Jetson Xavier NX 上的性能与资源占用
Table 3　Performance and resource usage on the Jetson Xavier NX embedded platform

配置	推理速度/(帧/s)	显存占用/MB
Baseline	14.2	1 236
缓存区大小10	12.5	1 342
缓存区大小20	10.3	1 415

2.5　消融实验

为全面评估所提方法中各组件的贡献, 本文进行了消融实验, 系统分析模板特征缓存区大小的影响以及模板特征缓存模块和自适应注意力在不同跟踪器和数据集上的有效性。

2.5.1　缓存区大小影响分析

缓存区大小是平衡记忆容量与计算效率的关键参数。为确定最优缓存区大小, 本文在 UAV20L 数据集上使用 STARK 跟踪器进行了不同缓存区大小的对比实验, 实验结果如表 4 所示。

表 4　模板特征缓存大小对跟踪性能和计算效率的影响
Table 4　Effect of template feature buffer size on tracking performance

缓存区大小	AUC/%	OP50/%	OP75/%	Precision/%	推理速度/(帧/s)
Baseline	62.24	77.72	57.23	80.85	39.2
5	65.78	79.64	61.05	83.27	38.0
10	67.45	80.52	62.47	84.36	36.5
20	68.56	81.31	63.23	85.01	34.9
30	68.94	81.58	63.51	85.33	33.7
40	69.12	81.72	63.68	85.47	32.6

通过对表 4 的分析可以看出，增加缓存区大小能够持续提升跟踪性能，但在缓存区大小超过 20 后收益递减。从跟踪性能来看，从基线 (无缓存区) 到缓存区大小 20 时性能提升最显著，AUC 提升 6.32 百分点 (从 62.24% 至 68.56%)。进一步增加缓存区大小虽能带来边际收益，但提升幅度有限 (从 20 增加到 40 时 AUC 仅提升 0.56 百分点)。

从计算效率来看，随着缓存区大小的增加，推理速度呈现下降趋势。基线 STARK 的推理速度为 39.2 帧/s。当缓存区大小为 5 时，推理速度为 38.0 帧/s，相比基线下降约 3.1%；缓存区大小为 20 时，推理速度为 34.9 帧/s，相比基线下降约 11.0%，但仍保持实时跟踪能力。进一步增大缓存区至 30 和 40 时，推理速度分别降至 33.7 帧/s 和 32.6 帧/s，下降幅度为 14.0% 和 16.8%，计算开销显著增加。

综合考虑跟踪性能与计算效率的平衡，本文选择 20 作为缓存区大小。在此配置下，跟踪器在保持高推理速度 (34.9 帧/s，满足实时性要求) 的同时，获得了显著的性能提升 (AUC 提升 6.32 百分点)，实现了跟踪精度与计算效率的最优平衡。

在实际应用中，模板特征缓存模块的长度 i 可根据具体场景需求进行调整。对于计算资源充足的场景，可选择较大的缓存长度 (如 20~30) 以保留更多历史信息；对于资源受限的嵌入式平台，可选择较小的缓存长度 (如 5~10) 以平衡性能和效率。此外，目标场景的复杂度也是重要考量因素：在目标频繁遮挡或快速变形等复杂场景中，较大的缓存长度有助于保持跟踪稳定性；而在目标外观相对稳定的简单场景中，较小的缓存长度已能满足需求。

2.5.2　模板特征缓存模块与注意力机制有效性验证

为了验证本文所提方法中各组件的贡献，本文比较了 3 种配置：1) 没有模板特征缓存模块的基线跟踪器；2) 具有模板特征缓存模块但使用人工设计权重的跟踪器 (记为-B)；3) 同时具有模板特征缓存模块和自适应注意力的完整方法 (记为-BA)。各种指标下的具体消融实验结果如表 5 和表 6 所示。

表 5　消融实验结果 (AUC 和 Precision)
Table 5　Ablation study results (AUC and Precision)　%

方法	UAV123		UAV20L		UAV123-L		UAV20L-L	
	AUC	Precision	AUC	Precision	AUC	Precision	AUC	Precision
ATOM	64.23	80.59	53.43	69.55	58.68	76.12	45.17	59.70
ATOM-B	65.56	82.65	60.73	78.63	61.41	79.32	55.43	72.67
ATOM-BA	**66.71**	**83.59**	**62.51**	**80.15**	**62.87**	**80.48**	**56.68**	**74.4**
DiMP50	65.32	84.42	59.83	78.28	60.38	78.60	51.63	70.26
DiMP50-B	66.67	86.89	64.35	82.28	62.91	80.55	59.77	76.97
DiMP50-BA	**67.13**	**88.24**	**65.01**	**84.42**	**64.43**	**81.54**	**60.63**	**78.97**
PrDiMP50	68.01	85.45	61.29	79.78	62.94	79.96	53.24	70.83
PrDiMP50-B	70.07	88.24	66.27	83.83	66.18	82.73	60.15	77.45
PrDiMP50-BA	**70.09**	**89.03**	**67.60**	**85.66**	**67.64**	**83.82**	**61.53**	**79.22**
ToMP	69.02	85.85	61.53	80.04	64.25	81.12	54.22	70.74
ToMP-B	70.71	87.48	66.95	84.04	67.85	83.01	60.28	78.21
ToMP-BA	**71.86**	**87.99**	**68.89**	**85.17**	**67.93**	**85.40**	**61.99**	**79.92**
STARK	71.53	85.84	62.24	80.85	67.92	82.38	56.63	72.85
STARK-B	73.72	88.79	68.56	85.01	69.45	84.32	63.55	79.77
STARK-BA	**74.03**	**89.65**	**70.53**	**86.98**	**71.83**	**85.65**	**65.76**	**81.00**

注：加粗表示效果最好。

表 6　消融实验结果 (OP50 和 OP75)
Table 6　Ablation study results (OP50 and OP75)　　%

方法	UAV123		UAV20L		UAV123-L		UAV20L-L	
	OP50	OP75	OP50	OP75	OP50	OP75	OP50	OP75
ATOM	78.45	61.51	65.41	51.91	73.13	55.88	56.08	43.49
ATOM-B	79.48	**64.14**	75.01	58.31	77.58	59.59	69.83	53.18
ATOM-BA	**79.70**	64.11	**76.74**	**59.67**	**78.93**	**60.58**	**71.02**	**54.26**
DiMP50	80.3	63.03	73.15	56.54	75.42	58.73	63.24	46.74
DiMP50-B	83.26	65.06	78.25	61.20	77.97	61.36	72.70	56.84
DiMP50-BA	**84.39**	**65.50**	**80.10**	**63.32**	**80.35**	**62.25**	**75.97**	**57.25**
PrDiMP50	81.2	64.34	76.10	56.08	76.60	59.33	67.77	47.83
PrDiMP50-B	83.51	65.71	79.67	62.42	79.71	63.49	73.29	57.76
PrDiMP50-BA	**84.82**	**66.18**	**80.27**	**64.03**	**82.03**	**65.26**	**76.60**	**58.01**
ToMP	81.65	66.01	77.02	56.03	76.06	60.34	68.43	47.11
ToMP-B	84.00	68.04	79.34	63.42	80.14	63.36	74.04	57.50
ToMP-BA	**84.36**	**69.47**	**81.05**	**64.41**	**81.19**	**64.58**	**76.73**	**58.98**
STARK	83.24	65.25	77.72	57.23	80.71	62.48	71.39	51.43
STARK-B	85.42	66.88	81.31	63.23	82.52	65.16	77.49	59.92
STARK-BA	**85.58**	**67.83**	**82.28**	**65.22**	**83.80**	**65.57**	**79.99**	**62.91**

注: 加粗表示效果最好。

通过对表 5 和表 6 的结果分析表明, 两个组件都对性能改进做出了显著贡献。模板特征缓存区 (-B) 在所有跟踪器上提供了大幅增益, 在 UAV123 上 AUC 提升了 1.33~2.19 百分点, 在 UAV20L 上提升了 5.93~7.30 百分点。当在模板特征缓存模块的基础上进一步融入自适应注意力 (-BA) 时, 相比单独使用缓存模块, 各评价指标均获得了额外提升。在 UAV123 上, AUC 额外增益 0.31~1.15 百分点, Precision 额外增益 0.70~1.35 百分点; 在 UAV20L 上, AUC 额外增益 0.66~1.97 百分点, OP50 额外增益 1.73~1.85 百分点。

在扩展数据集上, 所提方法对长期跟踪能力较弱的跟踪器改进效果更为显著。ATOM-BA 相比基线 ATOM 在 UAV123-L 上 AUC 提升 4.19 百分点, OP75 提升 4.70 百分点; 在 UAV20L-L 上 AUC 提升 11.51 百分点, OP50 提升 14.94 百分点。即使对已融入时间建模能力的 STARK, 本方法在 UAV123-L 和 UAV20L-L 上仍分别提供了 3.91 百分点和 9.13 百分点的 AUC 提升。

图 4 直观展示了 3 组方法在 4 个数据集上的平均 AUC 性能变化。加入缓存模块 (-B) 后, 性能在所有数据集上均显著提升, 尤其在 UAV20L-L 上从 52.2% 提升至 59.8%; 进一步融入自适应注意力 (-BA) 后, 性能再次提升至 61.3%。随着跟踪序列长度增加, 基线方法性能急剧下降, 而所提方法保持相对稳定, 凸显其长期跟踪优势。

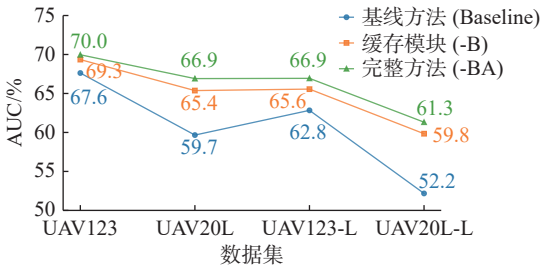

图 4　不同方法的平均 AUC 性能比较
Fig. 4　Average AUC performance comparison of different methods

为验证性能提升的统计显著性, 本文对基线方法与集成了模板特征缓存与自适应注意力 (-BA) 的方法在 4 个数据集上的表现进行了配对 t 检验, 结果如表 7 所示。从表中可见, 所提方法在所有评价指标上均取得了统计显著的性能提升 (p<0.001), AUC、Precision、OP50 和 OP75 指标分

别平均提升 5.71、5.90、6.14 和 6.40 百分点，进一步证实了所提方法的高度统计可靠性。

表 7　*t* 检验结果分析

Table 7　*t*-test analysis results

评价指标	平均提升/百分点	*t*值	*p*值
AUC	5.71	−8.82	<0.001
Precision	5.90	−8.31	<0.001
OP50	6.14	−7.57	<0.001
OP75	6.40	−8.46	<0.001

图 5 给出了不同跟踪场景下通道级自适应注意力的权重分布热力图。横轴表示特征通道(C1~C20, 从低层到高层特征), 纵轴表示连续时间步，颜色深浅表示权重大小。可以观察到：正常跟踪场景下权重集中在中低层特征通道 (C7~C11), 体现基本形状特征的优先使用；遮挡场景下权重明显向高层特征通道 (C13~C17) 转移，说明系统自动增强语义级特征的重要性以应对遮挡；外观变化场景下权重分布更为均匀，多个通道同时激活，表明系统能够综合利用多层次特征适应目标外观变化。这种动态通道权重分配能力使跟踪器能够智能选择最相关的历史特征，有效提升了在复杂环境下的跟踪鲁棒性。

此外，本文对典型跟踪场景进一步可视化分析。图 6 给出了 3 个模型 STARK、STARK-B 和 STARK-BA 在 UAV123 数据集 car-1 序列上的跟踪性能，可视化展示了本文的方法如何随时间改善跟踪鲁棒性。随着跟踪的进行，基线 STARK 模型表现出跟踪错误，逐渐从跟踪白色汽车转移到跟踪路边护栏。STARK-B 模型改进效果有一定改进，但仍然表现出偶尔的不稳定性。相比之下，STARK-BA 通过对历史模板特征进行自适应权重分配，在整个序列中保持对正确目标的一致跟踪。

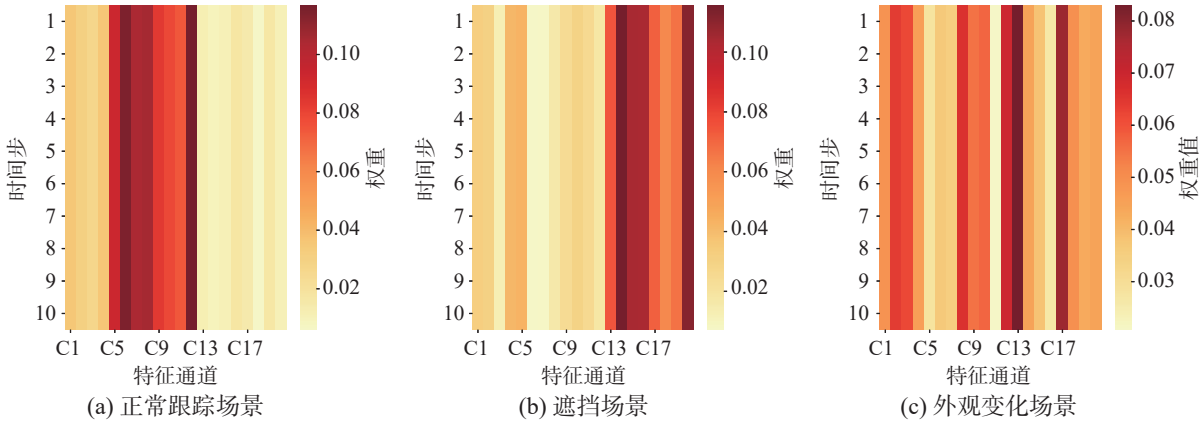

图 5　自适应通道注意力权重分布

Fig. 5　Distribution of adaptive channel attention weights

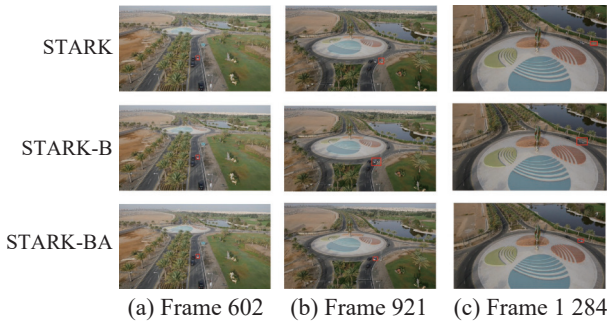

图 6　STARK、STARK-B 和 STARK-BA 在 UAV123 数据集 car-1 序列上的跟踪性能表现

Fig. 6　Tracking performance of STARK, STARK-B, and STARK-BA for the car-1 sequence in the UAV123

图 7 通过对 UAV123-L 数据集 truck-1-1 序列的跟踪，进一步证明了本文方法的有效性。这种对称跟踪场景专门测试跟踪器对先前遇到目标分布的记忆。基线 STARK 模型在序列后半部分遇到类似对象分布时无法保持一致跟踪，清楚地展示了历史目标信息丢失问题。STARK-B 模型显示了性能的提升但表现出一些漂移。而 STARK-BA 模型通过基于存储模板特征与当前帧的相关性自适应权重，成功地在整个序列中保持跟踪。

在长序列上更显著的性能提升表明，所提方法在长期目标跟踪场景中具有良好的适用性，特别是在历史信息保持困难的情况下。定量指标和定性可视化结果均证实，模板特征缓存模块能够有效保存历史目标信息，自适应注意力机制能够智能管理特征融合过程，两者结合对提升无人机长期目标跟踪性能具有重要意义。

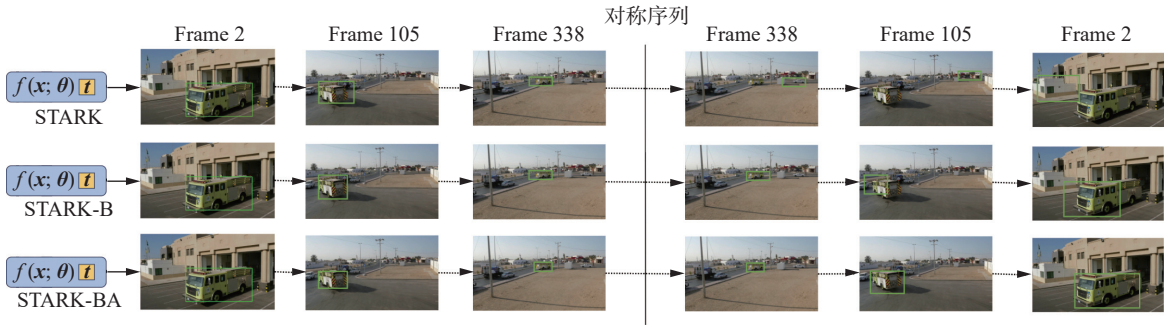

图 7　STARK、STARK-B 和 STARK-BA 在 UAV123-L 数据集 truck-1-1 序列上的跟踪性能表现

Fig. 7　Tracking performance of STARK, STARK-B, and STARK-BA for the truck-1-1 sequence in the UAV123-L

3　结束语

本文针对无人机目标跟踪中的目标信息丢失和历史信息有效利用问题，提出了基于模板特征缓存与自适应注意力的跟踪方法。该方法通过模板特征缓存模块系统性保存多样化的历史目标外观表示，结合自适应注意力机制动态评估存储特征的重要性，实现对历史模板信息的有效保存和自适应利用。此外，采用即插即用架构，可无缝集成到现有主流跟踪器中，增强了算法的实用性和通用性。实验结果表明，所提方法在多个基准数据集上均取得显著性能提升，在长期跟踪场景中效果更为突出，验证了方法的有效性和适用性。未来工作将继续探索无人机目标跟踪中的其他关键挑战，如小目标跟踪、目标尺度急剧变化、严重遮挡等问题，进一步增强跟踪系统在复杂环境中的可靠性。

参考文献：

[1] WU Xin, LI Wei, HONG Danfeng, 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 Nianyi, ZHAO Jin, SHI Qing, 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] 黄昱程, 肖子旺, 武丹凤, 等. 时空融合与判别力增强的孪生网络目标跟踪方法[J]. 智能系统学报, 2024, 19(5): 1218−1227.
HUANG Yucheng, XIAO Ziwang, WU Danfeng, et al. Spatiotemporal fusion and discriminative augmentation for improved Siamese tracking[J]. CAAI transactions on intelligent systems, 2024, 19(5): 1218−1227.

[4] WEI Xing, BAI Yifan, ZHENG Yongchao, et al. Autoregressive visual tracking[C]//2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Vancouver: IEEE, 2023: 9697−9706.

[5] BAI Yifan, ZHAO Zeyang, GONG Yihong, et al. ARTrackV2: prompting autoregressive tracker where to look and how to describe[C]//2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Seattle: IEEE, 2024: 19048−19057.

[6] 刘芳, 卢晨阳, 路言, 等. 基于自适应模板更新的 Transformer 无人机目标跟踪算法[J]. 航空学报, 2025, 46(16): 290−301.
LIU Fang, LU Chenyang, LU Yan, et al. Adaptive template update-based Transformer algorithm for UAV target tracking[J]. Acta aeronautica et astronautica sinica, 2025, 46(16): 290−301.

[7] ZHOU Xinyu, GUO Pinxue, HONG Lingyi, et al. Reading relevant feature from global representation memory for visual object tracking[EB/OL]. (2024−02−02)[2025−07−07]. https://arxiv.org/abs/2402.14392.

[8] CHEN Xiaotang, FENG Xiaokun, HU Shiyu, et al. MemVLT: vision-language tracking with adaptive memory-based prompts[C]//Advances in Neural Information Processing Systems 37. Vancouver: NeurIPS, 2024: 14903−14933.

[9] YUAN Xiaoying, XU Tingfa, LIU Xincong, et al. Multi-step temporal modeling for UAV tracking[J]. IEEE transactions on circuits and systems for video technology, 2024, 34(8): 7216−7230.

[10] ZHANG Tianshuo, ZHU Linlin, LIN Feng. A robust tracking method based on dynamic template updating[C]//2024 36th Chinese Control and Decision Conference. Xi'an: IEEE, 2024: 3338−3342.

[11] ZHANG Mingyang, VAN BEECK K, GOEDEMÉ T. Object tracking withMultiple dynamic templates updating [C]//Image and Vision Computing. Cham: Springer, 2023: 144−158.

[12] ZHENG Yang, XU Yijie, LIANG Jimin. Spatiotemporal memory network for UAV target tracking[C]//Proceed-

ings of 4th 2024 International Conference on Autonomous Unmanned Systems. Singapore: Springer, 2025: 79−88.

[13] ZHANG Huanlong, SONG Peipei, FU Weiqiang, et al. Brain-inspired memory network for visual tracking with recurrent meta-learning updater[J]. Digital signal processing, 2025, 162: 105159.

[14] CAI Wenrui, LIU Qingjie, WANG Yunhong. HIPTrack: visual tracking with historical prompts[C]//2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Seattle: IEEE, 2024: 19258−19267.

[15] WANG Yuanhui, YE Ben, CAI Zhanchuan. Dynamic template updating using spatial-temporal information in Siamese trackers[J]. IEEE transactions on multimedia, 2024, 26: 2006−2015.

[16] WU You, WANG Xucheng, YANG Xiangyang, et al. Learning occlusion-robust vision transformers for real-time UAV tracking[C]//2025 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Nashville: IEEE, 2025: 17103−17113.

[17] XUE Chaocan, ZHONG Bineng, LIANG Qihua, et al. Similarity-guided layer-adaptive vision transformer for UAV tracking[C]//2025 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Nashville: IEEE, 2025: 6730−6740.

[18] CAO Ziang, HUANG Ziyuan, PAN Liang, et al. Towards real-world visual tracking with temporal contexts[J]. IEEE transactions on pattern analysis and machine intelligence, 2023, 45(12): 15834−15849.

[19] XIE Jinxia, ZHONG Bineng, MO Zhiyi, et al. Autoregressive queries for adaptive tracking with spatio-temporal transformers[C]//2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Seattle: IEEE, 2024: 19300−19309.

[20] MUELLER M, SMITH N, GHANEM B. A benchmark and simulator for UAV tracking[C]//Computer Vision–ECCV 2016. Cham: Springer, 2016: 445−461.

[21] CHEN Xin, PENG Houwen, WANG Dong, et al. SeqTrack: sequence to sequence learning for visual object tracking[C]//2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Vancouver: IEEE, 2023: 14572−14581.

[22] 林淑彬, 吴贵山, 姚文勇, 等. 基于光照自适应动态一致性的无人机目标跟踪[J]. 智能系统学报, 2022, 17(6): 1093−1103.
LIN Shubin, WU Guishan, YAO Wenyong, et al. Unmanned aerial vehicles object tracking based on illumination adaptive dynamic consistency[J]. CAAI transactions on intelligent systems, 2022, 17(6): 1093−1103.

[23] DANELLJAN M, BHAT G, KHAN F S, et al. ATOM: accurate tracking by overlap maximization[C]//2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Long Beach: IEEE, 2020: 4655−4664.

[24] BHAT G, DANELLJAN M, VAN GOOL L, et al. Learning discriminative model prediction for tracking[C]//2019 IEEE/CVF International Conference on Computer Vision. Seoul: IEEE, 2020: 6181−6190.

[25] DANELLJAN M, VAN GOOL L, TIMOFTE R. Probabilistic regression for visual tracking[C]//2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Seattle: IEEE, 2020: 7181−7190.

[26] MAYER C, DANELLJAN M, BHAT G, et al. Transforming model prediction for tracking[C]//2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition. New Orleans: IEEE, 2022: 8721−8730.

[27] YAN Bin, PENG Houwen, FU Jianlong, et al. Learning spatio-temporal transformer for visual tracking[C]//2021 IEEE/CVF International Conference on Computer Vision. Montreal: IEEE, 2022: 10428−10437.

作者简介：

陈泷，硕士研究生，主要研究方向为电子数据取证。E-mail: 2023211395@stu.ppsuc.edu.cn。

丁锰，副教授，全国刑事技术标准化技术委员会委员，主要研究方向为电子数据取证，发表学术论文 20 余篇。E-mail: dingmeng@ppsuc.edu.cn。

石磊，副研究员，中国人工智能学会智能服务专委会委员，主要研究方向为智能信息处理、大数据分析与挖掘、社交网络搜索、人工智能。发表学术论文 40 余篇。E-mail: leiky_shi@cuc.edu.cn。

[责任编辑：丁锰]