

# 控制与决策

Control and Decision

## 基于可变形卷积的孪生网络目标跟踪算法

刘如浩, 张家想, 金辰曦, 卢先领

引用本文:

刘如浩, 张家想, 金辰曦, 卢先领. 基于可变形卷积的孪生网络目标跟踪算法[J]. *控制与决策*, 2022, 37(8): 2049–2055.

在线阅读 View online: <https://doi.org/10.13195/j.kzyjc.2021.0088>

---

## 您可能感兴趣的其他文章

### Articles you may be interested in

#### 特征响应权重自适应的IoU网络跟踪算法改进

Improvement of IoU network tracking with adaptive weighted characteristic responses

控制与决策. 2022, 37(7): 1752–1762 <https://doi.org/10.13195/j.kzyjc.2021.0148>

#### 自适应感受野网络的行人重识别

Adaptive receptive network for person re-identification

控制与决策. 2022, 37(1): 119–126 <https://doi.org/10.13195/j.kzyjc.2020.0505>

#### 基于紧凑混合网络的视网膜血管自动分割

Automatic segmentation of retinal vessel via compact mixed network

控制与决策. 2022, 37(2): 353–360 <https://doi.org/10.13195/j.kzyjc.2020.0883>

#### 基于条件对抗生成孪生网络的目标跟踪

Conditional generative adversarial siamese networks for object tracking

控制与决策. 2021, 36(5): 1110–1118 <https://doi.org/10.13195/j.kzyjc.2019.1215>

#### 尺度自适应的多特征融合相关滤波目标跟踪算法

Scale adaptation and multi-feature fusion correlation filtering object tracking algorithm

控制与决策. 2021, 36(2): 429–435 <https://doi.org/10.13195/j.kzyjc.2019.0445>

# 基于可变形卷积的孪生网络目标跟踪算法

刘如浩, 张家想, 金辰曦, 卢先领<sup>†</sup>

(江南大学 物联网工程学院, 江苏 无锡 214122)

**摘要:** 为解决多数基于孪生网络的跟踪算法存在骨干网络特征提取能力弱、模板不适应目标变化等问题, 在 SiamFC 的基础上提出基于可变形卷积的孪生网络算法(DCSiam). 首先, 采用可变形卷积模块在不同方向上学习多层特征数据的自适应偏移量, 增大卷积过程中的有效感受野; 然后, 通过多层可变形互相关融合得到最终响应图, 以增强骨干网络的深层语义特征提取能力; 最后, 采用一种高置信度的模板在线更新策略, 每隔固定帧计算响应图的峰值旁瓣比与最大值作为更新依据, 使用加权的方式融合特征以更新模板. 使用 OTB2013、OTB2015、VOT2016 和 VOT2017 四个公共基准数据集对所提出算法进行跟踪性能评估, 实验结果表明, 在 OTB2015 数据集上, DCSiam 算法整体精确率、成功率较基线分别提高 9.5% 和 7.5%, 很好地实现了复杂情况下的目标跟踪, 验证了所提出算法的有效性.

**关键词:** 目标跟踪; 孪生网络; 可变形卷积; 语义特征; 模板更新

中图分类号: TP273 文献标志码: A

DOI: 10.13195/j.kzyjc.2021.0088

引用格式: 刘如浩, 张家想, 金辰曦, 等. 基于可变形卷积的孪生网络目标跟踪算法[J]. 控制与决策, 2022, 37(8): 2049-2055.

## Target tracking based on deformable convolution siamese network

LIU Ru-hao, ZHANG Jia-xiang, JIN Chen-xi, LU Xian-ling<sup>†</sup>

(School of Internet of Things Engineering, Jiangnan University, Wuxi 214122, China)

**Abstract:** In order to solve the problem that most tracking algorithms based on siamese networks have weak feature extraction ability of backbone networks and template is not suitable for target transformation, this paper proposes the siamese network algorithm based on deformable convolution(DCSiam) on the basis of SiamFC. Firstly, the deformable convolution module is used to learn the adaptive offset of multi-layer feature data in different directions, and the effective receptive field in the convolution process is increased. The final response map is obtained by multi-layer deformable cross-correlation fusion to enhance the capability of deep semantic feature extraction of the backbone network. Finally, an online template updating strategy with high confidence is adopted, the peak sidelobe ratio and the maximum value of the response graph are calculated in every fixed frame as the basis for updating, and the features are fused in a weighted way to update the template. The tracking performance of the proposed algorithm is evaluated using four common benchmark datasets: OTB2013, OTB2015, VOT2016 and VOT2017. The experimental results show that the overall accuracy and success rate of the DCSiam algorithm are increased by 9.5% and 7.5% respectively compared with the baseline on the OTB2015 dataset, which well realizes the target tracking in complex situations and verifies the effectiveness of the proposed algorithm.

**Keywords:** target tracking; siamese network; deformable convolution; semantic feature; template update

## 0 引言

目标跟踪广泛应用于视频分析、智能交通、虚拟现实等领域, 其核心任务是给定目标在视频第 1 帧初始位置后, 自动估计后续帧中目标的位置. 但跟踪过程中容易受到遮挡、模糊、形变、光照变化等一系列复杂因素的干扰, 因此设计一个可以在现实复杂场景

下稳定高效运行的跟踪器仍是一个极具挑战性的任务.

目前, 大多数跟踪算法大致可以分为两类, 一类是以相关滤波为主的目标跟踪算法. Bolme 等<sup>[1]</sup>将相关滤波的思想应用到跟踪领域, 提出了 MOSEE 算法, 将跟踪任务转化为计算相似度任务, 通过求解最

收稿日期: 2021-01-16; 录用日期: 2021-05-12.

基金项目: 国家自然科学基金项目(61573167); 教育部科技发展中心“云数融合科教创新”基金项目(2017A13055).

<sup>†</sup>通讯作者. E-mail: jnluxl@jiangnan.edu.cn.

高响应值得到目标位置.为了解决MOSSE训练样本不足和只使用单一灰度特征限制跟踪性能的问题, Henriques等<sup>[2]</sup>在MOSSE的基础上提出了KCF算法,利用循环矩阵的特性减少运算并生成额外训练样本,将多通道特征融合来提高跟踪器的性能.虽然基于相关滤波的跟踪器具有高效性和可扩展性,但由于都是在线不断改进跟踪对象,一定程度上会限制跟踪器的表示能力.近年来的工作试图将相关滤波与深度学习相结合,引入深度特征以获得更强的表示能力. Danelljan等<sup>[3]</sup>提出了C-COT算法,使用VGGNet提取深度特征,再将HOG、CN与深度特征融合,自然整合多分辨率特征图,将位置估计转到时域上,在精度上取得了显著提高.此类跟踪算法还有ECO、CSR-DCF等<sup>[4-5]</sup>,虽然精度较好,但是存在速度瓶颈,难以达到实时.

为了取得精度与速度之间的平衡,另一类基于深度学习的孪生网络跟踪算法逐渐流行,孪生网络主要由特征提取过程中共享参数的模板分支与搜索分支组成. Bertinetto等<sup>[6]</sup>利用相似性学习的方法,提出了全卷积孪生网络跟踪(SiamFC)算法,通过计算目标模板与搜索区域深度特征之间的相似值,将跟踪问题转化为相似匹配问题.该算法可以使用大量数据进行端到端的离线训练,既可以获得较高的精度,也无需在线调整.此后在其基础上诞生了很多基于孪生网络的跟踪算法. DSiam算法<sup>[7]</sup>在孪生网络中添加了学习目标外观变化与抑制背景两个线性分支,使得模型拥有了在线学习更新的能力. TriSiam算法<sup>[8]</sup>利用三重损失替代交叉熵损失来训练孪生网络. SiamRPN算法<sup>[9]</sup>将区域建议网络与孪生网络结合,使用区分前景与背景、回归边界框的方法,实现了对目标尺度的灵活应对,同时得到了更为精确的边界框. DaSiamRPN算法<sup>[10]</sup>在SiamRPN的基础上通过添加增量学习,使跟踪器在面对干扰物时也具有好的跟踪性能. CGANSiam算法<sup>[11]</sup>在孪生网络中嵌入条件对抗生成网络模块以提升图像质量,使跟踪器在低分辨率情况下依旧拥有稳定的性能.虽然这些基于孪生网络的跟踪算法取得了不错的成果,但是它们依旧需要面对以下问题.首先,这些跟踪算法的识别能力在很大程度上受限于原SiamFC骨干网络较弱的特征提取能力;其次,这些算法并没有考虑有效在线更新策略,当跟踪过程中目标发生较大变化时,会造成原模板有用信息减少,跟踪器出现漂移得不到及时调整.

本文提出有效解决上述问题的端到端深度网络模型,名为孪生可变形卷积网络.首先,采用可变形卷积<sup>[12]</sup>(deformable convolution, DC)学习特征数据在不同方向上的自适应偏移量,以增强骨干网络的特征提取能力;然后,鉴于深层特征虽语义信息丰富,但缺少位置信息,通过多层可变形互相关融合以提高定位精度;最后,提出一种高置信度模板更新策略,即每隔固定帧计算一次响应图的峰值旁瓣比和最大值,以此为依据采用加权的方式融合特征以更新模板.该策略既能够保证有效更新,也能够防止在遮挡等特殊情况下更新模板,所提出算法在OTB2013、OTB2015<sup>[13]</sup>、VOT2016及VOT2017<sup>[14]</sup>这4个公共跟踪数据集上取得了较好的跟踪性能.

## 1 基于全卷积孪生网络跟踪算法

SiamFC算法将孪生网络框架在目标领域的应用推广开来,通过计算模板与搜索区域深度特征之间的相似性进行匹配,以达到跟踪的目的.其充分发挥了端到端训练的学习优势,不仅在精度上取得了不错的成效,而且在速度上也远超实时性.该算法使用大量标注目标位置信息的视频数据离线训练相似性度量函数.第1步,将用于训练的图像数据进行裁剪缩放,得到以目标为中心 $255 \times 255$ 大小的搜索图像与 $127 \times 127$ 模板图像.第2步,将模板图像与多个搜索图像组成的图像对,输入到端到端网络中进行训练.每对输入都会得到一个相似性度量函数计算出的响应图,其中视觉值高的区域为目标区域,其余为非目标区域.经过大量视频数据训练得到最优参数的相似性度量函数 $f(z, x)$ ,即

$$f(z, x) = \varphi(z) * \varphi(x) + b_1. \quad (1)$$

其中: $\varphi$ 为卷积特征提取函数,\*为卷积互相关运算, $b_1$ 为响应图每个位置的偏置值.该网络采用了二分类逻辑损失函数,其表达式如下:

$$l(y, v) = \log(1 + \exp(-yv)). \quad (2)$$

其中: $v$ 为图像对的最终得分; $y \in \{-1, +1\}$ 为该图像对的真实标签,计算模板与多个候选搜索区域的相似性.

## 2 基于可变形卷积孪生网络跟踪算法

DCSiam算法框架如图1所示,主要由输入、孪生骨干网络、自适应偏移生成模块、可变形互相关层与模板更新组成.本节所有输入图像、特征数据和卷积都是3D的,可变形互相关<sup>[15]</sup>在2D空间域上进行操作,并且该操作在整个通道维度上保持不变.

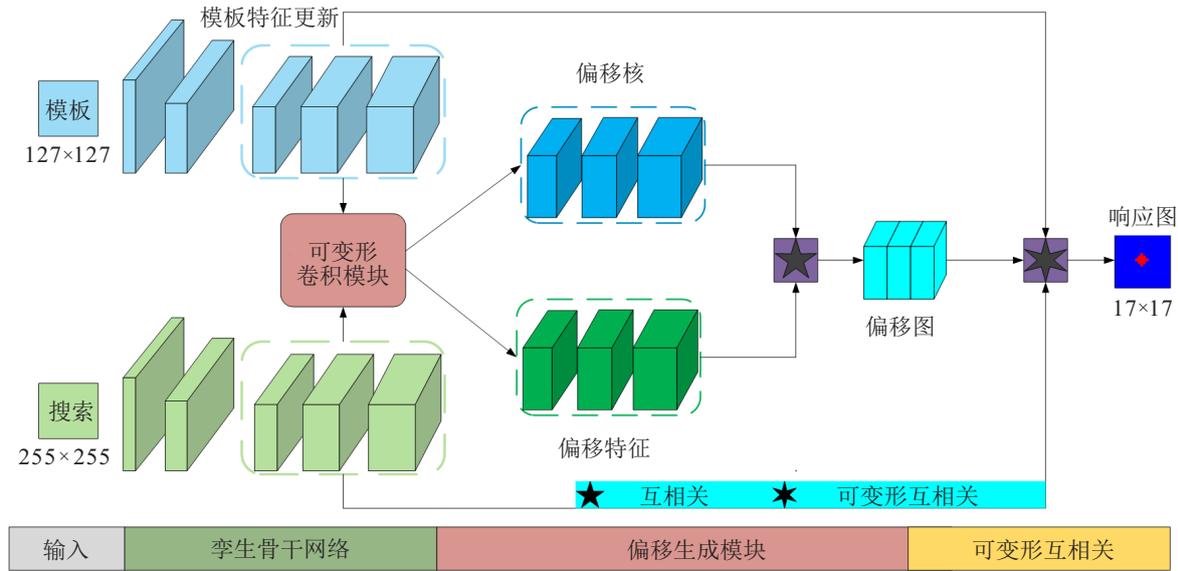


图1 基于可变形卷积的孪生网络框架

### 2.1 孪生骨干网络

在孪生骨干网络中,一个步幅为  $k$  的无填充全卷积神经网络定义如下:

$$f(I((x_a, y_a), (x_b, y_b))) = f(I)\left(\left(\frac{x_a}{k}, \frac{y_a}{k}\right), \left(\frac{x_b}{k}, \frac{y_b}{k}\right)\right). \quad (3)$$

其中:  $I$  为输入图像矩阵;  $I((x_a, y_a), (x_b, y_b))$  为  $I$  的子矩阵,以  $(x_a, y_a)$  和  $(x_b, y_b)$  为子矩阵的左上角与右下角;  $f(\cdot)$  为特征提取函数. 孪生骨干网络主要用于特征提取,由模板分支与搜索分支组成,模板分支与搜索分支分别接收模板图像与当前帧搜索图像作为输入、输出模板特征与搜索特征. 模板分支与搜索分支在特征提取的过程中共享参数.

### 2.2 自适应偏移生成模块

自适应偏移生成模块由一个可变形卷积模块和一个互相关层组成. 可变形卷积模块分别接收模板特征与搜索特征,选择孪生骨干网络中后3层特征数据作为输入,分别输出偏移核与偏移特征. 在提取偏移核与偏移特征的过程中,使用的都是满足式(3)的  $3 \times 3$  卷积. 输出偏移核的维度分别为  $64 \times 32 \times 200$ 、 $36 \times 32 \times 128$  与  $36 \times 32 \times 72$ , 偏移特征的维度分别为  $576 \times 32$ 、 $484 \times 32$  与  $400 \times 32$ . 偏移核与偏移特征进行互相关操作得到偏移图,记作  $c$ , 维度为  $17 \times 17 \times (200 + 128 + 72)$ . 在偏移图  $c$  中,两个相邻通道分别对应着模板特征中一个元素  $x$  与  $y$  方向的偏移,偏移图  $c$  将用于后面的可变形互相关运算.

### 2.3 可变形互相关层

SiamFC 中互相关层接受模板与搜索特征作为输入执行互相关运算,输出响应图  $y$ , 响应图  $y$  表现了

模板与搜索图像之间的相似性分数. 2D 互相关运算包括两个步骤: 1) 使用规则网格  $R$  在搜索特征  $x$  上采样; 2) 将模板特征作为卷积核  $w$  加权采样值总和. 网格  $R$  定义了感受野大小和膨胀,例如在 SiamFC 中

$$R = \{(0, 0), (0, 1), \dots, (k - 1, k), (k, k)\} \quad (4)$$

定义了一个膨胀为1, 填充为0的  $k \times k$  内核. 互相关操作中,对于输出响应图上的每个位置  $a$ , 计算如下:

$$y(a) = \sum_{p_n \in R} w(p_n) \cdot x(a + p_n), \quad (5)$$

其中  $p_n$  枚举了  $R$  上所有的位置. 与 SiamFC 中的互相关运算不同, 本文可变形互相关运算将模板特征  $w$ 、搜索特征  $x$  与偏移图  $c$  作为输入, 用偏移集  $\{\Delta p_n | n = 1, 2, \dots, N\} (N = |R|)$  扩张常规网格  $R$ , 在可变形互相关操作中,对于输出响应图  $y$  上的每个位置  $a$ , 计算如下:

$$y(a) = \sum_{p_n \in R} w(p_n) \cdot x(a + p_n + \Delta p_n), \quad (6)$$

其中所有的  $\Delta p_n$  的值来自输入偏移图  $c$  在通道维度上的位置  $a$ . 式(6)中, 采样在不规则且偏移的位置  $p_n + \Delta p_n$ , 偏移  $\Delta p_n$  通常为小数, 式(6)可通过如下双线性插值实现:

$$x(p) = \sum_q G(q^*, p) \cdot x(q^*). \quad (7)$$

其中:  $p$  为式(4)任意位置  $(a + p_n + \Delta p_n)$ ,  $q^*$  枚举搜索特征中所有整数空间位置,  $G(\cdot)$  为双线性插值核. 这里  $G$  是一个二维核, 它可以分为两个一维核, 即

$$G(q^*, p) = g(q_x^*, p_x) \cdot g(q_y^*, p_y), \quad (8)$$

其中  $g(a, b) = \max(0, 1 - |a - b|)$ .

## 2.4 模板更新

高置信度的固定间隔模板更新策略主要分为两步.

第1步,计算响应图的峰值旁瓣比<sup>[16]</sup>和最大值,峰值旁瓣比 $F$ 计算如下:

$$F = \frac{(A_{\max} - A_{\min})^2}{\text{mean}\left(\sum_{i,j} (A_{i,j} - A_{\min})^2\right)}, \quad (9)$$

其中 $A_{\max}$ 、 $A_{\min}$ 分别为响应图 $A$ 的最大值、最小值.将第1张响应图的峰值旁瓣比与最大值记为 $F_1$ 与 $M_1$ ,并作为阈值.每隔5帧重新计算第 $i$ 帧响应图的峰值旁瓣比和最大值,如果 $F_i$ 和 $M_i$ 均大于历史均值,则选择更新模板,并以历史峰值旁瓣比均值和最大值均值更新阈值.

第2步,当模板被判定更新时,第 $i+1$ 帧以第 $i$ 帧获取目标为中心截取 $127 \times 127 \times 3$ 大小的图像,图像通过孪生骨干网络提取特征,并将第 $i$ 帧图像特征与第1帧模板特征进行加权融合.之所以使用第1帧模板特征,是因为第1帧模板处于非遮挡情况下,其语义信息最为完整丰富.由于相邻帧信息差距很小,频繁的更新操作会造成时间代价过大,继而选择上述高置信度更新策略,只需要花费较少的时间资源,即可使得模板信息得到及时准确更新,在后续互相关运算中得到的响应图更为精确.模板更新公式为

$$f_i = (1 - \eta)f_1(z) + \eta f_i(z). \quad (10)$$

其中: $\eta$ 为超参数, $f_1(z)$ 为第1帧模板特征, $f_i(z)$ 为第 $i$ 帧图像提取的模板特征.

## 3 实验结果与分析

为了验证所提出算法的有效性,实验在OTB2013、OTB2015、VOT2016和VOT2017公开基准数据集上进行,并与多个先进算法进行比较,大量实验显示所提出算法具有较好的跟踪效果.

### 3.1 实验平台及训练细节

本文算法在Ubuntu16.04环境下使用Pytorch编程框架实现,并用GPU加速,平均帧率达到了35FPS.硬件平台为装有Intel(R) Xeon(R) Silver 4110 CPU、64 GB RAM, Nvidia Tesla P100 GPU的服务器.算法使用公开视觉跟踪数据集GOT-10K<sup>[17]</sup>、LaSOT<sup>[18]</sup>作为训练数据集,从其中提取模板图像与搜索图像离线训练网络结构.采用动量为0.9的随机梯度下降(SGD)训练整个网络模型,将权重衰减设置为 $10^{-2}$ ,学习率从 $10^{-2}$ 几何式下降到 $10^{-5}$ ,模型训练50个轮次,小批量规模为12,损失函数与SiamFC相同,式(4)分别为10、8、6.式(10)中超参数 $\eta$ 经实验测试设置为0.5.

## 3.2 数据集简介与评价指标

### 3.2.1 OTB2013/OTB2015数据集

OTB2013和OTB2015数据集分别包含50与100个视频序列,其中OTB2015是在OTB2013基础上拓展了50个序列.两个数据集针对在跟踪中11种复杂因素对不同序列标注了相应标签,包括光照变化、平面外旋转、尺度变化、快速运动、运动模糊、遮挡、背景干扰、视野外、平面内旋转、低分辨率、平面外旋转.本文在OTB2013与OTB2015数据上采用一次通过(one pass evaluation, OPE)得到的精确率与成功率评估跟踪器的性能.

### 3.2.2 VOT2016/VOT2017数据集

VOT2016与VOT2017数据集包含60个不同的挑战序列.VOT2017与VOT2016在最具挑战性的10个序列不同,其余相同.在评估跟踪器性能时,采用基准模式、非监督模式与实时模式3类评估模式,最后通过平均重叠率(expected average overlap, EAO)、精度( $A$ )、鲁棒性( $R$ )三个指标反映跟踪器性能.

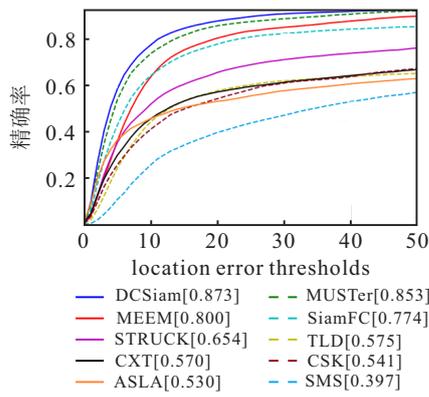
## 3.3 实验分析

### 3.3.1 OTB定量实验分析

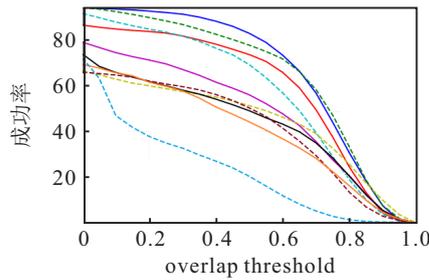
首先,选取目前先进的基于孪生网络跟踪算法,如SiamRPN、DaSiamRPN、SASiam<sup>[19]</sup>、StructSiam<sup>[20]</sup>、DACDSiam<sup>[21]</sup>、CGANSiam、TriSiam、SiamDW-FC<sup>[22]</sup>等与所提出算法DCSiam在OTB2015数据集上进行评估,具体见表1.由表1可见,所提出算法精确率与成功率分别取得了第1与第3的排名,其中精确率对比强基线SiamRPN提升了1.1个百分点,体现出所提出算法具有较强的竞争力.图2为DCSiam与基线算法SiamFC和一些具有经典代表性的算法相比,在不同阈值下OTB基准测试的精确率与成功率曲线.实验结果表明,所提出算法结果明显优于基线,在OTB2013数据集上,精确率与成功率分别较基线提升10.0%和7.2%.在OTB2015数据集上,精确率与成功率较基线分别提升9.5%和7.5%.

表1 OTB2015评估结果

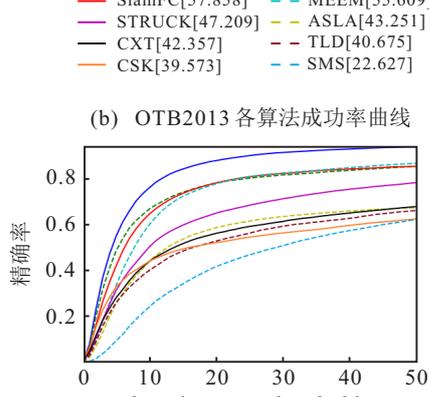
tracker	year	precision	success
SiamRPN	2018	<b>0.851</b>	0.637
DaSiamRPN	2018		<b>0.658</b>
SASiam	2018	<b>0.856</b>	<b>0.657</b>
StructSiam	2018	0.851	0.621
TriSiam	2018	0.780	0.590
SiamDW-FC	2019	0.850	0.640
CGANSiam	2020	0.760	0.580
DACDSiam	2020	0.778	0.638
DCSiam	2021	<b>0.862</b>	<b>0.648</b>



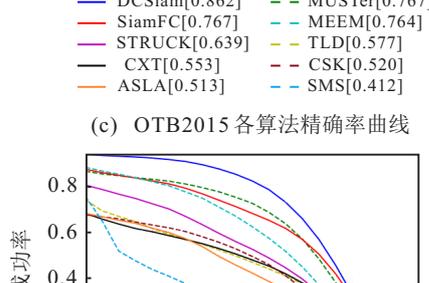
(a) OTB2013 各算法精确率曲线



(b) OTB2013 各算法成功率曲线



(c) OTB2015 各算法精确率曲线



(d) OTB2015 各算法成功率曲线

图2 OTB数据集上不同算法的精确率和成功率曲线

3.3.2 OTB定性分析

为了可视化实际的跟踪结果,从OTB2015数据集中选取5组具有代表性的视频序列,分别为

Basketball、Girl2、CarDark、MotorRolling、Skiing,代表相似干扰、遮挡、光照变化、快速运动、低分辨率等复杂因素。

图3为SiamFC、SiamRPN、DaSiamRPN与DCSiam在上述5组视频序列中可视化对比。可以看出,在面对相似物干扰时,所有跟踪器先后发生漂移,这也是基于孪生网络的跟踪算法的一个弊端。在遇到遮挡这种情况,所有跟踪器都会漂移,但是得益于本文更强的骨干网络可提取到更多的深层语义信息,当遮挡消失时,本文跟踪器迅速寻回目标。在目标光照变化、快速运动和低分辨等复杂情况下,所提算法跟踪效果依旧十分精确,与SiamRPN与DaSiamRPN相比,性能十分具有竞争力。从跟踪结果看,所提算法明显优于基准算法,与先进算法对比也有较强的竞争力。但是,在面对遮挡与相似物干扰时,所提算法也会出现模型漂移的问题。

3.3.3 VOT实验定量分析

对于VOT2016、VOT2017标准数据集的60组跟踪图像序列,选择与SiamRPN、SiamFC及VOT2016与VOT2017挑战赛前5名的跟踪器进行对比,这些跟踪器分别为Stape、STAPE+、SSKCF、DPT、SiamFC-R、CSRDCF+、ECOHC、KFebT、ASMS。算法评价指标采用EAO、A和R,其性能对比如表2所示。可以看出,本文跟踪算法在VOT2016和VOT2017数据集上分别取得了0.56和0.51的准确率,0.31和0.50的鲁棒性以及0.318和0.231的EAO分数。此外,图4还显示出上述跟踪算法在VOT2016与VOT2017数据集上的EAO分数排名,所提算法皆排在第2,仅次于SiamRPN。实验结果表明,不管与VOT2016或VOT2017中排名靠前的跟踪器相比,还是与一些基于孪生网络的先进跟踪器相比,所提算法都体现出较强的竞争力。

表2 VOT基准评估结果

tracker	VOT2016			VOT2017		
	A	R	EAO	A	R	EAO
STAPLE+	<b>0.56</b>	<b>0.37</b>	<b>0.296</b>	/	/	/
SSKCF+	0.55	0.37	0.277	/	/	/
DPT	0.49	0.49	0.236	/	/	/
SiamFC-R	0.55	0.38	0.277	/	/	/
Staple	0.54	0.37	0.295	<b>0.53</b>	0.69	0.170
CSRDCF+	/	/	/	0.46	<b>0.40</b>	<b>0.212</b>
ECO-HC	/	/	/	0.49	0.57	0.177
KFebT	/	/	/	0.45	0.684	0.169
ASMS	/	/	/	0.49	0.630	0.168
SiamFC	0.46	0.53	0.235	0.50	0.59	0.188
SiamRPN	<b>0.56</b>	<b>0.26</b>	<b>0.344</b>	<b>0.49</b>	<b>0.46</b>	<b>0.243</b>
DCSiam	<b>0.56</b>	<b>0.31</b>	<b>0.318</b>	<b>0.51</b>	<b>0.50</b>	<b>0.231</b>

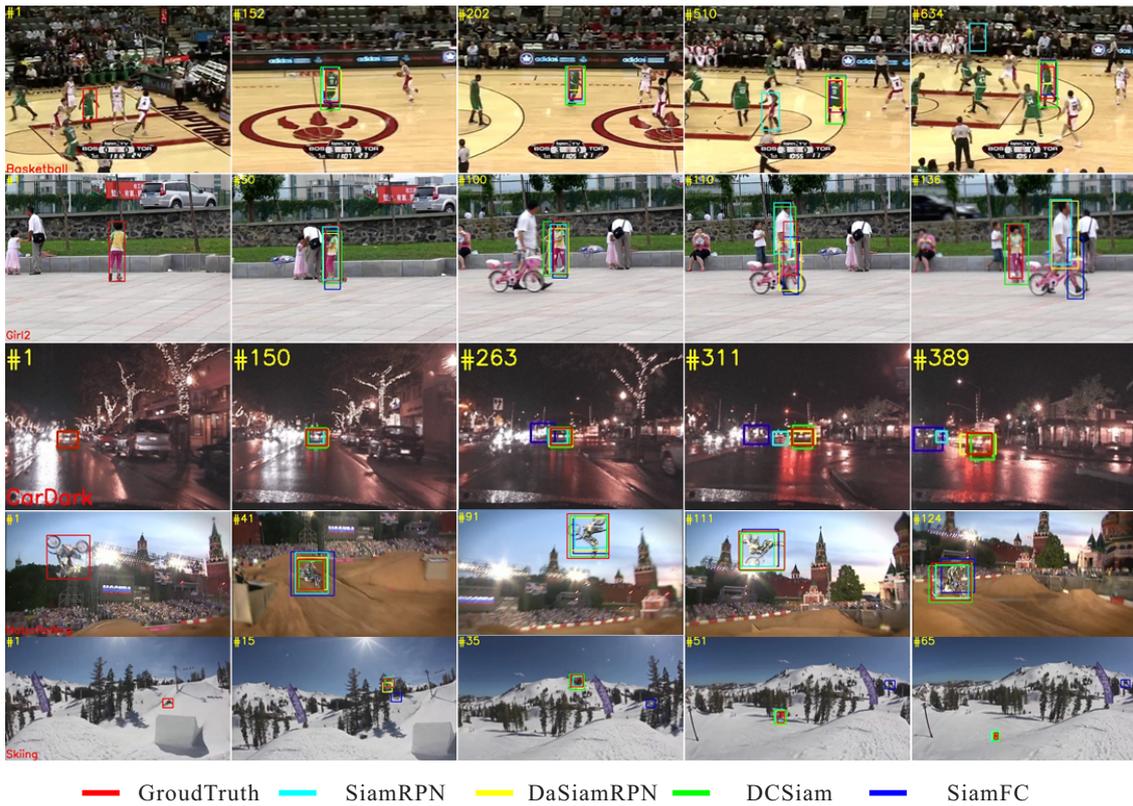


图3 OTB2015上不同算法的可视化对比

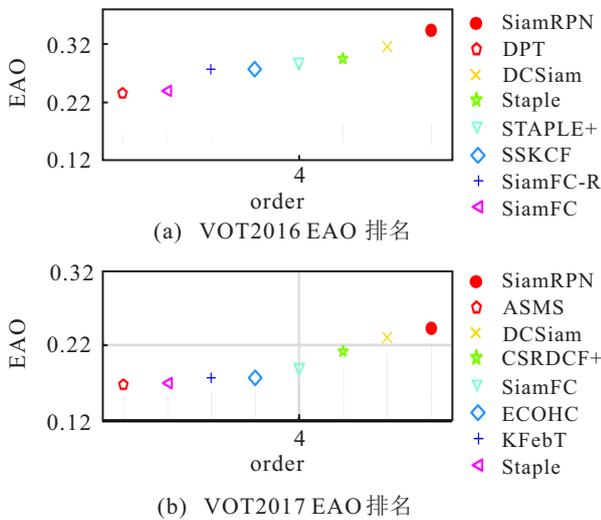


图4 VOT2016/VOT2017 EAO分数排名

### 4 结论

本文提出了一种基于孪生网络的可变形卷积目标跟踪算法,通过强化骨干网络的特征提取能力达到更好的识别效果.对多层数据学习并融合以提高目标的定位精度,引入高置信度的间隔模板更新策略,帮助孪生网络及时有效地更新模板信息.实验结果表明,所提出算法较基准SiamFC在性能提升上较为明显,也更为稳健.今后的工作着重关注相似物干扰识别的问题,以进一步提升算法在复杂场景下的跟踪能力.

### 参考文献(References)

- [1] Bolme D S, Beveridge J R, Draper B A, et al. Visual object tracking using adaptive correlation filters[C]. IEEE Computer Society Conference on Computer Vision and Pattern Recognition. San Francisco, 2010: 2544-2550.
- [2] Henriques J F, Caseiro R, Martins P, et al. High-speed tracking with kernelized correlation filters[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2015, 37(3): 583-596.
- [3] Danelljan M, Robinson A, Khan F S, et al. Beyond correlation filters: Learning continuous convolution operators for visual tracking[C]. Computer Vision—ECCV. Cham: Springer, 2016: 472-488.
- [4] Danelljan M, Bhat G, Khan F S, et al. ECO: Efficient convolution operators for tracking[C]. IEEE Conference on Computer Vision and Pattern Recognition. Honolulu, 2017: 6931-6939.
- [5] Lukežić A, Vojír T, Zajc L C, et al. Discriminative correlation filter with channel and spatial reliability[C]. IEEE Conference on Computer Vision and Pattern Recognition. Honolulu, 2017: 4847-4856.
- [6] Bertinetto L, Valmadre J, Henriques J F, et al. Fully-convolutional siamese networks for object tracking[C]. Computer Vision—ECCV 2016. Cham: Springer, 2016: 850-865.
- [7] Guo Q, Feng W, Zhou C, et al. Learning dynamic

- Siamese network for visual object tracking[C]. IEEE International Conference on Computer Vision. Venice, 2017: 1781-1789.
- [8] Dong X P, Shen J B. Triplet loss in siamese network for object tracking[C]. Computer Vision—ECCV 2018. Cham: Springer, 2018: 472-488.
- [9] Li B, Yan J J, Wu W, et al. High performance visual tracking with Siamese region proposal network[C]. 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City, 2018: 8971-8980.
- [10] Zhu Z, Wang Q, Li B, et al. Distractor-aware siamese networks for visual object tracking[C]. Computer Vision—ECCV 2018. Cham: Springer, 2018: 101-117.
- [11] 宋建辉, 张甲, 刘砚菊, 等. 基于条件对抗生成孪生网络的目标跟踪[J]. 控制与决策, 2021, 36(5): 1110-1118.  
(Song J H, Zhang J, Liu Y J, et al. Conditional generative adversarial siamese networks for object tracking[J]. Control and Decision, 2021, 36(5): 1110-1118.)
- [12] Dai J F, Qi H Z, Xiong Y W, et al. Deformable convolutional networks[C]. IEEE International Conference on Computer Vision. Venice, 2017: 764-773.
- [13] Wu Y, Lim J, Yang M H. Object tracking benchmark[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2015, 37(9): 1834-1848.
- [14] Kristan M, Leonardis A, Matas J, et al. The visual object tracking VOT2017 challenge results[C]. Proceedings of the IEEE International Conference on Computer Vision Workshops. Piscataway: IEEE, 2017: 1949-1972.
- [15] Zheng L Y, Chen Y Y, Tang M, et al. Siamese deformable cross-correlation network for real-time visual tracking[J]. Neurocomputing, 2020, 401: 36-47.
- [16] Wang M M, Liu Y, Huang Z Y. Large margin object tracking with circulant feature maps[C]. IEEE Conference on Computer Vision and Pattern Recognition. Honolulu, 2017: 4800-4808.
- [17] Huang L H, Zhao X, Huang K Q. GOT-10k: A large high-diversity benchmark for generic object tracking in the wild[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2021, 43(5): 1562-1577.
- [18] Fan H, Lin L T, Yang F, et al. LaSOT: A high-quality benchmark for large-scale single object tracking[C]. IEEE/CVF Conference on Computer Vision and Pattern Recognition. Long Beach, 2019: 5369-5378.
- [19] Zhang J M, Ma S G, Sclaroff S. MEEM: Robust tracking via multiple experts using entropy minimization[C]. Computer Vision—ECCV 2014. Cham: Springer, 2014: 188-203.
- [20] Zhang Y H, Wang L J, Qi J Q, et al. Structured Siamese network for real-time visual tracking[C]. Computer Vision—ECCV 2018. Cham: Springer, 2018: 355-370.
- [21] 许龙, 魏颖, 商圣行, 等. 基于异步相关判别性学习的孪生网络目标跟踪算法[J/OL]. 自动化学报, DOI: 10.16383/j.aas.c200237.  
(Xu L, Wei Y, Shang S H, et al. Design of asynchronous correlation discriminant single object tracker based on siamese network[J/OL]. Acta Automation Sinica, DOI: 10.16383/j.aas.c200237.)
- [22] Zhang Z P, Peng H W. Deeper and wider Siamese networks for real-time visual tracking[C]. IEEE/CVF Conference on Computer Vision and Pattern Recognition. Long Beach, 2019: 4586-4595.

### 作者简介

刘如浩(1995—), 男, 硕士生, 从事目标跟踪、目标检测的研究, E-mail: 6191918021@stu.jiangnan.edu.cn;

张家想(1996—), 男, 硕士生, 从事图像处理的研究, E-mail: 6191918029@stu.jiangnan.edu.cn;

金晨曦(1996—), 男, 硕士生, 从事人工智能的研究, E-mail: 6191918019@stu.jiangnan.edu.cn;

卢先领(1972—), 男, 教授, 博士, 从事图像处理、人工智能、传感器网络技术的研究, E-mail: jnluxl@jiangnan.edu.cn.

(责任编辑: 郑晓蕾)