

# 多尺度 Transformer 的在线更新无锚框工件跟踪方法研究<sup>\*</sup>

夏代洪 徐 健 郑自立 赵一剑 刘高峰  
(西安工程大学电子信息学院 西安 710048)

**摘 要:**针对工业场景目标工件跟踪任务精度低、失败率高的问题,提出了多尺度 Transformer 在线更新的工件跟踪算法。首先,采用 Transformer 特征金字塔结构,融合多层次特征信息,以实现鲁棒的对目标表观建模;其次,使用 Transformer 模块对高级语义信息进行特征融合,使得网络模型专注于目标工件本身;然后,提出了基于排序的交并化(IoU)损失函数优化策略,有效地抑制干扰物对跟踪器影响;最后,设计一种在线更新策略更新目标模板,增强网络的鲁棒性。实验结果表明,在 VOT-2018 上准确率和失败率分别比基准跟踪器提高 3.8% 和 4.1%,且能保持 53 fps 的实时跟踪速度;在 LaSOT 数据集上精度与成功率别为 0.578 和 0.573,均优于基准跟踪器。通过 CCD 工业相机采集视频序列验证算法可以准确且鲁棒的跟踪目标工件。

**关键词:**目标跟踪;特征金字塔;Transformer;损失函数;在线更新

**中图分类号:** TN911.73      **文献标识码:** A      **国家标准学科分类代码:** 520.604

## Research on online updating anchor-free frame workpiece tracking method for multiscale Transformer

Xia Daihong Xu Jian Zheng Zili Zhao Yijian Liu Gaofeng  
(School of Electronics and Information, Xi'an Polytechnic University, Xi'an 710048, China)

**Abstract:** A multi-scale Transformer online update workpiece tracking algorithm is proposed to address the issues of low accuracy and high failure rate in industrial scene target workpiece tracking tasks. Firstly, a Transformer feature pyramid structure is adopted to fuse multi-level feature information to achieve robust apparent modeling of the target; Secondly, using the Transformer module for feature fusion of advanced semantic information enables the network model to focus on the target artifact itself; Then, an IoU Loss function optimization strategy based on sorting is proposed to effectively suppress the influence of the interference on the tracker; Finally, design an online update strategy to update the target template and enhance the robustness of the network. The experimental results show that the accuracy and failure rate on VOT-2018 are 3.8% and 4.1% higher than the benchmark tracker, respectively, and can maintain a real-time tracking speed of 53 fps; The accuracy and success rate on the LaSOT dataset are 0.578 and 0.573, both of which are better than the benchmark tracker. The algorithm proposed in this paper can accurately and robustly track the target workpiece by capturing video sequences using a CCD industrial camera.

**Keywords:** target tracking; feature pyramid; Transformer; loss function; online update

## 0 引 言

近年来随着工业 4.0 和中国制造 2025 的提出,工业机器人通过搭载视觉技术,采用目标跟踪技术,帮助机器

人完成工件的分拣、装配等任务,提高工业生产效率。因此准确且鲁棒跟踪工件成为当前工业生产领域的热门话题<sup>[1-2]</sup>。

在实际的工业生产领域中,目标工件会随着旋转、光

收稿日期:2023-07-19

<sup>\*</sup> 基金项目:陕西省科技厅资助项目(2018GY-173)、西安市科技局项目(GXYD7.5)资助

照变化和遮挡等环境因素发生尺度变化和纵横比变化,一些主流的跟踪器采用多尺度搜索方案进行目标估计,例如 GradNet 算法<sup>[3]</sup>使用梯度参数更新模板来适应目标的变化,虽然拥有优秀的跟踪速度,但是算法会随着跟踪时间增长积累误差,最终导致跟踪失败。SA-Siam 算法<sup>[4]</sup>使用语义分支过滤背景信息,用外观分支评估目标的尺度变化,保证良好跟踪速度的同时提高了跟踪精度,但是未探究深度特征的融合,使得算法鲁棒性表现一般。虽然多尺度搜索方案有优秀的速度表现,但是在复杂工业场景不能准确估计目标尺度和纵横比变化。另外一些跟踪器利用锚框估计目标的变化状态,例如 SiamRPN 算法<sup>[5]</sup>引入 Faster R-CNN<sup>[6]</sup>中的区域候选网络(RPN)<sup>[7]</sup>,使用锚框评估目标边界框,大幅提高了跟踪精度与鲁棒性,虽然锚框的方法比多尺度搜索方案的准确度大大提升,但是使用锚框的方法需要对锚框参数进行预设置,导致大量的超参数需要人为调制,这与实际工业环境下工件跟踪要求相悖。而无锚框的方法可以摆脱锚框超参数带来的限制,可以更精确的评估目标状态,例如 SiamCAR 算法<sup>[8]</sup>通过计算目标中心到边界框的距离来预测下一帧的边界框,网络的准确性和泛化能力大幅提高。Ocean 算法<sup>[9]</sup>采用和 SiamCAR 类似的无锚框思路预测真实目标的边界框,但是其分类和回归网络采用非对称的采样策略,提高了算法在复杂场景下的鲁棒性。SiamBAN 算法<sup>[10]</sup>通过对图像像素点精细划分正负样本的无锚框方法,生成目标回归框,完成跟踪任务,进一步提高了跟踪精度与鲁棒性。但是上述跟踪算法缺乏全局建模能力,容易陷入局部最优问题,其次工件抓取过程中会发生剧烈的尺度变化和相似物干扰,这很有可能导致边界框预测不准确。

最近,Transformer 结构因其优秀的全局依赖建模能力开始在跟踪领域得到应用,例如 TrDiMP 算法<sup>[11]</sup>和 TransT 算法<sup>[12]</sup>使用 Transformer 结构进行特征增强和特征融合,跟踪器展现出优秀的精度表现。

综上所述,针对工业生产环境工件跟踪精度低、鲁棒性差的问题,提出多尺度 Transformer 在线更新的工件跟踪算法,以 SiamBAN 作为基准跟踪器,首先使用特征金字塔融合浅层和深层特征信息,提高骨干网络的语义表征能力;其次采用 Transformer<sup>[13]</sup>替换 SiamBAN 网络中互相关操作进行特征融合,提高跟踪任务的定位准确性;然后在分类与回归子网络中提出基于排序的损失函数的优化策略,优化无锚框的正负样本细致度,提高跟踪的精确度;最后采用一种在线更新策略,保证跟踪任务的鲁棒性。经实验验证,改进后的算法模型可以准确且鲁棒地跟踪目标工件。

## 1 无锚框跟踪网络概述

SiamBAN 算法由孪生骨干网络和多个自适应头部组成,骨干网络由 ResNet-50 构成,在最后两个卷积层删除下采样操作解决因连续卷积带来的特征分辨率降低问题,

在第 3、4、5 卷积层引入空洞卷积同时设置不同的扩张率和步长来增大感受野。

自适应头部包含分类和回归两个子模块,每个模块使用深度互相关层组合特征得到分类特征图和回归特征图,如式(1)所示。

$$\begin{cases} P_{w \times h \times 4}^{cls} = [\varphi(x)]_{cls} * [\varphi(z)]_{cls} \\ P_{w \times h \times 4}^{reg} = [\varphi(x)]_{reg} * [\varphi(z)]_{reg} \end{cases} \quad (1)$$

式中:  $\varphi(x)$ 、 $\varphi(z)$  分别表示搜索分支特征和模板分支特征; \* 表示卷积运算。分类模块的特征图用于前景与背景分类,回归模块的特征图用于边界框的预测。由于 ResNet-50 引入空洞卷积,第 3、4、5 卷积层获得感受野各不相同,为充分利用多层次特征信息,采用多个自适应头部加权融合分类特征图和回归特征图预测目标位置,如式(2)所示。

$$\begin{cases} P_{w \times h \times 2}^{cls-all} = \sum_{l=3}^5 \alpha_l P_l^{cls} \\ P_{w \times h \times 4}^{reg-all} = \sum_{l=3}^5 \beta_l P_l^{reg} \end{cases} \quad (2)$$

式中:  $\alpha_l$ 、 $\beta_l$  表示特征图对应权重;  $P_l^{cls}$  表示分类特征图;  $P_l^{reg}$  表示回归特征图。SiamBAN 算法根据搜索图像的目标真实框,对正负样本区域进行划分为椭圆  $E_1$ 、 $E_2$ ,像素点落在  $E_2$  之外的为负样本,  $E_1$  内的为正样本,落在  $E_1$ 、 $E_2$  之间的样本集则忽略。根据落在正样本标签位置  $(p_i, p_j)$  计算预测框:

$$\begin{cases} p_{x_1} = p_i - d_l^{reg} \\ p_{y_1} = p_j - d_t^{reg} \\ p_{x_2} = p_i + d_r^{reg} \\ p_{y_2} = p_j + d_b^{reg} \end{cases} \quad (3)$$

式中:  $d_l^{reg}$ 、 $d_t^{reg}$ 、 $d_r^{reg}$ 、 $d_b^{reg}$  表示从该位置到预测框的四边的距离;  $(p_{x_1}, p_{y_1})$ 、 $(p_{x_2}, p_{y_2})$  表示预测框左上角和右下角坐标。生成预测框后,使用余弦窗口和尺度变化惩罚项来平滑目标的移动和变化,然后选择得分最好的预测框。

## 2 多尺度 Transformer 的无锚框跟踪模型

针对复杂多变的工件生产环境,能够精确且鲁棒的跟踪工件,本文提出多尺度 Transformer 在线更新的无锚框目标工件跟踪算法。整体框架如图 1 所示,首先模型输入  $127 \times 127$  模板图像和  $255 \times 255$  搜索区域图像,通过基于 Transformer 的特征金字塔融合不同层次的特征图信息获得混合特征信息,以适应工件和背景的变化;然后使用 Transformer 对模板特征图和搜索区域特征图进行相关性计算,生成高质量的特征响应图,精确定位工件位置;最后采用基于排序的损失函数优化策略,提高分类模块对正负样本划分的细致度,并且增强分类与回归模块之间的关联性,生成精确的跟踪结果;同时引入在线更新策略,根据工件预测得分阈值对模板进行动态更新,保证工件跟踪过程的鲁棒性。

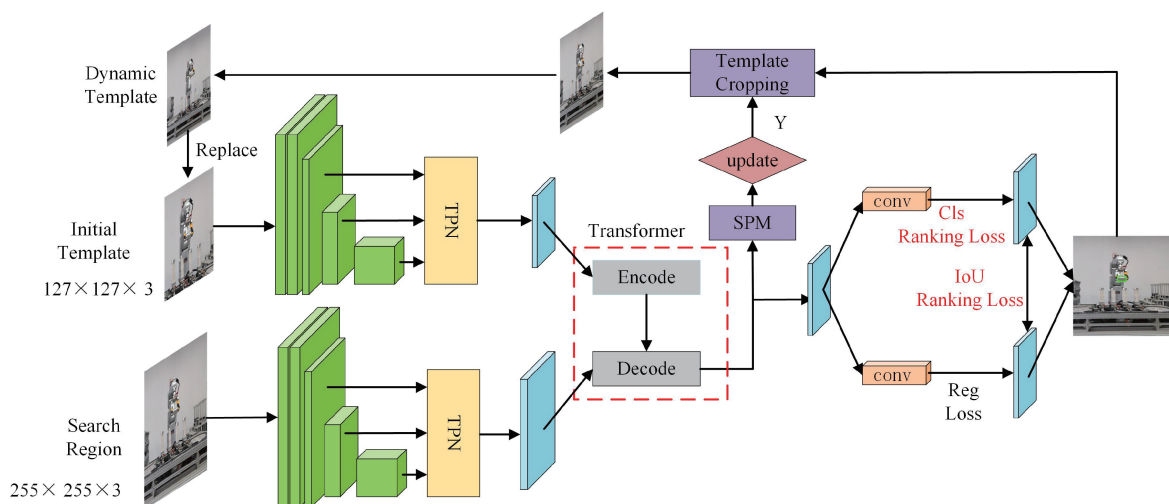


图1 多尺度 Transformer 在线更新的无锚框跟踪模型

## 2.1 基于 Transformer 的特征金字塔

工件跟踪任务需要丰富的特征信息,浅层的特征可以获取到目标颜色、纹理等表现信息,有利于精确定位;而深层的特征包含丰富的语义信息,对目标外观变化更具有鲁棒性。因此采用基于 Transformer 的特征金字塔结构

(transformer pyramid network, TPN),融合浅层与深层特征信息,得到同时具有表现信息和语义信息的高级混合工件特征信息,增强跟踪器表征能力,更鲁棒地跟踪工件。TPN 结构如图 2 所示。

特征金字塔构建如图 2(a)所示,分为自下而上路径

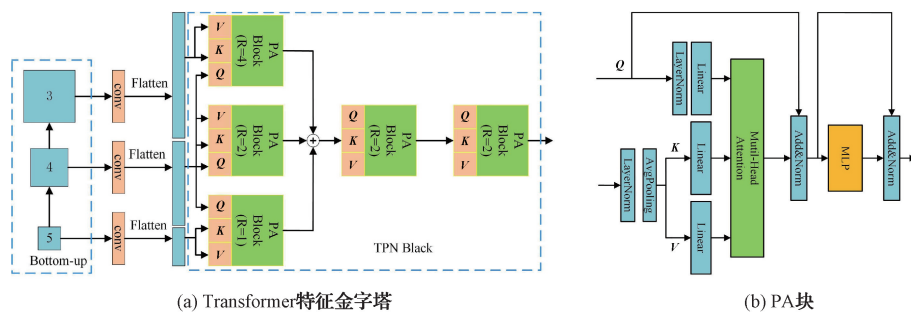


图2 TPN 结构

和融合路径,自下而上的路径是来自主干架构的前馈卷积,得到第 3、4、5 层网络的特征层次,再通过  $1 \times 1$  卷积和扁平化操作得到特征向量  $Q$ 、 $K$ 、 $V$ ,其中第 4 层特征作为所有特征层次的特征向量  $Q$ ,然后将特征向量  $Q$ 、 $K$ 、 $V$  送入 3 个不同池化尺度的并行池化自注意力块(pooling self-attention black, PA),融合后的输出直接馈送到两个串联 PA 块中,获得最终的语义特征。

PA 块如图 2(b)所示,采用 Transformer 编码器结构来组成。引入池化操作减少  $K$ 、 $V$  的空间维度,同时删除原有的 Transformer 结构位置编码操作,在减少参数数量的同时不降低 Transformer 性能。将改进后的  $K$ 、 $V$  和  $Q$  送入多头注意力层得到注意力矩阵:

$$\text{Attention}(Q, K, V) = \text{softmax}\left(\frac{QK^T}{\sqrt{C}}\right)V \quad (4)$$

式中:  $C$  表示向量  $K$  的维度,最后将注意力矩阵送入 Norm 和 MLP 层进行数据增强。

## 2.2 Transformer 模块

当前主流孪生网络采用简单的线性匹配互相关方式计算目标位置,无法满足实际的工业场景中的跟踪要求,因此采用 Transformer 结构替代互相关操作。该架构使用自注意力和交叉注意力机制来聚合全局和丰富的时空上下文依赖关系,可关注全局信息,在搜索特征图上自适应地寻找与模板特征图最相关的区域,获得更加精准的响应图。具体结构如图 3 所示,为了提升结构的推理速度,Transformer 的编码—解码部分只包含一层编码块和解码块。Transformer 首先将特征图压缩为一维特征序列输入进编码—解码器中,同时与加权矩阵进行线性变换得到  $Q$ 、 $K$ 、 $V$  3 个特征序列,再送入注意力块,得到注意力特征图:

$$A_{ij} = \frac{e^{\frac{1}{\sqrt{d}}Q_i^T K_j}}{\sum_{j=1}^{N_{kv}} e^{\frac{1}{\sqrt{d}}Q_i^T K_j}} \quad (5)$$

式中:  $Q_i, K_j$  表示第  $i$  个和第  $j$  个  $Q, K$  特征向量;  $d'$  表示通道维度。然后多头注意力层将  $M$  个注意力块进行级联和线性变换,进一步使用残差连接和归一化层进行归一化。最后使用由 ReLU 激活函数组成的前反馈网络 (FFN) 将特征维度还原至输入前。

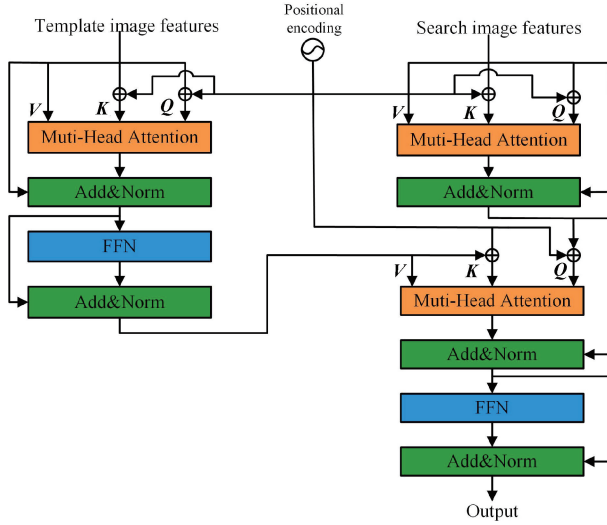


图3 用于跟踪任务的Transformer体系结构

### 2.3 基于排序的损失函数优化策略

跟踪器分类分支在处理正样本和负样本的划分时,未对正负样本进行信息交互,导致跟踪器容易被干扰物信息欺骗;其次分类和回归分支用独立的目标函数训练,这可能会导致分类和定位之间的预测不一致,本文提出一种基于排序的损失函数优化策略,来解决上述问题。首先将目标置信度得分小于0.5的划分为干扰物样本集,然后将处于分类边界线附近的低置信度正样本忽略,其余构建成正样本集,正负样本计算如下:

$$\begin{cases} P_+ = \sum_{j_+}^n w_{j_+} p_{j_+} \\ P_- = \sum_{j_-}^n w_{j_-} p_{j_-} \end{cases} \quad (6)$$

式中:  $p_{j_+}, p_{j_-}$  表示正负样本的置信度得分;  $w_{j_+}, w_{j_-}$  是正负样本的期望值。分类分支的损失函数约束如下:

$$L_{\text{rank-cls}} = \frac{1}{\lambda_1} \log(1 + \exp(\lambda_1 \cdot (P_- - P_+ + \lambda_2))) \quad (7)$$

式中: 常量  $\lambda_1, \lambda_2$  分别是控制损失值和正负样本排名裕度的超参数,本文默认为4和0.5。在  $L_{\text{rank-cls}}$  的监督下,调整决策超平面,将干扰物样本划分为负样本。效果可视化如图4所示,可以看出干扰物目标信息被显著抑制。

为了解决分类置信度和定位之间的不匹配,在分类与回归分支之间添加交并比 (IoU) 引导的排序损失函数,以增加分类和回归分支之间的关联性。IoU 引导的排序损失函数定义如下:

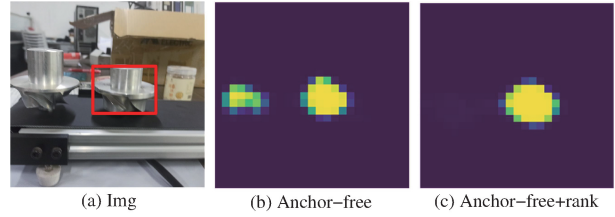


图4 基于排序的损失函数效果对比

$$L_{\text{rank-iou}} = \frac{1}{N_{\text{pos}}} \sum_{i,j \in A_{\text{pos}}, p_i > p_j} \exp(-\gamma \cdot (p_i - p_j)) + \frac{1}{N_{\text{pos}}} \sum_{i,j \in A_{\text{pos}}, v_i^{\text{iou}} > v_j^{\text{iou}}} \exp(-\gamma \cdot (v_i^{\text{iou}} - v_j^{\text{iou}})) \quad (8)$$

式中:  $N_{\text{pos}}$  表示正样本集  $A_{\text{pos}}$  的数目; 常量  $\gamma$  设置为3,控制着损失函数对模型的优化;  $p_i, p_j$  表示正样本置信度得分;  $v_i^{\text{iou}}, v_j^{\text{iou}}$  表示 IoU 预测值,保证 IoU 引导的排名损失可以将分类分数与相关联的 IoU 对齐,缩小分类和回归分支之间的误差,使得分类分支所得高置信正样本可以正确表示预测目标信息。

### 2.4 在线更新策略

由于目标工件的外观可能随着生产过程的推移而显著变化,因此捕获目标的最新状态以进行跟踪可以提高网络的鲁棒性。本文提出得分预测模块 (score prediction module, SPM) 对模板在线更新, SPM 如图5所示, SPM 由两个注意力块和一个三层感知器组成。首先,可学习 Score Token 用作参与搜索 RoI Token 的查询向量,能够对挖掘的目标信息进行编码。接下来, Score Token 关注初始目标 Token 的所有位置,将挖掘的目标与初始目标进行比较。最后,由 MLP 层和 Sigmoid 激活函数产生预测得分。当在线模板的预测分数高于0.5时,该模板将被视为正样本,并对模板更新。

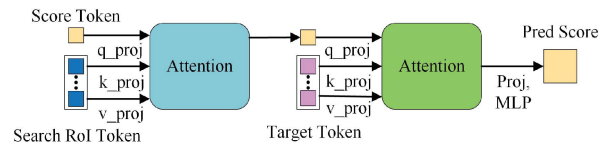


图5 SPM 模块

SPM 模块采用标准交叉熵损失函数训练,如式(9)所示。

$$L_{\text{score}} = y_i \log(p_i) + (1 - y_i) \log(1 - p_i) \quad (9)$$

式中:  $y_i$  表示目标真实值标签;  $p_i$  表示预测置信度得分。

### 3 实验及结果分析

实验在 Ubuntu18.04 系统下使用 Pytorch 框架,硬件设备为 CPU i5-13600KF 5.1 GHz, 64 G RAM, Nvidia 2080Ti  $\times$  2 22 G 的深度学习机下实现。采用 SGD (stochastic gradient descent) 训练模型,使用 YouTube-BoundingBoxes, COCO, GOT10K 和 LaSOT 作为训练集,单次训练样本个数为28,总共20个epoch,在前5个epoch中使用0.001~0.005的预热学习率,在最后15个



epoch 中学习率从 0.005~0.000 05 呈指数衰减,权重和动量衰减参数设置为 0.000 1~0.9。

### 3.1 消融实验

本文通过构建 Transformer 特征金字塔,引入 Trans-

former 模块以及采用基于排序的损失函数和在线更新策略等方法,优化 SiamBAN 跟踪器更加适合工业生产环境。为了确认改进方法的有效性,通过 VOT-2018 数据集进行消融实验,实验结果如表 1 所示。

表 1 消融实验结果

实验	Baseline	TPN	Transformer	IoU 优化	在线更新	EAO	$\Delta$ EAO/%
1	✓	×	×	×	×	0.452	0
2	✓	✓	×	×	×	0.464	1.2
3	✓	×	✓	×	×	0.492	4.0
4	✓	×	×	✓	×	0.477	2.5
5	✓	×	×	×	✓	0.471	1.9
6	✓	✓	✓	✓	✓	<b>0.514</b>	<b>6.2</b>

注:加粗表示结果最优

平均重叠期望(expected average overlap, EAO)是 VOT-2018 数据集评估跟踪器综合性能的重要评价指标,平均重叠期望越高,表明跟踪器性能越好。实验 2 引入 TPN 结构,获得混合高级语义信息,有效表征目标工件,与实验 1 相比 EAO 提高 1.2%。实验 3 使用 Transformer 将高级语义信息充分融合使得跟踪器有效判别前景与背景信息,EAO 提高了 4.0%,实验 4 使用 IoU 损失优化策略,有效抑制干扰物影响,使得 EAO 提升了 2.5%,实验 5 使用在线更新策略对模板进行动态更新,增强对目标外观变化的鲁棒性,EAO 提高 1.9%,实验 6 将所有组件组合,EAO 达到 0.514,证明各模块互补构建了高效的目标工件跟踪算法。

### 3.2 定量分析

为充分验证本文方法的有效性,将在 VOT-2018 和 LaSOT<sup>[14]</sup>数据集上进行测试验证,并选取当前主流跟踪算法进行对比。VOT-2018 数据集上对比算法包括有 DRT<sup>[15]</sup>、RCO<sup>[16]</sup>、DiMP<sup>[17]</sup>、Ocean、D3s<sup>[18]</sup>、SiamBAN。LaSOT 数据集上对比算法有 Dsiam<sup>[19]</sup>、SiamGAT<sup>[20]</sup>、SiamCAR、Ocean、StructSiam<sup>[21]</sup>、SiamRPN++<sup>[22]</sup>、SiamBAN。

跟踪精度(Accuracy)、失败率(Rubustness)、EAO 作为 VOT-2018 数据集主要评价指标,其中,Accuracy 通过计算跟踪视频序列所产生 IoU 的平均值得到,Accuracy 越大表示跟踪器精度越高;Robustness 计算平均跟踪失败帧数,越大表示跟踪器鲁棒性越差;EAO 则是在视频序列中同时计算精确值和失败率对跟踪器进行整体评估。VOT-2018 数据集对比结果如表 2 所示,可以看出本文方法拥有优秀的准确率(0.635),最低的失败率(0.137),获得了最好的 EAO(0.514)。

对于精度来说,与基准算法 SiamBAN 相比提升 3.8%,归功于本文采用特征金字塔结构聚合不同层次特征信息,有效的表征目标工件,随后又采用 Transformer 对高级特征语义信息进行充分融合,准确的定位目标位置

表 2 VOT-2018 数据集对比结果

算法	EAO ↑	Accuracy ↑	Robustness ↓
DRT	0.355	0.518	0.201
RCO	0.376	0.507	0.155
DiMP	0.376	0.590	0.150
Ocean	0.327	0.590	0.376
D3s	0.489	<b>0.640</b>	0.151
SiamBAN	0.452	0.597	0.178
Ours	<b>0.514</b>	0.635	<b>0.137</b>

注:加粗表示最优结果;下划线表示次优结果

区域,提高了跟踪的准确率,与 D3s 算法采用分割的方法跟踪目标相比准确率略有不足。对于失败率来说,与基准算法 SiamBAN 相比提升 4.1%,得益于本文方法采用基于排序的 IoU 优化策略,有效抑制干扰物对跟踪器的影响,同时提出一种在线更新策略,防止目标工件随时间变化导致跟踪出现漂移的情况,提升了网络的鲁棒性。

此外跟踪速度也是工业环境的一项重要要求,在 VOT-2018 数据集上进行跟踪速度对比,实验结果如图 6 所示,本文方法相比其他主流算法既保持着优秀的性能,又能满足工业的实时要求,跟踪速度达到 53 fps。

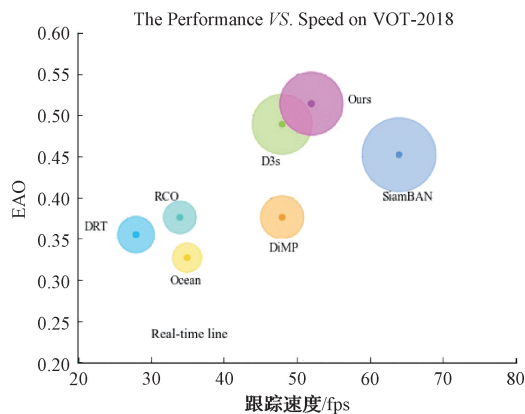


图 6 跟踪速度对比

LaSOT数据集包含训练集和测试集,共1400个视频序列,本文选取测试集作为实验验证,对比结果如图7所示。

从图7可以看出,本文算法成功率为0.573,比SiamGAT跟踪器高3.4%,比基准跟踪器SiamBAN高6.0%,得益于本文使用基于排序的IoU优化策略极大的抑制了干扰物对模型的影响,其次在线更新策略的使用在长视频

序列上展现出良好的自适应能力。对于精确率来说,本文方法精确率达到0.578,比SiamGAT跟踪器高3.2%,比基准跟踪器SiamBAN高6.1%,归功于特征金字塔可以获得多尺度语义信息,提高了骨干网络的表征能力,其次使用Transformer结构优秀的长距离建模能力,准确区分目标与背景信息,提高了跟踪器的精度。

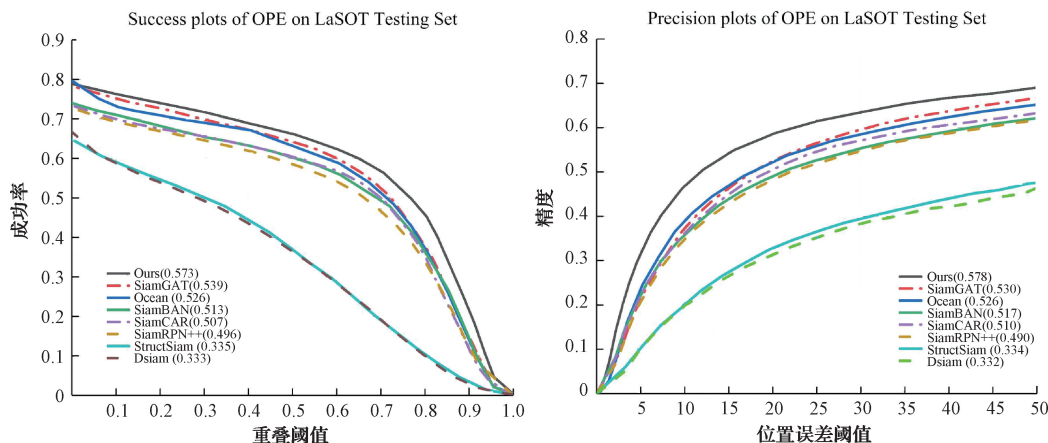


图7 LaSOT测试集算法对比

### 3.3 定性分析

为进一步验证本文方法在工业场景下的跟踪效果,使用三菱RV-3SD示教平台搭建仿真实验平台,传送带速度设置为0.7 m/s,采用CCD工业相机采集工件视频序列进

行定性分析,实验结果如图8所示,3个视频序列包括背景干扰、快速移动、尺度变化、平面内旋转、相似物干扰等视觉挑战。黄色、红色、黑色分别表示SiamRPN++、SiamBAN、SiamCAR算法,绿色表示本文算法。

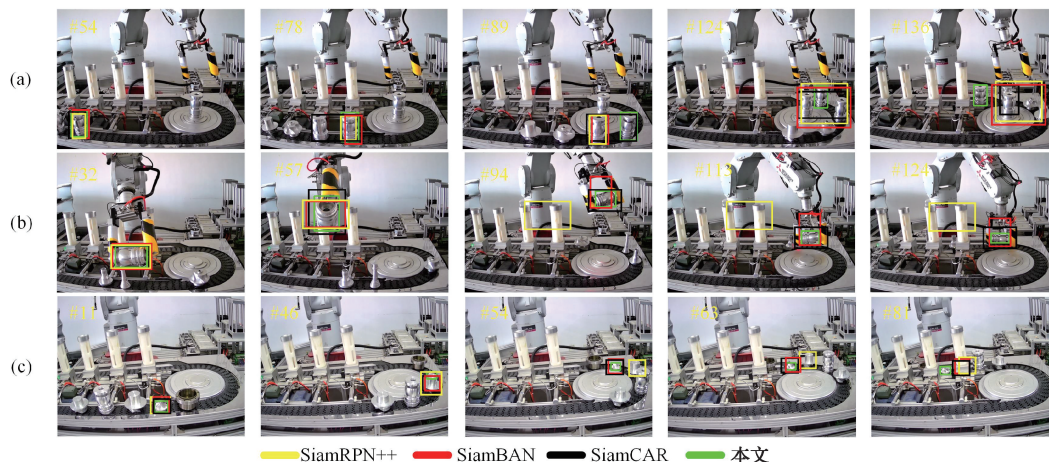


图8 目标工件视频序列跟踪结果

图5(a)目标工件在快速移动过程中受到相似物干扰,SiamRPN++等跟踪器相继出现漂移的问题,而本文在IoU优化策略指导下可以有效抑制相似物干扰,从而可以准确跟踪目标工件。图5(b)视频序列存在着平面内旋转、尺度变化等视觉属性场景挑战,在第57帧工件发生旋转时SiamRPN++、SiamBAN、SiamCAR出现小幅的跟踪漂移问题,在随后的94帧工件出现尺度变化,SiamRPN++则完全失去跟踪目标,SiamBAN、SiamCAR虽然还能跟踪目标工件,但是目标中心定位和尺度估计发生巨大

偏移,本文算法采用特征金字塔聚合多尺度信息,可以有效表征目标,同时采用Transformer对目标位置定位,提高了跟踪精度。图5(c)视频序列是对小目标工件跟踪,存在着遮挡、背景干扰等影响,SiamRPN++、SiamBAN、SiamCAR上述孪生网络未采用在线更新对目标模板进行更新,导致工件出现遮挡未能及时更新模板,最终使得跟踪器模型被污染导致跟踪失败,本文采用在线更新策略,可以及时动态更新模板,使得跟踪面对遮挡以及长时跟踪过程中依然可以准确跟踪目标工件。

## 4 结 论

本文以无锚框孪生网络框架为基础,提出了多尺度Transformer在线更新的目标工件跟踪算法,首先采用基于Transformer的特征金字塔结构显著地提升算法面对尺度变化、平面内旋转等干扰因素的鲁棒性,其次采用Transformer对特征信息融合,准确定位目标工件位置,然后提出基于排序的IoU优化策略极大地抑制了干扰物对跟踪的欺骗,提高了网络的准确性,最后使用在线更新策略,对未出现目标工件和遮挡等因素引发的跟踪偏移的问题发挥了巨大作用。在VOT-2018和LaSOT数据集上的实验结果表明本文算法的有效性和先进性。在未来工作中考虑采用轻量型跟踪算法,可以更好的在工业实际场景中部署。

## 参 考 文 献

- [1] 刘秀平,袁皓,李梦璐,等. ECO 多特征融合目标工件跟踪方法研究[J]. 电子测量与仪器学报, 2021, 35(10):161-167.
- [2] 张林耀,徐健,夏代洪,等. 基于孪生网络目标工件精确分割跟踪方法研究[J]. 国外电子测量技术, 2022, 41(9):66-72.
- [3] LI P, CHEN B, OUYANG W, et al. GradNet: Gradient-guided network for visual object tracking[C]. Proceedings of the IEEE/CVF International Conference on Computer Vision, 2019: 6162-6171.
- [4] HE A, LUO C, TIAN X, et al. A twofold Siamese network for real-time object tracking[C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018: 4834-4843.
- [5] LI B, YAN J, WU W, et al. High performance visual tracking with siamese region proposal network[C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018: 8971-8980.
- [6] 向宽,李松松,梁明慧,等. 基于改进 Faster RCNN 的铝材表面缺陷检测方法[J]. 仪器仪表学报, 2021, 42(1):191-198.
- [7] 李睿康,焦新泉,陈建军. 样本重采样小目标检测算法的改进[J]. 电子测量技术, 2021, 44(13):41-47.
- [8] GUO D, WANG J, CUI Y, et al. SiamCAR: Siamese fully convolutional classification and regression for visual tracking[C]. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2020: 6269-6277.
- [9] ZHANG Z, PENG H, FU J, et al. Ocean: Object-aware anchor-free tracking [C]. Computer Vision-ECCV 2020: 16th European Conference. Springer, 2020: 771-787.
- [10] CHEN Z, ZHONG B, LI G, 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.
- [11] CHEN X, YAN B, ZHU J, et al. Transformer tracking [C]. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2021: 8126-8135.
- [12] WANG N, ZHOU W, WANG J, et al. Transformer meets tracker: Exploiting temporal context for robust visual tracking [C]. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2021: 1571-1580.
- [13] ZHAO M, OKADA K, INABA M. TrTr: Visual tracking with transformer[J]. Computer Science, 2021, doi:10.48550/arXiv.2105.03817.
- [14] FAN H, LIN L, YANG F, et al. Lasot: A high-quality benchmark for large-scale single object tracking [C]. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2019: 5374-5383.
- [15] SUN C, WANG D, LU H, et al. Correlation tracking via joint discrimination and reliability learning[C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018: 489-497.
- [16] KRISTAN M, LEONARDIS A, MATAS J, et al. The sixth visual object tracking vot2018 challenge results[C]. Proceedings of the European Conference on Computer Vision (ECCV) Workshops, 2018.
- [17] BHAT G, DANELLJAN M, GOOL L V, et al. Learning discriminative model prediction for tracking [C]. Proceedings of the IEEE/CVF International Conference on Computer Vision, 2019: 6182-6191.
- [18] LUKEZIC A, MATAS J, KRISTAN M. D3s-a discriminative single shot segmentation tracker[C]. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2020: 7133-7142.
- [19] GUO Q, FENG W, ZHOU C, et al. Learning dynamic Siamese network for visual object tracking [C]. Proceedings of the IEEE International Conference on Computer Vision, 2017: 1763-1771.
- [20] GUO D, SHAO Y, CUI Y, et al. Graph attention tracking [C]. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2021: 9543-9552.
- [21] ZHANG Y, WANG L, QI J, et al. Structured siamese network for real-time visual tracking [C]. Proceedings of the European Conference on Computer Vision (ECCV), 2018: 351-366.
- [22] LI B, WU W, WANG Q, et al. Siamrpn++: Evolution of Siamese visual tracking with very deep networks [C]. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2019: 4282-4291.

## 作 者 简 介

夏代洪, 硕士研究生, 主要研究方向为机器视觉、目标跟踪。

E-mail: xdh18716632497@163.com

徐健(通信作者), 教授, 硕士生导师, 主要研究方向为机器视觉、图像处理。

E-mail: xujian@xpu.edu.cn