中图法分类号:TP391　　**文献标识码**: A　　**文章编号**: 1006-8961(XXXX)XX-0001-17

论文引用格式:Jiang Yanji, Zong Yali, Dong Hao, Zhang Haiyang, Liu Daqian, Fei Bowen, Chen Pengda. Nighttime object tracking framework via fusing darkness perception prompts[J/OL]. Journal of Image and Graphics, XXXX:1-17. DOI: 10.11834/jig.260034. (姜彦吉, 宗亚利, 董浩, 张海洋, 刘大千, 费博雯, 陈鹏达. 融合黑暗感知提示的夜间目标跟踪框架[J/OL]. 中国图象图形学报, XXXX:1-17. DOI: 10.11834/jig.260034.)[DOI:10.11834/jig.260034]

融合黑暗感知提示的夜间目标跟踪框架

姜彦吉[1,4],宗亚利[1,4],董浩[2,3,4],张海洋[5],刘大千[1],费博雯[1],陈鹏达[1,4]

1. 辽宁工程技术大学软件学院,辽宁　葫芦岛　125105; 2. 苏州工学院汽车工程学院,江苏　苏州　215104; 3. 清华大学苏州汽车研究院(相城),江苏　苏州　215100; 4. 优策(江苏)安全科技有限公司Opensafe实验室,江苏　苏州　215100; 5. 西交利物浦大学计算机科学与技术学院,江苏　苏州　215104

摘　要: **目的**　针对夜间低光照环境下目标特征退化与背景噪声导致跟踪器特征匹配失效及模板更新漂移的问题。提出一种融合黑暗感知提示的夜间目标跟踪框架(ProDAPT),不同于现有方法仅在输入端进行简单的像素级叠加或浅层特征注入,ProDAPT创新性地以冻结的Transformer为基础,构建全链路黑暗感知提示机制。**方法**　首先,提出跨层层级一致性提示生成器(CTCP),利用迭代反向投影与跨层语义约束,在深层特征空间中强制恢复被噪声稀释的目标结构;其次,设计提示语义校准注意力(PSCA),通过提示特征的结构先验作为显式偏置校正Transformer的注意力分布,有效抑制夜间相似干扰物导致的注意力弥散;最后,提出能量感知双重门控更新策略(EDGU),利用提示能量作为独立于分类分数的结构完整性的度量指标,实现更加可靠的动态模板更新。**结果**　在NAT2021、LLOT和UAVDark135三个主流基准数据集及自采的4K分辨率自动驾驶跟车数据集上进行实验。该算法在三个公开数据集上的成功率分别达到0.557、0.585和0.608,其中在NAT2021上相比同类提示学习方法DCPT提升了3.1%;在NVIDIA A100硬件上的推理速度达到76.8FPS,参数量仅为全量微调的7.03%,并在自采数据集的真实夜间场景中也展示了良好跟踪效果。**结论**　该方法通过全链路黑暗感知机制,有效增强了夜间特征鉴别力与更新可靠性,为解决夜间视觉跟踪任务中的感知与决策断层问题提供了新的有效途径。

关键词: 视觉目标跟踪;夜间跟踪;视觉提示学习;Transformer;跨层一致性;注意力校准;自动驾驶

Nighttime object tracking framework via fusing darkness perception prompts

Jiang Yanji[1,4], Zong Yali[1,4], Dong Hao[2,3,4], Zhang Haiyang[5], Liu Daqian[1], Fei Bowen[1], Chen Pengda[1,4]

1. School of Software, Liaoning Technical University, Liaoning Huludao 125105, China; 2. School of Automotive Engineering, Suzhou University of Technology, Jiangsu Suzhou 215104, China; 3. Suzhou Automobile Research Institute, Tsinghua University, Jiangsu Suzhou 215100, China; 4. OpenSafe Lab, Utcer (Jiangsu) Safety Technology Co., Ltd., Jiangsu Suzhou 215100, China; 5. The School of Computer Science, Technology of Xi'an Jiaotong-Liverpool University, Jiangsu Suzhou 215104, China

Abstract: **Objective** Visual Object Tracking serves as a fundamental capability in computer vision, underpinning critical applications such as autonomous driving systems and unmanned aerial vehicle surveillance. While trackers based on the Transformer architecture have demonstrated robust performance in standard illumination conditions, their efficacy declines

收稿日期:2026-01-15;**修回日期**:2026-02-28

基金项目:国家自然科学基金青年基金项目(62302509,62303477);国家自然科学基金面上项目(52274205);浙江省自然科学基金面上项目(LMS25G010003);2025年中央引导地方科技发展专项资金(STKJ2025107);广东省科技创新战略专项市县科技创新支撑项目(STKJ2023071)

Supported by: Project supported by the National Natural Science Foundation of China (Grant Nos. 62302509, 62303477, 52274205); the Natural Science Foundation of Zhejiang Province, China (Grant No. LMS25G010003); the Central Guidance on Local Science and Technology Development Fund of China (Grant No. STKJ2025107); the Science and Technology Innovation Strategy Fund of Guangdong Province (Grant No. STKJ2023071)

markedly in nighttime environments. This performance degradation arises primarily from two factors. First, low illumination conditions lead to the loss of texture details and edge information, a phenomenon known as feature degradation, which impedes the ability of the feature extractor to distinguish the target object from the background. Second, nighttime scenes are frequently characterized by high-frequency background noise and intense interference sources, such as streetlights and vehicle headlights. These distractors often exhibit visual characteristics similar to the target, resulting in a phenomenon termed High-Score Drift. In this scenario, the tracker incorrectly identifies a background light source as the target with high classification confidence, leading to the contamination of the dynamic template and irreversible tracking failure. Existing approaches to mitigate these issues fall into two main categories: low-light enhancement and domain adaptation. Enhancement-based methods introduce additional image processing modules, which increase computational latency and may amplify noise artifacts. Domain adaptation methods attempt to align nighttime features with daytime domains but often fail to preserve fine-grained geometric structures required for precise localization. Although recent visual prompt learning methods, such as DCPT, attempt to inject darkness cues into trackers, they predominantly operate at the input level. As the network depth increases in deep Transformers, these prompt features undergo feature dilution, losing their semantic guidance capability in deeper layers. Furthermore, current methods lack specific mechanisms to rectify attention dispersion caused by background distractors. To address these limitations, this study proposes a framework named Nighttime Object Tracking via Cross-layer Consistent Darkness Perception Prompts (ProDAPT). The primary objective is to bridge the disconnect between feature extraction and decision-making in low-light scenarios by constructing a full-link darkness perception mechanism within a frozen Transformer backbone. **Method** The ProDAPT framework is constructed upon the ProContEXT baseline, utilizing a standard ViT-Base backbone pre-trained on daytime datasets. To ensure generalization capabilities and maintain parameter efficiency, the backbone parameters remain frozen, and only the proposed prompt-related modules are subjected to fine-tuning. The methodology comprises three innovative components designed to enhance feature extraction, attention allocation, and template updating. First, a Cross-layer Consistent Prompt Generator is introduced to address the issue of feature dilution. Unlike methods that rely on simple pixel-level superposition, this generator adopts an iterative back-projection mechanism. It consists of emphasizing blocks and undermining blocks that iteratively estimate the residual difference between the target structure and background noise. Crucially, a cross-layer semantic constraint is applied. This mechanism enforces a consistency loss between the prompts generated in deep layers and the semantic projections of prompts from shallow layers. This constraint compels the network to retain fine-grained geometric structures, such as edges and shapes captured in shallow layers, and transmit them to deep semantic layers, thereby facilitating the recovery of target details submerged in noise. Second, a Prompt Semantic Calibration Attention module is designed to resolve the attention dispersion issue. In low-light conditions, the standard self-attention mechanism often assigns high weights to bright background distractors. This module integrates the generated prompts into the Transformer self-attention calculation by computing the cosine similarity between the prompts of the template and the prompts of the search region. The resulting structural correlation matrix serves as an explicit structural bias which is injected into the original Query-Key attention map. By increasing the attention weights of regions that share similar prompt structures, the module effectively calibrates the attention distribution, reducing the response of background distractors even when they exhibit high pixel intensity. Third, an Energy-aware Dual Gating Update strategy is proposed to address the unreliability of template updates. Traditional update mechanisms rely solely on classification scores, which are prone to drift at night. This strategy introduces Prompt Energy, calculated via the L_2 norm of the prompt features, as an orthogonal metric to evaluate structural integrity. The theoretical basis is that a true target retains high prompt energy due to structural consistency across layers, whereas background noise exhibits low energy after the filtering process of the prompt generator. A dual-gating logic is established wherein updates are permitted only when both the classification score and the prompt energy meet specific thresholds. This mechanism prevents the template bank from being contaminated by high-confidence background noise. **Result** The experimental evaluation was conducted on three mainstream nighttime tracking benchmarks: NAT2021, LLOT, and UAVDark135. To further verify engineering feasibility and generalization in real-world scenarios, a dataset was collected using autonomous driving test vehicles in Suzhou, China. This self-collected dataset features videos recorded with high-specification on-board cameras at 4K Ultra-HD resolution (3840 by 2160 pixels) and a frame rate of 30 frames per sec-

ond. The testing sequences cover challenging scenarios including glare from oncoming vehicles, rain reflection on the road surface, and temporary occlusion during car-following tasks. Experimental results indicate that ProDAPT achieves competitive performance across the tested benchmarks. On the challenging NAT2021 dataset, the Success Rate (AUC) reaches 0.557, representing an increase of 3.1% compared to the prompt-learning method DCPT and outperforming domain adaptation methods. On the LLOT and UAVDark135 datasets, the AUC scores reach 0.585 and 0.608, respectively. Ablation studies confirm that the Cross-layer Consistent Prompt Generator, Prompt Semantic Calibration Attention, and Energy-aware Dual Gating Update modules contribute incrementally to the performance, with the full model achieving the best results in terms of both center precision and overlap accuracy. In terms of computational efficiency, tests conducted on an NVIDIA A100 GPU show that ProDAPT maintains an inference speed of 76.8 frames per second, which exceeds the real-time requirement of 30 frames per second. Although there is a minor decrease in speed compared to the baseline due to the additional modules, the trade-off is justified by the performance gains. The model requires training 6.51 million parameters, accounting for 7.03% of the total model parameters, indicating a balance between tracking accuracy and computational cost compared to full fine-tuning methods. Qualitative results on the 4K self-collected dataset demonstrate that ProDAPT maintains stable bounding boxes on the leading vehicle in the presence of severe glare interference, whereas baseline trackers exhibit drift towards nearby streetlights. **Conclusion** This study presents ProDAPT, a unified framework that addresses the critical challenges of feature dilution and decision unreliability in nighttime tracking. Through the introduction of the Cross-layer Consistent Prompt Generator, the framework ensures that darkness cues are preserved throughout the deep network layers. The Prompt Semantic Calibration Attention and Energy-aware Dual Gating Update mechanisms further support adaptability against background distractors and prevent erroneous template updates. The extensive experiments on public benchmarks and high-quality self-collected 4K data validate the effectiveness and generalization capability of the method. ProDAPT provides a feasible and efficient solution for visual perception in nighttime autonomous driving and unmanned aerial vehicle applications. Future work will focus on optimizing the model for edge deployment on embedded devices and exploring the integration of multi-modal data, such as infrared or thermal imaging, into the prompt generation process to further enhance tracking reliability in extreme darkness.

Key words: Visual object tracking; Nighttime tracking; Visual prompt learning; Transformer; Cross-layer consistency; Attention calibration; Autonomous driving

0 引　言

视觉目标跟踪(visual object tracking,VOT)旨在仅凭首帧状态预测后续轨迹与尺度。近年来,随着深度神经网络的演进,尤其是Transformer架构的引入,VOT技术取得了显著进展(Kugarajeevan等,2023)。变换器跟踪(transformer tracking,TransT)(Chen等,2021)、时空架构跟踪(spatio-temporal architecture for tracking,STARK)(Yan等,2021)、混合注意力变换器(mixed attention transformer,MixFormer)(Cui等,2022)、单流跟踪(one-stream tracking,OSTrack)(Ye等,2022)、渐进式上下文变换器(progressive context transformer,ProContEXT)(Lan等,2023)以及Xu等通过增强特征融合(Xu等,2025),在大规模单目标跟踪数据集(large-scale single object tracking,LaSOT)(Fan等,2021)、跟踪网络数据集(tracking network,TrackingNet)(Muller等,2018)和通用目标跟踪-10k数据集(generic object tracking-10k,GOT-10k)(Huang等,2021)等上均获最先进的(state-of-the-art,SOTA)性能。此外,自回归跟踪(autoregressive tracking,ARTrack)(Wei等,2023)和序列到序列跟踪(sequence-to-sequence tracking,SeqTrack)(Chen等,2023)等方法进一步探索了将跟踪建模为序列生成任务,而层级交互提示跟踪(hierarchical interaction prompt tracking,HIPTrack)(Cai等,2024)则通过层级交互进一步提升了复杂场景下的表征能力。

然而,上述方法在夜间低光照场景泛化能力受限。极低照度导致纹理丢失与高噪点干扰,直接迁移日间SOTA模型往往因特征提取失效导致性能骤降(Yi等,2024),且现有的视觉智能评估研究也指出,在光照变化等复杂环境下,算法的鲁棒性与人类视觉能力相比仍存在差距(Hu等,2024)。

针对夜间跟踪的困境，现有研究主要遵循几条技术路线，但均存在局限性。第一类是先增强后跟踪范式，利用零参考深度曲线估计（zero-reference deep curve estimation，Zero-DCE）（Guo等，2020）、黑暗增亮器（dark lighter，DarkLighter）（Ye等，2021）、高光网络（highlight network，HighlightNet）（Fu等，2022）以及融合正则化的图像增强方法（Shao等，2025）来改善画质，但级联结构增加延迟且易放大噪声，会破坏目标的边缘结构，导致跟踪器在增强后的图像上依然难以锁定目标。第二类是抗干扰注意力设计。为了缓解夜间特征模糊的问题，抗黑跟踪（anti-dark tracking，ADTrack）（Li等，2021）提出了目标感知的双滤波器学习机制，而孪生注意力聚合网络++（siamese attentional aggregation network++，SiamAPN++）（Cao等，2021）则通过注意力聚合网络增强了对微小目标的捕捉能力。虽增强抗扰性，但在极暗条件下难解语义丢失。第三类是无监督域自适应（unsupervised domain adaptation，UDA）方法。无监督域适应跟踪-相关对齐规约（unsupervised domain adaptation tracking-correlation alignment reduction，UDAT-CAR）（Ye等，2022）和变换器特征融合与相关对齐规约（transformer feature fusion with correlation alignment reduction，TransffCAR）（Wei等，2024）试图通过对抗学习或风格迁移将夜间特征对齐到日间域，或者利用Transformer桥接层来缩小域偏移。但难以保留细微几何结构，且生成对抗网络的训练往往不够稳定，难以适应多变的夜间动态场景。第四类是近期兴起的视觉提示学习（visual prompt learning，VPL）。该范式旨在通过注入少量可学习参数来激发冻结的预训练模型对夜间特征的感知能力，避免了全参数微调的高昂成本。视觉提示跟踪（visual prompt tracking，ViPT）（Zhu等，2023）初步探索了多模态提示的有效性，而黑暗线索提示跟踪（darkness clue-prompted tracking，DCPT）（Zhu等，2024）是这一方向在夜间单模态跟踪中的代表性工作。DCPT通过反向投影机制在输入端注入黑暗线索提示，在参数效率和性能上均优于传统方法。随后，历史黑暗线索提示跟踪（historical darkness clue-prompted tracking，H-DCPT）（Zhong等，2026）进一步引入了历史信息提示，以增强时序上的连贯性。

然而，DCPT类方法存在浅层化与交互弱的本质缺陷，提示信息多在网络浅层注入，随着Transformer层数的加深，微弱的暗光线索极易被主干网络平滑或稀释；此外，提示仅作为补充特征参与运算，未能深度干预Transformer核心的自注意力计算与模板更新决策。在夜间复杂场景下，背景中的高亮干扰物极易产生高分漂移现象，即跟踪器错误锁定背景光亮，输出了极高的分类置信度，导致错误的特征被更新进动态模板库，最终造成不可逆的跟踪失败。

为了突破上述局限，提出了融合黑暗感知提示的夜间目标跟踪框架（ProDAPT）。通过提示微调，有效提取、保持并利用跟踪对象在暗区的特征线索作为指导信号。以ProContEXT为基线，构建了一套贯穿特征提取至决策更新的全链路感知机制。

首先通过跨层层级一致性提示生成器（cross-layer consistent prompt generator，CTCP），将基于迭代反向投影的提示生成逻辑扩展为跨层约束的机制，有效防止了暗光线索在ViT深层网络中的稀释；随后，针对匹配漂移问题，设计了提示语义校准注意力（prompt semantic calibration attention，PSCA），利用提示特征的结构先验作为显式偏置来修正Transformer的QK矩阵；最后，针对模板更新问题，提出了能量感知双重门控更新策略（energy-aware dual gating update，EDGU），通过引入提示能量作为目标结构性指标，建立了双重门控逻辑，解决了传统更新机制在夜间的不可靠问题。同时在夜间空中跟踪2021数据集（nighttime aerial tracking 2021，NAT2021）（Ye等，2022）、低光照目标跟踪数据集（low-light object tracking，LLOT）（Zhong等，2026）和无人机夜间跟踪135数据集（unmanned aerial vehicle dark tracking 135，UAVDark135）（Li等，2023）具有挑战性的夜间跟踪基准数据集上进行了对应的实验评估。

1 方 法

1.1 总体框架

针对夜间场景特征鉴别力下降与时序决策不可靠问题，提出一种融合黑暗感知提示的夜间目标跟踪框架（ProDAPT）。该框架以ProContEXT为基线，保持预训练的ViT（vision transformer）主干参数冻结，引入了孪生提示注入机制，即CTCP生成的提示不仅注入搜索区域分支，同时也注入静态与动态模板分支。这种共享权重的注入方式确保了模板与搜索区域在统一的暗光增强特征空间中进行交互。

如图1所示，CTCP在特征提取阶段挖掘并维持暗光线索；PSCA在特征融合阶段利用结构先验校正注意力分布；EDGU在决策阶段基于提示能量验证目标质量，防止错误更新。

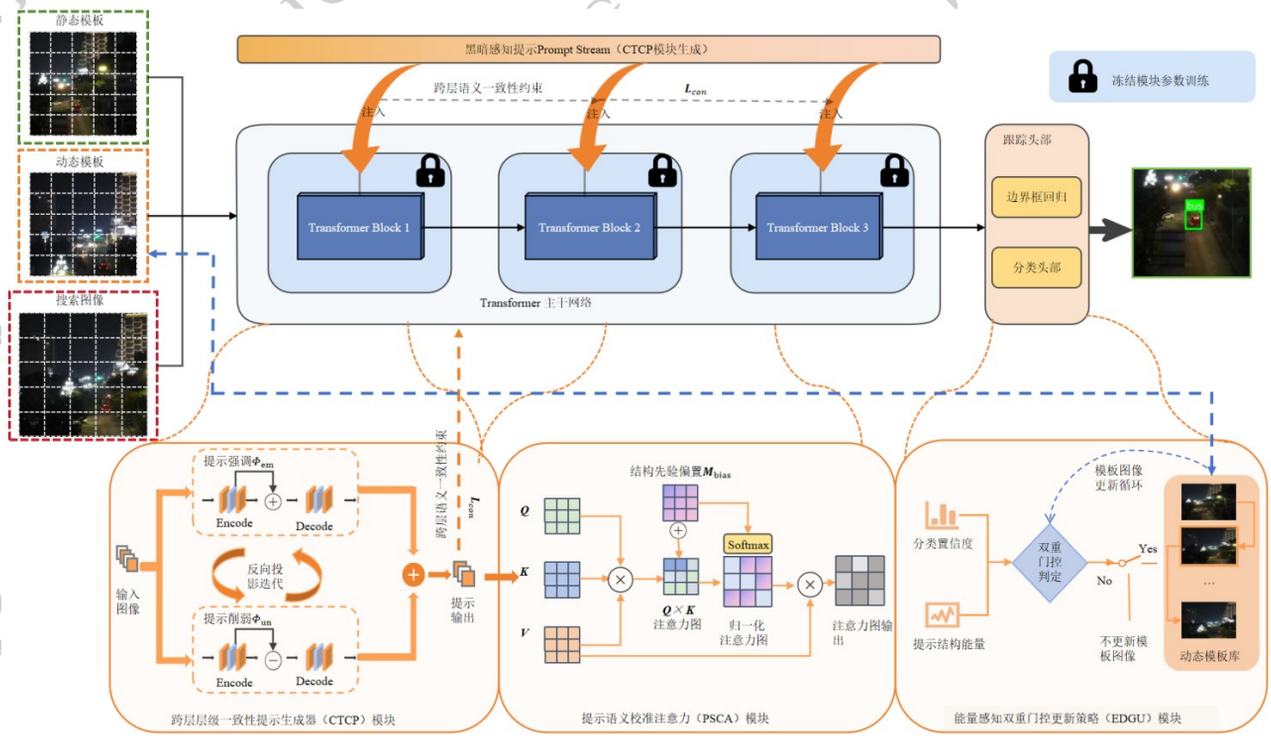


图1　ProDAPT总体架构图

Figure1　Overall Architecture Diagram of ProDAPT

1.2　跨层层级一致性提示生成器(CTCP)

针对夜间微弱目标易在深层网络被噪声淹没的问题，设计了CTCP模块。如图2所示，该模块通过迭代反向投影与跨层约束，在特征空间强制恢复被稀释的几何结构。

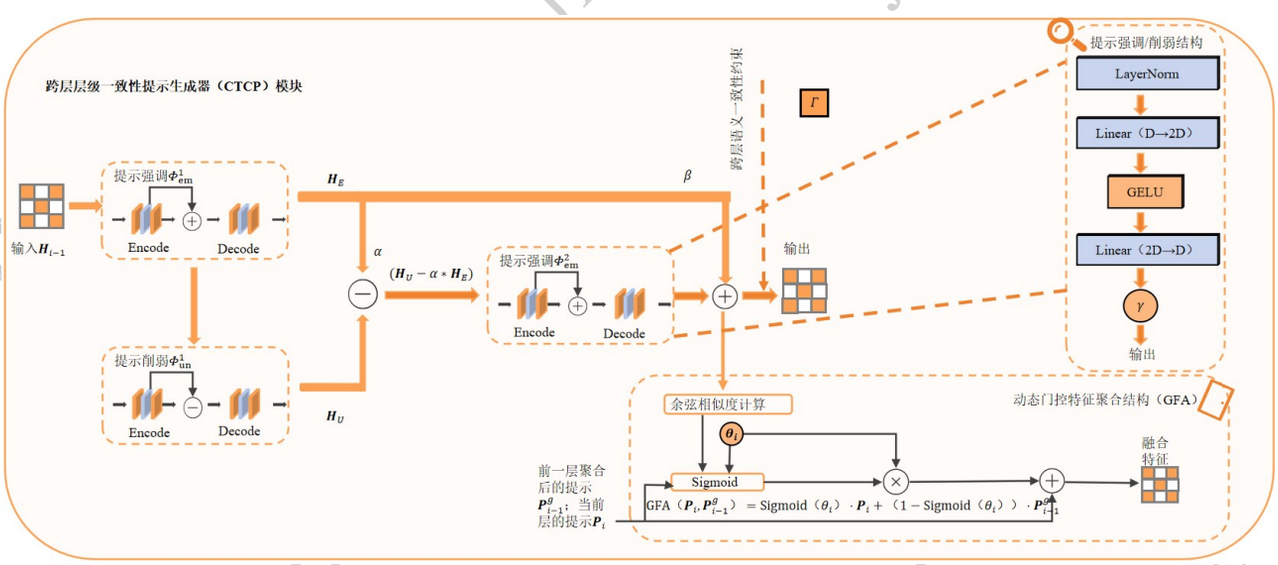


图2　CTCP框架图

Figure2　CTCP Framework Diagram

1.2.1　基于反向投影的提示生成器

为了从低光照特征中挖掘潜在模式，采用了受DCPT(Zhu等，2024)反向投影在超分辨率领域的成功启发，构建了一个残差式的提示生成结构，在ViT主干的第i层编码器前嵌入一个轻量级的提示生成器G_i，不同于处理原始图像，第i层的提示生成器则以第$i-1$层ViT编码器输出的特征图$\boldsymbol{H}_{i-1}$作为输入；然后通过一系列的提示强调与削弱操作，从特征中挖掘暗光线索。

生成器$\boldsymbol{G}_i$核心是由三个结构相同但参数独立的倒残差MLP(multi-layer perceptron)块构成的迭代处理流程。包含层归一化(Layer Norm)、通道扩展投影(D → 2D)、GELU(Gaussian error linear unit)激活以及通道压缩投影(2D → D)，通过非线性映射挖掘深层特征中的隐含模式。处理流程如下：

首先，第$i-1$层输入特征$\boldsymbol{H}_{i-1}$通过第一个强调模块$\boldsymbol{\Phi}^1_{\text{em}}$，生成突出目标轮廓与纹理的强调特征$\boldsymbol{H}_E$，公式为：

$$\boldsymbol{H}_E=\boldsymbol{\Phi}^1_{\text{em}}\left(\boldsymbol{H}_{i-1}\right) \tag{1}$$

式中，$\boldsymbol{H}_{i-1}$为主干网络第$i-1$层的输入特征图；$\boldsymbol{\Phi}^1_{em}$为第一个强调模块；$\boldsymbol{H}_E$为强调后的特征。

接着，为了验证提取特征的可靠性，将$\boldsymbol{H}_E$输入参数独立地削弱模块$\boldsymbol{\Phi}^1_{\text{un}}$。该模块从强调特征中抹除已提取的显著特征，生成残差基准$\boldsymbol{H}_U$，公式为：

$$\boldsymbol{H}_U=\boldsymbol{\Phi}^1_{\text{un}}\left(\boldsymbol{H}_E\right) \tag{2}$$

式中，$\boldsymbol{\Phi}^1_{\text{un}}$为第一个削弱模块；$\boldsymbol{H}_U$为削弱后的特征。

最后，计算反向投影残差($\boldsymbol{e}_U=\boldsymbol{H}_U-\alpha\cdot\boldsymbol{H}_E$)，并通过第二个强调模块$\boldsymbol{\Phi}^2_{\text{em}}$将其映射回提示空间，与原始强调特征融合。这一过程迫使网络通过残差学习区分真实的暗光细节与随机噪声。最终生成的原始提示$\boldsymbol{P}_i^{\text{raw}}$表示为：

$$\boldsymbol{P}_i^{\text{raw}}=\beta\boldsymbol{H}_E+\boldsymbol{\Phi}^2_{\text{em}}\left(\boldsymbol{H}_U-\alpha\boldsymbol{H}_E\right) \tag{3}$$

式中，α和β为平衡残差项的可学习系数，$\boldsymbol{\Phi}^2_{\text{em}}$为第二个强调模块；$\boldsymbol{P}_i^{\text{raw}}$为第$i$层的原始特征。

1.2.2　跨层语义一致性约束

为防止深层抽象特征丢失微弱边缘信息，引入跨层约束机制。不同于独立生成各层提示，该机制并不直接约束主干网络的特征图，而是在提示空间内进行对齐。将第i层的提示与第$i-k$层的提示建立约束关系，强制深层提示在几何结构上与经过语义映射的浅层提示保持一致，最终的提示$\boldsymbol{P}_i$计算如下。

$$\boldsymbol{P}_i=\boldsymbol{P}_i^{\text{raw}}+\gamma\cdot\mathcal{T}\left(\boldsymbol{P}_{i-k}\right) \tag{4}$$

式中，$i\leqslant k$，k=3为跨层步长；$\mathcal{T}$为1×1卷积实现的语义映射投影；γ为可学习系数；$\boldsymbol{P}_{i-k}$为第$i-k$层的提示特征；$\boldsymbol{P}_i$为第i层最终生成的提示特征。

为了强制深层黑暗感知提示在几何结构上记忆浅层的细粒度边缘信息，通过一致性损失$\boldsymbol{L}_{\text{con}}$显式地防止暗光线索在ViT中消失，$\boldsymbol{L}_{\text{con}}$被定义为两者的L2距离，计算公式如下。

$$\boldsymbol{L}_{\text{con}}=\frac{1}{\mathbf{N}}\sum_{j=1}^{N}\left|\left|\boldsymbol{P}_{ij}-\mathcal{T}\left(\boldsymbol{P}_{i-kj}\right)\right|\right|_2^2 \tag{5}$$

式中，$|\cdot|_2^2$表示对所有token计算均方误差(mean squared error，MSE)；j为token的索引；N为特征图中的token的总数量。

该约束迫使网络记忆浅层丰富的细粒度纹理，并将其传递至深层语义空间。

1.2.3　动态注入

在提示与特征的融合环节，改进传统固定权重注入方式，设计提示—特征相似度驱动的token级权重，动态调整提示对不同空间位置特征的贡献。

首先计算$\boldsymbol{P}_i$与输入特征$\boldsymbol{H}_{i-1}$在通道维度上的逐token余弦相似度，量化提示与特征的语义匹配度；随后生成token级注入权重，公式为：

$$\boldsymbol{p}_i=\frac{1}{1+\mathrm{e}^{-\gamma_i}}\cdot\boldsymbol{S}_{\text{sim}} \tag{6}$$

式中，γ_i为可学习参数，控制Sigmoid函数的灵敏度；$\boldsymbol{S}_{\text{sim}}$为余弦相似度；$\boldsymbol{p}_i$为token级注入权重。

最后通过门控特征聚合(gated feature aggregation，GFA)实现提示与特征的融合，公式为：

$$\boldsymbol{H}_i^g=\boldsymbol{H}_{i-1}+\boldsymbol{p}_i\cdot\text{GFA}\left(\boldsymbol{P}_i,\boldsymbol{P}_{i-1}^g\right) \tag{7}$$

式中，$\boldsymbol{P}_{i-1}^g$为前一层聚合后的提示；$\boldsymbol{H}_i^g$为第i层融合后的特征。GFA机制为基于门控的可学习线性插值，公式为：

$$\text{GFA}\left(\boldsymbol{P}_i,\boldsymbol{P}_{i-1}^g\right)=\sigma(\boldsymbol{\theta}_i)\cdot\boldsymbol{P}_i+(1-\sigma(\boldsymbol{\theta}_i))\cdot\boldsymbol{P}_{i-1}^g \tag{8}$$

式中，$\sigma(\cdot)$表示Sigmoid激活函数，$\boldsymbol{\theta}_i$为可学习的门控系数。

该机制通过自动平衡当前层提示与上一层聚合提示的权重，确保了黑暗特征在深层网络中的层间连续性。

1.2.4　CTCP算法

为了明确结构细节，算法1简要说明了跨层层级一致性提示(CTCP)生成算法的前向推理过程。

该算法以当前层特征、跨层浅层提示以及上一层聚合提示为输入，依次经过基于反向投影的特征强调与削弱、残差重构、跨层一致性约束，最终输出增强后的特征与当前层的聚合提示。具体执行步骤如算法1所示。

1.3 提示语义校准注意力（PSCA）

在获得融合了暗光线索的特征后，传统的自注意力机制在暗处的背景干扰物往往与目标具有相似的模糊特征，导致注意力机制错误分配权重。PSCA模块将生成的黑暗感知提示视为显式结构先验，用于校正Transformer的注意力分布。结构框架图如图3所示。

1.3.1 提示相似度计算

ProContEXT采用拼接视角的自注意力机制。为纠正夜间纹理缺失导致的Query-Key匹配偏差，分别提取最新动态模板的提示分量和搜索区域的提示分量。由于黑暗感知提示本质上是经过强调和削弱筛选后的结构化高频信息，其在目标区域具有高度的语义一致性。因此利用余弦相似度量化两者在暗光线索空间中的结构匹配程度，构建注意力先验矩阵$\boldsymbol{M}_{\text{bias}}$。计算如下：

$$\boldsymbol{S}_{\text{zx}}=\frac{\boldsymbol{P}_{\text{z}}\cdot\left(\boldsymbol{P}_{\text{x}}\right)^{T}}{\left\|\boldsymbol{P}_{\text{z}}\right\|\cdot\left\|\boldsymbol{P}_{\text{x}}\right\|} \tag{9}$$

$$\boldsymbol{M}_{\text{bias}}=Co\left(\left[\boldsymbol{S}_{\text{zz}},\boldsymbol{S}_{\text{zx}}\right],\left[\boldsymbol{S}_{\text{xz}},\boldsymbol{S}_{\text{xx}}\right]\right) \tag{10}$$

式中，$\boldsymbol{P}_{\text{z}}$为动态模板的提示分量，$\boldsymbol{P}_{\text{x}}$为搜索区域的提示分量，$\boldsymbol{S}_{\text{zx}}$数值限制在[-1,1]，显式地量化了模板与搜索区域在暗光线索空间中的相似度；Co为拼接操作；$\boldsymbol{M}_{\text{bias}}$为暗光注意力先验。

由于数值被限制在[-1,1]之间，正值表示提示特征在几何结构上高度一致，能够增强对目标区域的关注；负值则意味着结构冲突，有助于抑制背景噪声。通过引入这一偏置，即使原始图像特征模糊，只要两者共享相似的暗光结构，注意力权重就会被强化，从而实现语义纠偏。

1.3.2 提示引导的注意力校正

将计算出的相似度矩阵$\boldsymbol{M}_{\text{bias}}$作为偏置项注入Transformer的注意力计算中。为了在保留原始语义与利用暗光先验之间找到平衡，引入可学习系数λ来动态控制校正强度。校正后的注意力计算公式为：

$$At\left(\boldsymbol{Q},\boldsymbol{K},\boldsymbol{V}\right)=So\left(\frac{\boldsymbol{Q}\boldsymbol{K}^{T}}{\sqrt{d_k}}+\lambda\cdot\boldsymbol{M}_{\text{bias}}\right)\boldsymbol{V} \tag{11}$$

式中，So为Softmax，d_k为特征维度；λ为可学习标量。

在训练初期将λ初始化为0来保证梯度稳定性。通过这种显式的语义引导，跟踪器能够聚焦于由黑暗感知提示点亮的目标区域，显著减少了夜间相似干扰物的影响。

1.4 能量感知双重门控更新策略（EDGU）

针对夜间跟踪常发的高分漂移现象，即跟踪器错误锁定高亮背景干扰物并输出高置信度，单纯依赖分类分数进行模板更新已不可靠。EDGU策略引入提示能量作为独立的结构完整性度量。结构框架图如图4所示。

1.4.1 提示能量计算

将生成的黑暗感知提示视为一种用于补偿低光退化的语义信号。基于1.2节所述的反向投影机制，提示生成器包含强调和削弱的对抗过程。将范数L_2作为衡量特征激活强度的标准度量，这符合信号处理中能量的经典定义。尽管背景噪声与目标纹理在原始图像中均表现为高频信号，但两者在深层

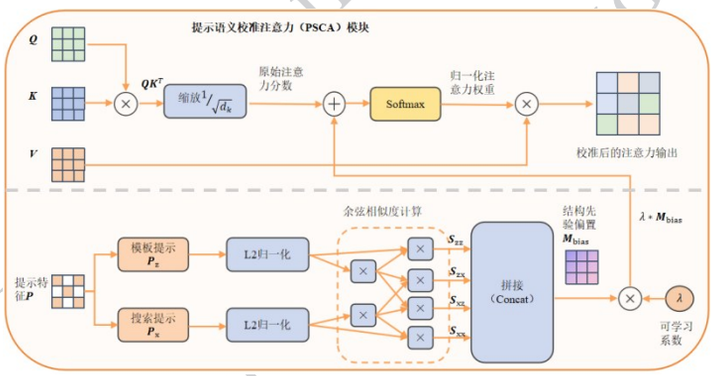


图3 PSCA框架图

Figure3 PSCA Framework Diagram

算法1 跨层层级一致性提示(CTCP)生成算法

行号	算法内容
输入：	当前层特征 $\boldsymbol{H}_{i-1}$；跨层浅层提示 $\boldsymbol{P}_{i-k}$；上一层聚合提示 $\boldsymbol{P}_{i-1}^{g}$
参数：	独立参数的MLP增强/削弱块 $\boldsymbol{\Phi}$；语义映射投影 $\mathcal{T}$，门控参数 $\boldsymbol{\theta}_i$，$\boldsymbol{\gamma}_i$；可学习系数 α、β、γ
输出：	增强后的特征 $\boldsymbol{H}_i^{g}$；当前层聚合提示 $\boldsymbol{P}_i^{g}$
1:	阶段1：基于反向投影的特征强调
2:	$\boldsymbol{H}_E \leftarrow \boldsymbol{\Phi}_{\text{em}}^{1}(\text{LayerNorm}(\boldsymbol{H}_{i-1}))$//突出潜在目标结构
3:	阶段2：基于反向投影的特征削弱
4:	$\boldsymbol{H}_U \leftarrow \boldsymbol{\Phi}_{\text{un}}^{1}(\text{LayerNorm}(\boldsymbol{H}_E))$//抹除显著特征生成基准
5:	阶段3：基于反向投影的残差提炼与重构
6:	$\boldsymbol{e}_U \leftarrow \boldsymbol{H}_U - \alpha \cdot \boldsymbol{H}_E$//计算反向投影残差
7:	$\boldsymbol{R} \leftarrow \boldsymbol{\Phi}_{\text{em}}^{2}(\text{LayerNorm}(\boldsymbol{e}_U))$//将残差映射回提示空间
8:	阶段4：基于反向投影的原始提示生成
9:	$\boldsymbol{P}_{\text{raw}} = \beta \boldsymbol{H}_E + \boldsymbol{R}$//生成原始提示
10:	阶段5：跨层语义一致性约束
11:	If $i > k$ then
12:	$\boldsymbol{P}_i \leftarrow \boldsymbol{P}_i^{\text{raw}} + \gamma \cdot \mathcal{T}(\boldsymbol{P}_{i-k})$//注入浅层边缘细节
13:	else
14:	$\boldsymbol{P}_i \leftarrow \boldsymbol{P}_i^{\text{raw}}$
15:	endif
16:	阶段6：门控聚合与动态注入
17:	$\boldsymbol{P}_i^{g} \leftarrow \text{Sigmoid}(\boldsymbol{\theta}_i) \cdot \boldsymbol{P}_i + (1 - \text{Sigmoid}(\boldsymbol{\theta}_i)) \cdot \boldsymbol{P}_{i-1}^{g}$//GFA聚合
18:	$\boldsymbol{S}_{\text{sim}} \leftarrow \text{CosineSimilarity}(\boldsymbol{P}_i, \boldsymbol{H}_{i-1})$//计算通道相似度
19:	$\boldsymbol{p}_i \leftarrow \frac{1}{1+\text{e}^{-\gamma_i}} \cdot \boldsymbol{S}_{\text{sim}}$//生成注入权重
20:	$\boldsymbol{H}_i^{g} \leftarrow \boldsymbol{H}_{i-1} + \boldsymbol{p}_i \cdot \boldsymbol{P}_i^{g}$//最终特征注入
21:	return $\boldsymbol{H}_i^{g}, \boldsymbol{P}_i^{g}$

特征空间存在本质差异，目标结构具备跨层一致性与语义连续性，而随机噪声因缺乏语义支撑在深层传递中被逐渐过滤。

得益于CTCP模块引入的跨层一致性约束，提示生成器被训练为仅对具备时空连续性的几何结构产生高响应。真实目标在迭代反向投影中被显著增强，而背景高光虽像素值高，却本质上属于缺乏跨层支撑的非语义杂波，在一致性正则化下被层级式抑制。因此，特征图的L_2范数有效实现了从像素亮度到结构置信度的语义转化，高能量E_{p}代表模型提取到了清晰、稳定的目标结构，低能量E_{p}则意味着当前区域仅包含被抑制的无效信息。

所以将当前帧生成的最终层提示$\boldsymbol{P}_{\text{f}}$的$L_2$范数定义为“提示能量”：

$$E_{\text{p}} = \left|\left|\boldsymbol{P}_{\text{f}}\right|\right|_2 \tag{12}$$

式中，E_{p}为提示能量；$\boldsymbol{P}_{\text{f}}$为当前帧生成的最终层提示特征。

1.4.2 双重门控更新逻辑

基于提示能量与分类分数的互补性，将模板更新的决策逻辑设计为双重门控机制。该机制不再盲目信任分类分数，而是根据分数的置信区间引入能量约束：

$$Up = \begin{cases} \text{True}, & \text{if} \begin{pmatrix} S_{\text{cls}} > \tau_{\text{h}} \\ \wedge E_{\text{p}} > \tau_{\text{b}} \end{pmatrix} \\ \text{True}, & \text{if} \begin{pmatrix} \tau_{\text{l}} < S_{\text{cls}} \leqslant \tau_{\text{h}} \\ \wedge E_{\text{p}} > \tau_{\text{s}} \end{pmatrix} \\ \text{False}, & \text{otherwise} \end{cases} \tag{13}$$

式中，S_{cls}为分类分支输出的置信度分数，τ_{h}、τ_{l}、τ_{b}和τ_{s}分别为置信度高阈值、低阈值、基础阈值、限制阈值。

该逻辑的物理含义为，当分类置信度极高时，需通过基础能量阈值τ_{b}排除纯光斑干扰；当置信度中等时（夜间常态），必须要求极高的提示能量τ_{s}来确认目标结构的存在。这有效解决了夜间跟踪中跟丢不更新与误跟乱更新的矛盾。

1.5 联合函数

为了实现端到端的提示微调，将跟踪任务损失与跨层一致性损失相结合，总损失函数定义为：

$$\boldsymbol{L}_{\text{tot}} = \boldsymbol{L}_{\text{tra}} + \eta_{\text{con}} \boldsymbol{L}_{\text{con}} \tag{14}$$

式中，跟踪任务损失$\boldsymbol{L}_{\text{tra}}$包含分类损失和回归损失；$\boldsymbol{L}_{\text{con}}$用于约束提示生成的层级一致性；$\eta_{\text{con}}$为平衡系数。

2 实 验

为了全面评估提出的ProDAPT框架的有效性，进行了一系列详尽的实验。

2.1 实现细节

本方法基于ProContEXT构建，ViT-Base主干加载日间预训练权重并冻结。仅对新引入的CTCP、

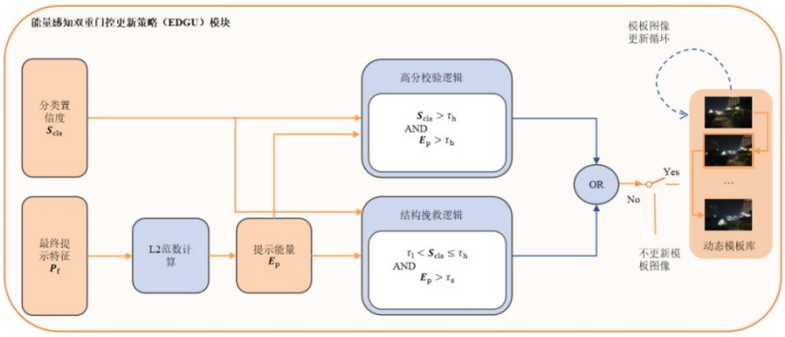


图4　EDGU框架图

Figure4　EDGU Framework Diagram

PSCA模块在LLOT和UAVDark135混合训练集进行参数微调。

超参数设置：微调过程持续60Epochs。优化器采用AdamW，权重衰减设置为10^{-4}，初始学习率设置为4×10^{-4}，并在第50个Epoch后下降10。EDGU阈值设置为τ_h=0.8，τ_l=0.5、τ_b = 0.3和τ_s = 0.7。

输入尺寸：搜索区域为256×256，模板区域为128×128。

设备：实验在搭载NVIDIA A100-PCIE（40G）服务器进行，采用PyTorch框架进行推理。

2.2　数据集与评价指标

在三个主流的夜间跟踪基准数据集上对ProDAPT进行了全面评估。

LLOT（Zhong等，2026）数据集包含269个序列，涵盖低环境照度属性；UAVDark135（Li等，2023）数据集包含135个无人机夜间序列，涉及快速运动与光照变化，这两个数据集按照8∶2随机划分为训练集和测试集，用来训练和作为测试基准评估。

NAT2021（Ye等，2022）数据集包含180个测试序列，存在强噪声干扰，由于训练集是无标签，使用其测试集作为测试基准来评估。

评价指标采用目标跟踪领域通用的单次通过评估（one pass evaluation，OPE）准则下的成功率（Success/AUC），精确率（Precision）和归一化精度（P_norm）。

2.3　与SOTA方法对比

将ProDAPT与近年来先进的跟踪器进行对比，主要包括基于提示学习的DCPT（Zhu等，2024）、H-DCPT（Zhong等，2026）；基于域自适应的UDAT-CAR（Ye等，2022）、TransffCAR（Wei等，2024）；基于渐进式上下文的ProContEXT（Lan等，2023）以及其他SOTA方法：HIPTrack（Cai等，2024）、ARTrack（Wei等，2023）、OSTrack（Ye等，2022）、ADTrack（Li等，2021）、SiamAPN++（Cao等，2021）。

2.3.1　NAT2021数据集评估

如表1所示，ProDAPT在各项指标均获最优，AUC达到0.557，相比DCPT提升3.1%，主要归功于CTCP模块，它克服了DCPT仅在输入端注入提示导致的深层特征稀释问题，确保了在NAT2021这种高噪声环境下目标语义的连续性，且优于域适应方法TransffCAR。得益于冻结主干策略，模型仅需训练6.51M参数，推理速度达76.8FPS，实现了高精度与实时性的平衡。图5展示了ProDAPT在NAT2021数据集上涵盖大型车辆、快速运动目标、极低照度微小行人及密集人群等典型挑战序列中的综合跟踪结果，ProDAPT凭借PSCA模块对背景噪声的有效抑制以及CTCP模块对深层语义的恢复，在所有测试序列中能紧密贴合真值框（绿框），而SiamAPN++等方法易发生漂移。

2.3.2　LLOT数据集评估

如表2所示，ProDAPT在LLOT上取得AUC 0.585及P_norm为0.746的SOTA性能，显著优于基线ProContEXT（AUC 0.557）。可视化结果图6表明，在雨天反光、极低照度等复杂场景下，EDGU策略保证了模板纯净度，使跟踪框能精准贴合目标边缘，而其他算法跟踪框多少都有一些偏移。

2.3.3　UAVDark135数据集评估

如表3所示，ProDAPT在UAVDark135上AUC与精确率达0.608和0.735，较基线ProContEXT提升2.4%和2.3%；相比H-DCPT，凭借能量感知机制仍保持1.0%的AUC优势，验证了其极暗环境下的噪声区分力。图7显示，面对尺度剧变、低分辨及背景

表1　NAT2021-test数据集在不同跟踪器的评估对比表

Table1　Comparison of NAT2021-test dataset evaluation across different trackers

跟踪器(Tracker)	来源(Venue)	NAT2021(AUC)	NAT2022(Prec)	NAT2021(P_norm)	参数量/可学习参数量	推理速度FPS
ProDAPT	Proposed	**0.557**	**0.728**	**0.676**	92.6M/6.51M	76.8
HIPTrack(Cai等,2024)	CVPR′24	0.545	0.708	0.660	120.4M/–	65.0
ProContEXT(Lan等,2023)	ICASSP′23	0.539	0.702	0.655	92.6M/–	85.2
ARTrack(Wei等,2023)	CVPR′23	0.532	0.696	0.642	92.1M/–	42.0
DCPT(Zhu等,2024)	ICRA′24	0.526	0.690	0.635	93.0M/**3.03M**	81.5
OSTrack(Ye等,2022)	ECCV′22	0.525	0.685	0.635	92.1M/–	105.1
H-DCPT(Zhong等,2026)	TIP′24	0.522	0.715	0.668	93.0M/**3.03M**	78.5
TransffCAR(Wei等,2024)	Access′24	0.511	0.721	0.647	–	–
UDAT-CAR(Ye等,2022)	CVPR′22	0.483	0.687	0.564	35.6M/–	45.0
ADTrack(Li等,2021)	ICRA′21	0.445	0.592	0.510	**24.5M**/–	80.1
SiamAPN++(Cao等,2021)	ICRA′21	0.377	0.431	0.461	28.0M/–	**120.2**

注:加粗字体为最优值,–为原文无对应数据

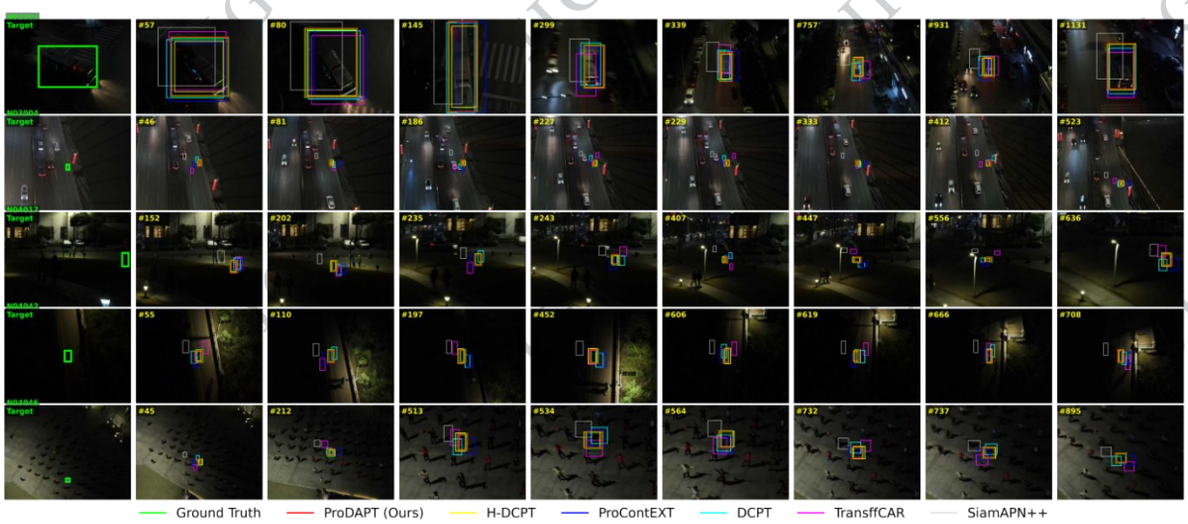


图5　NAT2021-test数据集跟踪可视化对比图

Figure5　Visual Comparison of NAT2021-test Dataset Tracking

干扰,ProDAPT分别利用EDGU适应边缘、CTCP防漂移及PSCA抑制光斑,实现紧密跟踪 。

2.3.4　总体可视化分析

如图8所示,图中展示了ProDAPT三个数据集上的成功率图(Successplots)、精确率图(Precisionplots)和归一化精确率图(NormalizedPrecisionplots)。红色实线代表ProDAPT方法。可以看出,ProDAPT在所有基准测试的各项指标曲线中均处于高位。

2.4　消融实验

为了全面评估ProDAPT框架中各核心模块的有效性,在具有挑战性且未参与微调的NAT2021数据集上进行了定量的消融研究,实验结果见表4。

为了验证核心模块的原创性与有效性,本文在NAT2021-test数据集上结合表4定量数据与图9定性结果进行了联合分析。针对同类提示方法DCPT(Zhu等,2024)及H-DCPT(Zhong等,2026)的对比表明,上述方法虽引入提示机制,但因各层提示独立生成,忽略了深层特征传递的语义连续性,导致其成功率仅为0.526和0.522(见表1)。相比之下,CTCP模块的核心创新在于引入跨层语义一致性约束,强制

表2　LLOT-test数据集在不同跟踪器的评估对比表

Table2　Comparison of LLOT-test dataset performance across different trackers

跟踪器(Tracker)	来源(Venue)	LLOT (AUC)	LLOT (Prec)	LLOT (P_norm)	参数量/可学习参数量	推理速度FPS
ProDAPT	Proposed	**0.585**	**0.692**	**0.746**	92.6M/6.51M	76.8
H-DCPT(Zhong等,2026)	TIP′24	0.576	0.684	0.739	93.0M/**3.03M**	78.5
HIPTrack(Cai等,2024)	CVPR′24	0.560	0.662	0.716	120.4M/–	65.0
ProContEXT(Lan等,2023)	ICASSP′23	0.557	0.660	0.708	92.6M/–	85.2
ARTrack(Wei等,2023)	CVPR′23	0.553	0.656	0.707	92.1M/–	42.0
DCPT(Zhu等,2024)	ICRA′24	0.527	0.614	0.665	93.0M/**3.03M**	81.5
OSTrack(Ye等,2022)	ECCV′22	0.521	0.613	0.663	92.1M/–	105.1
TransffCAR(Wei等,2024)	Access′24	0.485	0.580	0.505	–	–
UDAT-CAR(Ye等,2022)	CVPR′22	0.409	0.513	0.538	35.6M/–	45.0
ADTrack(Li等,2021)	ICRA′21	0.325	0.420	0.355	**24.5M**/–	80.1
SiamAPN++(Cao等,2021)	ICRA′21	0.216	0.300	0.285	28.0M/–	**120.2**

注:加粗字体为最优值,–为原文无对应数据。

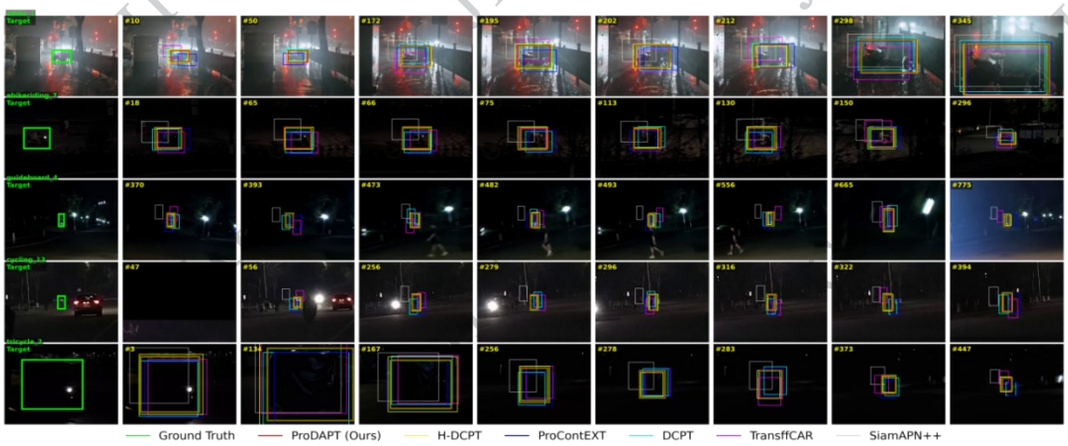


图6　LLOT-test数据集跟踪可视化对比图

Figure6　Visual Comparison of LLOT-test Dataset Tracking

深层提示在几何结构上与浅层保持一致,从本质上对抗了特征稀释。如表4所示,在相同特征提取条件下,仅引入带有跨层约束的CTCP,成功率即从基线ProContEXT的0.539显著提升至0.543。图9可视化结果进一步佐证了这一点,相比Baseline因特征迷失导致的大幅漂移(蓝框),引入CTCP后(黄框)模型能够重新锁定目标区域,有力证明了其优越的暗光特征挖掘能力。

在此基础上,集成PSCA模块利用提示结构偏置显式校准Transformer注意力,有效抑制背景噪声,使成功率进一步升至0.550;图9中引入PSCA后(紫框),跟踪框中心精度显著优于仅使用CTCP,验证了其抗干扰效能。最终结合EDGU策略,通过提示能量校验拒止高分漂移,模型达到了0.557的最优性能,如图9红框所示,完整模型始终紧密包围目标,解决了边缘拟合瑕疵。表4最后两列数据进一步显示,得益于冻结主干的提示微调策略,模型可学习参数量从全量微调的92.63M大幅降低至6.51M(占比仅7.03%),且在叠加三个模块后推理速度仍保持76.8FPS的超实时水平,充分证实了本文提出的三个模块在实现性能跃升的同时,达成了计算效率的平衡。

表3　UAVDark135数据集在不同跟踪器的评估对比表

Table3　Comparison of UAVDark135 Dataset Performance Across Different Trackers

跟踪器(Tracker)	来源(Venue)	UAVDark135(AUC)	UAVDark135(Prec)	UAVDark135(P_norm)	参数量/可学习参数量	推理速度FPS
ProDAPT	Proposed	**0.608**	**0.735**	**0.729**	92.6M/6.51M	76.8
H-DCPT(Zhong等,2026)	TIP′24	0.598	0.725	0.720	93.0M/**3.03M**	78.5
HIPTrack(Cai等,2024)	CVPR′24	0.589	0.718	0.712	120.4M/–	65.0
ProContEXT(Lan等,2023)	ICASSP′23	0.584	0.712	0.708	92.6M/–	85.2
DCPT(Zhu等,2024)	ICRA′24	0.577	0.703	0.701	93.0M/**3.03M**	81.5
ARTrack(Wei等,2023)	CVPR′23	0.570	0.705	0.695	92.1M/–	42.0
OSTrack(Ye等,2022)	ECCV′22	0.565	0.690	0.688	92.1M/–	105.1
UDAT-CAR(Ye等,2022)	CVPR′22	0.512	0.672	0.663	35.6M/–	45.0
TransffCAR(Wei等,2024)	Access′24	0.495	0.620	0.610	–	–
ADTrack(Li等,2021)	ICRA′21	0.469	0.605	0.601	**24.5M**/–	80.1
SiamAPN++(Cao等,2021)	ICRA′21	0.337	0.428	0.421	28.0M/–	**120.2**

注:加粗字体为最优值,–为原文无对应数据。

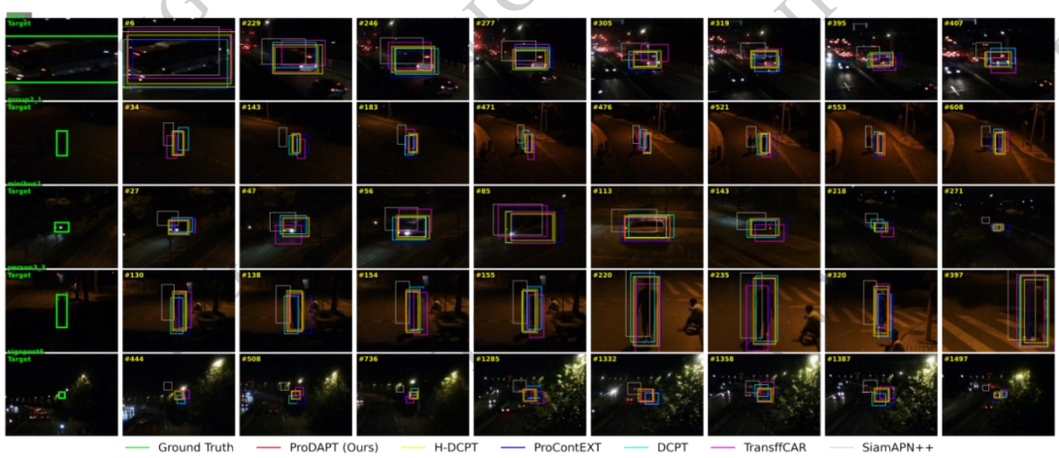


图7　UAVDark135数据集跟踪可视化对比图

Figure7　Visual Comparison of UAVDark135 Dataset Tracking

为了深入验证EDGU模块中“提示能量”作为结构完整性度量的物理可靠性及其阈值设置的合理性,在NAT2021-test数据集上进行了统计分析。图10展示了目标区域与高频背景噪声区域的提示特征L2范数概率密度分布。图10的能量分布统计呈现显著的双峰特性,物理上验证了提示能量作为结构完整性度量的可靠性,确保了噪声样本被有效拒绝。

为了再次深入探究各核心模块在特征提取与注意力分配中的物理作用,图11可视化展示了Transformer注意力热力图在消融实验不同阶段的逐级演变过程。选取了涵盖有光照、极低照度、微小目标及运动模糊等典型夜间挑战场景。从左至右的演变趋势揭示了性能提升,图11的热力图演变表明,PSCA与EDGU成功将注意力从背景光斑拉回目标中心,解决了夜间感知断层问题。验证了黑暗感知提示机制在夜间环境下的良好性能。

2.5　参数效率分析

为了全面评估算法的计算开销,将ProDAPT与传统的全量微调方法及同类提示学习方法DCPT进行了对比。表5展示了不同训练策略的参数效率对比。相比于全量微调(92.63M),ProDAPT采用冻结

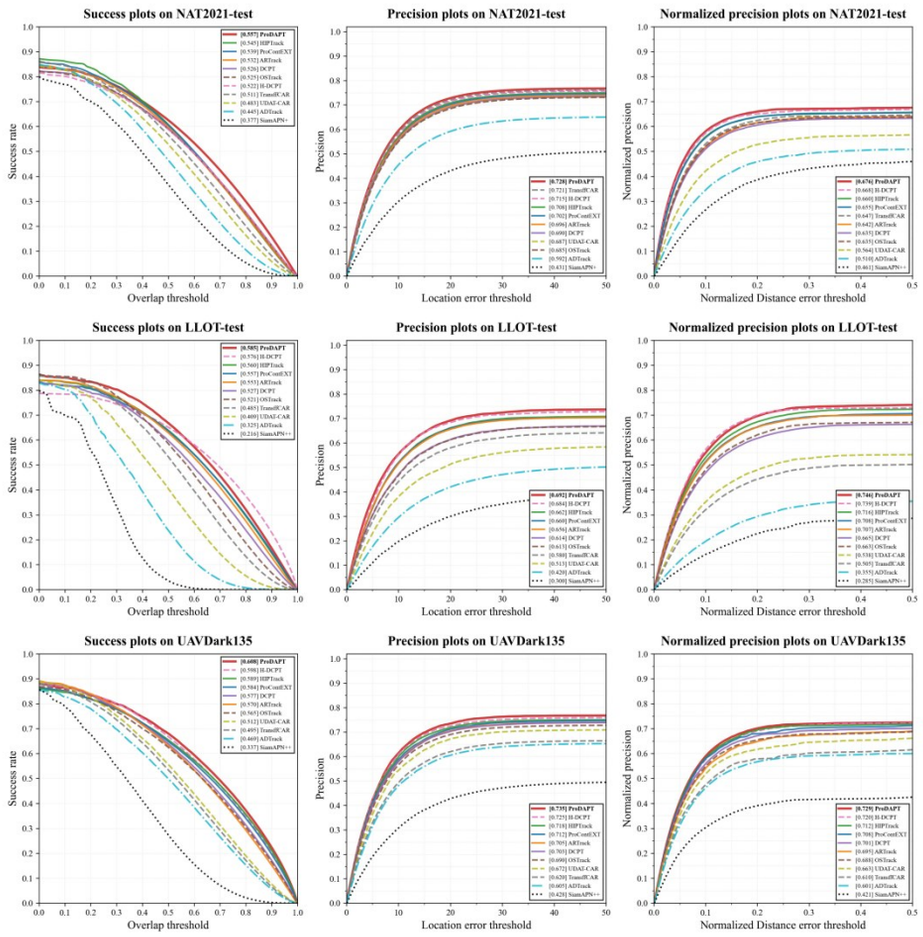


图8　ProDAPT与最先进(SOTA)跟踪器在三个数据集上的可视化比较

Figure8　Visual comparison of ProDAPT and state-of-the-art(SOTA) trackers across three datasets

表4　在NAT2021-test数据集上的消融实验结果

Table4　Results of Ablation Experiments on the NAT2021-test Dataset

编号	模块	成功率(Success/AUC)	精确率(Precision)	可学习参数量	推理速度FPS
A	Baseline	0.539	0.702	92.63M	**85.2**
B	A+CTCP	0.543	0.711	**3.01M**	79.5
C	B+PSCA	0.550	0.721	6.51M	77.2
D	C+EDGU	**0.557**	**0.728**	6.51M	76.8

注:加粗字体为最优值。

主干策略,仅需训练6.51M参数(占比7.03%)。虽然参数量略高于DCPT(3.03M),但这部分增量主要用于CTCP模块在深层网络维持语义一致性,以及PSCA和EDGU模块的注意力校准与更新决策,克服了DCPT仅在输入端注入导致的特征稀释缺陷。实验结果表明,这3.48M的额外参数投入换来了NAT2021上AUC从0.526至0.557的显著提升,证明了该框架在保持轻量化特性的同时,实现了优于现有方法的性能与开销平衡。

2.6　自采数据验证

为了进一步验证算法的泛化能力,本文选取了苏州市真实夜间自动驾驶场景进行测试(4K超高清分辨率,30FPS),重点针对跟车任务进行持续跟踪,由安装在自动驾驶测试车辆上的高规格视觉采集设备采集。该测试序列涵盖了典型的城市夜间复杂光照场景。不同于完全无光的实验室环境,该测试序

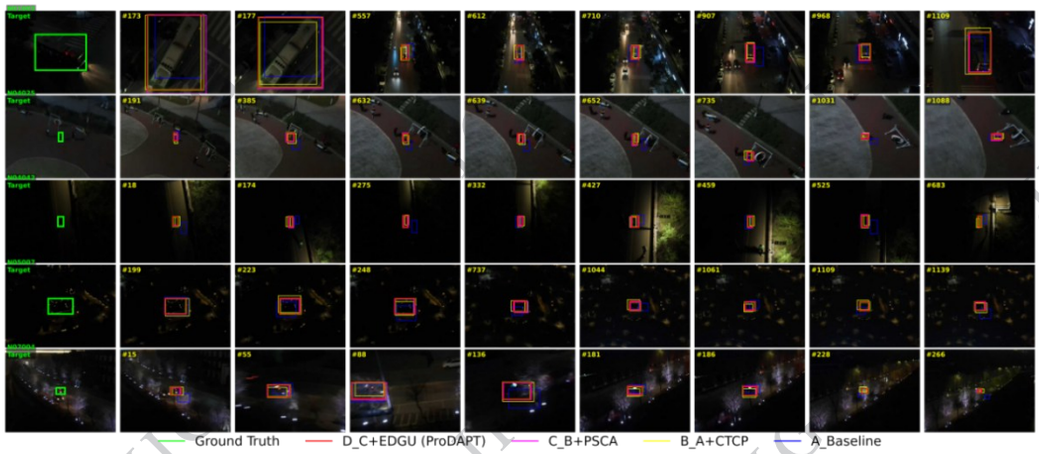


图9　消融实验的可视化图

Figure9　Visualization of ablation experiments

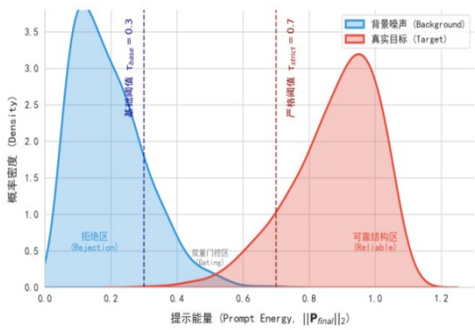


图10　NAT2021数据集的提示能量分布统计特性

Figure10　Statistical characteristics of the hint energy distribution in the NAT2021 dataset

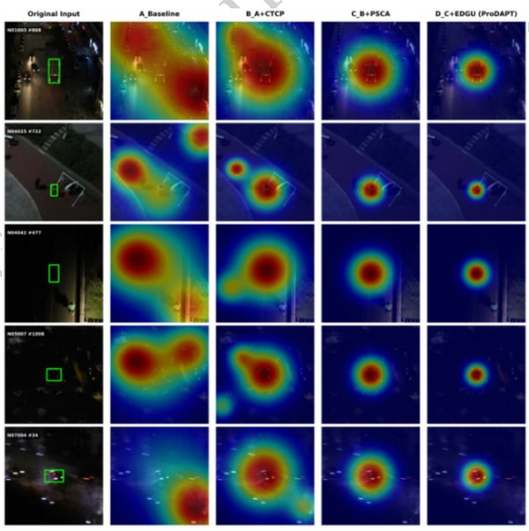


图11　热力图的可视化消融分析

Figure11　Visual Ablation Analysis of Heat Map

列涵盖了路面反光、强光干扰及目标短暂消失等情况。这种明暗交替与强光噪声共存的环境，更能检验跟踪器在真实自动驾驶夜间路况下的抗干扰能力。

为了更好地展示算法性能，逐帧性能用CLE图表示。绿色虚线（CLE=20px）以下的误差通常被认为是可以接受的。

如图12（a）所示，基线ProContEXT等方法受光影影响出现中心漂移，而ProDAPT得益于全链路黑暗感知机制，始终紧密锁定目标。中心位置误差（center location error，CLE）曲线进一步量化了该结果（图12（b）），ProDAPT全序列平均误差仅为2.65像素，远低于20像素成功阈值。特别是在“目标无真值区域（GT Missing）”的视野外场景中，相比其他算法的高置信度误跟踪，本方法有效避免了对背景噪声的错误更新，这一实测结果有力证明了ProDAPT在真实夜间自动驾驶场景下的性能。

2.7　局限性分析

尽管ProDAPT在多数夜间挑战场景下表现优异，但在面对极端环境时仍存在局限性。图13展示了算法在测试数据集某具有挑战性序列中的失败案例。

如图13所示，当目标处于大部分被遮挡或像素亮度极低的黑暗区域时，CTCP模块无法从输入中提取到任何有效的几何结构信息，导致跟踪发生偏移。

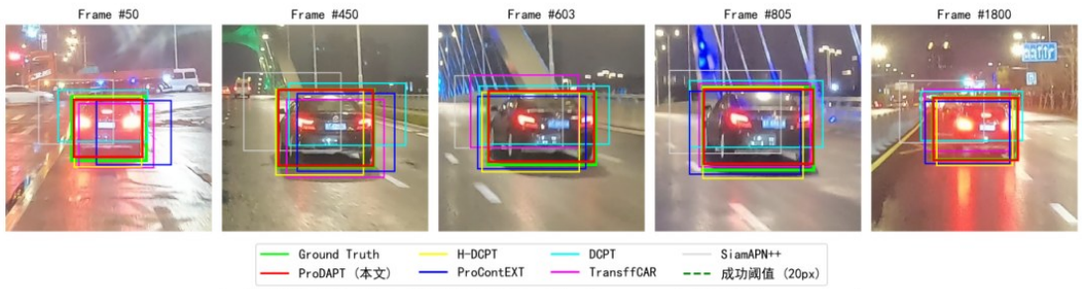
上述局限性表明，仅依赖单一的可见光模态在极端夜间场景下存在物理感知边界。未来的改进方向将集中在多模态融合框架的研究，通过引入红外热成像或事件相机数据，利用红外光的热辐射特性补充可见光在极暗条件下的纹理缺失，从而构建全

表5　不同训练策略的参数效率对比

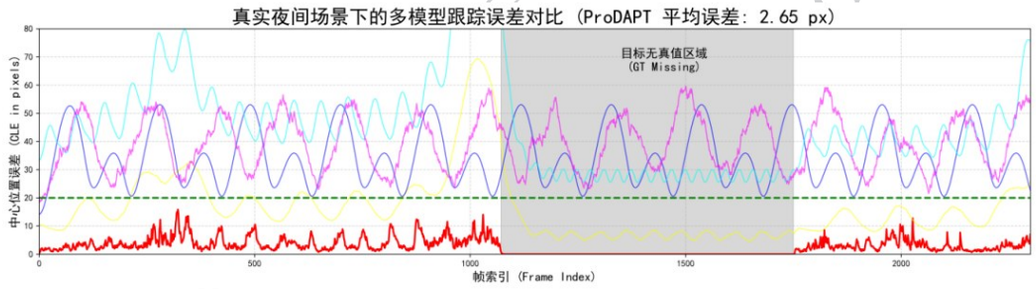
Table5　Comparison of parameter efficiency of different training strategies

方法	训练策略	可学习参数量	参数占比	NAT2021成功率	推理速度FPS
Baseline(Lan等,2023)	全部训练	92.63M	100%	0.539	**85.2**
DCPT(Zhu等,2024)	提示学习	**3.03M**	**3.03%**	0.526	81.5
ProDAPT	提示学习	6.51M	7.03%	**0.550**	76.8

注:加粗字体为最优值。



（a）　自采数据集跟踪可视化对比图

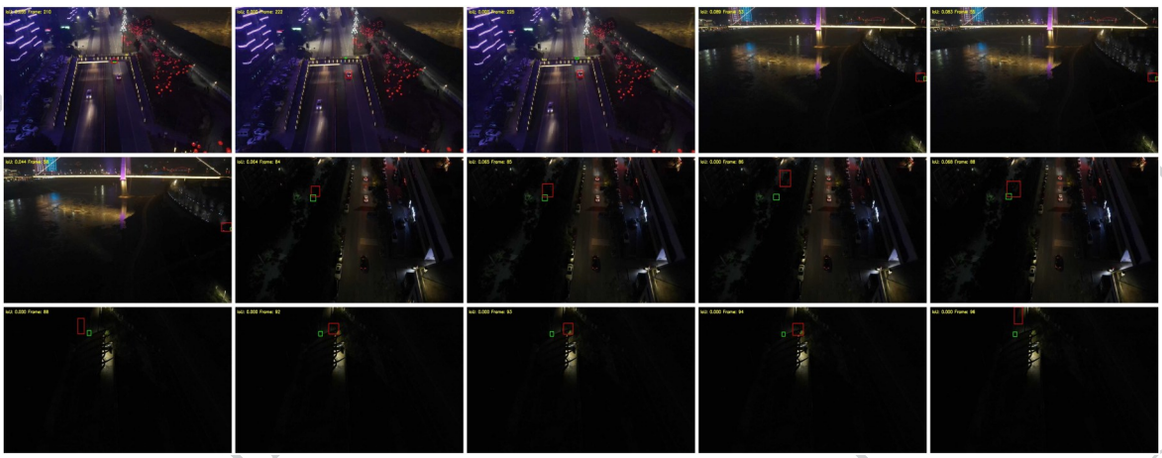


（b）　自采数据集跟踪误差对比图

((a)Visualized comparison of self-collected datasets tracking;(b)Comparison of tracking error of self-collected data set)

图12　自动驾驶自采数据集跟踪测试结果

Figure12　Tracking Test Results of Autonomous Driving Self-collected Data Sets



绿色框为真实标注框;红色为跟踪框

图13　ProDAPT跟踪失败结果图

Figure13　ProDAPT tracking failure result chart

天候、高可靠的夜间跟踪系统。

3 结 论

针对夜间跟踪面临的特征退化与决策漂移难题，提出了一种融合黑暗感知提示的夜间目标跟踪框架（ProDAPT）。跨层一致性提示生成器（CTCP）创新性地结合反向投影与跨层语义约束，在ViT深层特征空间中强制恢复了被噪声淹没的目标几何结构，确立了语义连续性；提示语义校准注意力（PSCA）通过引入可学习的结构先验偏置，显式校正了Transformer的注意力分布，有效抑制了夜间强光与相似干扰物引起的注意力弥散；能量感知双重门控策略（EDGU）则开创性地利用提示能量作为独立于分类分数的结构完整性度量，建立了双重校验逻辑，解决了夜间高分漂移引发的动态模板污染问题。

实验结果在NAT2021、LLOT和UAVDark135等公开数据集上的测试中，ProDAPT在保持仅占全量微调的7.03%低参数量的同时，在成功率与精确率等关键指标上均优于现有SOTA方法。同时在自采的苏州市夜间4K自动驾驶跟车数据集中，该算法仍能保持稳定的跟踪轨迹。

尽管引入提示学习模块在理论上增加了少量的模型参数，但在极端环境下的性能增益证明了该设计的价值。未来的工作将致力于算法的轻量化移植与推理加速，并探索将红外、热成像等多模态信息引入提示生成机制，以进一步突破完全无光场景下的视觉感知极限。

参考文献（References）

Cai W, Liu Q and Wang Y. 2024. Hiptrack: Visual tracking with historical prompts//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Seattle, USA: IEEE: 19258-19267 [DOI: 10.1109/CVPR52733.2024.01822.]

Cao Z, Fu C, Ye J, Li B and Li Y. 2021. SiamAPN++: Siamese attentional aggregation network for real-time UAV tracking//2021 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS). Prague, Czech Republic: IEEE: 3086-3092 [DOI: 10.1109/IROS51168.2021.9636309.]

Chen X, Peng H, Wang D, Lu H and Wu H. 2023. SeqTrack: Sequence to sequence learning for visual object tracking//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Vancouver, Canada: IEEE: 14572-14581 [DOI: 10.1109/CVPR52729.2023.01400]

Chen X, Yan B, Zhu J, Wang D, Yang X and Lu H. 2021. Transformer tracking//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Nashville, USA: IEEE: 8126-8135

Cui Y, Jiang C, Wang L and Wu G. 2022. Mixformer: End-to-end tracking with iterative mixed attention//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. New Orleans, USA: IEEE: 13608-13618

Fan H, Bai H, Lin L, Yang F, Chu P, Deng G, et al. 2021. LaSOT: A High-quality Large-scale Single Object Tracking Benchmark. Int J Comput Vis129, 439 – 461. https://doi.org/10.1007/s11263-020-01387-y

Fu C, Dong H, Ye J, Li Z, Duan Y and Lu G. 2022. HighlightNet: Highlighting low-light potential features for real-time UAV tracking//2022 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS). Kyoto, Japan: IEEE: 12146-12153 [DOI: 10.1109/IROS47612.2022.9981070.]

Guo C, Li C, Guo J, Loy C C, Hou J, Kwong S, et al. 2020. Zero-reference deep curve estimation for low-light image enhancement//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Seattle, USA: IEEE: 1777-1786 [DOI: 10.1109/CVPR42600.2020.00185.]

Huang L, Zhao X and Huang K. 2021. GOT-10k: A large high-diversity benchmark for generic object tracking in the wild. IEEE Transactions on Pattern Analysis and Machine Intelligence, 43(5): 1562-1577 [DOI: 10.1109/TPAMI.2019.2957464.]

Hu S Y, Zhao X and Huang K Q. 2024. Visual intelligence evaluation techniques for single object tracking: a survey. Journal of Image and Graphics, 29(08): 2269-2302 (胡世宇, 赵鑫, 黄凯奇. 2024. 单目标跟踪中的视觉智能评估技术综述. 中国图象图形学报, 29(08): 2269-2302) [DOI: 10.11834/jig.230498]

Kugarajeevan J, Kokul T, Ramanan A and Pinidiyaarachchi A. 2023. Transformers in single object tracking: An experimental survey. IEEE Access, 11: 80297-80326 [DOI: 10.1109/ACCESS.2023.3298440.]

Lan J P, Cheng Z Q, He J Y, Li C, Luo B, Bao X, et al. 2023. ProContEXT: Exploring progressive context transformer for tracking//ICASSP 2023-2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). Rhodes Island, Greece: IEEE: 1-5 [DOI: 10.1109/ICASSP49357.2023.10094971]

Li B, Fu C, Ding F, Zhu J and Lu G. 2021. ADTrack: Target-aware dual filter learning for real-time anti-dark UAV tracking//2021 IEEE International Conference on Robotics and Automation (ICRA). Xi′an, China: IEEE: 496-502 [DOI: 10.1109/ICRA48506.2021.9561564.]

Li B, Fu C, Ding F, Zhu J and Lu G. 2023. All-day object tracking for unmanned aerial vehicle. IEEE Transactions on Mobile Computing, 22(8): 4515-4529 [DOI: 10.1109/TMC.2022.3162892.]

Muller M, Bibi A, Giancola S, Alsubaihi S and Ghanem B. 2018. TrackingNet: A large-scale dataset and benchmark for object tracking in the wild//Proceedings of the European Conference on Computer Vision (ECCV). Munich, Germany: Springer: 300-317

Shao Y H, Chen H L, Fu G, Wu Y D and Ren Z W. 2025. Fuse image enhancement with a regularized correlation filter for target tracking of UAVs. Journal of Image and Graphics, 30(10): 3302-3318 (邵延华, 陈慧玲, 付贵, 吴亚东, 任珍文. 2025. 融合图像增强的正则化相关滤波无人机目标跟踪. 中国图象图形学报, 30(10): 3302-3318) [DOI: 10.11834/jig.240576]

Wei H, Fu Y, Wang D and Pan H. 2024. Unsupervised nighttime object tracking based on transformer and domain adaptation fusion network.IEEE Access, 12: 130896-130913 [DOI: 10.1109/ACCESS.2024.3378117.]

Wei X, Bai Y, Zheng Y, Shi D and Gan Y. 2023. Autoregressive visual tracking//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Vancouver, Canada: IEEE: 9697-9706 [DOI: 10.1109/CVPR52729.2023.00935.]

Xu H, Dong S H, Zhang J W and Zheng Y H. 2025. Context-aware attention fused Transformer tracking. Journal of Image and Graphics, 30(01): 0212-0224 (徐晗, 董仕豪, 张家伟, 郑钰辉. 2025. 融合上下文感知注意力的Transformer目标跟踪方法. 中国图象图形学报, 30(01): 0212-0224) [DOI: 10.11834/jig.240084]

Yan B, Peng H, Fu J, Wang D and Lu H. 2021. Learning spatio-temporal transformer for visual tracking//Proceedings of the IEEE/CVF International Conference on Computer Vision. Montreal, Canada: IEEE: 10428-10437 [DOI: 10.1109/ICCV48922.2021.01028.]

Ye B, Chang H, Ma B, Shan S and Chen X. 2022. Joint feature learning and relation modeling for tracking: A one-stream framework//European Conference on Computer Vision. Tel Aviv, Israel: Springer: 341-357

Ye J, Fu C, Zheng G, Cao Z and Li B. 2021. DarkLighter: Light up the darkness for UAV tracking//2021 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS). Prague, Czech Republic: IEEE: 3079-3085 [DOI: 10.1109/IROS51168.2021.9636680.]

Ye J, Fu C, Zheng G, Paudel D P and Chen G. 2022. Unsupervised domain adaptation for nighttime aerial tracking//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. New Orleans, USA: IEEE: 8886-8895 [DOI: 10.1109/CVPR52688.2022.00869.]

Yi A and Anantrasirichai N. 2024. A comprehensive study of object tracking in low-light environments.Sensors, 24(13): 4359 [DOI: 10.3390/s24134359]

Zhong P, Guo X, Huang D, Wang B and Zhang R. 2026. Low-light object tracking: A benchmark. IEEE Transactions on Intelligent Vehicles, 11(1): 220-235 [DOI: 10.1109/TIV.2025.3621205]

Zhu J, Lai S, Chen X, Wang D and Lu H. 2023. Visual prompt multi-modal tracking//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Vancouver, Canada: IEEE: 9516-9526 [DOI: 10.1109/CVPR52729.2023.00918.]

Zhu J, Tang H, Cheng Z Q, He J Y, Luo B, Qiu S, et al. 2024. DCPT: Darkness clue-prompted tracking in nighttime UAVs//2024 IEEE International Conference on Robotics and Automation (ICRA). Yokohama, Japan: IEEE: 7381-7388 [DOI: 10.1109/ICRA57147.2024.10610544.]

作者简介

姜彦吉，男，博士，副教授，主要研究方向为目标检测，目标追踪，预期功能安全和自动驾驶视觉感知。E-mail：jyjvip@126.com

董浩，通讯作者，男，博士，高级工程师，主要研究方向为预期功能安全。E-mail：eason@utcer.com

宗亚利，女，硕士研究生，主要研究方向为目标跟踪、自动驾驶。E-mail：1074911795@qq.com

张海洋，女，博士，主要研究方向为弱监督学习，推荐体统，大模型。E-mail：Haiyang.zhang@xjltu.edu.cn

刘大千，男，博士，校聘教授，主要研究方向为智能无人系统，图像与视觉信息计算。E-mail：liudaqian@lntu.edu.cn

费博雯，女，博士，副教授，主要研究方向为分布式资源组织与优化，智能数据处理。E-mail：feibowen@lntu.edu.cn

陈鹏达，男，硕士研究生，主要研究方向为目标检测、自动驾驶。E-mail：18424244614@163.com