

文章编号 1004-924X(2021)02-0363-11

基于显著性区域加权的相关滤波目标跟踪

朱均安^{1,2}, 陈 涛^{1*}, 曹景太¹

(1. 中国科学院 长春光学精密机械与物理研究所, 吉林 长春 130033;

2. 中国科学院大学, 北京 100049)

摘要: 为了提高跟踪过程中目标位置的定位精度, 提出了基于显著性区域加权的相关滤波目标跟踪算法。本文在高效卷积算子跟踪算法(Efficient Convolution Operators for Tracking, ECO)的跟踪框架基础上, 首先采用预训练的改进残差网络 SE-ResNet 来提取不同层的多分辨率特征, 充分利用浅层和深层特征的不同特性来增强特征表达, 通过因式分解的卷积求出相关滤波的响应图; 然后采用背景对像模型来获取目标的显著性图, 并使用显著性图来对相关滤波的响应图进行加权, 提高定位精度; 最后, 在视觉目标跟踪(Visual Object Tracking, VOT)竞赛上与 8 种流行的跟踪算法进行对比, 在 VOT2016 和 VOT2017 竞赛上的平均重叠期望(Expected Average Overlap, EAO)得分分别达到了 0.415 7 和 0.341 2, 均优于其他算法。实验表明本算法可以有效提升目标跟踪中的定位精度, 改善跟踪性能。

关 键 词: 卷积神经网络; 相关滤波; 显著性区域; 目标跟踪

中图分类号: TP391 **文献标识码:** A **doi:** 10.37188/OPE.20212902.0363

Salient region weighted correlation filter for object tracking

ZHU Jun-an^{1,2}, CHEN Tao^{1*}, CAO Jing-tai¹

(1. Changchun Institute of Optics, Fine Mechanics and Physics Chinese Academy of Sciences,
Changchun 130033, China;

2. University of Chinese Academy of Sciences, Beijing 100049, China)

* Corresponding author, E-mail: chent@ciomp.ac.cn

Abstract: To improve the positioning accuracy of target positions in object tracking, an object tracking algorithm based on a salient region weighted correlation filter is proposed in this study. Using the tracking framework of efficient convolution operators (ECO) for tracking, we first apply SE-ResNet, which is a pre-trained improved residual network, to extract the multi-resolution features of different layers and fully utilize the different characteristics of the shallow and deep features to enhance feature expression. Next, a background object model is used to obtain a saliency map of the target. The saliency map is then applied to weight the response map of the correlation filter to improve positioning accuracy. Finally, compared with eight popular tracking algorithms employed at the Visual Object Tracking (VOT) challenge, the expected average overlap scores of VOT2016 and VOT2017 are determined to be 0.415 7 and 0.341 2, respectively, which are better than those of the other algorithms. Experimental results show that the proposed algorithm can effectively improve positioning accuracy and tracking performance.

收稿日期: 2020-07-13; 修订日期: 2020-09-08.

基金项目: 国家自然科学基金资助项目 (No. 61601195)

Key words: convolutional neural network; correlation filter; salient region; object tracking

1 引言

视觉目标跟踪是计算机视觉中的一个重要的研究问题,是高层语义分析的基础,例如场景理解、行为理解等。视觉目标跟踪要解决的问题可以表述为:在视频序列中,给出第一帧中目标的位置及大小(通常是一个矩形边界框),需要在后续帧中预测出目标的位置及大小。视觉目标跟踪中存在许多挑战^[1-2],包括光照变化、遮挡、旋转、形变等。研究人员提出了许多跟踪算法来解决这些问题,主要分为生成模型^[3]和判别模型^[4-8]两类。基于生成模型的目标跟踪算法,采用历史帧的结果来生成用于描述目标特征的统计模型,能够有效处理跟踪过程中目标丢失的情况,但是基于生成模型的方法忽略了目标周围的背景信息,在面对背景混乱时容易丢失目标。与生成模型不同,判别模型主要是学习出一个决策边界,利用这个决策边界来区分目标区域与背景区域。

近年来,基于判别模型的相关滤波跟踪算法取得了良好的跟踪性能。2010年,误差最小平方和(Minimum Output Sum of Squared Error, MOSSE)^[9]跟踪算法首次将相关滤波器引入目标跟踪领域。通过最小化均方误差,在后续帧中找到目标的最佳位置。2012年,循环结构的检测跟踪(Circulant Structure of Tracking-by-detection with Kernels, CSK)^[10]算法提出了一种使用循环移位对数据进行密集采样的方法,并通过快速傅立叶变换(Fast Fourier Transformation, FFT)快速训练出分类器。2014年,核相关滤波器(Kernelized Correlation Filter, KCF)^[11]跟踪算法对CSK进行了扩展,由单通道特征拓展到多通道特征,并引入核方法提高了基于相关滤波的目标跟踪算法的精度。为了更好地应对尺度变化,区分尺度空间跟踪算法(Discriminative Scale Space Tracking, DSST)^[12]和多特征的尺度自适应跟踪算法(Scale Adaptive with Multiple Features tracker, SAMF)^[13]分别提出了不同的解决方案。DSST将目标跟踪看成目标中心平移和目标尺度变化两个独立的问题,通过添加尺度滤波器来更好地应对尺度变化。SAMF则对候选

区域的目标采用固定的七个尺度,同时检测目标平移变化和尺度变化,快速确定目标的位置和尺度。2015年,为了抑制边界效应并扩大搜索范围,空间正则化的判别相关滤波器(Spatially Regularized Discriminative Correlation Filter, SRDCF)^[14]跟踪算法采用了更大的检测区域,同时加入空域正则化。2016年,连续卷积算子跟踪算法(Continuous Convolution Operator Tracker, C-COT)^[15]在连续域中应用多分辨率特征,并在VOT2016竞赛上实现了较好的跟踪性能。模板与像素互补学习(Sum of Template and Pixel-wise Learners, Staple)^[16]跟踪算法采用梯度方向直方图(Histogram of Oriented Gradients, HOG)特征^[17]和颜色直方图来建立目标的外观模型,这两种特征具有一定的互补性,通过分别求解他们的响应图,然后对响应图进行加权融合获得了较好的跟踪效果。2017年,在C-COT跟踪算法的基础上,高效卷积算子跟踪算法(Efficient Convolution Operators for Tracking, ECO)^[18]将手工设计的特征和卷积特征相结合,使用因式分解的卷积来减少特征的维数,并压缩学习模型中的训练样本以提高跟踪速度和鲁棒性。

基于判别模型的相关滤波跟踪算法虽然在精度和速度上取得了较好的效果,但是这些算法目标位置的定位精度依然有待提升。为了解决这个问题,本文在ECO的基础上提出了基于显著性区域加权的相关滤波目标跟踪算法,首先使用改进的残差网络(SE-ResNet)^[19]来提取多分辨率特征,可以充分利用浅层和深层特征的不同特性。然后,提出了一种显著性区域加权策略,通过对相关滤波器的响应进行重新加权来提高定位精度。最后,在视觉目标跟踪(Visual Object Tracking, VOT)^[1-2]竞赛上进行评估,实验结果表明本文算法取得了较好的效果,具有良好的跟踪性能。

2 基于相关滤波的目标跟踪

相关性是用来描述两个信号的联系,分为互相关和自相关,在基于相关滤波的目标跟踪中,

相关指的是互相关。

假设有两个信号 f 和 g ,当他们分别为连续信号和离散信号时的相关性计算方法如下:

$$(f \otimes g)(\tau) \stackrel{\text{def}}{=} \int_{-\infty}^{\infty} f^*(t)g(t+\tau)dt, \quad (1)$$

$$(f \otimes g)(n) \stackrel{\text{def}}{=} \sum_{m=-\infty}^{\infty} f^*[m]g(m+n), \quad (2)$$

其中: $\otimes$ 代表卷积运算, f^* 表示 f 的共轭。互相关就是衡量两个信号在某个时刻 τ 时的相似程度。因此,将相关滤波应用到目标跟踪领域最基本的思想就是:设计一个滤波模板,使得该模板与跟踪目标的感兴趣区域(Region of Interest, ROI)做相关运算,得到最大的输出响应。该思想用数学语言可以描述为:

$$g = f \otimes h, \quad (3)$$

其中: g 表示相关滤波的输出响应, f 表示输入原始图片的特征, h 表示滤波模板。在相关滤波的跟踪框架中,只需要不断地修正滤波模板,得到最大的输出响应即可。

由于卷积运算的计算量较大,采用快速傅里叶变换FFT的性质将空间域的卷积运算转化为频率域的乘法运算,可以极大地提高运算速度。公式(3)可以转化为:

$$F(g) = F(f \otimes h) = F(f) \odot F(h)^*, \quad (4)$$

其中: $\odot$ 表示逐元素相乘, F 表示FFT变换,采用对应字母的大写表示其FFT变换后的结果,公式(4)可以简化为:

$$G = F \odot H^*. \quad (5)$$

因此,相关滤波器可以采用如式(6)求出:

$$H^* = G/F. \quad (6)$$

3 本文算法

3.1 算法总体结构

本文提出的跟踪算法的整体结构如图1所示,该算法由两部分构成,即相关滤波响应的计算和显著性图的计算。对于相关滤波的响应图的计算,本文算法在ECO的跟踪框架基础上,采用SE-ResNet-50来提取多分辨率特征, R_c 是求出的相关滤波的响应;对于显著性图的计算,采用背景对象模型^[20]来获取目标的显著性图, R_s 是求出的显著性图。最后,将相关滤波的响应图与显著性图的响应相乘,即可得到最终的响应图,最终的响应图 R_{final} 可以通过如式(7)计算出:

$$R_{\text{final}} = R_c \odot R_s. \quad (7)$$

把响应图 R_{final} 取得最大值时的位置映射到原图中就可以求得在后续帧中目标的位置。

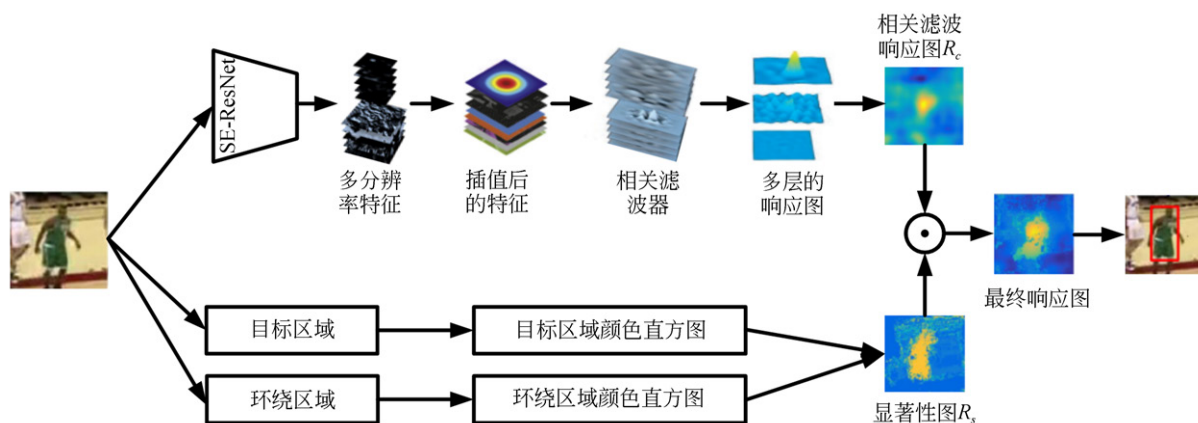


图1 本文算法的跟踪框架

Fig. 1 Framework of proposed algorithm

3.2 分层卷积特征

近年来,手工设计的特征,如HOG特征和颜色名(Color Name, CN)^[21]特征在目标跟踪领域取得了良好的效果,但手工设计的特征已经成为了跟踪精度提升的瓶颈。Yosinski等人^[22]研究发现,在

卷积神经网络中低层的部分学习的是一些颜色、纹理等简单信息,而高层部分学习的是综合的语义信息。卷积神经网络中的不同卷积层的特征图的可视化的结果如图2所示。卷积神经网络特征在图像分类和目标检测领域都取得了较高的精度。因

此,本文采用卷积神经网络来提取浅层和深层的特 征,充分利用卷积神经网络不同层的特性。

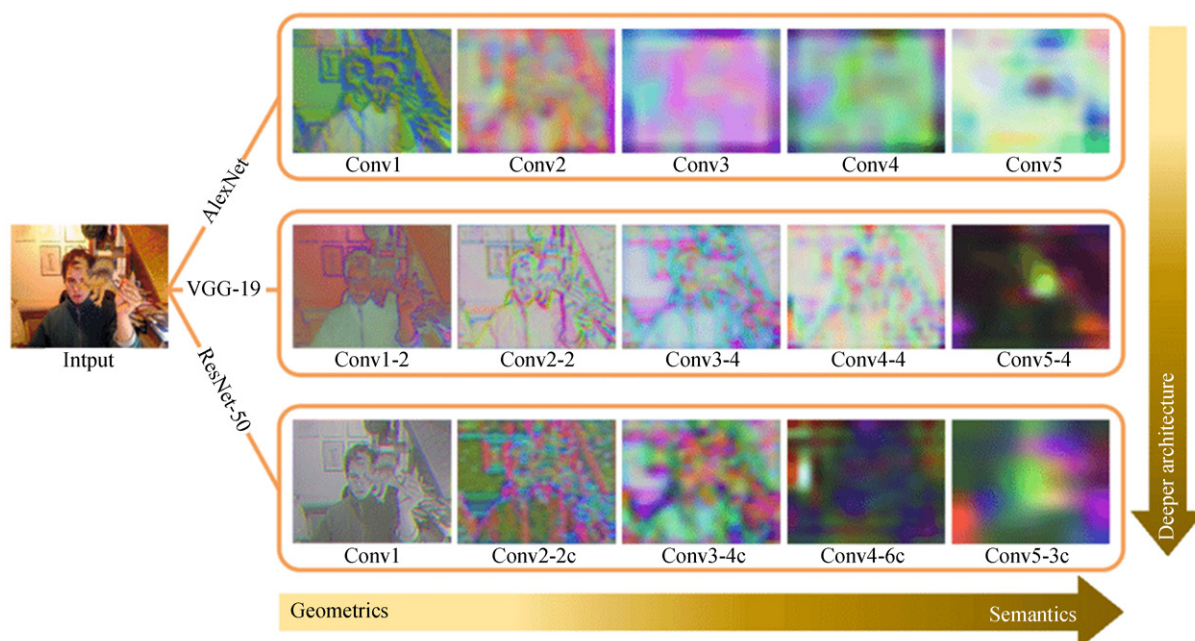


图2 卷积神经网络中的不同卷积层的特征图的可视化^[23]

Fig. 2 Visualization of deep feature maps from different convolutional layers of different CNN architectures

3.3 相关滤波的响应图的计算

为了进一步提升跟踪算法的性能,在计算相关滤波的响应图时,采用了ECO的跟踪框架,并使用了改进的残差网络SE-ResNet-50网络来提取多分辨率特征,充分利用不同层的特征信息,并结合高层抽象的特征和低层基础的图像特征的特点。本文算法选择 *Conv1x*, *Res3d* 和 *Res4f* 层作为多分辨率特征图,输出的特性在ReLU激活之前,提取的特征图的大小为 112×112 , 28×28 和 14×14 。由于提取的特征的分辨率不同,采用C-COT中的隐式插值模型。每个样本 x_j 都包含 D 维的特征通道,从同一图像块中提取的 D 维通道的特征为 $x_j^1, x_j^2, \dots, x_j^D$ 。 N_d 表示特征通道 x 中空间采样点的数目, $d \in \{0, 1, 2, \dots\}$ 。特征通道 $x \in R^{N_d}$ 可以看作一个关于离散空间变量 $n \in \{0, \dots, N_d-1\}$ 的函数 $x[n]$ 。为了学习连续的卷积算子,特征图被转换到连续的空间域 $[0, T) \in R$ 。其中常数 T 代表支持区域的大小。插值算子 J_d 被构造为:

$$J_d\{x^d\}(t) = \sum_{n=0}^{d-1} x^d[n] b_d\left(t - \frac{T}{N_d}n\right), \quad (8)$$

其中, $b_d \in L^2(T)$, 它表示插值函数。

$$b_d(t) = \begin{cases} (a+2)|t|^3 - (a+3)t^3 + 1, & |t| \leq 1 \\ a|t|^3 - 5at^3 + 8a|t| - 4a, & 1 < |t| \leq 2 \\ 0, & |t| > 2 \end{cases} \quad (9)$$

式中 a 是固定系数。通过插值函数将不同分辨率的特征插值到相同尺寸,然后采用ECO中的因式分解的卷积来减小特征的维度,使用一组较小的基滤波器 $f^1, f^2, \dots, f^C$, 在这里 $C < D$ 。然后使用一组学习的系数 $P_{d,c}$ 将特征层 d 的滤波器构造为一个线性组合 $\sum_{c=1}^C p_{d,c} f^c$ 的滤波器 f^c , 系数 $P_{d,c}$ 可以紧凑地表示为 $D \times C$ 的矩阵 $P = (P_{d,c})$ 。新的多通道滤波器可以写成矩阵向量乘积 Pf , 因式分解的卷积算子可以表示为:

$$R_c = S_{pf}\{x\} = Pf \otimes J\{x\} = \sum_{c,d} p_{d,c} f^c J_d\{x^d\} = f \otimes P^T J\{x\} \quad (10)$$

公式(10)的最后一步是卷积的线性组合,可以将因式分解卷积过程分为两步操作,其中每个位置 t 的特征向量 $J\{x\}(t)$ 首先与矩阵 P^T 相乘,然后将生成的 C 维的特征图与滤波器 f 进行卷积。

因此,矩阵 P^T 的作用类似于线性的降维算子。 R_c 即为相关滤波的响应图。

考虑从单个训练样本 x 中学习因式分解的卷积算子公式(10),为了简化表示,使用 $\hat{z}^d[k] = X^d[k] \hat{b}_d[k]$ 表示插值特征图 $z = J\{x\}$ 的傅立叶系数。傅立叶域中的相应的损失函数可推导为:

$$E(f, P) = \|\hat{z}^T P \hat{f} - \hat{y}\|_{l^2}^2 + \sum_{c=1}^C \|\hat{w} \otimes \hat{f}^c\|_{l^2}^2 + \lambda \|P\|_F^2. \quad (11)$$

公式(11)相比于C-COT中的损失函数,添加了 P 的Frobenius范数作为正则化项,其约束效果由参数 λ 控制。损失函数公式(10)是一个非线性最小二乘问题,为了解决这个非线性最小二乘问题,首先采用了高斯-牛顿方法将上述非线性最小二乘问题转化为线性最小二乘问题,然后再通过共轭梯度的方法来求解。

相关滤波的响应图的求解过程总结如下:

Step1:采用公式(8)对不同分辨率的特征图进行双三次插值操作,将不同分辨率的特征图转换到连续空间域;

Step2:通过采用公式(11)最小化损失函数,求出相关滤波器;

Step3:采用公式(10)进行因式分解的卷积求出相关滤波的响应图 R_c 。

3.4 显著性区域提取

视觉显著性(Visual Saliency)^[24]是指人眼可以自动地从真实世界的场景中识别出感兴趣区域,并对感兴趣的区域进行处理而忽略掉不感兴趣的区域。作为对人类视觉注意机制的模拟,显著性检测算法的目标是将输入图像转化为一幅显著图,显著图表现为将图像中可能的感兴趣区域进行高亮显示,并抑制背景区域的显示。通过在计算机视觉任务中引入视觉显著性,可以为视觉信息处理任务带来帮助和改善。通常,跟踪任务中的目标是运动中的前景物体,因此,在目标跟踪中采用视觉显著性可以帮助快速定位目标,提高定位的准确率,显著性检测算法的计算结果对于目标跟踪任务具有重要的指导作用。

本文采用背景对象模型来获取目标的显著

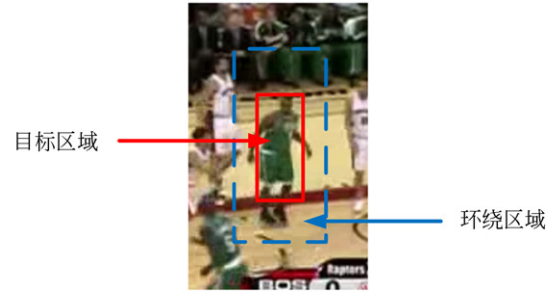


图3 目标区域和环绕区域

Fig. 3 Object region and surrounding region

性图。假设输入图像为 I ,为了从背景中分离出目标像素 $x \in \theta$,采用基于颜色直方图的贝叶斯分类。如图3所示,给出一个目标的矩形框区域 O 和它的环绕区域 S 。在 x 处的像素属于目标像素的概率为:

$$P(x \in \theta | O, S, b_x) \approx \frac{P(b_x | x \in O) P(x \in O)}{\sum_{\Omega \in \{O, S\}} P(b_x | x \in \Omega) P(x \in \Omega)}, \quad (12)$$

其中 b_x 表示分配给输入图像 $I(x)$ 的颜色分量,由于是从颜色直方图直接估算,因此颜色分量属于目标区域和环绕区域的概率可以分别表示为:

$$P(b_x | x \in O) \approx H_O^I(b_x) / |O|, \quad (13)$$

$$P(b_x | x \in S) \approx H_S^I(b_x) / |S|. \quad (14)$$

$H_O^I(b)$ 表示在区域 $\Omega \in I$ 上计算的非标准化直方图 H 的第 b 个计算区间,先验概率可以近似为:

$$P(x \in O) \approx |O| / (|O| + |S|). \quad (15)$$

根据公式(15),公式(12)可以被简化为:

$$P(x \in \theta | b_x) = P(x \in \theta | O, S, b_x) \approx \begin{cases} \frac{H_O^I(b_x)}{H_O^I(b_x) + H_S^I(b_x)}, & I(x) \in I(O \cup S). \\ 0.5, & \text{otherwise} \end{cases} \quad (16)$$

分配给背景的像素值的最大熵为0.5,采用背景对象模型可以从背景像素中区分出目标像素。通过搜索前一帧目标位置的一个矩形区域 O_{t-1} ,当前帧的显著性图 R_t 的计算公式如式(17)所示:

$$R_s = s_v(O_{t,i}) s_d(O_{t,i}), \quad (17)$$

其中: $s_v(O_{t,i})$ 表示基于目标模型的概率分数, $s_d(O_{t,i})$ 是基于到前一个目标中心的欧式距离中心 c_{t-1} 的距离分数, 它们的计算公式如下:

$$s_v(O_{t,i}) = \sum_{x \in O_{t,i}} P_{1:t-1}(x \in \mathcal{D}|b_x), \quad (18)$$

$$s_d(O_{t,i}) = \sum_{x \in O_{t,i}} \exp\left(-\frac{\|x - c_{t-1}\|^2}{2\sigma^2}\right). \quad (19)$$

在跟踪阶段, 由于目标的外观是不断变化的, 所以需要不断地更新目标外观模型, 采用线性插值的方式来更新目标外观模型, 公式如下:

$$P_{1:t}(x \in \mathcal{D}|b_x) = \eta P(x \in \mathcal{D}|b_x) + (1 - \eta) P_{1:t-1}(x \in \mathcal{D}|b_x), \quad (20)$$

其中 η 是学习率。

3.5 模型更新策略

在相关滤波响应图和显著性图的计算中都需要更新模型, 并且采用了不同的更新策略。

在相关滤波响应的计算过程中, 本文采用了相关滤波的跟踪框架, 如果对每一帧都更新模型, 由于图像帧间样本特征的差异较小, 所以基于相关滤波方法的图像帧间的损失函数变化也很小, 每一帧都更新使得整个样本集内存在大量的冗余样本信息, 给目标跟踪算法带来很大的计算负担。为了解决上述逐帧更新模型的策略带来的内存与算力问题, 采用与 ECO 相同的更新策略, 使用稀疏的模型更新方法, 直接设置每 N_s 帧图像进行一次模型更新。

在显著性图的计算中, 本文采用背景对象模型来获取目标的显著性图, 由于目标的外观是不断变化的, 需要不断的更新目标外观模型。因此在显著性图的计算过程中, 采用公式(20)在获取每一帧的目标后都更新目标外观模型。

3.6 目标跟踪过程

在目标跟踪过程中, 给定第一帧的目标位置, 跟踪算法在后续帧中根据前一帧中目标的位置, 在其周围进行搜索, 从而预测目标在后续帧中最佳的位置和尺度。为了更好的应对尺度变化, 本文算法采用了与 SAMF 中相同的尺度策略, 采用了七个尺度。本文算法的跟踪过程可以

作如下描述:

算法: 基于显著性区域加权的相关滤波目标跟踪算法

输入: 图像序列和第一帧图像的目标位置和尺寸大小

输出: 图像序列后续帧中的目标位置和尺寸大小

Begin

if 第一帧

手动选定需跟踪的目标, 提取目标区域的多层卷积特征, 最小化式损失函数得到初始滤波器 f ;

else

Step1: 根据上一帧目标的位置和尺寸大小裁剪出七个不同尺度的候选区域, 并提取预测目标区域的多层卷积特征;

Step2: 利用公式(10)计算相关滤波的响应图 R_t ;

Step3: 利用公式(17)计算显著性图 R_s ;

Step3: 利用公式(7)计算得到最终的响应图, 得到的目标位置和尺度;

Step4: 通过公式(11)更新滤波器 f 和公式(20)更新外观模型;

Step5: 如果不是最后一帧, 返回 Step1。

End

4 实验结果与分析

4.1 实验设置

始于 2013 年的 VOT 是一项目标跟踪算法的竞赛, 本文算法在 VOT2016^[1] 和 VOT2017^[2] 上进行评估, 它们都包含 60 个手动标记的图像序列, 但 VOT2017 用 10 个困难的视频序列代替了 VOT2016 中 10 个简单的视频序列, 并且重新标记了一些视频序列, 使 VOT2017 比 VOT2016 更具挑战性。VOT 竞赛中, 跟踪性能按三个主要指标进行排序, 即精度 (Accuracy)、鲁棒性 (Robustness) 和平均重叠期望 (Expected Average Overlap, EAO), 使用 VOT 竞赛的评估工具包 (VOT toolkit) 来评估本文算法。

本文使用一台搭载 Intel i7-8700K CPU 和 GTX 1080 Ti 显卡的电脑对提出的跟踪算法进行了评估, 使用的软件是 MATLAB 2016a 和它的工具包 MatConvNet^[25] (版本为 MatConvNet-1.0-beta25), 改进残差网络 SE-ResNet-50 模型可以公开下载 (<http://www.robots.ox.ac.uk/~albanie/models/se-nets/SE-ResNet-50-mcn.mat>)。实验使用的参数如下: 在相关滤波部分搜索区域范围

设置在 200×200 和 250×250 之间,尺度数量为7,尺度因子为1.03,固定更新频率 N_s 为5;在显著性图的计算中,学习率 η 设置为0.05。根据 MatConvNet 的官方说明,不同版本的 MatConvNet 会影响运行结果,为了避免采用不同版本导致的结果差异,其他跟踪算法的测试结果来自作者个人主页或者 VOT 竞赛的官方网站 (<https://www.votchallenge.net/vot2016/results.html>; <https://www.votchallenge.net/vot2017/results.html>)。

4.2 消融实验

为了验证本文提出的显著区域加权策略的有效性,在 VOT2016 和 VOT2017 上进行了消融实验,将本文算法与没有显著区域加权策略的算法进行了比较。Ours_N 代表没有显著区域加权策略,其他与本文算法相同。实验结果如表1和表2所示,从这两个表中可以看出,本文算法的 EAO、准确性和鲁棒性方面都优于没有显著区域加权策略的算法。实验结果表明,显著区域加权策略能够有效地改善跟踪性能。此外,与本文的基准算法 ECO 相比,本文算法即使没有显著区域加权策略,在 EAO 的准确性和鲁棒性方面仍然比 ECO 好。实验结果表明,利用改进的残差网络 SE-ResNet-50 网络在相关滤波框架中提取多分辨率特征的策略是有效的,有助于提升跟踪性能。

表1 VOT2016上的消融实验结果对比

Tab. 1 Ablation study results on VOT2016

	EAO	Accuracy	Robustness
Ours	0.413 3	0.556 9	0.154
Ours_N	0.383 9	0.554 3	0.168
ECO	0.373 8	0.552 6	0.200

表2 VOT2017上的消融实验结果对比

Tab. 2 Ablation study results on VOT2017

	EAO	Accuracy	Robustness
Ours	0.340 4	0.504 9	0.192
Ours_N	0.304 2	0.498 0	0.234
ECO	0.280 9	0.484 8	0.276

4.3 VOT2016 实验结果

本文算法与8种先进的跟踪算法(包括 ECO^[17], Staple^[15], DeepSRDCF^[26], SRDCF^[13],

SiamFC^[27], KCF^[10], DSST^[11] 和 SAMF^[12]) 在 VOT2016 上进行比较。实验结果如表3所示,表中最好的三个结果分别用红色、绿色和蓝色表示(彩图见期刊电子版)。从表中可以看出本文提出的算法的平均重叠期望、准确性和鲁棒性均优于其他跟踪算法,获得了比其他算法更好的精度和鲁棒性,具有良好的跟踪性能。

表3 多种算法在 VOT2016 上的对比

Tab. 3 Comparison of trackers on VOT2016

	EAO	Accuracy	Robustness
Ours	0.415 7	0.559 9	0.154
ECO	0.373 8	0.552 6	0.200
Staple	0.294 2	0.544 2	0.378
DeepSRDCF	0.275 7	0.528 3	0.326
SRDCF	0.245 9	0.533 6	0.419
SiamFC	0.235 8	0.534 1	0.461
KCF	0.193 5	0.493 5	0.567
SAMF	0.185 1	0.506 3	0.587
DSST	0.180 6	0.532 0	0.704

VOT2016 测试数据集上的精度-鲁棒性图和平均期望重叠率图如图4和图5所示,精度-鲁棒性图中横坐标为指数化处理后的鲁棒性值,纵坐标为精度,算法在图中的位置越靠近右上总体性能越好;平均期望重叠率图中横坐标为算法排名,纵坐标为平均期望重叠率,算法在图中的位置越靠近右上总体性能越好。从图4和图5可以看出本文算法在图中都处于右上的位置,这表示本文算法在9个跟踪算法中总体跟踪性能最好。

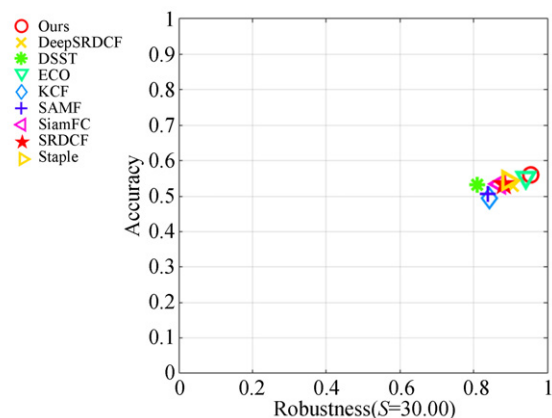


图4 VOT2016测试数据集上的精度-鲁棒性排名

Fig. 4 AR plot for experiment baseline on VOT2016

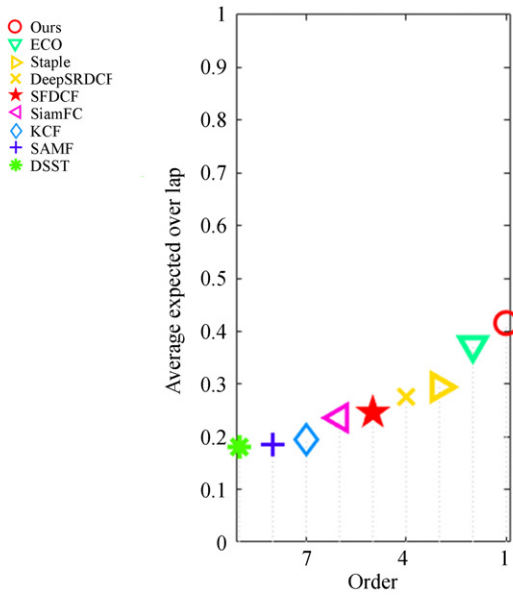


图5 VOT2016测试数据集上的平均期望覆盖率排名
Fig. 5 Expected overlap scores for baseline on VOT2016

4.4 VOT2017 实验结果

本文算法与 7 个先进的跟踪算法 (包括 ECO^[17], CFNet^[28], DCFNet^[29], Staple^[15], SRDCF^[13], KCF^[10] 和 DSST^[11]) 在 VOT2017 上进行了比较, 实验结果如表 4 所示, 最好的三个结果分别用红色、蓝色和绿色标注 (彩图见期刊电子版)。从表中可以看出, 在 8 个跟踪算法中, 本文提出的跟踪算法在 EAO 和鲁棒性上都排名第一, 精度略低于 Staple 排名第二。本文算法在 VOT2017 上也取得了较好的结果, 具有良好的跟踪性能。本文算法在 VOT2017 数据集上的精度-鲁棒性图和平均期望重叠率图如图 6 和图 7 所示, 本文算法在图中都处于右上的位置, 这代表本文算法的总体跟踪性能较好。

表 4 多种算法在 VOT2017 上的对比
Tab. 4 Comparison of trackers on VOT2017

	EAO	Accuracy	Robustness
Ours	0.341 2	0.505 9	0.197
ECO	0.280 9	0.484 8	0.276
CFNet	0.187 5	0.501 0	0.585
DCFNet	0.182 7	0.470 0	0.543
Staple	0.168 8	0.527 8	0.688
KCF	0.135 1	0.448 0	0.773
SRDCF	0.117 9	0.486 7	0.974
DSST	0.079 3	0.396 9	1.452

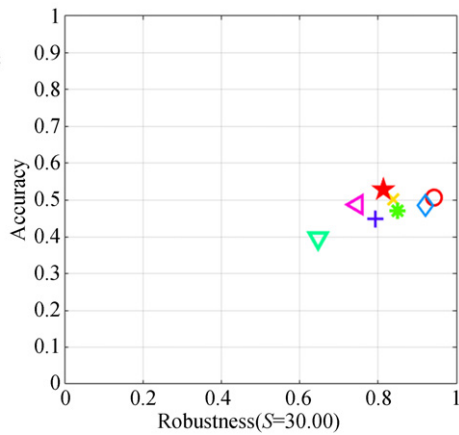


图6 VOT2017测试数据集上的精度-鲁棒性排名
Fig. 6 AR plot for experiment baseline on VOT2017

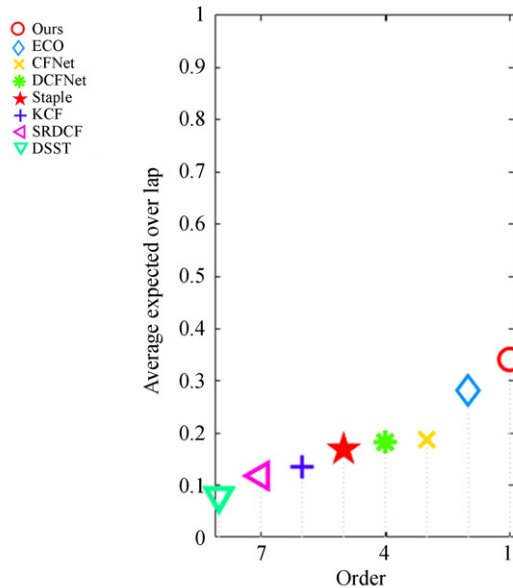


图7 VOT2017测试数据集上的平均期望覆盖率排名
Fig. 7 Expected overlap scores for baseline on VOT2017

4.5 定性分析实验

为了直观地对比本文跟踪算法的效果, 从 VOT 数据集中选择了 5 个有代表性的图像序列 (Bag, Bmx, Butterfly, Fish1 和 Matrix) 进行定性分析实验。这些图像序列几乎包含了跟踪任务中所有的挑战, 不同跟踪算法预测的目标边界框如图 8 所示。为了更好地展示不同算法的结果, 图 8 只展示了本文提出的跟踪算法和本文算法的基准算法 ECO 的对比。如图 8 所示, 本文算法的跟踪框与基准算法 ECO 的跟踪框相比位置更准确, 具有更好的跟踪性能。甚至, 在某些 ECO 丢失目标的情况下, 本文提出的算法依然可以很好的跟踪目标。

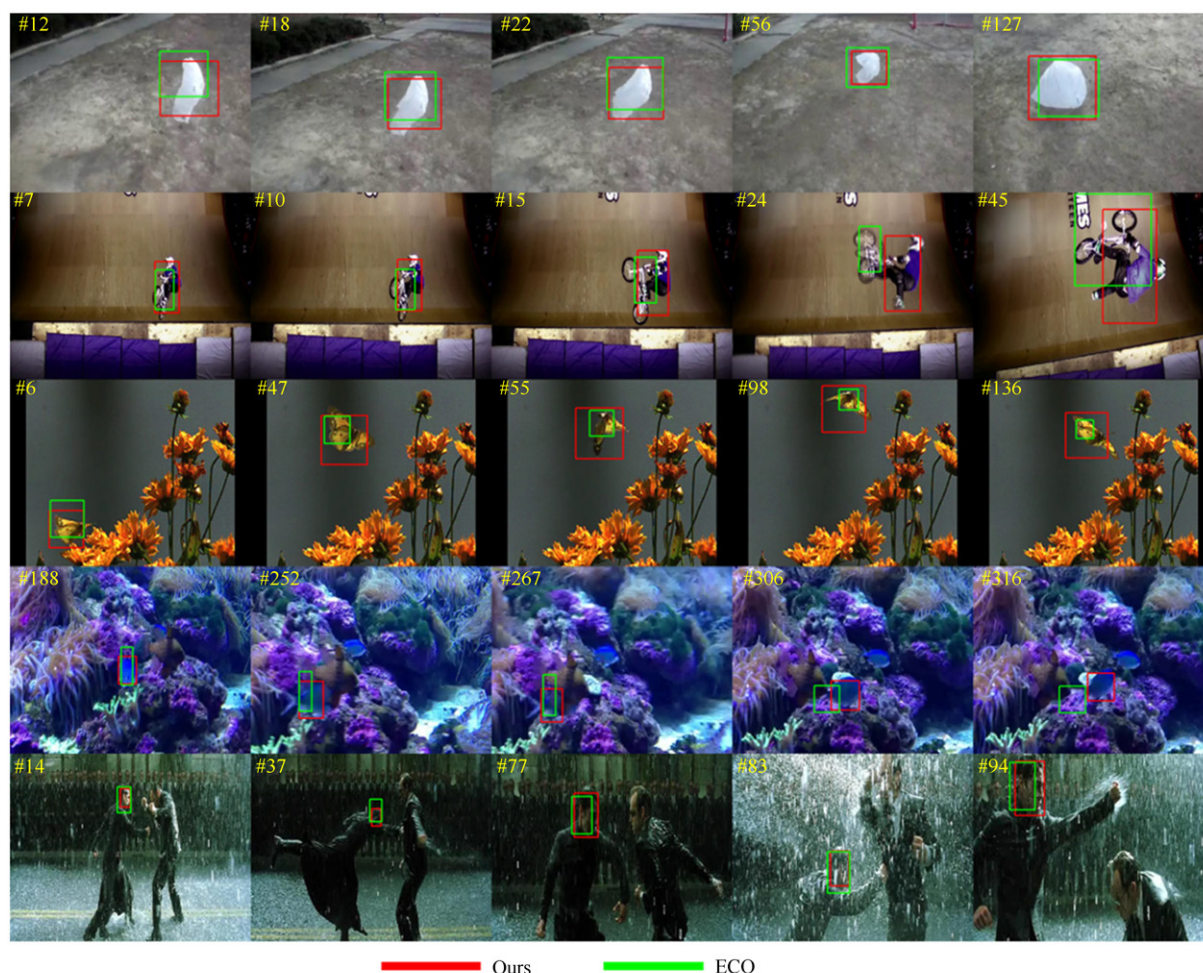


图8 与基准算法ECO相比,在五个具有挑战性的序列(从上到下:Bag,Bmx,Butterfly,Fish1和Matrix)上对所提出算法的定性评估

Fig. 8 Qualitative evaluation of the proposed algorithm compared with ECO on five challenging sequences (from top to bottom: Bag, Bmx, Butterfly, Fish1, and Matrix)

4.6 运行速度对比实验

运行速度是目标跟踪算法评价的重要指标,本文算法在Intel i7-8700K CPU和GTX 1080 Ti显卡的电脑上的运行速度为平均8帧速率(Frames Per Second, FPS),与另外三种跟踪算法的运行速度对比如表5所示。从表中可以看出本文算法的运行速度与采用卷积特征的跟踪算法ECO和DeepSRDCF相比相差不大,但是与采用手工设计特征的Staple跟踪算法相比运行速度有一定差距,这是由于本文算法采用改进的残差网络SE-ResNet-50来提取多分辨率特征,在这个特征提取的过程中计算量较大,导致在跟踪过程中即使采用GPU进行加速仍然不能实时运行。

表5 与3个跟踪算法的速度对比

Tab. 5 Speed comparisons with three trackers

算法	Ours	ECO	DeepSRDCF	Staple (CPU)
平均速度/(FPS)	8	7	5	80

5 结论

本文在ECO的跟踪框架的基础上,提出了基于显著区域加权的相关滤波目标跟踪算法。首先采用SE-ResNet-50来提取多分辨率特征,增强特征表达;然后采用显著性图来对相关滤波的响应图进行加权,提高定位精度。最后,在VOT

数据集上进行了实验,VOT2016和VOT2017的EAO得分达到了0.4157和0.3412,实验表明采用特征表达更强的卷积神经网络来提取特征对

跟踪算法的精度有较大提升。而且,采用视觉显著性来加权相关滤波的响应图也可以有效地改善跟踪精度。

参考文献:

- [1] KRISTAN M, LEONARDIS A, MATAS J, *et al.*. The visual object tracking VOT2016 challenge results[C]. *European Conference on Computer Vision*, 2016:777-823.
- [2] KRISTAN M, LEONARDIS A, MATAS J, *et al.*. The visual object tracking VOT2017 challenge results[J]. *IEEE International Conference on Computer Vision*, 2017:1949-1972.
- [3] COLLINS R T. Mean-shift blob tracking through scale space[C]. *International Conference on Computer Vision and Pattern Recognition*, 2003: 234-240.
- [4] 董超, 郑兵, 李彬, 等. 改进核相关滤波器的海上船只目标跟踪[J]. *光学精密工程*, 2019, 27(4): 911-921.
DONG CH, ZHENG B, LI B, *et al.*. Ship target tracking with improved kernelized correlation filters [J]. *Opt. precision Eng.*, 2019, 27(4): 911-921. (in Chinese)
- [5] 刘教民, 郭剑威, 师硕. 自适应模板更新和目标重定位的相关滤波器跟踪[J]. *光学精密工程*, 2018, 26(8):270-281.
LIU J M, GUO J W, SHI SH. Correlation filter tracking based on adaptive learning rate and location refiner [J]. *Opt. precision Eng.*, 2018, 26(8): 2100-2111. (in Chinese)
- [6] 张博, 江沸菠, 刘刚. 利用视觉显著性和扰动模型的上下文感知跟踪[J]. *光学精密工程*, 2018, 26(8):282-291.
ZHANG B, JIANG F B, LIU G. Context-aware tracking based on a visual saliency and perturbation model [J]. *Opt. precision Eng.*, 2018, 26(8): 2112-2121. (in Chinese)
- [7] 梁慧慧, 何秋生, 贾伟振, 等. 一种多特征融合的目标跟踪算法[J]. *液晶与显示*, 2020, 35(6): 583-594.
LIANG H H, HE Q SH, JIA W ZH, *et al.*. Multi-feature fusion target tracking algorithm[J]. *Chinese Journal of Liquid Crystals and Displays*, 2020, 35(6):583-594. (in Chinese)
- [8] 韩亚君, 杨德东, 李勇, 等. 基于多特征融合相关滤波的红外目标跟踪[J]. *液晶与显示*, 2019, 34(2):177-187.
HAN Y J, YANG D D, LI Y, *et al.*. Infrared target tracking based on correlation filter with multi-features fusion[J]. *Chinese Journal of Liquid Crystals and Displays*, 2019, 34(2):177-187. (in Chinese)
- [9] BOLME D, BEVERIDGE J R, DRAPER B A, *et al.*. Visual object tracking using adaptive correlation filters [C]. *International Conference on Computer Vision and Pattern Recognition. IEEE*, 2010: 2544-2550.
- [10] HENRIQUES J F, CASEIRO R, MARTINS P, *et al.*. Exploiting the circulant structure of tracking-by-detection with kernels [C]. *European Conference on Computer Vision*, 2012:702-715.
- [11] 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: 583-596.
- [12] DANELLJAN M, HAGER G, KHAN F S, *et al.*. Discriminative scale space tracking[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2017, 39:1561-1575.
- [13] LI Y, ZHU J. A scale adaptive kernel correlation filter tracker with feature integration[C]. *European Conference on Computer Vision*, 2014: 254-265.
- [14] DANELLJAN M, HAGER G, KHAN F S, *et al.*. Learning spatially regularized correlation filters for visual tracking [J]. *IEEE International Conference on Computer Vision*, 2015:4310-4318.
- [15] DANELLJAN M, ROBINSON A, KHAN F S, *et al.*. Beyond correlation filters: Learning continuous convolution operators for visual tracking [C]. *European Conference on Computer Vision*, 2016: 472-488.
- [16] BERTINETTO L, VALMADRE J, GOLODETZ S, *et al.*. Staple: Complementary learners for real-time tracking [J]. *International Conference on Computer Vision and Pattern Recognition*,

- 2016, (12):1401-1409.
- [17] DALAL N, TRIGGS B. Histograms of oriented gradients for human detection [C]. *International Conference on Computer Vision and Pattern Recognition*, 2005:886-893.
- [18] DANELLJAN M, BHAT G, SHAHBAZ KHAN F, *et al.*. ECO: Efficient convolution operators for tracking[J]. *IEEE Conference on Computer Vision and Pattern Recognition*, 2017:6931-6939.
- [19] HU J, SHEN L, SUN G. Squeeze and excitation networks[J]. *IEEE Conference on Computer Vision and Pattern Recognition*, 2018:7132-7141.
- [20] POSSEGER H, MAUTHNER T, BISCHOF H. In defense of color-based model-free tracking [C]. *International Conference on Computer Vision and Pattern Recognition*, 2015: 2113-2120.
- [21] DANELLJAN M, KHAN F S, FELSBERG M, *et al.*. Adaptive color attributes for real-time visual tracking[J]. *International Conference on Computer Vision and Pattern Recognition*, 2014: 1090-1097.
- [22] YOSINSKI J, CLUNE J, NGUYEN A, *et al.*. Understanding neural networks through deep visualization [J]. *arXiv preprint arXiv: 1506. 06579*, 2015.
- [23] GAO P, ZHANG Q, WANG F, *et al.*. Learning reinforced attentional representation for end to end visual tracking [J]. *Information Sciences*, 2020, 517:52-67.
- [24] ITTI L, KOCH C, NIEBUR E. A model of saliency-based visual attention for rapid scene analysis [J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 1998, 20(11):1254-1259.
- [25] VEDALDI A, LENC K. MatConvNet: Convolutional neural networks for MATLAB [C]. *ACM International Conference on Multimedia*, 2015: 689-692.
- [26] DANELLJAN M, HAGER G, KHAN F S, *et al.*. Convolutional features for correlation filter based visual tracking[C]. *IEEE International Conference on Computer Vision*, 2015:621-629.
- [27] BERTINETTO L, VALMADRE J, HENRIQUES J F, *et al.*. Fully-convolutional siamese networks for object tracking[C]. *European Conference on Computer Vision*, 2016:850-865.
- [28] VALMADRE J, BERTINETTO L, HENRIQUES J, *et al.*. End to end representation learning for correlation filter based tracking[J]. *International Conference on Computer Vision and Pattern Recognition*, 2017:5000-5008.
- [29] WANG Q, GAO J, XING J, *et al.*. DCFNet: Discriminant correlation filters network for visual tracking [J]. *arXiv preprint arXiv: 1704.04057*, 2017.

作者简介:



朱均安(1992—),男,湖北黄冈人,博士研究生,2015年于东北大学获得学士学位,主要从事目标检测、识别与跟踪方面的研究。E-mail: zhujunan15@mails.ucas.edu.cn

导师简介:



陈涛(1965—),男,内蒙古赤峰人,博士,研究员,博士生导师,2007年于中科院长春光学精密机械与物理研究所获得博士学位,主要从事光电精密跟踪测量技术的研究。E-mail: chent@ciomp.ac.cn