

# 融合注意力机制与自适应模板更新的目标跟踪算法

仝卫国 寇德龙 李茂冉 石宗锦

(华北电力大学自动化系 保定 071003)

**摘要:** 针对基于孪生网络的跟踪算法在目标快速移动、发生较大形变、处于复杂背景等情况下容易出现跟踪性能下降的问题,提出一种融合注意力机制与自适应模板更新的目标跟踪算法。跟踪算法以 SiamRPN 为基础,通过在特征提取网络融合通道注意力机制与空间注意力机制,抑制图像中的干扰信息,补充目标特征在通道空间中的信息,更好地对目标进行定位。将对象在不同时刻的模板,包括初始模板、累积模板和预测模板作为帧残差模块的输入,采用残差学习策略,充分利用初始模板语义信息,自适应更新当前帧所需的模板,减少了跟踪漂移的现象。在 OTB100 数据集上的实验结果表明,本文跟踪算法与其他跟踪算法相比取得了更高的跟踪成功率和准确率。

**关键词:** 目标跟踪;孪生网络;SiamRPN;注意力机制;自适应模板更新

**中图分类号:** TP391.41 **文献标识码:** A **国家标准学科分类代码:** 520.604

## Target tracking algorithm combining attention mechanism and adaptive template updating

Tong Weiguo Kou Delong Li Maoran Shi Zongjin

(Department of Automation, North China Electric Power University, Baoding 071003, China)

**Abstract:** Aiming at the problem that the tracking algorithm based on twin network is prone to degradation under the condition of fast moving target, large deformation and complex background, a target tracking algorithm based on the integration of attention mechanism and adaptive template updating is proposed. Based on SiamRPN, the tracking algorithm combines the channel attention mechanism and the spatial attention mechanism in the feature extraction network to suppress the interference information in the image, supplement the target feature information in the channel space, and better locate the target. The template of the object at different times, including the initial template, the accumulated template and the predicted template, is taken as the input of the residual module. The residual learning strategy is adopted to make full use of the semantic information of the initial template and adaptively update the template needed for the current frame, which reduces the phenomenon of tracking drift. Experimental results on the OTB100 dataset show that the proposed tracking algorithm achieves higher tracking success rate and accuracy compared with other tracking algorithms.

**Keywords:** target tracking; siamese network; SiamRPN; attention mechanism; adaptive template update

## 0 引言

近几年,随着计算机视觉的发展,目标跟踪算法在人机交互、自动驾驶以及视频监控等领域有着广泛的应用<sup>[1]</sup>。然而在实际跟踪场景中,因目标物体快速移动、形变、运动模糊、遮挡和复杂背景等因素,目标跟踪仍然存在诸多挑战和困难。

目标跟踪可以分为基于生成式模型的算法和基于判别式模型的算法<sup>[2]</sup>。基于生成式模型的算法通过学习一个代表目标的模型,利用它去搜索图像区域,在图像中找到和这

个模型最匹配的区域,从而得到目标的预测位置。基于判别式模型的算法将目标跟踪问题看成一个二分类问题,找到目标和背景的决策边界,通过分类将目标与非目标分别开来,从而预测目标位置。

随着深度学习的发展,基于孪生网络的目标跟踪算法取得了很大的进展。Bertinetto 等<sup>[3]</sup>提出全卷积孪生网络目标跟踪算法 SiamFC,将模板图像和搜索图像输入到特征提取网络获得图像特征,使两者进行互相关操作后得到响应图,寻找峰值区域从而得到预测的目标位置。Valmadre 等<sup>[4]</sup>在 SiamFC 网络框架模板分支上融合相关滤波

(correlation filter, CF), 提出基于孪生网络的反向传播滤波器 CFNet, 将卷积神经网络 CNN 与 CF 的在线学习效率的优势结合起来, 实现端到端的训练。Li 等<sup>[5]</sup>将区域候选网络(region proposal network, RPN)引入孪生网络目标跟踪中, 提出孪生区域候选网络跟踪算法 SiamRPN, 改善了传统的多尺度检测方法。

为了提高跟踪效果, 注意力机制被频繁使用。Wang 等<sup>[6]</sup>在网络结构融合了残差注意力和通道注意力, 提高了跟踪性能。Chen 等<sup>[7]</sup>将注意力模块与长短时记忆网络模块结合, 以增强表示对象的能力。王淑贤等<sup>[8]</sup>在孪生网络特征提取阶段融合多尺度通道特征, 整合了空间信息和运动信息, 但却忽略了通道特征对整体特征的重要贡献。Yang 等<sup>[9]</sup>在 RPN 网络中增加注意力分支, 通过加权融合分类分支和注意力分支的得分来区分前景与背景, 但分类与回归的模板语义特征并未被充分提取。

基于孪生网络的跟踪算法模板为不更新机制, 无法更新模板可能会导致跟踪器过早失效。Guo 等<sup>[10]</sup>提出在 FFT 域封闭形式的转换学习模型, 能够有效地在在线学习目标外观变化, 但却忽略了原始模板语义信息。Yan 等<sup>[11]</sup>设置当算法满足动态模板更新间隔的阈值和置信阈值时, 将替换当前模板, 但却没有与累计模板特征信息建立联系。李雪等<sup>[12]</sup>引入在线更新机制, 当置信度小于阈值时进行模板线性特征融合, 但是简单的线性更新通常不足以应付不断变化的更新需求。

上述算法虽然取得了一定的效果, 然而当目标快速移动、处于复杂背景等情况时, 上述方法难以提取目标语义位置信息, 且跟踪算法模板不更新或简单的线性更新会带来跟踪器过早失效的问题。因此本文基于 SiamRPN, 通过在特征提取网络融合通道注意力机制与空间注意力机制, 更好地对目标进行定位; 将对象在不同时刻的模板, 包括初始模板、累积模板和预测模板作为帧残差模块的输入, 自适应更新当前帧所需的模板。实验表明本文算法在 OTB100 数据集上具有良好的成功率与准确率。

## 1 SiamRPN 目标跟踪算法

SiamRPN 结合了跟踪中的孪生网络和检测中的区域候选网络; 孪生网络让算法可以利用被跟踪目标的信息, 完成检测器的初始化; 区域候选网络让算法可以对目标位置进行更精准的预测。经过两者的结合, SiamRPN 可以进行端到端的训练。

孪生网络由两个分支组成。一个是模板分支, 接收初始帧中的目标作为输入  $z$ , 另一个是检测分支, 接收当前帧中的目标作为输入  $x$ , 这两个分支在特征提取网络中共享参数。 $\varphi(z)$  和  $\varphi(x)$  表示为孪生网络的输出特征。

区域候选网络有两个分支, 一个是分类分支进行前景背景分类, 另一个是回归分支进行位置回归。首先通过两个卷积层将  $\varphi(z)$  分成两个分支  $[\varphi(z)]_{cls}$  和  $[\varphi(z)]_{reg}$ , 它们

分别有  $2k$  和  $4k$  的通道。再通过两个卷积层将  $\varphi(x)$  也分成了两个分支  $[\varphi(x)]_{cls}$  和  $[\varphi(x)]_{reg}$ , 将两者分别在分类分支和回归分支上进行卷积运算:

$$A_{w \times h \times 2k}^{cls} = [\varphi(x)]_{cls} * [\varphi(z)]_{cls} \quad (1)$$

$$A_{w \times h \times 4k}^{reg} = [\varphi(x)]_{reg} * [\varphi(z)]_{reg} \quad (2)$$

式中:  $*$  表示卷积操作, 模板特征  $[\varphi(z)]_{cls}$  和  $[\varphi(z)]_{reg}$  用作卷积核。 $A_{w \times h \times 2k}^{cls}$  包含  $2k$  个通道, 每组通道表示原始图片上对应位置的每个锚框的前景和背景。 $A_{w \times h \times 4k}^{reg}$  包含  $4k$  个通道, 每组通道表示锚框和真实目标框之间的坐标修正值  $dx, dy, dw, dh$ 。根据下列公式修正锚框:

$$x^{pro} = x^{an} + dx * w^{an} \quad (3)$$

$$y^{pro} = y^{an} + dy * h^{an} \quad (4)$$

$$w^{pro} = w^{an} * e^{dw} \quad (5)$$

$$h^{pro} = h^{an} * e^{dh} \quad (6)$$

式中:  $x^{an}, y^{an}, w^{an}, h^{an}$  表示锚框的中心坐标以及宽度、高度。

将  $A_{w \times h \times 2k}^{cls}$  每奇数个代表正样本的通道进行 softmax 处理, 寻找得分最高的锚框, 并得到与之对应的修正后的锚框位置信息, 从而完成目标跟踪。

## 2 本文跟踪算法

本文跟踪算法框架如图 1 所示, 包含以下 3 个步骤:

1) 首先在 SiamRPN 特征提取网络的后面融合混合注意力模块, 更好地提取图像特征信息。

2) 将初始帧目标区域经过 SiamRPN 的特征提取网络之后, 得到初始模板特征  $T_0^{GT}$ 。由于初始帧之前没有任何帧, 此时将当前帧模板特征  $T_i$  和累计模板特征  $\tilde{T}_{i-1}$  设置为  $T_0^{GT}$ 。将累计模板  $\tilde{T}_{i-1}$  (即  $T_0^{GT}$ ) 分别进行卷积操作得到分类分支与回归分支的卷积核, 与下一帧的检测模板在分类分支与回归分支的特征分别进行互相关操作, 得到目标位置, 将此时预测的目标区域输入到跟踪算法的特征提取网络即可得  $T_i$ , 将  $T_0^{GT}, \tilde{T}_{i-1}, T_i$  输入到自适应模板更新网络 UpdateNet 中, 经过拼接以及一系列卷积操作和残差连接, 得到预测新的累计模板特征  $\tilde{T}_i$ 。

3) 将  $\tilde{T}_i$  与下一帧的检测模板重复之前的操作, 即使模板不断的更新从而提高跟踪算法的跟踪性能。

### 2.1 注意力机制

孪生网络特征提取在快速移动、复杂背景等干扰因素的影响下, 网络的鲁棒性较差。因为不同通道、不同区域的特征图对整个跟踪过程的贡献是有差异的, 所以引入混合注意力机制(convolutional block attention module, CBAM)<sup>[13]</sup>, 其包含两个独立的子模块, 分别是通道注意力模块和空间注意力模块, 如图 2 所示。

计算公式如下:

$$F' = M_C(F) \otimes F \quad (7)$$

$$F'' = M_S(F') \otimes F' \quad (8)$$

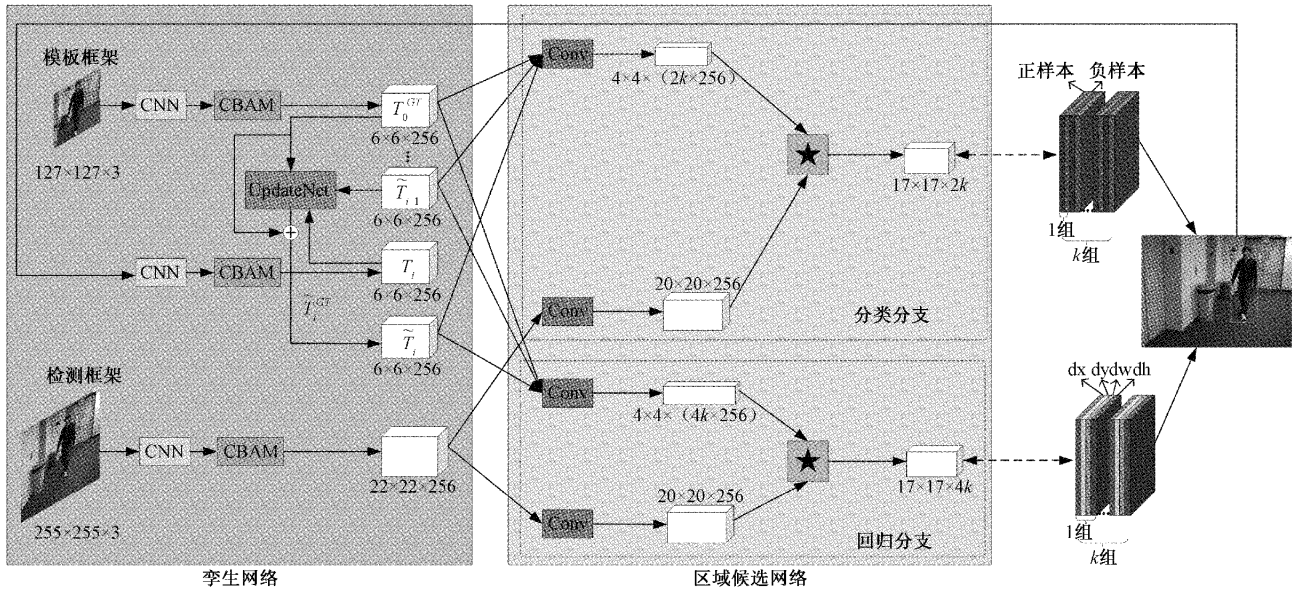


图 1 本文算法框架图

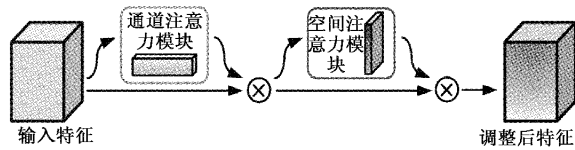


图 2 混合注意力模块

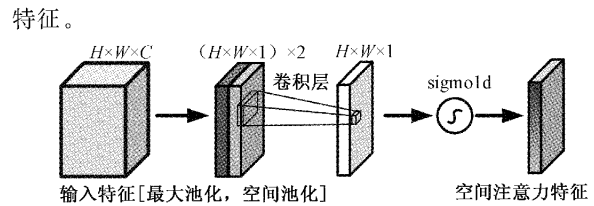


图 4 空间注意力模块

式中： $F$  为输入特征， $M_c$  为通道注意力机制， $M_s$  为空间注意力机制。

通道注意力关注哪个通道上的特征是有意义的，如图 3 所示。首先对输入大小为  $H \times W \times C$  的特征层分别进行空间维度的最大池化和平均池化，得到两个大小为  $1 \times 1 \times C$  的特征层，然后将其送入到多层感知机 (multilayer perceptron, MLP) 学习通道维度的特征，再经过 Sigmoid 函数得到通道注意力特征。

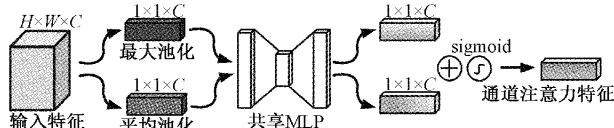


图 3 通道注意力模块

计算公式如下：

$$M_c(F) = \sigma(MLP(AvgPool(F)) + MLP(MaxPool(F))) = \sigma(W_1(W_0(F_{avg}^C)) + W_1(W_0(F_{max}^C))) \quad (9)$$

式中： $\sigma$  是 Sigmoid 函数， $W_0, W_1$  为 MLP 的权重， $F_{avg}^C$  为平均池化后的特征， $F_{max}^C$  为最大池化后的特征。

空间注意力机制注重目标图像在局部位置的特征信息，如图 4 所示。首先对输入大小为  $H \times W \times C$  的特征层分别进行通道维度的最大池化和平均池化操作，得到两个大小为  $H \times W \times 1$  的特征层，将其拼接之后再经过一个  $7 \times 7$  的卷积层，将结果通过 Sigmoid 函数得到空间注意力

计算公式如下：

$$M_s(F) = \sigma(f^{7 \times 7}([AvgPool(F); MaxPool(F)])) = \sigma(f^{7 \times 7}([F_{avg}^S; F_{max}^S])) \quad (10)$$

式中： $f^{7 \times 7}$  为与  $7 \times 7$  大小的卷积核进行卷积操作。

如图 1 所示，本文在对初始帧目标、预测目标以及检测目标的处理上，结合 CBAM 设计混合注意力模块，形成 AlexNet\_CBAM 特征提取网络。即对模板、检测分支经过原特征提取网络 (AlexNet) 所提取到的目标特征，再通过混合注意力机制，形成初始模板 ( $T_0^{GT}$ )、预测模板 ( $T_j$ ) 以及检测模板，使模型能够在端到端训练中自适应地对特征通道、空间加权。其中初始模板与预测模板的大小为  $6 \times 6 \times 256$ ，检测模板的大小为  $22 \times 22 \times 256$ 。

### 2.2 自适应模板更新

一般的孪生跟踪算法使用初始帧计算目标模板，在后续帧中进行追踪。但是视频中目标的外观变化很大时，初始帧模板保持固定或者仅使用线性更新不足以满足模板更新的需要。因此使用 UpdateNet 自适应模板更新网络合<sup>[11]</sup>，给出初始模板、累积模板和当前帧的模板，从而估计下一帧的最佳模板。

UpdateNet 网络结构图如图 5 所示，在本文所提算法中其更新模板主要经过以下 3 个步骤：

1) 首先使用与本文所提算法相同的特征提取网络

(AlexNet\_CBAM),从初始帧中的目标真实位置提取特征  $T_0^{GT}$ ,使用从所有先前帧生成的累积模板  $\tilde{T}_{i-1}$  来预测第  $i$  帧中目标的位置并从该区域中提取特征获得当前帧的  $T_i$ 。

2)连接提取的特征  $T_0^{GT}$ 、 $T_i$  和累积特征  $\tilde{T}_{i-1}$ ,形成 UpdateNet 的输入。经过  $1 \times 1$  大小的卷积进行通道降维,使用 Relu 激活函数后,再经过  $1 \times 1$  的卷积进行通道升维,减少了计算量,完成通道信息融合。

3)UpdateNet 唯一使用的真实信息是初始帧中给定的目标位置,所有其他输入均基于预测位置。因此,采用了残差学习策略,通过从  $T_0^{GT}$  增加一个到 UpdateNet 的输出的跳跃连接。将两者融合输出预测的新的累积模板  $\tilde{T}_i$ 。

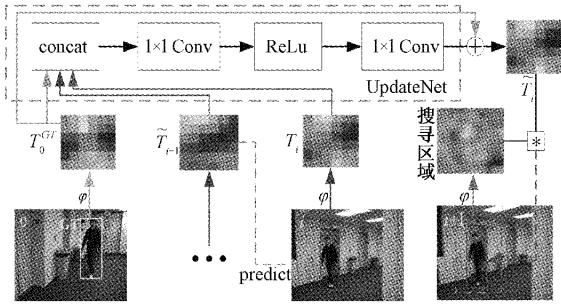


图5 UpdateNet 网络结构图

为了降低训练计算量,使用分阶段来进行网络训练。第一个阶段,在训练集上运行原始的跟踪器,利用标准线性更新策略:

$$\tilde{T}_i^0 = (1 - \gamma)\tilde{T}_{i-1}^0 + \gamma T_i^0 \quad (11)$$

式中: $\gamma$  为更新率, $T_i^0$  为当前帧的模板特征, $\tilde{T}_{i-1}^0$  为累计模板特征。将此时获得的当前模板特征和累计模板特征作为的网络输入。训练过程如图6所示。

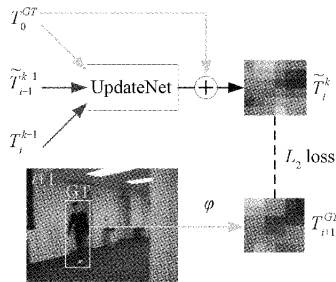


图6 UpdateNet 网络训练结构图

在之后每一次训练阶段,利用每上一个阶段训练出来的 Updatenet 来获得累计模板和预测的目标位置:

$$\tilde{T}_i^k = \phi^k(T_0^{GT}, \tilde{T}_{i-1}^{k-1}, T_i^{k-1}) \quad (12)$$

式中: $\phi^k$  为 UpdateNet 网络。

训练 UpdateNet 来预测下一帧的目标模板,预测的应该和从下一帧的真实位置截取的目标模板越相近越好。

因此通过将更新的模板和下一帧真实位置之间的欧氏距离作为训练的损失函数,公式定义如下:

$$L_2 = \|\phi(T_0^{GT}, \tilde{T}_{i-1}, T_i) - T_{i+1}^{GT}\|_2 \quad (13)$$

在本文中使用  $k=2$  即先利用标准线性更新方法获得训练数据,再通过自适应更新网络迭代一次得到训练的网络。

### 2.3 损失函数

本文跟踪算法使用区域候选网络实现目标的分类和回归。分类损失采用交叉熵损失函数,回归损失采用归一化坐标的平滑  $L_1$  损失  $smooth_{L_1}$ 。首先计算锚框与实际标注间的归一化距离:

$$\begin{aligned} \delta[0] &= \frac{T_x - A_x}{A_w}, \delta[1] = \frac{T_y - A_y}{A_h} \\ \delta[2] &= \ln \frac{T_w}{A_w}, \delta[3] = \ln \frac{T_h}{A_h} \end{aligned} \quad (14)$$

式中: $T_x, T_y, T_w, T_h$  为真实样本的中心点坐标以及宽度、高度; $A_x, A_y, A_w, A_h$  为锚框的中心点坐标以及宽度、高度。

然后通过  $smooth_{L_1}$  损失函数计算损失值,数学表达式如下:

$$smooth_{L_1}(x, \sigma) = \begin{cases} 0.5\sigma^2 x^2, & |x| < \frac{1}{\sigma} \\ |x| - \frac{1}{2\sigma^2}, & |x| \geq \frac{1}{\sigma} \end{cases} \quad (15)$$

式中: $x$  为模型的预测值, $\sigma$  为超参数。

回归分支的损失计算公式:

$$L_{reg} = \sum_{i=0}^3 smooth_{L_1}(\delta[i], \sigma) \quad (16)$$

网络整体的损失函数公式:

$$loss = L_{cls} + \lambda L_{reg} \quad (17)$$

式中: $\lambda$  为超参数, $L_{cls}$  为分类损失。

## 3 实验结果及分析

### 3.1 实验环境

本文的实验环境如表1所示。

表1 实验环境

名称	标准
操作系统	ubuntu20.04
CPU	Intel (R) Core i7-8700K 3.70 GHz
GPU	NVIDIA GeForce RTX 2080Ti
内存	16 GB
Python/PyTorch	3.8/1.6

### 3.2 数据集及评价指标

训练集使用 GOT-10K<sup>[15]</sup> 数据集,基于 WordNet 的大型目标跟踪数据集,广泛覆盖了 560 类户外常见的移动物

体,边界框数量超过 150 万。从训练数据集中选取间隔小于 100 帧的图像对进行裁剪,模板帧为  $127 \times 127$ ,检测帧为  $255 \times 255$ ,使用随机梯度下降算法(SGD)对网络进行训练,学习率从 0.001 线性衰减到 0.000 01,共进行 30 轮迭代。

测试集采用 OTB100<sup>[16]</sup>数据集,包含 100 个视频序列、多种挑战属性,包括快速运动、尺度变化、平面内旋转、变形等情况。本文沿用 OTB100 数据集的标准,采用一次通过评价标准 OPE 来评估跟踪算法的性能。用目标真实的位置初始化第一帧,计算目标跟踪算法在跟踪目标过程中的精确率与成功率。

3.3 消融实验

为了验证注意力机制、自适应模板更新机制在算法中的有效性,在 OTB100 数据集上进行消融实验。分别是使用注意力机制的算法,使用自适应模板更新机制的算法,以及同时使用两种机制的算法。通过表 2 可知,引入注意力机制后成功率提升了 1.1%,准确率提升了 1.2%;引入自适应模板更新后成功率提升了 2.0%,准确率提升了 2.3%。本文所提出跟踪算法在准确率与成功率上均有所提高,每个模块都能一定程度提高跟踪器的性能。

表 2 消融实验结果

注意力 机制	自适应 模板更新	OTB100		
		成功率	准确率	FPS
✓		0.643	0.862	53
	✓	0.652	0.873	46
✓	✓	0.665	0.887	42

3.4 定量分析

为了验证本文算法的有效性,将提出的算法与 KCF<sup>[17]</sup>、CFNet、SiamFC、SiamRPN、DaSiamRPN<sup>[18]</sup> 5 种跟踪算法在 OTB2015 数据集上对比,如图 7、8 所示。本文所提算法在上述 6 个算法中获得最优跟踪性能。本文所提算法的成功率为 0.665,准确率为 0.887,与基准算法 SiamRPN 相比,成功率提高了 3.3%,准确率提高了 3.7%。

表 3 6 种跟踪算法在 OTB100 数据集五种挑战属性上的成功率和准确率

跟踪算法	光照变化		尺度变换		遮挡		变形		运动模糊	
	S	P	S	P	S	P	S	P	S	P
本文	<b>0.639</b>	<b>0.873</b>	<b>0.664</b>	0.867	<b>0.641</b>	<b>0.878</b>	<b>0.671</b>	<b>0.894</b>	<b>0.632</b>	0.868
CFNet	0.564	0.769	0.582	0.786	0.579	0.763	0.553	0.757	0.591	0.734
SiamRPN	0.612	0.853	0.636	0.847	0.619	0.839	0.608	0.816	0.613	<b>0.869</b>
KCF	0.503	0.663	0.506	0.672	0.522	0.699	0.530	0.702	0.516	0.687
DaSiamRPN	0.634	0.870	0.652	<b>0.882</b>	0.637	0.841	0.659	0.874	0.617	0.862
SiamFC	0.577	0.764	0.563	0.783	0.591	0.755	0.582	0.747	0.563	0.739

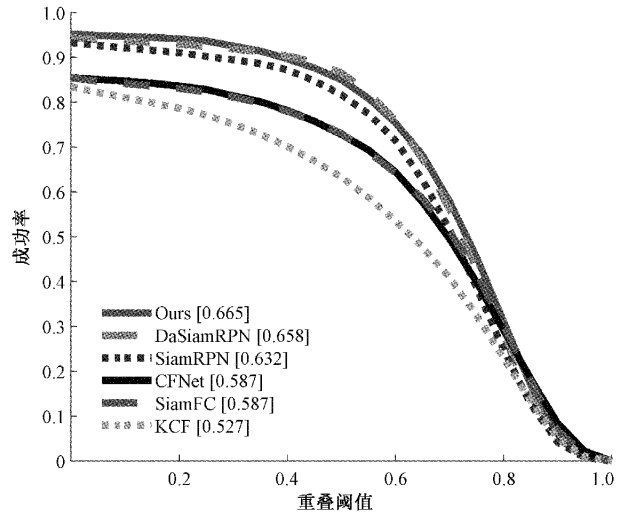


图 7 在 OTB100 上的跟踪成功率

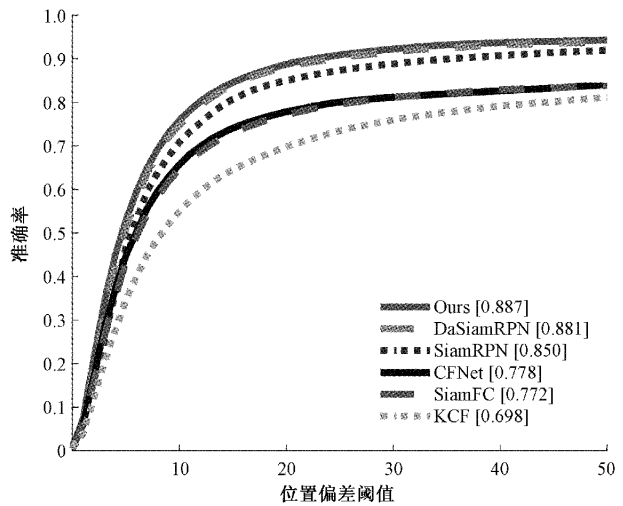


图 8 在 OTB100 上的跟踪准确率

为进一步分析所提算法的跟踪性能,将本文算法与 KCF、SiamRPN 等 5 种跟踪器在具有多种挑战属性的 OTB100 数据集上进行对比,如表 3 所示。所提算法在光照变化、尺度变化、遮挡、变形、运动模糊等属性下视觉跟踪成功率与准确率大都为最优。

### 3.5 定性分析

为了更加直观的观察本文所提算法的跟踪性能,从OTB100数据集中选取具有代表性的5段视频图像序列对

跟踪结果进行分析,并与其他5种跟踪器进行比较。图9展示了6种不同的跟踪器在尺度变换、遮挡、复杂背景、平面内旋转等挑战因素干扰下的跟踪结果。

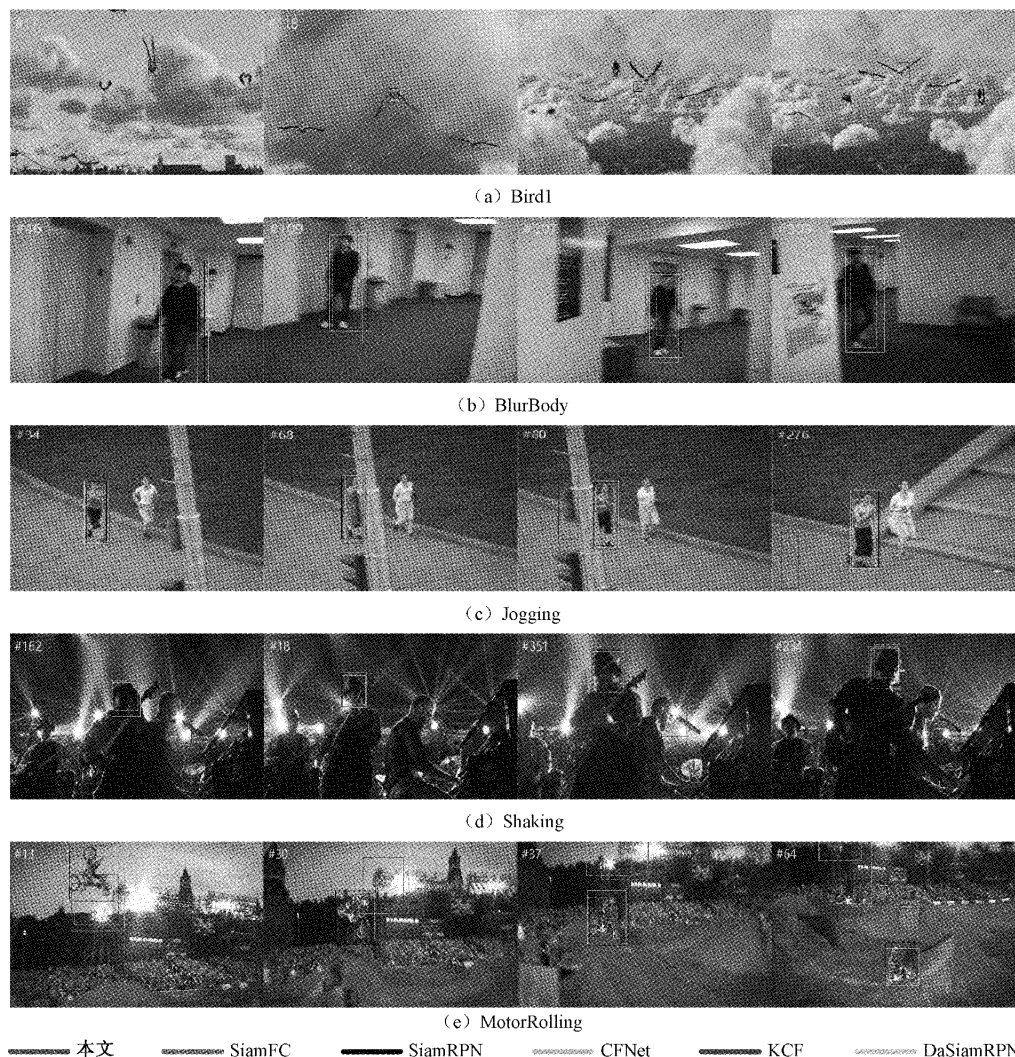


图9 6种算法在5个序列中的跟踪结果

1)尺度变换。在视频帧序列 Bird1、BlurBody中,目标在运动过程中由近及远产生了尺度大小变化的现象。KCF、CFNet算法产生的预测目标框大小无法跟随目标尺寸的变化而变化,SiamRPN等算法出现跟丢现象,由于本文所提算法引入了空间注意力机制,能够充分提取目标在空间维度的信息,从而更加准确的定位目标位置。

2)遮挡。在视频帧序列 Jogging中,在第68帧目标开始部分遮挡,此时所有算法均可跟踪目标,在第80帧目标完全出现时,KCF算法跟丢目标,SiamFC、CFNet等算法目标框只框到物体的一部分,只有本文所提算法可以完美跟踪目标。

3)复杂背景。在视频帧序列 Shaking中,演唱会场景中光线昏暗,无法区分目标和周围的环境,背景信息很容易误导跟踪,其他算法均出现跟踪飘移的现象。由于本文

所提算法引入了通道注意力机制,抑制了干扰信息,从而更加准确的框选目标。

4)平面内旋转。在视频帧序列 MotorRolling中,由于摩托车表演时多次旋转,目标外观与位置变化很快。KCF、SiamFC、CFNet跟丢目标物体,SiamRPN、DaSiamRPN仅跟踪到目标的局部,而本文所提算法由于自适应更新模板信息,所选模板最适合当前帧,从而可以准确的跟踪目标。

## 4 结论

本文算法在基于 SiamRPN 的基础上,提出了一个融合注意力机制与自适应模板更新的孪生网络跟踪算法。通过在特征提取网络融合通道注意力机制与空间注意力机制,有效区分有用特征与无用特征,提升了跟踪的准确

率。将对象在不同时刻的模板,包括初始模板、累积模板和预测模板作为帧残差模块的输入,动态更新当前帧所需的模板,提升了跟踪的成功率。在 OTB100 数据集上的实验结果表明,本文跟踪算法面对各种干扰因素都取得了较好的跟踪准确率和成功率。接下来将在保证跟踪性能的前提下提高跟踪的实时性。

### 参考文献

- [1] 付谱平,叶俊.融合语义特征网络的孪生网络目标跟踪算法[J].电子测量技术,2022,45(8):136-142.
- [2] 孟球,杨旭.目标跟踪算法综述[J].自动化学报,2019,45(7):1244-1260.
- [3] BERTINETTO L, VALMADRE J, HENRIQUES J F, et al. Fully-convolutional siamese networks for object tracking [ C ]. Computer Vision-ECCV 2016 Workshops: Amsterdam, The Netherlands, 2016: 850-865.
- [4] VALMADRE J, BERTINETTO L, HENRIQUES J, et al. End-to-end representation learning for correlation filter based tracking [ C ]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017: 2805-2813.
- [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] WANG Q, TENG Z, XING J, et al. Learning attentions: residual attentional siamese network for high performance online visual tracking [ C ]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018: 4854-4863.
- [7] CHEN B, LI P, SUN C, et al. Multi attention module for visual tracking [ J ]. Pattern Recognition, 2019, 87: 80-93.
- [8] 王淑贤,葛海波,李文浩.多尺度通道注意与孪生网络的目标跟踪算法[J].计算机工程与应用,2023,59(14):142-150.
- [9] YANG K, HE Z, ZHOU Z, et al. Siamatt: Siamese attention network for visual tracking [ J ]. Knowledge-based systems, 2020, 203: 106079.
- [10] 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.
- [11] YAN B, PENG H, FU J, et al. Learning spatio-temporal transformer for visual tracking [ C ]. Proceedings of the IEEE/CVF International Conference on Computer Vision, 2021: 10448-10457.
- [12] 李雪,李晓艳,王鹏,等.结合双注意力与特征融合的孪生网络目标跟踪[J].北京邮电大学学报,2022,45(4):116-122.
- [13] WOO S, PARK J, LEE J Y, et al. Cbam: Convolutional block attention module [ C ]. Proceedings of the European Conference on Computer Vision (ECCV), 2018: 3-19.
- [14] ZHANG L, GONZALEZ-GARCIA A, WEIJER J, et al. Learning the model update for siamese trackers [ C ]. Proceedings of the IEEE/CVF International Conference on Computer Vision, 2019: 4010-4019.
- [15] HUANG L, ZHAO X, HUANG K. Got-10k: A large high-diversity benchmark for generic object tracking in the wild [ J ]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2019, 43(5): 1562-1577.
- [16] WU Y, LIM J, YANG M H. Object tracking benchmark [ J ]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2015, 37 ( 9 ): 1834-1848.
- [17] 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, 2014, 37(3): 583-596.
- [18] ZHU Z, WANG Q, LI B, et al. Distractor-aware siamese networks for visual object tracking [ C ]. Proceedings of the European conference on computer vision (ECCV), 2018: 101-117.

### 作者简介

**全卫国**,博士,副教授,主要研究方向为数字图像处理及模式识别、流量参数测量与检测新技术。

E-mail: tongwg@ncepu.edu.com

**寇德龙**(通信作者),硕士研究生,主要研究方向为数字图像处理及模式识别。

E-mail: 1440418429@qq.com

**李茂冉**,硕士研究生,主要研究方向为气液两相流参数测量。

E-mail: 1535848257@qq.com

**石宗锦**,硕士研究生,主要研究方向为现代测控技术与系统。

E-mail: 1535006344@qq.com