

DOI: 10.19783/j.cnki.pspc.240250

## 基于掩码记忆的无人机电力设备分割跟踪方法研究

张焕龙<sup>1</sup>, 周钊燕<sup>1</sup>, 王延峰<sup>1</sup>, 田杨阳<sup>2</sup>, 翟登辉<sup>3</sup>, 李哲<sup>2</sup>, 许丹<sup>4</sup>, 刘恒<sup>5</sup>

(1. 郑州轻工业大学电气信息工程学院, 河南 郑州 471200; 2. 国网河南省电力公司电力科学研究院, 河南 郑州 450052; 3. 中国电气装备集团科学技术研究院有限公司, 上海 200000; 4. 许继电气股份有限公司, 河南 许昌 461000; 5. 平高集团有限公司, 河南 平顶山 467001)

**摘要:** 无人机巡检是保障新型电力系统安全运行的重要手段, 视频目标跟踪技术能够有效提高巡检精度。针对实际电力场景中背景复杂、目标形变大等挑战, 提出一种基于掩码记忆的无人机电力设备分割跟踪方法。首先, 设计了基于掩码记忆的 L2 匹配分割方法, 使用负平方欧几里得距离来计算帧间亲和度, 利用构建的历史掩码信息进行目标匹配, 提升设备的粗分割精度。其次, 为了减少记忆冗余, 提出变化感知的记忆更新机制, 通过图像和掩码联合评估目标变化程度, 设计变化置信度阈值, 以此决策是否进行记忆更新。最后, 将分割技术融入判别式相关滤波的跟踪框架中, 实现电力设备鲁棒分割跟踪。该方法分别在通用数据集 VOT2018、无人机数据集 UAV123 和实际电力场景数据集集中进行测试。实验结果表明所提方法能有效提升无人机巡检精度, 并为电力设备安全运行提供技术参考。

**关键词:** 目标跟踪; 目标分割; 记忆更新; 电力设备跟踪; 无人机巡检

### A segmentation and tracking method for UAV power equipment based on mask memory

ZHANG Huanlong<sup>1</sup>, ZHOU Keyan<sup>1</sup>, WANG Yanfeng<sup>1</sup>, TIAN Yangyang<sup>2</sup>, ZHAI Denghui<sup>3</sup>, LI Zhe<sup>2</sup>, XU Dan<sup>4</sup>, LIU Heng<sup>5</sup>  
(1. College of Electrical and Information Engineering, Zhengzhou University of Light Industry, Zhengzhou 471200, China; 2. State Grid Henan Electric Power Research Institute, Zhengzhou 450052, China; 3. China Electric Equipment Group Science and Technology Research Institute Co., Ltd., Shanghai 200000, China; 4. XJ Electric Co., Ltd., Xuchang 461000, China; 5. Pinggao Group Co., Ltd., Pingdingshan 467001, China)

**Abstract:** UAV inspection is vital for ensuring the safe operation of new power systems, and visual object tracking technology can effectively improve inspection accuracy. Given the challenges of complex background and large target deformation in an actual electrical power scene, a segmentation and tracking method for UAV power equipment based on mask memory is proposed. First, an L2 mask memory matching method improves segmentation accuracy by calculating frame affinities using the negative squared Euclidean distance and uses historical mask information for target matching. Second, to reduce memory redundancy, a change-aware memory updating mechanism is introduced. The degree of target change is jointly assessed by image and mask, and a confidence threshold-based memory update strategy is designed. Finally, the segmentation strategy is combined with the DCF tracking framework to realize robust segmentation tracking of power equipment. The proposed method is evaluated on VOT2018 (generalized dataset), UAV123 (UAV dataset), and a real power scenario dataset. Experimental results demonstrate that the proposed method effectively enhances UAV inspection accuracy and provides valuable technical references for the safe operation of power equipment.

This work is supported by the National Natural Science Foundation of China (No. 62272423, No. 62072416, No. 62006213, and No. 62102373).

**Key words:** object tracking; target segmentation, memory update; power equipment tracking; UAV inspection

## 0 引言

随着新型电力系统的快速发展, 对电力设备运

行安全性和稳定性的要求日益提高。无人机巡检作为一种高效、灵活且极具成本效益的检测手段, 已经广泛应用于电力系统的日常维护和故障排查中, 变电、输电、配电等环节的监控与巡检任务急剧增长<sup>[1-3]</sup>。目前, 无人机和机器人<sup>[4-6]</sup>成为替代人力的主要手段, 配合深度学习方法广泛应用于各项任务中, 包括但不限于电网巡检<sup>[7]</sup>、违规穿戴检测<sup>[8]</sup>、

**基金项目:** 国家自然科学基金项目资助(62272423, 62072416, 62006213, 62102373); 河南省杰出青年科学基金项目资助(2300421055); 河南省重点研发专项资助(241111210400)

输电线路防外破检测<sup>[9]</sup>、绝缘子故障诊断<sup>[10-11]</sup>和输电线路鸟巢识别<sup>[12]</sup>等。其中, 计算机视觉技术, 尤其是目标跟踪技术起着重要作用。例如, 无人机精准降落过程中<sup>[13]</sup>, 需要对待降落目标区域进行持续跟踪, 通过不断调整无人机姿态以满足下降条件。电塔是电力系统的重要组成部分<sup>[14]</sup>, 使用无人机对电塔进行自主检测时, 需要利用目标跟踪技术, 使目标电塔始终位于拍摄场景内。在基于移动机器人的电力系统智能巡检中<sup>[15]</sup>, 读取仪表盘的指示灯号需要利用目标跟踪技术连续跟踪仪表盘, 直到仪表显示画面清晰。因此, 视觉目标跟踪算法在无人机巡检中发挥着重要作用, 其对于电力设备的安全运行具有重要意义。

目前, 目标跟踪算法主要包括基于相关滤波的方法和基于深度学习的方法。

1) 基于相关滤波的方法。该方法将跟踪问题转化为滤波器的训练和响应计算问题, 通过学习目标和背景之间的差异准确定位跟踪目标。文献[16]基于最小输出平方误差之和设计滤波模板, 跟踪速度达到 600 fps。但该方法使用灰度特征, 对目标的表征能力有限, 影响跟踪性能。文献[17]通过提取方向梯度直方图(histogram of oriented gradients, HOG)特征进行目标跟踪, 并引入高斯核函数将非线性问题转换为高维空间中的线性问题, 使得算法更具一般性, 但很难适应目标的尺度变化。文献[18]提出高效卷积算子降低算法复杂度, 但在复杂场景中的鲁棒性也随之降低。文献[19]通过限制响应图的变化率, 减小了跟踪场景中目标遮挡的影响。基于相关滤波的方法一般具有较快的跟踪速度, 能满足无人机电力巡检的实时性要求, 但该类方法在面对复杂电力场景中目标遮挡、相似干扰等挑战时, 适应性较差。

2) 基于深度学习的方法。该方法使用深度神经网络提取目标特征, 其中具有代表性的是基于孪生网络的跟踪方法, 该类方法将跟踪问题转化为相关性匹配问题, 在第一帧中提取目标模板特征, 随后与当前帧搜索区域进行相似性比较确定目标位置。文献[20]训练了一个端到端的全卷积网络, 能够处理任意大小的输入图像, 但难以应对目标尺度变化。文献[21]将区域建议网络(region proposal network, RPN)引入孪生跟踪框架, 通过预定义的锚框进行多尺度分类和回归更好地预测目标大小。文献[22]通过增加网络的深度和宽度提高对目标特征的提取能力, 并引入残差模块避免梯度消失等问题出现。文献[23-24]去除了预定义的锚框, 降低了模型的参数量, 提升了跟踪速度。

基于孪生的方法提升了跟踪的准确性, 但该方法使用矩形框进行目标表征, 会不可避免地引入背景干扰信息, 难以应对电力场景中背景复杂等问题。相比之下, 像素级分割掩码能更加精确地估计目标状态。文献[25]在孪生跟踪框架<sup>[21]</sup>中引入分割分支, 实现回归边界框和分割目标的联合训练。文献[26]在相关滤波跟踪框架中集成分割技术, 使用判别式相关滤波器(discriminative correlation filter, DCF)作为分类分支, 用基于几何不变模板的模型进行对象分割。文献[27]使用注意力机制使得算法在跟踪过程中自适应地关注目标的关键区域, 从而提高跟踪鲁棒性。文献[28]使用无锚的跟踪框架, 学习对象感知特征, 帮助目标和背景分类。仅利用目标初始帧的信息, 不能很好地应对目标形变、跟踪背景复杂等情况。文献[29]建立分割记忆库, 提出了一种结合相关性和注意力机制的特征融合方法, 同时利用位置和语义信息进行目标状态估计。文献[30]在判别式分割跟踪器中引入背景特征更新方法以抵抗背景干扰, 获得更精确的目标估计状态。历史帧信息的利用使得跟踪算法能更好地适应目标外观变化, 但不必要的历史信息会造成记忆冗余。

因此, 本文提出一种无人机巡检中的电力设备分割跟踪方法。首先, 为了应对复杂的跟踪场景, 设计基于掩码记忆的 L2 目标匹配模块, 利用构建的历史掩码信息进行目标匹配, 提高算法目标表征能力, 以适应目标外观变化。另外, 为了避免不必要的历史信息堆积, 提出一种能够感知图像变化的记忆更新机制, 根据目标变化程度判断是否进行记忆更新, 减少记忆冗余。最后, 使用通用数据集、无人机数据集和实际电力场景数据集对该方法进行测试。结果表明, 本文方法对于电力设备无人机巡检具有较强的泛化能力和适用性。

## 1 基于掩码记忆的分割跟踪框架

与其他经典的目标跟踪方法相比, 基于分割的跟踪方法由于利用了像素级信息而具有良好的性能, 具有广阔的发展前景。然而大多数现有的跟踪器仅参考初始帧中的目标特征, 不能很好地应对目标外观变化大、背景复杂等情况。因此, 本文提出了一种基于匹配的像素级记忆更新方法, 来提高基于分割跟踪器的性能。如图 1 所示, 本文所提出的跟踪架构主要包括: 特征提取网络、基于 DCF 的目标定位分支、基于掩码记忆的 L2 目标匹配模块、分割分支和变化感知的记忆更新机制。

给定当前帧和初始帧图像, 特征提取网络(ResNet50<sup>[31]</sup>)对其逐层提取特征, 产生搜索特征和

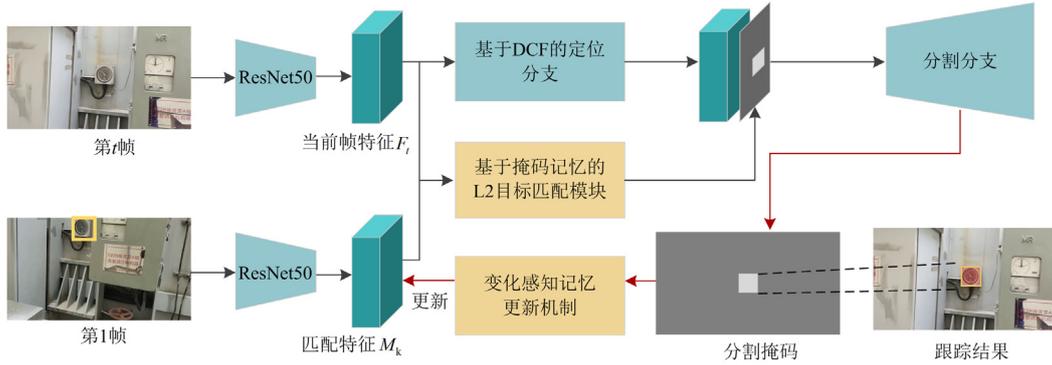


图 1 基于掩码记忆的电力设备分割跟踪方法框架图

Fig. 1 Overall architecture of a segmentation tracking method for power equipment based on mask memory

参考特征。不同层特征的功能不同，特征提取网络同时保留高中低 3 个不同层次的特征作为输出。基于 DCF 的目标定位分支<sup>[32]</sup>用于产生定位信息，最深层的语义特征通过卷积层和激活函数处理，响应图最大位置即为跟踪目标位置。基于掩码记忆的 L2 目标匹配模块通过掩码和历史信息进行目标匹配，在 2.1 节中详细介绍。分割分支将像素级记忆匹配结果和目标定位响应沿通道连接在一起作为输入，通过一系列上采样操作生成最终的分割跟踪结果。变化感知的记忆更新机制通过掩码和图像联合判断目标变化程度，并自适应地进行记忆更新。

## 2 基于掩码记忆的 L2 匹配分割方法

为了解决电力场景目标跟踪中背景复杂、目标形变等问题，本文提出基于 L2 匹配的像素级更新方法，并将其应用于基于 DCF 的目标跟踪网络。该方法包括基于 L2 的目标匹配模块和变化感知的记忆更新机制。

基于 L2 的匹配分割方法利用掩码信息和历史信息提高算法对复杂跟踪场景的适应能力。引入像素级分割掩码提高算法的目标表征能力，从而提升算法性能。另外，将历史信息通过键和值嵌入匹配过程，提高算法的分辨能力和对目标形变的适应能力。为了避免历史信息积累造成记忆冗余，提出变化感知的记忆更新机制，在感知到明显帧间变化时触发更新机制，有选择地对动态像素进行内存更新和匹配，减少了不必要的计算。

### 2.1 基于掩码记忆的 L2 匹配分割方法

目标跟踪方法常用矩形框来表示和定位目标。但实际电力跟踪场景存在背景复杂、目标形变大等特点，使用矩形框进行目标表征不能很好地近似目标，会引入干扰信息。另外，初始帧仅包含有限且固定的目标结构信息，其不能适应视频序列中目标

外观随时间的变化。因此，本文设计了基于 L2 的匹配分割方法来解决以上问题，该方法利用像素级掩码信息和历史信息提高跟踪算法对电力复杂场景的适应能力。半监督分割任务的输出由指示像素是否属于目标的二进制分割掩码组成，能够更好地近似目标。历史信息有助于适应目标形变。

该匹配网络由查询、键和值组成<sup>[33]</sup>，网络结构如图 2 所示。其中，查询特征  $F_t \in \mathbb{R}^{h \times w \times c}$  指当前帧特征。将目标从复杂背景中精确分离需要有效的参考信息，初始帧真实值具有准确和完整的目标描述。因此，本文使用该帧信息进行初始化。值得注意的是，传统注意力机制中键和值是当前帧特征的不同变换，而本文提出的方法中，键初始化为第一帧特征  $F_1 \in \mathbb{R}^{h \times w \times c}$ ，值初始化为第一帧前景和背景分割掩码 ( $Y_1^f \in \mathbb{R}^{h \times w \times 1}$  和  $Y_1^b \in \mathbb{R}^{h \times w \times 1}$ )。由此，本文直接在帧之间建立起对应关系，而不需要在特征提取阶段依赖于目标掩码。

固定的初始信息并不能适应目标外观变化，因此本文通过在键和值的匹配过程中嵌入历史信息来提高匹配质量。对于第  $t-1$  帧图像  $I_{t-1}$ ，获得对应的目标掩码(前景掩码  $Y_{t-1}^f$  和背景掩码  $Y_{t-1}^b$ )后，该帧的特征  $F_{t-1}$  和目标掩码 ( $Y_{t-1}^f$  和  $Y_{t-1}^b$ ) 中的有效信息将被选择性地补充进记忆键 ( $M_k \in \mathbb{R}^{T \times h \times w \times c}$ ) 和记忆值中 (前景  $M_{vf} \in \mathbb{R}^{T \times h \times w \times 1}$  和背景  $M_{vb} \in \mathbb{R}^{T \times h \times w \times 1}$ )。

记忆键和查询特征之间的相似性由相似性矩阵  $S$  衡量。在计算相似性矩阵之前，本文首先对记忆键  $M_k$  和查询特征  $F_t$  沿着通道维度进行每像素 L2 归一化处理，使其更好地进行匹配。方便起见，处理后的特征仍写作  $M_k$  和  $F_t$ 。相似性计算公式如式 (1) 所示。

$$S = c(M_k, F_t) \quad (1)$$

式中： $c$  表示相似性计算函数； $S$  为相似性矩阵。

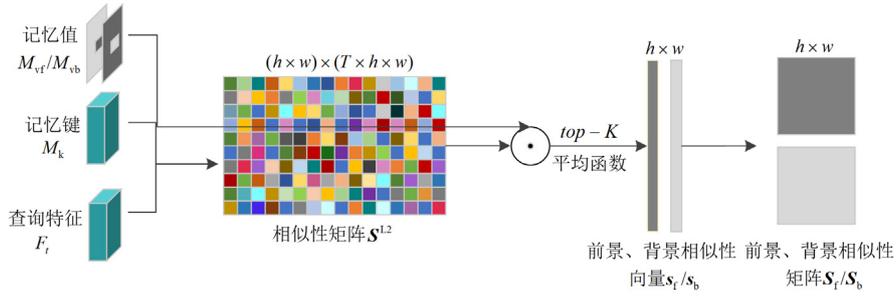


图2 基于掩码记忆的 L2 匹配分割方法

Fig. 2 Architecture of the L2 matching segmentation method based on mask memory

在大多数匹配方法中, 常采用点积或余弦相似度方法作为相似性函数, 但本文采用负平方欧几里得距离(下文简称 L2 方法)来计算相似性矩阵。因为点积等方法更多地反映两个向量之间的方向相似性, 而 L2 方法直接反映了向量在特征空间中的几何距离, 对于帧间特征的微小变化更为敏感。这种敏感性使得网络能够更精确地捕捉到帧与帧之间特征的动态变化, 尤其是在视频序列中目标对象发生形变、移动或受到的光照发生变化时<sup>[34]</sup>, L2 方法如式(2)所示。

$$\mathbf{S}^{L2} = -\|M_k - F_i\|_2^2 \quad (2)$$

式中,  $\mathbf{S}^{L2}$  表示查询帧与记忆键之间的 L2 相似性,  $\mathbf{S}^{L2} \in \mathbb{R}^{(h \times w) \times (T \times h \times w)}$ 。

在视频序列中, 目标对象会随着时间发生形变, 通过相似性计算建立查询帧和记忆键之间的联系。相似性矩阵反映哪些特征对于当前查询帧的分割更为重要, 然后, 网络可以选择性地聚合这些重要的特征, 从而提高分割的准确性和鲁棒性。但使用 L2 方法也有局限性, 该方法需要对矩阵中的元素进行逐元差分计算, 然后对其进行平方和求和, 这个过程比点积慢得多, 并且需要性能更好的硬件设备, 这会给电力图像采集处理设备带来计算负担。本文采用式(3)所示的分解方法实现计算过程的简化<sup>[35]</sup>。

$$\mathbf{S}^{L2} = -\|M_k - F_i\|_2^2 = 2M_k \cdot F_i - \|M_k\|_2^2 - \|F_i\|_2^2 \quad (3)$$

简化后的计算量与点积相差不大, 且只需使用矩阵运算来实现。相似性矩阵反映了查询特征  $F_i$  和记忆键  $M_k$  之间每个像素的相似性, 需要进一步在记忆值中检索信息, 以获得准确的目标匹配。首先将前景记忆值  $M_{vf}$  和背景记忆值  $M_{vb}$  整形为  $T \times h \times w \times 1$  的矢量  $\mathbf{m}_{vf}$  和  $\mathbf{m}_{vb}$ 。对于  $i \in [1, h \times w]$ , 相似性向量  $\mathbf{a}^i \in \mathbb{R}^{T \times h \times w}$  通过点积从记忆值向量  $\mathbf{m}_{vf}$  和  $\mathbf{m}_{vb}$  中检索有效信息, 如式(4)、式(5)所示。

$$\hat{\mathbf{s}}_f^i = \mathbf{a}^i \cdot \mathbf{m}_{vf} \quad (4)$$

$$\hat{\mathbf{s}}_b^i = \mathbf{a}^i \cdot \mathbf{m}_{vb} \quad (5)$$

式中,  $\hat{\mathbf{s}}_f^i$  和  $\hat{\mathbf{s}}_b^i$  分别表示  $i$  处的前景、背景相似性得分,  $\hat{\mathbf{s}}_f^i$ 、 $\hat{\mathbf{s}}_b^i \in \mathbb{R}^{T \times h \times w}$ 。

匹配得分的置信度越高, 目标匹配越准确。因此, 采用  $top-K$  平均函数从检索到的向量  $\hat{\mathbf{s}}_f^i$  和  $\hat{\mathbf{s}}_b^i$  中提取目标分数  $\mathbf{s}_f^i$  和  $\mathbf{s}_b^i$ 。具体来说, 该函数被用于从当前帧的所有特征中选择与目标模板最匹配的  $K$  个特征。这  $K$  个特征被认为是最有可能属于目标对象的, 因此它们在后续的分割过程中会被赋予更高的权重或更多的关注。选择较小的  $K$  值可能会导致信息不足, 使得算法难以准确地分割出目标对象; 而选择较大的  $K$  值虽然可以包含更多的信息, 但也会增加计算复杂度和内存消耗。如式(6)、式(7)所示。

$$\mathbf{s}_f^i = \frac{1}{K} \sum_{\hat{\mathbf{s}}_f^{ij} \in Top(\hat{\mathbf{s}}_f^i, K)} \hat{\mathbf{s}}_f^{ij} \quad (6)$$

$$\mathbf{s}_b^i = \frac{1}{K} \sum_{\hat{\mathbf{s}}_b^{ij} \in Top(\hat{\mathbf{s}}_b^i, K)} \hat{\mathbf{s}}_b^{ij} \quad (7)$$

式中: 集合  $Top(\hat{\mathbf{s}}_f^i, K)$  表示向量  $\hat{\mathbf{s}}_f^i$  中前  $K$  个匹配得分;  $j$  为前  $K$  个匹配分数中每个分数的索引。此外  $\mathbf{s}_f^i$  是一个标量。在本文的方法中, 设置  $K$  为 3 以达到最好的效果。目标背景匹配和前景匹配采用相同的方法。完成所有点积和  $top-K$  平均运算后, 得到前景和背景相似度向量  $\mathbf{s}_f \in \mathbb{R}^{h \times w}$  和  $\mathbf{s}_b \in \mathbb{R}^{h \times w}$ 。转换其维度得到前景和背景相似性矩阵  $\mathbf{S}_f \in \mathbb{R}^{h \times w}$  和  $\mathbf{S}_b \in \mathbb{R}^{h \times w}$ 。然后, 通过  $\text{soft max}(\cdot)$  函数生成粗分割掩码  $M'$ , 如式(8)所示。

$$M' = \text{soft max}(\mathbf{S}_f, \mathbf{S}_b) \quad (8)$$

## 2.2 变化感知的记忆更新机制

为了利用历史帧信息, 以往的匹配方法提出许多解决方案。文献[36-37]使用时间轴上所有历史帧来构建参考帧的模型, 且文献[36-38]均采用全帧建模与匹配方法。这在时间和空间上都造成了大量的计算负荷, 给电力图像采集和处理设备带来挑战。丰富的历史信息可以有效提高目标相似性匹配的准

确性。在时间上, 受限于计算设备的存储容量, 不可能将所有历史帧存储到记忆键中, 特别是对于长时间视频。为了减少计算, 文献[36,39]以预定义的速度对历史帧进行采样。但是人为设定的固定的采样间隔虽然减少了冗余, 却降低了算法对目标变化的适应能力, 从而使精确度的下降。在空间上, 实际上能表示目标变化的像素只占图像的一小部分, 且在相近的帧中, 目标往往变化不大。因此在记忆

更新时进行全帧建模, 重复储存变化不大的目标信息会导致计算冗余, 且对算法的性能提升并无益处。

因此, 为了减少时间和空间上的记忆冗余, 本文提出变化感知的记忆更新机制, 首先评估帧间像素变化程度, 当变化达到一定阈值时, 触发像素级记忆更新, 只将变化量补充进记忆特征中。该模块在减少计算量的同时, 最大程度地利用了目标历史信息。该模块的工作过程如图 3 所示。

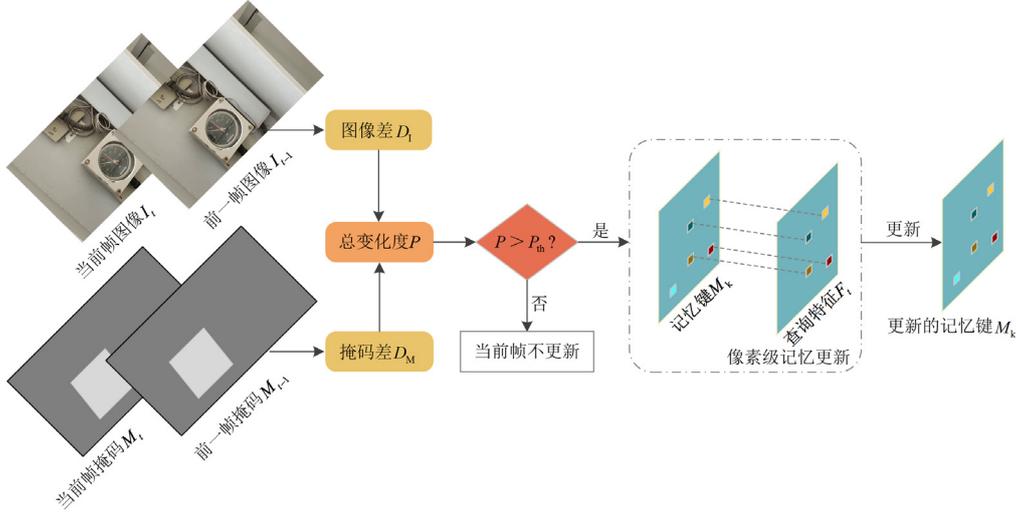


图 3 变化感知的记忆更新机制

Fig. 3 Diagram of the change-aware memory updating mechanism

本文通过图像和掩码联合感知目标变化, 给定当前帧图像  $I_t$ 、掩码  $M_t$  和前一帧图像  $I_{t-1}$ 、掩码  $M_{t-1}$ , 则图像差和掩码差的计算公式分别为式(9)、式(10)所示。

$$D_I^n = \sum_{x \in \{R,G,B\}} |I_t^{n,x} - I_{t-1}^{n,x}| / 255 \quad (9)$$

$$D_M^n = |M_t^n - M_{t-1}^n| \quad (10)$$

式中:  $I_t^{n,x}$  表示第  $t$  帧图像中位置  $n$  在通道  $x$  上的灰度值;  $D_I^n$  表示图像位置  $n$  处的图像差值;  $R$ 、 $G$ 、 $B$  表示彩色图像的 3 个颜色分量;  $M_t^n$  表示第  $t$  帧图像位置  $n$  的掩码值;  $D_M^n$  表示表示图像位置  $n$  处的掩码差值。对于每个像素  $i$ , 总变化度  $P$  通过式(11)更新, 当  $P$  超过变化度阈值  $P_{th}$  时, 证明目标在连续两帧之间发生了相对明显的变化, 因此触发记忆键和值的更新。

$$P = \begin{cases} P+1, & \text{if } D_I^n > th_1 \text{ or } D_M^n > th_m \\ P, & \text{otherwise} \end{cases} \quad (11)$$

式中:  $th_1$  表示图像差阈值;  $th_m$  表示掩码差阈值。为达到最佳效果, 本文将  $P_{th}$ 、 $th_1$  和  $th_m$  分别设定为

200、1 和 0。

记忆键更新时, 首先需在当前帧特征  $F_t$  中找到变化显著的像素, 然后将其增量式地更新入记忆键。为了发现不同的像素, 将  $F_t$  与记忆键  $M_k$  进行比较。将两者展开, 并计算余弦相似性矩阵  $S^{cos}$ , 如式(12)所示。

$$S^{cos} = \frac{F_t^i \cdot M_k^j}{\|F_t^i\|_2 \times \|M_k^j\|_2} \quad (12)$$

对于  $S^{cos}$  中的每一行, 将其中的最大得分作为查询帧  $F_t$  中的像素  $i$  与记忆键  $M_k$  中的像素  $j$  之间的特征相似性。像素相似性向量  $s_i^{cos}$  计算公式为

$$s_i^{cos} = \underset{j}{\operatorname{argmax}} S^{cos}[i, :] \quad (13)$$

通过相似性得分对  $s_i^{cos}$  进行排序, 并保持原始索引不变, 则排名越靠前的像素表现出越明显的特征变化。本文选择前  $\beta$  个像素用于记忆键更新,  $\beta$  是一个超参数, 用来平衡记忆更新方法的效率和全面性。为了获得最佳性能, 实验中将  $\beta$  设置为 10%。

根据  $s_i^{cos}$  中的索引, 从  $F_t$  中找到所选像素集合的特征向量, 然后直接将它们添加到记忆键  $M_k$  中,

前景记忆值和背景记忆值也通过同样的方法更新。这种自适应的更新策略使得算法能够根据目标的实际变化情况进行灵活调整, 既保证了分割的精度, 又避免了不必要的计算开销。

### 3 实验结果与分析

#### 3.1 实验设置

本文采用 Intel Core i7-10700 CPU @ 2.90 GHz, NVIDIA GeForce GTX 1650 4G 硬件平台实现所提跟踪算法。跟踪器的骨干网络为 ResNet50 的前 4 阶段, 将其在 ImageNet 上进行预训练以提取图像特征。第 4 阶段的特征通过  $1 \times 1$  卷积层、 $3 \times 3$  卷积层和 ReLU 激活函数, 将特征维度减少至 64, 用于目标匹配。前 3 个阶段的特征用于在分割上采样过程中补充目标的空间细节信息。为了测试算法性能, 本文分别在通用数据集 VOT2018、无人机数据集 UAV123 和实际电力场景视频中进行实验验证。

#### 3.2 通用数据集评估

本文通过在通用数据集 VOT2018 上与先进算法进行对比, 验证所提算法的有效性。

视觉目标跟踪数据集(vedio object tracking, VOT)是目前最常用的视觉目标评价基准之一。其包含 60 个视频序列, 具有多种属性挑战。该数据集采用跟踪失败后直接重置的机制以提高数据利用率。官方的 VOT 工具包<sup>[31]</sup>提供了 3 个评估标准, 即平均期望重叠率(expected average overlap, EAO)、精度和鲁棒性。

EAO 综合考虑了跟踪器的准确性和鲁棒性, 其值越高算法性能越好, 一般通过 EAO 分数对跟踪器进行排名。精度反映跟踪器对于目标尺度估计的精确程度, 其值越高算法性能越好。鲁棒性反映跟踪器的稳定程度, 其值越低则跟踪器失败的次数越少。

本文将所提方法与基于孪生网络的跟踪器和分割跟踪器进行对比。这些方法在数据集 VOT2018 中的跟踪结果如表 1 所示。基于孪生网络的方法包括经典孪生跟踪器 SiamFC<sup>[20]</sup>、引入区域建议网络的方法 SiamRPN<sup>[21]</sup>、多尺度特征方法和多尺度特征方法 SiamRPN++<sup>[40]</sup>、基于数据增强的长距离跟踪方法 DaSiamRPN<sup>[41]</sup>和无锚方法 SiamBAN<sup>[23]</sup>。分割跟踪方法包括 SiamMask<sup>[25]</sup>、SiamAttn<sup>[42]</sup>和 D3S<sup>[26]</sup>。通过比较可以看到, 由于引入了像素级信息, 基于分割的跟踪方法具有更好的表现。另外, 由于引入了历史帧信息和记忆更新机制, 本文所提算法获得了更准确、更鲁棒的跟踪结果。

表 1 不同算法在 VOT2018 数据集上的实验结果

Table 1 Experimental results of different algorithms

on VOT2018 datasets			
跟踪器	精度	鲁棒性	平均期望重叠率
SiamFC	0.506	0.585	0.188
SiamRPN	0.588	0.270	0.384
SiamBAN	0.597	0.178	0.452
DaSiamRPN	0.610	0.220	0.411
SiamMask	0.610	0.276	0.386
SiamAttn	0.630	0.160	0.470
SiamRPN++	0.640	0.200	0.464
D3S	0.640	0.150	0.489
Ours	0.645	0.212	0.481

#### 3.3 无人机数据集评估

UAV123 数据集集中的视频序列都由无人机采集, 考虑到实际电力场景跟踪任务中的图像大多由无人机采集, 因此采用此数据集来验证算法在无人机视角下的有效性。该数据集通过一次性评估方法获得精确度和成功率曲线图。对于精确度图, 首先计算跟踪预测的边界框的中点与目标真实框中点之间的距离, 小于阈值的帧占整个视频帧的百分比即为精确度, 根据不同阈值下的不同百分比得到整个精度曲线图。精确度得分是阈值设置为 20 像素时的百分比。对于成功率图, 重合率分数  $V$  表示为

$$V = \frac{|A_t^G \cap A_t^T|}{|A_t^G \cup A_t^T|} \quad (14)$$

式中:  $A_t^G$  表示第  $t$  帧的真实目标框;  $A_t^T$  表示第  $t$  帧的预测框。当视频帧的重叠率分数大于设定阈值时认为跟踪成功, 则成功率为成功帧数占总视频帧数的百分比。成功图下的面积值即为精确度评分。

本文主要比较基于孪生的方法 SiamDW<sup>[22]</sup>、SiamRPN<sup>[21]</sup>、无锚方法 SiamBAN<sup>[23]</sup>、SiamCAR<sup>[24]</sup>、Ocean<sup>[28]</sup>及分割跟踪方法 SiamMask<sup>[26]</sup>。总体的跟踪结果如图 4 中第 1 列所示, 本文所提算法在成功率和精确度方面取得了更好的结果。另外, 从图 4 中第 2—4 列可以看出, 本文所提算法在面对部分遮挡和相机运动的情况时, 取得了最好的结果, 证明该算法可以有效面对复杂电力场景。另外, 由于该算法充分利用了像素级掩码信息, 因此在面对低分辨率挑战时, 也获得了最好的成功率和精确度。

#### 3.4 实际电力场景评估

为了测试算法在真实电力场景中的性能, 本文将某省电力科学研究院提供的实际检测场景视频制作成数据集。如图 5 所示, 获取的图像包含电力场景中的各种挑战, 如背景复杂、相似目标干扰、遮挡和小目标等。

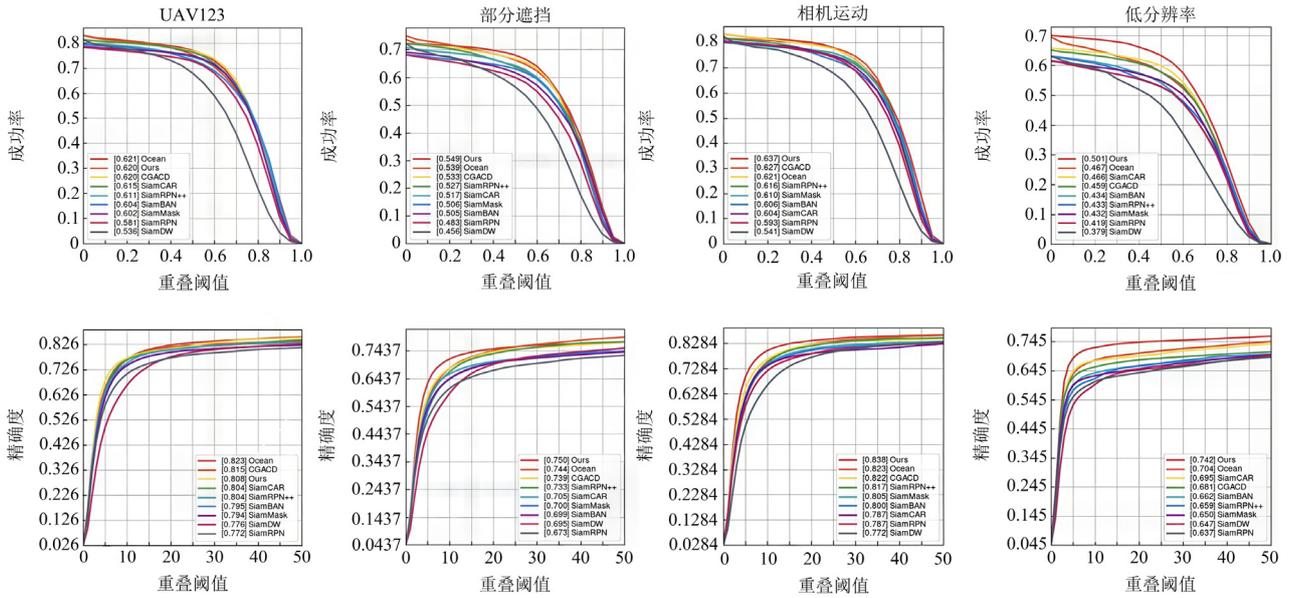


图 4 UAV123 数据集上的总体性能

Fig. 4 Overall performance on the UAV123 dataset



图 5 不同电力场景下的跟踪目标

Fig. 5 Tracking targets with diverse power scenes

该数据集的评价方法与 UAV123 数据集相同，算法整体性能如图 6 所示。本文算法与基于孪生网络的经典算法、分割跟踪算法 SiamMask<sup>[26]</sup>进行比较。SiamRPN<sup>[21]</sup>将跟踪视为一次性检测问题，并不适用于复杂电力场景。SiamCAR<sup>[24]</sup>和 Ocean<sup>[28]</sup>利用无锚的方法简化跟踪时的参数调整。但其均未利用像素级掩码信息。SiamMask<sup>[26]</sup>同本文算法均利用了像素级掩码信息，但其仅使用当前帧掩码，并未充分利用历史信息。可以看到本文算法取得了最好的成功率和精确度。这表明充分利用像素级掩码信息和历史信息可以有效提升跟踪器的性能。为了更好地显示跟踪效果，图 7 展示了不同跟踪算法在实际跟踪场景的可视化结果，其中包含视角变化、尺度变化和小目标等挑战。从第 2 行和第 5 行可以看出，

基于检测的算法 SiamRPN<sup>[21]</sup>不适用于此任务。从第 3、4 行可以看出，由于没有利用历史信息，Tadt<sup>[43]</sup>和 SiamMask<sup>[26]</sup>难以获得更高的跟踪精度。

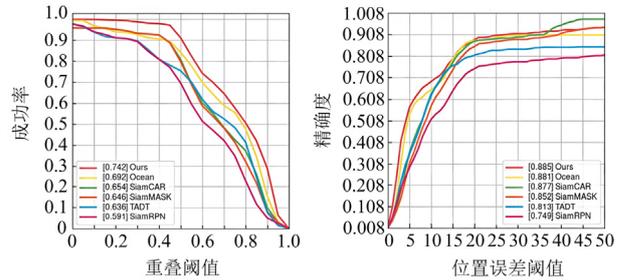


图 6 在实际电力场景视频中的成功率和精确度图

Fig. 6 Success rate and precision plots on actual inspection scene videos

### 3.5 速度分析

除了跟踪精度外，跟踪速度也是评价算法性能的重要指标，在实际跟踪任务中，帧率达到 30 fps(帧每秒)以上通常被认为满足实时性要求<sup>[44]</sup>。本文比较了不同类别的算法在实际电力视频中的得分和跟踪速度，如表 2 所示。基于相关滤波的跟踪器 CSK<sup>[44]</sup>和 KCF<sup>[17]</sup>的跟踪速度分别达到 2118 fps 和 969 fps，但由于基于孪生的跟踪方法 SiamRPN<sup>[21]</sup>和 SiamMask<sup>[26]</sup>利用了深度特征，跟踪性能显著提高，且由于 SiamMask 引入了分割掩码信息改善目标表征形式，获得了更明显的性能提升。其跟踪速度与基于相关滤波的方法相比有所下降，但仍能满足实时要求。由于引入历史信息 and 掩码信息，本文

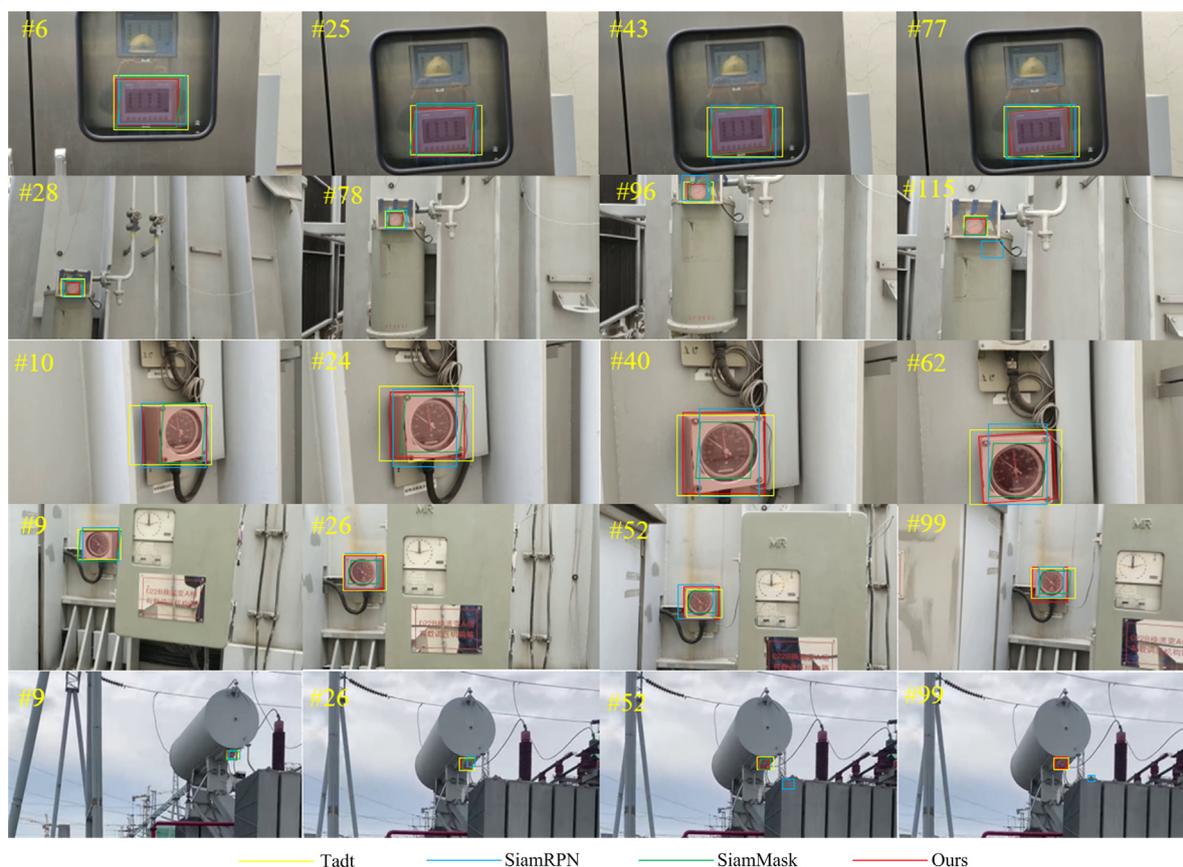


图 7 不同跟踪算法在实际跟踪场景的可视化结果

Fig. 7 Visualization results of different tracking algorithms in real tracking scenarios

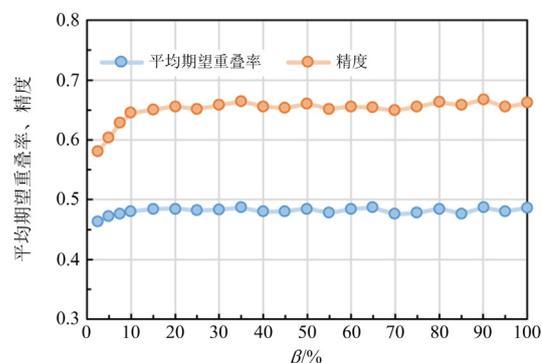
表 2 跟踪速度比较

Table 2 Comparison of tracking speed

视频序列	跟踪器	得分	帧率
Instrument1	CSK	0.596	357.2
	KCF	0.313	190.1
	SiamRPN	0.631	67.8
	SiamMask	0.893	50.3
	Ours	0.915	30.2
Insulator1	CSK	0.417	610.6
	KCF	0.417	34.5
	SiamRPN	0.465	70.2
	SiamMask	0.654	54.3
	Ours	0.687	31.7
liquidometer	CSK	0.338	2118.2
	KCF	0.113	969.5
	SiamRPN	0.557	98.5
	SiamMask	0.804	56.4
	Ours	0.843	32.3

所提方法跟踪精度有效提高, 且速度都达到 30 fps 以上, 满足实时性要求。另外, 超参数的设置对跟

踪器的速度也有一定影响, 本文在 VOT2018 数据集上探索超参数  $\beta$  对跟踪算法的影响并对其结果进行分析。精度、EAO 和帧率(fps)随  $\beta$  值的变化情况如图 8、图 9 所示。由图 8 可以看出, 随着  $\beta$  值的增加, 精度和 EAO 得分首先呈现上升趋势然后趋于稳定, 而帧率稳定下降。因此, 为了平衡精确度和速度, 选择  $\beta$  值为 10%。

图 8 VOT2018 中不同  $\beta$  值下跟踪结果对比Fig. 8 Comparison of tracking results with different  $\beta$  values on VOT2018

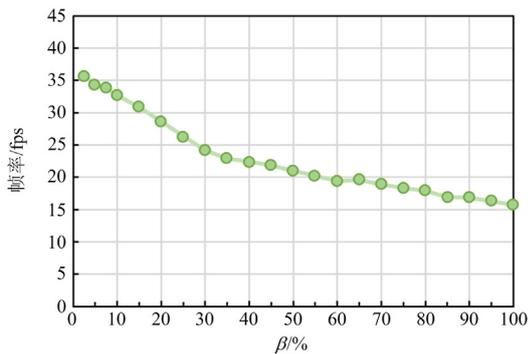


图9 VOT2018 中不同  $\beta$  值下跟踪速度对比

Fig. 9 Comparison of tracking speed with different  $\beta$  values on VOT2018

## 4 结论

无人机巡检是维护电力设备安全运行的重要手段，其中目标跟踪算法应用广泛。本文提出一种无人机巡检中的电力设备分割跟踪算法。

首先，为了解决现有跟踪算法在电力场景中适应性较差的问题，提出基于掩码记忆的 L2 匹配分割方法，该匹配方法使用负平方欧几里得距离作为相似性函数，充分利用像素级掩码信息和丰富的历史信息进行目标匹配。其次，为了减小计算负担，本文采用变化感知的记忆更新方法，通过图像和掩码联合感知目标的变化程度，以此来判断是否更新，减少了由记忆更新产生的计算冗余。最后，将该匹配更新方法集成在 DCF 跟踪框架中，利用定位信息和掩码信息联合确定跟踪结果。通过不同类型数据集的测试，证明本文方法与其他类型的跟踪器相比，具有更好的跟踪性能，且能较好地适应电力复杂场景，有效提升无人机巡检精度，为电力设备安全运行提供技术参考。

## 参考文献

- [1] 杜觉晓, 赵留学, 莽修伟, 等. 基于边缘端目标识别与跟踪算法的输电线路航拍视频流诊断技术研究[J]. 智慧电力, 2023, 51(6): 106-113.  
DU Juexiao, ZHAO Liuxue, MANG Xiwei, et al. Aerial video stream diagnosis technology for transmission lines based on edge target recognition and tracking algorithm[J]. Smart Power, 2023, 51(6): 106-113.
- [2] 叶翔, 孙嘉兴, 甘永叶, 等. 改进 YOLOv3 模型在无人机巡检输电线路部件缺陷检测中的应用研究[J]. 电测与仪表, 2023, 60(5): 85-91.  
YE Xiang, SUN Jiaying, GAN Yongye, et al. Application of improved YOLOv3 model in defect detection of transmission line components in UAV patrol inspection[J]. Electrical Measurement & Instrumentation, 2023, 60(5): 85-91.
- [3] 王红星, 陈玉权, 张欣, 等. 基于离线高斯模型的输电线路无人机巡检缺陷智能识别方法研究[J]. 电测与仪表, 2022, 59(3): 92-99.  
WANG Hongxing, CHEN Yuquan, ZHANG Xin, et al. Research on intelligent recognition method of transmission line UAV inspection defects based on offline Gaussian model[J]. Electrical Measurement & Instrumentation, 2022, 59(3): 92-99.
- [4] 吴嘉琦, 冉术兰. 输电线路巡线中的无人机技术应用[J]. 电子技术, 2023, 52(9): 242-243.  
WU Jiaqi, RAN Shulan. Application of UAV technology in transmission line inspection[J]. Electronic Technology, 2023, 52(9): 242-243.
- [5] 罗显跃, 周敬余, 禹天润, 等. 基于四旋翼无人机的二次屏柜巡检系统[J]. 电测与仪表, 2022, 59(7): 175-182.  
LUO Xianyue, ZHOU Jingyu, YU Tianrun, et al. Secondary screen cabinet inspection system based on quadrotor UAV[J]. Electrical Measurement & Instrumentation, 2022, 59(7): 175-182.
- [6] 常安, 陈振辉, 付明, 等. 无人机电力巡检航线智能规划及自主巡检研究[J]. 电网与清洁能源, 2023, 39(7): 61-66.  
CHANG An, CHEN Zhenhui, FU Ming, et al. Research on intelligent planning of uav power inspection routes and autonomous patrol inspection[J]. Power System and Clean Energy, 2023, 39(7): 61-66.
- [7] 罗潇, 於锋, 彭勇. 基于深度学习的无人机电网巡检缺陷检测研究[J]. 电力系统保护与控制, 2022, 50(10): 132-140.  
LUO Xiao, YU Feng, PENG Yong. UAV power grid inspection defect detection based on deep learning[J]. Power System Protection and Control, 2022, 50(10): 132-140.
- [8] 杨乐, 郭一鸣, 霍勇博, 等. 改进 YOLOv5 在电力生产违规穿戴检测中的应用[J]. 电力系统保护与控制, 2023, 51(14): 160-168.  
YANG Le, GUO Yiming, HUO Yongbo, et al. Application of improved YOLOv5 for illegal wearing detection in electric power construction[J]. Power System Protection and Control, 2023, 51(14): 160-168.
- [9] 魏贤哲, 卢武, 赵文彬, 等. 基于改进 Mask R-CNN 的输电线路防外破目标检测方法研究[J]. 电力系统保护与控制, 2021, 49(23): 155-162.  
WEI Xianzhe, LU Wu, ZHAO Wenbin, et al. Target detection method for external damage of a transmission line based on an improved Mask R-CNN algorithm[J].

- Power System Protection and Control, 2021, 49(23): 155-162.
- [10] 王道累, 张正刚, 张世恒, 等. 基于密集连接网络的航拍绝缘子旋转目标精准定位方法[J]. 电力系统保护与控制, 2024, 52(1): 35-43.  
WANG Daolei, ZHANG Zhenggang, ZHANG Shiheng, et al. Accurate positioning method of insulator rotating target in aerial photography based on dense connection network[J]. Power System Protection and Control, 2024, 52(1): 35-43.
- [11] 何宁辉, 王世杰, 刘军福, 等. 基于深度学习的航拍图像绝缘子缺失检测方法研究[J]. 电力系统保护与控制, 2021, 49(12): 132-140.  
HE Ninghui, WANG Shijie, LIU Junfu, et al. Research on infrared image missing insulator detection method based on deep learning[J]. Power System Protection and Control, 2021, 49(12): 132-140.
- [12] 张焕龙, 齐企业, 张杰, 等. 基于改进 YOLOv5 的输电线路鸟巢检测方法研究[J]. 电力系统保护与控制, 2023, 51(2): 151-159.  
ZHANG Huanlong, QI Qiye, ZHANG Jie, et al. Bird nest detection method for transmission lines based on improved YOLOv5[J]. Power System Protection and Control, 2023, 51(2): 151-159.
- [13] 陈菲雨. 无人机自主精准定点降落中图像处理技术的研究[D]. 济南: 山东大学, 2020.  
CHEN Feiyu. Research on image processing technology of unmanned aerial vehicle autonomous and accurate fixed-point landing[D]. Jinan: Shandong University, 2020.
- [14] MARTINEZ C, SAMPEDRO C, CHAUHAN A, et al. Towards autonomous detection and tracking of electric towers for aerial power line inspection[C] // International Conference on Unmanned Aircraft Systems, May 27-30, 2014, Orlando, USA: 284-295.
- [15] ZHANG Jianwei, GUO Zeng, ZHANG Huanlong, et al. Target-aware deep feature compression for power intelligent inspection tracking[J]. Journal of Electrical and Computer Engineering, 2022, 2022(1): 1-10.
- [16] BOLME D S, BEVERIDGE J R, DRAPER B A, et al. Visual object tracking using adaptive correlation filters[C] // Computer Society Conference on Computer Vision and Pattern Recognition, June 13-18, 2010, San Francisco, USA: 2544-2550.
- [17] JOAO F H, RUI C, PEDRO M, et al. High-speed tracking with kernelized correlation filters[J]. Transactions on Pattern Analysis and Machine Intelligence, 2015, 37(3): 583-596.
- [18] DANELLJAN M, BHAT G, KHAN F S, et al. ECO: efficient convolution operators for tracking[C] // Computer Vision and Pattern Recognition, July 21-26, 2017, Honolulu, USA: 6931-6939.
- [19] HUANG Z, FU C, LI Y, et al. Learning aberrance repressed correlation filters for real-time UAV tracking[C] // International Conference on Computer Vision, October 27-November 2, 2019, Seoul, Korea: 2891-2900.
- [20] BERTINETTO L, VALMADRE J, HENRIQUES J F, et al. Fully-convolutional siamese networks for object tracking[C] // European Conference on Computer Vision, October 10-16, 2016, Amsterdam, Netherlands: 850-865.
- [21] LI Bo, YAN Junjie, WU Wei, et al. High performance visual tracking with Siamese region proposal network[C] // Computer Vision and Pattern Recognition, June 18-22, 2018, Salt Lake City, USA: 8971-8980.
- [22] ZHANG Zhipeng, PENG Houwen. Deeper and wider Siamese network for real-time object tracking[C] // Computer Vision and Pattern Recognition, June 16-20, 2019, Long Beach, USA: 4586-4595.
- [23] CHEN Zedu, ZHONG Bineng, LI Guorong, et al. SiamBAN: target-aware tracking with Siamese box adaptive network[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2022, 45(4): 5158-5173.
- [24] GUO Dongyan, WANG Jun, CUI Ying, et al. SiamCAR: Siamese fully convolutional classification and regression for visual tracking[C] // Computer Vision and Pattern Recognition, June 14-19, 2020, Seattle, USA: 6269-6277.
- [25] HU Weiming, WANG Qiang, ZHANG Li, et al. Siammask: a framework for fast online object tracking and segmentation[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2023, 45(3): 3072-3089.
- [26] LUKEZIC A, MATAS J, KRISTAN M, et al. D3s-a discriminative single shot segmentation tracker[C] // Computer Vision and Pattern Recognition, June 14-19, 2020, Seattle, USA: 7133-7142.
- [27] YU Yuechen, XIONG Yilei, HUANG Weilin, et al. Deformable Siamese attention networks for visual object tracking[C] // Computer Vision and Pattern Recognition, June 14-19, 2020, Seattle, USA: 6728-6737.
- [28] ZHANG Zhipeng, PENG Houwen, FU Jianlong, et al. Ocean: object-aware anchor-free tracking[C/OL] // European Conference on Computer Vision, August 23-28, 2020: 771-787.
- [29] YANG Yijin, GU Xiaodong. Joint correlation and attention based feature fusion network for accurate visual tracking[J]. IEEE Transactions on Image Processing, 2023, 32: 1705-1715.
- [30] YANG Yijin, GU Xiaodong. Learning edges and adaptive surroundings for discriminant segmentation tracking[J].

- Digital Signal Processing, 2022, 121: 103309.
- [31] HE Kaiming, ZHANG Xiangyu, REN Shaoqing, et al. Deep residual learning for image recognition[C]// Computer Vision and Pattern Recognition, June 26-July 1, 2016, Las Vegas, USA: 770-778.
- [32] DANELLJAN M, BHAT G, KHAN F S, et al. Atom: accurate tracking by overlap maximization[C]// Computer Vision and Pattern Recognition, June 16-20, 2019, Long Beach, USA: 4660-4669.
- [33] VASWANI A, SHAZEER N, PARMAR N, et al. Attention is all you need[C]// Annual Conference on Neural Information Processing Systems, December 4-9, 2017, Red Hook, USA: 5998-6008.
- [34] CHENG H, TAI Y, TANG C. Rethinking space-time networks with improved memory coverage for efficient video object segmentation[J]. Advances in Neural Information Processing Systems, 2021, 34: 11781-11794.
- [35] KIM H, PAPAMAKARIOS G, MNIH A. The lipschitz constant of self-attention[C]// International Conference on Machine Learning, July 18-24, 2021, New York, USA: 5562-5571.
- [36] OH S W, LEE J Y, XU N, et al. Video object segmentation using space-time memory networks[C]// International Conference on Computer Vision, October 27-November 2, 2019, Seoul, Korea: 9226-9235.
- [37] ZHOU Qiang, HUANG Zilong, HUANG Lichao, et al. Motion-guided spatial time attention for video object segmentation[C]// International Conference on Computer Vision, October 27-November 2, 2019, Seoul, Korea: 1-5.
- [38] YANG Zongxin, WEI Yunchao, YANG Yi. Collaborative video object segmentation by foreground-background integration[C/OL]// European Conference on Computer Vision, August 23-28, 2020: 332-348.
- [39] ROBINSON A, LAWIN F J, DANELLJAN M, et al. Learning fast and robust target models for video object segmentation[C]// Computer Vision and Pattern Recognition, June 14-19, 2020, Seattle, USA: 7406-7415.
- [40] LI Bo, WU Wei, WANG Qiang, et al. Siamrpn++: Evolution of Siamese visual tracking with very deep networks[C]// Computer Vision and Pattern Recognition, June 16-20, 2019, Long Beach, USA: 4282-4291.
- [41] ZHU Zheng, WANG Qiang, LI Bo, et al. Distractor-aware Siamese networks for visual object tracking[C]// European Conference on Computer Vision, September 8-14, 2018, Munich, Germany: 101-117.
- [42] YU Yuechen, XIONG Yilei, HUANG Weilin, et al. Deformable Siamese attention networks for visual object tracking[C]// Computer Vision and Pattern Recognition, June 14-19, 2020, Seattle, USA: 6728-6737.
- [43] LI Xin, MA Chao, WU Baoyuan, et al. Target-aware deep tracking[C]// Computer Vision and Pattern Recognition, June 16-20, 2019, Long Beach, USA: 1369-1378.
- [44] FU Changhong, LU Kunhan, ZHENG Guangze, et al. Siamese object tracking for unmanned aerial vehicle: a review and comprehensive analysis[J]. Artificial Intelligence Review, 2023, 56(1): 1417-1477.

收稿日期：2024-03-05； 修回日期：2024-08-01

作者简介：

张焕龙(1981—)，男，博士，副教授，研究方向为模式识别、机器学习、图像处理、计算机视觉和电力人工智能；E-mail: zhl\_lit@163.com

周钊燕(1999—)，女，硕士，研究方向为计算机视觉、目标跟踪；E-mail: 1045390155@qq.com

王延峰(1973—)，男，通信作者，博士，教授，研究方向为智能电网通信、信息化电器、智能计算与电力系统智能化。E-mail: wangyanfeng@zzuli.edu.cn

(编辑 许威)