

### 基于自适应搜索范围调整的视觉目标跟踪

王超, 王凯

引用本文

王超, 王凯. 基于自适应搜索范围调整的视觉目标跟踪[J]. 计算机科学, 2023, 50(11A): 221000172-6.

WANG Chao, WANG Kai. Visual Object Tracking Based on Adaptive Search Range Adjustmen[J]. Computer Science, 2023, 50(11A): 221000172-6.

---

### 相似文章推荐 (请使用火狐或 IE 浏览器查看文章)

Similar articles recommended (Please use Firefox or IE to view the article)

#### [多特征感知的时空自适应相关滤波目标跟踪](#)

Multi-feature-aware Spatiotemporal Adaptive Correlation Filtering Target Tracking  
计算机科学, 2023, 50(11A): 230200096-9. <https://doi.org/10.11896/jsjcx.230200096>

#### [注意力特征融合的孪生网络目标跟踪方法](#)

Attentional Feature Fusion Approach for Siamese Network Based Object Tracking  
计算机科学, 2023, 50(6A): 220300237-9. <https://doi.org/10.11896/jsjcx.220300237>

#### [基于孪生注意力网络的建设用地遥感影像变化检测](#)

Remote Sensing Image Change Detection of Construction Land Based on Siamese AttentionNetwork  
计算机科学, 2023, 50(6A): 220500040-5. <https://doi.org/10.11896/jsjcx.220500040>

#### [融合注意力特征的无锚框视觉目标跟踪方法](#)

AFTM:Anchor-free Object Tracking Method with Attention Features  
计算机科学, 2023, 50(1): 138-146. <https://doi.org/10.11896/jsjcx.211000083>

#### [多轮对话技术及其在电网数据查询中的应用](#)

Multi-turn Dialogue Technology and Its Application in Power Grid Data Query  
计算机科学, 2022, 49(10): 265-271. <https://doi.org/10.11896/jsjcx.200600078>

# 基于自适应搜索范围调整的视觉目标跟踪

王超<sup>1,2</sup> 王凯<sup>1</sup>

1 河北工程大学信息与电气工程学院 河北 邯郸 056038

2 河北省安防信息感知与处理重点实验室 河北 邯郸 056000

**摘要** 当前主流的视觉目标跟踪算法检测目标时,其搜索范围是以前一帧目标位置为中心设定的。然而目标可能由于运动而偏离设定的搜索中心,其在当前帧的检测响应易受到余弦窗惩罚机制的抑制,导致跟踪失败。为解决上述问题,提出了自适应搜索范围调整(Adaptive Search Range Adjustment, ASRA)方法。该方法采用了基于循环神经网络的运动预测模型来预测当前帧目标位置,并与相关滤波响应相结合来对搜索中心进行调整,进一步根据目标的运动矢量对搜索范围尺寸进行调整。将ASRA方法应用于当前先进的基于孪生网络的目标跟踪算法,在OTB2015和VOT2018数据集上进行的实验结果表明ASRA方法可以改善跟踪算法的准确率和鲁棒性。

**关键词** 视觉目标跟踪;搜索范围;运动预测;相关滤波;孪生网络

**中图分类号** TP391

## Visual Object Tracking Based on Adaptive Search Range Adjustment

WANG Chao<sup>1,2</sup> and WANG Kai<sup>1</sup>

1 School of Information and Electrical Engineering, Hebei University of Engineering, HeBei, Handan 056000, China

2 Hebei Key Laboratory of Security & Protection Information Sensing and Processing, HeBei, Handan 056000, China

**Abstract** The mainstream visual object tracking algorithms generally set the position of object that tracked in the last frame as the center of a search range, which is used to detect the object in current frame. However, the tracking object may deviate from the center of search range due to its motion, thus its detection response in current frame can be easily inhibited by the cosine window penalty mechanism, which leads to tracking failure. To solve this problem, an adaptive search range adjustment (ASRA) method is proposed. In this method, a motion prediction model based on recurrent neural network (RNN) is used to predict the object position in current frame, and it is combined with the correlation filtering response to adjust the center of search range. The size of search range is further adjusted according to the motion vector of the tracking object. The proposed ASRA method is applied to current state-of-the-art object tracking algorithms based on Siamese networks. Experiments on OTB2015 and VOT2018 datasets show that ASRA can improve the accuracy and robustness of these algorithms.

**Keywords** Visual object tracking, Search range, Motion prediction, Correlation filtering, Siamese networks

## 1 引言

视觉目标跟踪是计算机视觉领域的一个基础研究方向,在安全监控、智慧交通和医疗诊断等方面有着重要的应用前景<sup>[1]</sup>。但是在实际应用场景中经常存在光照变化、复杂背景干扰、目标的快速运动和形变,以及遮挡等复杂情况,现有的目标跟踪算法研究仍面临着巨大的挑战。

视觉目标跟踪根据任务不同可以分为多目标跟踪和单目标跟踪两种,其中单目标跟踪是当前的研究热点。主流的单目标跟踪算法一般采用“检测-跟踪”框架,通常需要在图像序列的第一帧指定待跟踪目标区域(参考目标),然后在后续帧设定的搜索范围内检测该目标,并估计其位置和边界<sup>[2-3]</sup>。由于目标检测局限在给定的搜索范围内,因此搜索范围的设定在一定程度上决定了算法的准确性。然而,主流目标跟踪算法的搜索范围通常直接以前一帧的目标位置为中心来设定,

还存在较大的改进空间。

本文针对主流目标跟踪算法提出自适应搜索范围调整(ASRA)方法,并将其应用于几种基于孪生网络的目标跟踪算法,结果表明该方法能够改善跟踪器的准确率和鲁棒性。

## 2 视觉目标跟踪的搜索范围

出于计算效率的考虑,视觉目标跟踪算法通常不在整个图像上检测目标,而是局限在一个有限的搜索范围内进行。通常,搜索范围对应图像中的一个矩形区域,可以由式(1)所示的像素坐标集合来定义。

$$S(c_x, c_y, s_w, s_h) = \{(x, y)\} \quad (1)$$

其中,

$$c_x - \left\lfloor \frac{s_w}{2} \right\rfloor \leq x \leq c_x + \left\lfloor \frac{s_w}{2} \right\rfloor$$

基金项目:国家自然科学基金(62071071);邯郸市科技计划项目(21422031251)

This work was supported by the National Natural Science Foundation of China(62071071) and Science and Technology Plan Project of Handan(21422031251).

通信作者:王超(wangchao@hebeu.edu.cn)

$$c_y - \left\lfloor \frac{s_h}{2} \right\rfloor \leq y \leq c_y + \left\lfloor \frac{s_h}{2} \right\rfloor$$

其中,  $(x, y)$  是搜索范围中像素的坐标,  $(c_x, c_y)$  是搜索范围的中心坐标,  $s_w$  和  $s_h$  分别是其宽和高,  $\lfloor \cdot \rfloor$  表示向下取整运算。

### 2.1 搜索范围的设定

各类主流的跟踪算法在图像中检测目标时,其搜索范围均以前一帧跟踪得到的目标位置为中心来划定。设前一帧跟踪得到的目标位置为  $P = (x, y)$ , 则当前帧搜索范围中心由式(2)决定。

$$(c_x, c_y) = (x, y) \tag{2}$$

搜索范围的尺寸由前一帧目标的大小决定,但各类算法的设定方法各有不同。假定前一帧目标的宽和高分别为  $w, h$ 。在基于相关滤波的跟踪算法中,搜索范围的尺寸与目标大小成正比,如式(3)所示:

$$s_w = \rho \cdot w, s_h = \rho \cdot h \tag{3}$$

其中,  $\rho$  是比例因子。在基于孪生网络的跟踪算法和判别性跟踪算法中,搜索范围通常为正方形区域,如式(4)所示:

$$s_w = s_h = \rho \cdot \sqrt{\left(w + \frac{w+h}{2}\right) \cdot \left(h + \frac{w+h}{2}\right)} \tag{4}$$

### 2.2 搜索范围的调整

设定搜索范围后,主流目标跟踪算法仅在该范围内检测目标,并将检测响应的最大值所在位置作为目标位置。为了抑制目标检测的边界效应,主流跟踪算法在确定目标位置之前,普遍引入了余弦窗惩罚机制<sup>[4-5]</sup>,使检测响应图上偏离搜索中心的响应被抑制。然而,由于目标的运动,当前帧的目标位置可能已经偏离前一帧目标位置,此时如果仍将前一帧目标位置设为搜索范围中心,则真正的目标响应必然会受到余弦窗惩罚的抑制。

图1给出了Football1测试序列上不同搜索范围内的目标检测响应图和跟踪结果示例。

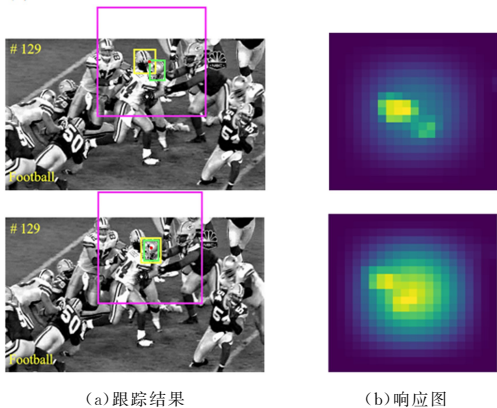


图1 不同搜索范围下的目标跟踪(电子版为彩图)  
Fig.1 Object tracking with different search ranges

图1中,紫色框对应搜索范围,红色点代表搜索中心,绿色框是真实目标框,黄色框是目标跟踪算法预测的目标框。本例中目标跟踪算法采用了SiamRPN。图1(a)上图的搜索范围是以前一帧目标位置为中心划定的,但是当前帧真实目标已经偏离该中心,导致余弦窗惩罚后的检测响应最大值落在真实目标左上方的干扰目标上,而真实目标响应被抑制,如图1(b)上图所示,结果目标跟踪算法定位到错误的目标。图1(a)下图将搜索范围中心移动到真实目标上,采用调整后的搜索范围得到的最终响应如图1(b)下图所示,响应最大值落在真实目标上,此时能够正确跟踪目标。

由上例可见,对主流目标跟踪算法搜索范围的位置进行适当的调整,有助于改善跟踪结果。为防止余弦窗惩罚机制对真实目标的错误抑制,搜索范围的中心应尽量选在真实目标上,这就需要为目标位置进行预测。

一些早期的基于模板匹配的目标跟踪方法<sup>[6-7]</sup>利用卡尔曼滤波原理预估下一帧目标可能出现的位置,并将该位置作为搜索中心设置搜索范围。另外一些工作<sup>[8-9]</sup>通过构建一阶或二阶运动预测模型对目标位置进行预测。然而,这些方法均采用经验模型来预测目标运动,对复杂运动模式的预测能力较弱。近年来,随着循环神经网络的引入,运动轨迹预测技术得到了快速发展<sup>[10-12]</sup>,这种采用数据驱动的运动预测模型比单纯线性预测模型具有更强的适应能力,但还未有相关工作将其应用于视觉目标跟踪的搜索范围调整。

还有一些工作<sup>[13-14]</sup>对搜索范围的大小进行调整以适应目标快速运动、尺度变化以及遮挡等复杂情况,并取得了一定效果。这些方法按照一定方式扩大搜索范围来避免跟踪目标超出范围后无法被检测到。

本文针对主流目标跟踪算法提出一种自适应搜索范围调整方法,采用基于循环神经网络的运动预测模型来预测目标位置,以此为中心重新设定搜索范围,并根据目标运动信息对搜索范围大小进行自适应调整。

## 3 自适应搜索范围调整(ASRA)

本文提出的自适应搜索范围调整方法主要包含决策模块  $\Psi$ 、搜索中心预测模块  $\Theta$  和搜索范围尺寸调整模块  $\Phi$  3 个组成部分。在使用特定的目标跟踪算法(如 SiamRPN)进行目标跟踪的过程中,对每一跟踪帧设定初始的搜索范围:首先,使用决策模块  $\Psi$  进行判断,如果需要调整,则通过搜索中心预测模块  $\Theta$  对搜索范围位置进行调整;然后,再次使用决策模块  $\Psi$  对新的搜索范围进行判断,如果还需要进一步调整,则使用搜索范围尺寸调整模块  $\Phi$  对搜索范围的大小进行调整;最后,使用跟踪算法在最终的搜索范围内跟踪目标。整体的处理流程如图2所示。

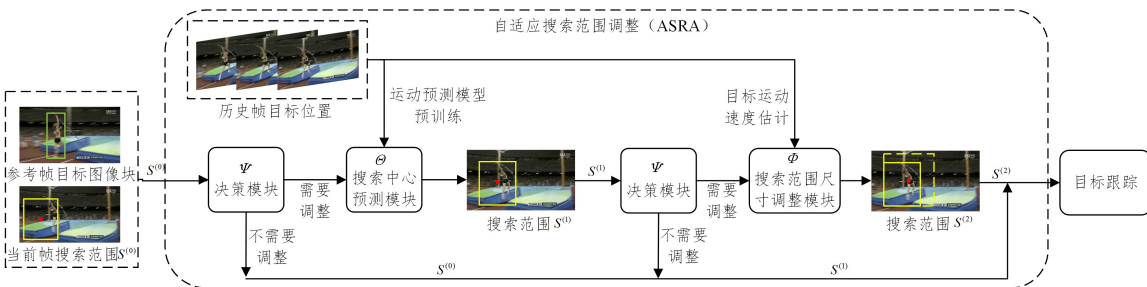


图2 自适应搜索范围调整处理流程

Fig.2 Adaptive search range adjustment process

### 3.1 决策模块 $\Psi$

由于当前主流的目标跟踪算法通常只在设定的搜索范围内检测目标,因此如果能够提前对当前搜索范围的情况做出判断,则可以预先进行必要的调整,以提高后续跟踪算法的性能。

决策模块  $\Psi$  将参考目标图像块与当前搜索范围图像块做相关滤波后,加余弦窗惩罚得到响应图,并根据响应图的峰值分布情况来判断搜索范围内目标的显著性情况。如果目标响应不够显著,则可能是搜索范围中心存在干扰目标,而真实目标偏离搜索中心受到余弦窗惩罚机制的抑制所致<sup>[8]</sup>。此时使用主流目标跟踪器在该搜索范围内检测目标时也会受到同样的影响,需要对当前搜索范围进行调整。

传统上,一般采用峰值旁瓣比(Peak Sidelobe Ratio, PSR)指标来描述响应图的显著性情况<sup>[15-16]</sup>。但是 PSR 指标对存在部分遮挡情况的场景鲁棒性较差<sup>[17]</sup>,不具有普适性。最近 Wang 等<sup>[17]</sup>提出了更鲁棒的峰间相关能(Peak-to-Correlation Energy, PCE)指标。本文选用 PCE 作为标准进行判断。

假设第  $k$  帧的当前搜索范围为  $S_k$ ,参考目标图像块与当前帧搜索范围内的图像块的相关滤波响应图为  $G_k$ ,PCE 指标计算如下:

$$PCE = \frac{|G_k^*|^2}{\sum_{(i,j) \in S_k} |G_k(i,j)|^2} \quad (5)$$

其中, $G_k(i,j)$ 是  $G_k$  在当前帧 $(i,j)$ 位置处的响应值, $G_k^*$ 是  $G_k$  的最大值。 $G_k$  的计算与 KCF<sup>[5]</sup>跟踪器的计算过程相同。

当搜索范围内目标较为显著时,响应图  $G_k$  应该包含单一陡峭、尖锐的“山峰”,此时 PCE 取值较大;当搜索范围内干扰目标较多时,响应图  $G_k$  内可能会没有明显单一“高峰”,此时 PCE 取值较小。图 3 给出了上述两种情况的示例,其中紫色框为搜索范围,绿色框为目标框。设定一个阈值  $\lambda$ ,通过 PCE 与  $\lambda$  的关系来判断所设搜索范围内目标的显著性。当  $PCE \geq \lambda$  时,认为当前设定的搜索范围有利于跟踪器定位目标,不需要调整;当  $PCE < \lambda$  时,认为搜索范围内的干扰目标容易影响跟踪器的判断,应该进行调整。

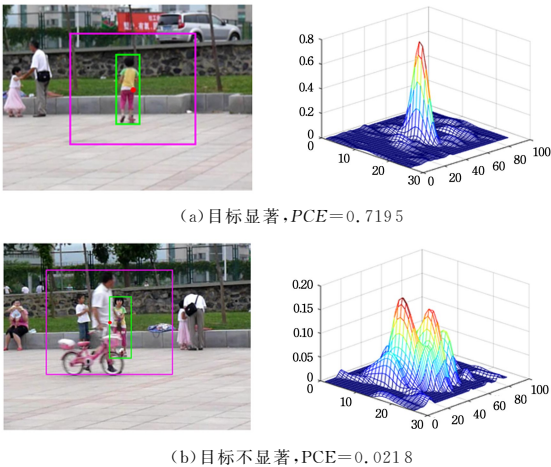


图 3 不同目标显著性下的峰间相关能(电子版为彩图)

Fig. 3 PCE of object with different significance

### 3.2 搜索中心预测模块 $\Theta$

当决策模块  $\Psi$  判定当前帧搜索范围需要调整时,搜索中心预测模块  $\Theta$  用基于 RNN 的运动预测模型对当前帧位置

进行预测,并结合相关滤波响应图联合确定新的搜索中心,进而对搜索范围进行调整。

运动预测模型由  $N$  个 RNN 单元组成,有  $N$  个输入,1 个输出。设当前帧为第  $k$  帧,则输入序列为过去  $N$  帧图像的目标位置组成的序列  $P_{K-N}, P_{K-N+1}, \dots, P_{K-1}$ ,输出为对当前帧目标位置的预测值  $\hat{P}_k = (u_k, v_k)$ ,这里  $(u_k, v_k)$  是  $\hat{P}_k$  在图像中的像素坐标。运动预测模型在大量运动轨迹数据上训练得到。实际上,从 VOT2018<sup>[18]</sup>训练数据集的真实目标位置标签采集获得目标运动轨迹数据。

单纯根据历史运动信息预测得到的位置并不完全可靠,还需要结合当前帧的空间纹理信息对预测的目标位置进行修正。假设  $S_k^{(0)} = S(c_{x,k}^{(0)}, c_{y,k}^{(0)}, s_{w,k}^{(0)}, s_{h,k}^{(0)})$  代表第  $k$  帧图像的初始搜索范围。其中, $(c_{x,k}^{(0)}, c_{y,k}^{(0)})$  是  $S_k^{(0)}$  的中心坐标, $s_{w,k}^{(0)}$  和  $s_{h,k}^{(0)}$  分别是  $S_k^{(0)}$  的宽和高,其取值如式(2)一式(4)所示。得到第  $k$  帧图像的目标位置  $\hat{P}_k$  后,首先以此位置为中心在  $S_k^{(0)}$  范围内构建一个二维高斯矩阵  $D_k$ ,如式(6)所示:

$$D_k(i,j) = \frac{1}{2\pi\sigma^2} e^{-\frac{(i-u_k)^2 + (j-v_k)^2}{2\sigma^2}} \quad (6)$$

其中, $(i,j) \in S_k^{(0)}$  为搜索范围中的点在图像中的像素坐标。然后,将决策模块  $\Psi$  中得到的当前帧相关滤波响应图  $G_k$  与  $D_k$  加权求和,得到高斯联合置信度图  $U_k$ ,如式(7)所示:

$$U_k(i,j) = \alpha D_k(i,j) + \beta G_k(i,j) \quad (7)$$

其中, $\alpha, \beta$  为加权系数。最后,取  $U_k$  的最大值所在位置作为第  $k$  帧新的搜索中心,如式(8)所示:

$$(c_{x,k}^{(1)}, c_{y,k}^{(1)}) = \arg \max_{(i,j) \in S_k^{(0)}} (U_k(i,j)) \quad (8)$$

第  $k$  帧调整位置后的新搜索范围为  $S_k^{(1)} = S(c_{x,k}^{(1)}, c_{y,k}^{(1)}, s_{w,k}^{(0)}, s_{h,k}^{(0)})$ 。

### 3.3 搜索范围尺寸调整模块 $\Phi$

当目标快速运动或者发生尺度变化时,仍有可能偏离新的搜索中心甚至超出搜索范围  $S_k^{(1)}$ ,此时还需要对搜索范围的大小进行调整。使用决策模块  $\Psi$  再次对  $S_k^{(1)}$  进行判断,如果仍需要继续调整,则对目标偏离搜索中心的方向进行估计,然后将搜索范围向该方向进行扩展。

为了估计目标偏离搜索中心的方向,在  $G_k$  中寻找响应最大的局部区域,则该区域相对于搜索中心的方向即为目标偏离中心的方向。实际过程中,首先将  $G_k$  沿水平和垂直中线平均分为 4 个局部区域,分别记为  $A_1, A_2, A_3, A_4$ 。调整前搜索范围和响应图分割示例如图 4(a)和图 4(b)所示,其中紫色框对应当前搜索范围,绿色框是真实目标框。然后,分别计算各个区域的相关响应均值,如式(9)所示:

$$\bar{G}_{k,n} = \frac{\sum_{(i,j) \in A_n} G_k(i,j)}{s_{w,k}^{(0)}/2 \cdot s_{h,k}^{(0)}/2}, n \in \{1, 2, 3, 4\} \quad (9)$$

并找到最大的一个对应的序号,即

$$n^* = \arg \max_{n \in \{1, 2, 3, 4\}} \bar{G}_{k,n} \quad (10)$$

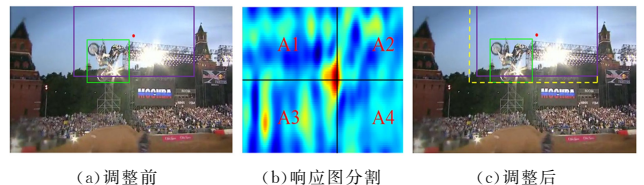


图 4 响应图分割及搜索范围尺寸调整(电子版为彩图)

Fig. 4 Response graph segmentation and search range resizing

搜索范围尺寸的调整根据目标运动速度来决定。当前帧目标运动速度  $\vec{V}_k$  可以利用 Li 等<sup>[20]</sup> 提出的帧间速度公式计算,如下所示:

$$\vec{V}_k = (\Delta x_k, \Delta y_k) = (u_{k-1} - u_{k-2}, v_{k-1} - v_{k-2}) \quad (11)$$

其中,  $\Delta x_k$  和  $\Delta y_k$  分别代表  $\vec{V}_k$  的水平 and 垂直速度,  $(u_{k-1}, v_{k-1})$  和  $(u_{k-2}, v_{k-2})$  是前两帧的目标位置。这里假定相邻帧间目标速度不变,则搜索范围变为  $S_k^{(2)} = S(c_{x,k}^{(2)}, c_{y,k}^{(2)}, s_{w,k}^{(2)}, s_{h,k}^{(2)})$ , 其中:

$$c_{x,k}^{(2)} = \begin{cases} c_{x,k}^{(1)} - |\Delta x_k|/2, & n^* = 1, 3 \\ c_{x,k}^{(1)} + |\Delta x_k|/2, & n^* = 2, 4 \end{cases} \quad (12)$$

$$c_{y,k}^{(2)} = \begin{cases} c_{y,k}^{(1)} - |\Delta y_k|/2, & n^* = 1, 2 \\ c_{y,k}^{(1)} + |\Delta y_k|/2, & n^* = 3, 4 \end{cases} \quad (13)$$

$$s_{w,k}^{(2)} = s_{w,k}^{(1)} + |\Delta x_k| \quad (14)$$

$$s_{h,k}^{(2)} = s_{h,k}^{(1)} + |\Delta y_k| \quad (15)$$

尺寸调整后的搜索范围如图 4(c) 中黄色虚线框所示。

## 4 实验

### 4.1 实验设置

本文将提出的自适应搜索范围调整方法(ASRA)应用在目前先进的几种跟踪算法上,分别在 OTB100<sup>[19]</sup> 和 VOT2018<sup>[18]</sup> 数据集上比较了使用 ASRA 前后各算法的跟踪性能。

表 1 OTB2015 上的平均准确-成功率对比实验结果

Table 1 Average precision-success rate comparison on OTB2015

Method	Base		Base+ASRA w/o $\emptyset$		Base+ASRA	
	Success	Precision	Success	Precision	Success	Precision
SiamRPN	0.624	0.840	0.642(+0.018)	0.854(+0.014)	0.657(+0.033)	0.868(+0.028)
SiamDW	0.642	0.889	0.657(+0.015)	0.903(+0.014)	0.664(+0.022)	0.912(+0.023)
SiamMask	0.641	0.831	0.652(+0.011)	0.848(+0.017)	0.662(+0.021)	0.859(+0.028)
SiamBAN	0.672	0.874	0.681(+0.009)	0.887(+0.013)	0.689(+0.017)	0.903(+0.029)
SiamCar	0.675	0.873	0.682(+0.007)	0.884(+0.011)	0.690(+0.015)	0.901(+0.028)
SiamRCNN	0.678	0.866	0.688(+0.010)	0.876(+0.010)	0.693(+0.015)	0.882(+0.016)
ECO	0.684	0.881	0.685(+0.001)	0.881(+0.000)	0.685(+0.001)	0.881(+0.000)
DiMP	0.670	0.878	0.671(+0.001)	0.878(+0.000)	0.672(+0.002)	0.878(+0.000)

### 4.2.2 VOT2018 数据集上的测试结果比较

各种对比方法在 VOT2018 数据集 baseline 模式下的期望平均重叠分数(EAO)、准确率(A)和鲁棒性(R) 3 个指标如表 2 所列。这里仅给出了 ASRA 对几种孪生网络类跟踪算法的性能改变情况。由表 2 可见,使用 ASRA 方

法的参数配置为:决策模块  $\Psi$  中的阈值根据文献[17] 设定为  $\lambda=0.2$ ,联合置信度图加权系数和循环神经元数根据实验选择为  $\alpha=\beta=0.5, N=3$ 。所有实验均在 PyTorch 深度学习框架下实现,实验平台硬件环境配置为: i5218R CPU, 126GB 内存, NVIDIA RTX 3090 GPU。

### 4.2 不同数据集上的性能比较

#### 4.2.1 OTB2015 数据集上的测试结果比较

在 OTB2015 数据集的 100 个视频序列上分别比较了当前先进的目标跟踪算法在使用 ASRA 方法前后的平均成功率和平均准确率,结果如表 1 所列。比较的跟踪算法包括 SiamRPN<sup>[21]</sup>, SiamDW<sup>[22]</sup>, SiamMask<sup>[23]</sup>, SiamBAN<sup>[24]</sup>, SiamCar<sup>[25]</sup>, SiamRCNN<sup>[26]</sup> 等孪生网络类算法。表中 Base 表示原始算法, Base+ASRA 表示使用 ASRA 方法后的算法。表中还单独给出了只做一次搜索范围调整,即不含搜索范围尺寸调整模块  $\emptyset$  时的结果,用 Base+ASRA w/o  $\emptyset$  来标记。

由表 1 可见,6 种孪生网络类跟踪算法在使用 ASRA 方法后,其平均成功率和平均准确率相较于基准算法约有 0.02~0.03 的提升,说明引入 ASRA 方法可以有效改善该类跟踪器的性能。由 Base+ASRA w/o  $\emptyset$  的结果可见,不含搜索范围尺寸调整模块  $\emptyset$  时,性能相比完整算法略有下降但仍优于基准算法,说明 ASRA 方法中的两次搜索范围调整模块  $\Theta$  和  $\Phi$  均对改善跟踪算法性能有一定贡献。

表 2 VOT2018 上的对比实验结果

Table 2 Experiment results on VOT2018

Method	Base			Base+ASRA		
	EAO( $\uparrow$ )	A( $\uparrow$ )	R( $\downarrow$ )	EAO( $\uparrow$ )	A( $\uparrow$ )	R( $\downarrow$ )
SiamRPN	0.343	0.572	0.304	0.378(+0.035)	0.586(+0.014)	0.276(-0.028)
SiamDW	0.270	0.538	0.398	0.317(+0.047)	0.585(+0.047)	0.376(-0.022)
SiamMask	0.373	0.584	0.272	0.394(+0.021)	0.590(+0.006)	0.203(-0.069)
SiamBAN	0.434	0.630	0.210	0.447(+0.013)	0.590(-0.040)	0.178(-0.032)
SiamCar	0.414	0.600	0.234	0.423(+0.009)	0.578(-0.022)	0.198(-0.036)
SiamRCNN	0.393	0.580	0.300	0.405(+0.012)	0.612(+0.032)	0.220(-0.080)

### 4.3 结果分析和讨论

#### 4.3.1 跟踪结果直观比较

为了直观地观察比较 SiamRPN+ASRA 与 SiamRPN 方法的跟踪性能,图 5 分别给出了两种算法在 Bird1 序列上的

跟踪结果。其中,上排为 SiamRPN 的跟踪结果,紫色框代表搜索范围,红色点是搜索中心,绿色框是真实目标框,黄色框是两种算法分别得到的跟踪框。

由图 5 可见, SiamRPN 在第 24 帧时,由于目标相关匹

配失败,导致跟踪框大幅度偏离真实框;在第25帧时,其搜索中心由前一帧跟踪框的中心来设定,导致其跟踪结果仍然落在该错误位置附近,进而影响后续帧的跟踪结果。而 SiamRPN+ASRA 尽管在第24帧出现同样的跟踪

错误,但是由于其搜索中心预测模块 $\Theta$ 利用了目标的历史运行信息对搜索中心进行调整,使得第25帧的搜索中心被校正到真实目标附近,因此后续帧能够继续正确跟踪目标。



图5 Bird1序列上的跟踪结果

Fig. 5 Tracking results on Bird1 sequence

#### 4.3.2 不同场景下定性分析

为了更为细致的分析所提 ASRA 方法的性能,表3和表4分别列出了改进算法 SiamRPN+ASRA, SiamRPN+ASRA w/o  $\Phi$  与基准算法 SiamRPN 在 OTB2015 数据集中的11种不同跟踪场景下的平均成功率和准确率。由结果可见,两种改进算法相比基准算法在全部11个跟踪场景下性能均有提升,充分说明本文 ASRA 方法能够改善跟踪器性能。以平均成功率为例, SiamRPN+ASRA 相对于 SiamRPN 性能提升较明显的跟踪场景为 LR(低分辨率)、FM(快速移动)和 SV(尺度变化),平均成功率分别提升了6.4%,5.7%和4.8%。下面对不同场景下 ASRA 方法能够改进基准算法性能的原因展开定性分析。

1)LR场景:该场景指将初始参考目标图像块的像素数降到400以下,从而给跟踪算法提供了低分辨率的参考目标。由于 SiamRPN 算法主要依赖于空间信息相似度来检测和定位目标,而参考目标的低分辨率处理破坏了原始跟踪目标的空间信息,从而导致跟踪过程中出现目标漂移、丢失等现象的可能性增加,影响跟踪算法的性能。而 SiamRPN+ASRA

方法除了考虑跟踪目标空间信息相似度以外,还充分利用了目标的运动信息,从而获取了更加丰富有效的判断依据,因此能够实现明显的性能提升。

2)FM场景:由于 SiamRPN 算法的搜索中心就是前一帧的跟踪结果的目标中心,因此当前帧的搜索范围还是停留在前一帧目标周围。如果目标快速移动而超出当前设定的搜索范围,则跟踪算法就无法在搜索范围内找到跟踪目标,跟踪性能受损。而本文提出的 ASRA 方法依靠目标运动信息来自适应调整搜索范围的位置和大小,从而在目标快速运动的情况下也可以大概率将目标包含在搜索范围内,从而实现跟踪性能的提升。

3)SV场景:该场景指跟踪目标尺寸相对于第一帧参考目标变化超过2倍的视频序列。目标尺寸的变化也破坏了目标的空间信息,因此与 LR 场景一样,额外利用了目标运动信息的 SiamRPN+ASRA 方法能够显著提升跟踪性能。此外,对于因目标尺寸扩大而只有部分落在原搜索范围内的情况,由于本文 ASRA 方法能够自适应地扩大搜索范围,因此能够将原搜索范围不能覆盖的目标部分也包含在内,从而提升了跟踪性能。

表3 OTB2015中不同跟踪场景下的成功率对比

Table 3 Comparison of success rates in different tracking scenarios on OTB2015

	SV	FM	BC	MB	DEF	IV	IPR	LR	OCC	OPR	OV
SiamRPN	0.619	0.605	0.596	0.630	0.602	0.651	0.631	0.585	0.573	0.618	0.544
SiamRPN+ASRA w/o $\Phi$	0.652	0.644	0.612	0.659	0.614	0.683	0.635	0.613	0.585	0.621	0.596
SiamRPN+ASRA	0.667	0.662	0.618	0.664	0.618	0.691	0.667	0.649	0.598	0.633	0.586

表4 OTB2015中不同跟踪场景下的准确率对比

Table 4 Comparison of precision rates in different tracking scenarios on OTB2015

	SV	FM	BC	MB	DEF	IV	IPR	LR	OCC	OPR	OV
SiamRPN	0.843	0.800	0.807	0.831	0.807	0.862	0.857	0.867	0.763	0.841	0.728
SiamRPN+ASRA w/o $\Phi$	0.861	0.843	0.804	0.850	0.832	0.886	0.873	0.880	0.770	0.844	0.794
SiamRPN+ASRA	0.876	0.866	0.806	0.864	0.842	0.900	0.902	0.898	0.793	0.853	0.786

**结束语** 本文针对当前主流跟踪算法中搜索范围设定不准确导致跟踪性能受损的问题,提出了自适应搜索范围调整方法 ASRA。该方法可以与当前主流的基于孪生网络的跟踪算法相结合以提升其性能。在 OTB2015 和 VOT2018 两个数据集上的实验结果表明,所提 ASRA 方法能够提升各孪生网络跟踪器的准确率和成功率。通过在不同场景下与 SiamRPN 跟踪性能的比较发现,采用 ASRA 方法改进后的跟踪算法在运动场景和跟踪目标空间信息变化的场景下相对更有优势。本文仅使用简单的循环神经网络构建目标运动模型,

无法精准地预测目标运动趋势,后续工作将重点研究如何更好地将目标的运动信息与当前主流的基于空间特征匹配的跟踪算法相结合,以提升跟踪算法性能。

#### 参考文献

- [1] MENG Y, YANG X. Overview of target tracking algorithms [J]. Journal of Automation, 2019, 45(7): 1244-1260.
- [2] ZHANG S P, YAO H X, SUN X, et al. Sparse coding based visual tracking: review and experimental comparison [J]. Pattern

- Recognition, 2013, 46(7):1772-1788.
- [3] ZHE C, HONG Z, TAO D. An Experimental Survey on Correlation Filter-based Tracking [J]. Computer Science, 2015, 53(6025):68-83.
- [4] LI B, WU W, WANG Q, et al. SiamRPN++: Evolution of Siamese Visual Tracking With Very Deep Networks[C] // 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition(CVPR). Long Beach, CA, USA, 2019:4277-4286.
- [5] FAN H, LING H B. Siamese cascaded region proposal networks for real-time visual tracking[C] // Proceedings of the 2019 IEEE Conference on Computer Vision and Pattern Recognition. Long Beach, USA; IEEE, 2019:7952-7961.
- [6] WANG S, WANG H, ZHOU Y, et al. Automatic Laser Profile Recognition and Fast Tracking for Structured Light Measurement Using Deep Learning and Template Matching[J]. Measurement, 2020, 169:108362.
- [7] ZHANG Y, WANG L, WANG D, et al. Learning Regression and Verification Networks for Robust Long-term Tracking[J]. International Journal of Computer Vision, 2021, 129:2536-2547.
- [8] LI J, YAN B, LIN C, et al. JROTM: Jointly Reinforced Object Tracking with Temporal Content Reference and Motion Guidance[J]. Neurocomputing, 2021, 434:285-294.
- [9] ALAHI A, GOEL K, RAMANATHAN V, et al. Social LSTM: Human Trajectory Prediction in Crowded Spaces[C] // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition(CVPR). 2016:961-971.
- [10] BHUJEL N, TEOH E K, YAU W Y. Pedestrian Trajectory Prediction Using RNN Encoder-Decoder with Spatio-Temporal Attentions[C] // 2019 IEEE 5th International Conference on Mechatronics System and Robots(ICMSR). IEEE, 2019.
- [11] FEREDÉ S, XIE X, ZHANG C, et al. Small Ball Tracking with Trajectory Prediction[C] // 2020 IEEE 5th International Conference on Signal and Image Processing(ICSIP). IEEE, 2020.
- [12] ZHU Z, WANG Q, LI B, et al. Distractor-Aware Siamese Networks for Visual Object Tracking[C] // Computer Vision(ECCV 2018). Lecture Notes in Computer Science, Cham: Springer, 2018:103-119.
- [13] LI S, CHU J, ZHONG G, et al. Robust Visual Tracking with Occlusion Judgment and Re-Detection[J]. IEEE Access, 2020, 8:122772-122781.
- [14] WANG J, YANG H, XU N, et al. Long-term target tracking combined with re-detection[J]. EURASIP Journal on Advances in Signal Processing, 2021, 2021(1):1-16.
- [15] CHEN Y, WANG P, ZHONG B, et al. Coarse-to-fine visual tracking with PSR and scale driven expert-switching[J]. Neurocomputing, 2018, 275(JAN. 31):1456-1467.
- [16] WANG Y, DING L, LAGANIERE R. Real-Time UAV Tracking Based on PSR Stability[C] // 2019 IEEE/CVF International Conference on Computer Vision Workshop (ICCVW). IEEE, 2019.
- [17] WANG M, SU D, SHI L, et al. Real-time 3D human tracking for mobile robots with multisensors[C] // 2017 IEEE International Conference on Robotics and Automation (ICRA). 2017:5081-5087.
- [18] KRISTAN M, MATAS J, LEONARDIS A, et al. A Novel Performance Evaluation Methodology for Single-Target Trackers [C] // IEEE Transactions on Pattern Analysis and Machine Intelligence. 2016:2137-2155.
- [19] WU Y, LIM J, YANG M. Object Tracking Benchmark [C] // IEEE Transactions on Pattern Analysis and Machine Intelligence. 2015:1834-1848.
- [20] LAO Y W, ZHUJ D, ZHENG Y F. Sequential particle generation for visual tracking [J]. IEEE Transactions on Circuits and Systems for Video Technology, 2009, 19(9):1365-1378.
- [21] LI B, YAN J J, WU W, et al. High performance visual tracking with siamese region proposal network[C] // Proceedings of the 2018 IEEE Conference on Computer Vision and Pattern Recognition. Salt Lake City, USA; IEEE, 2018:8971-8980.
- [22] ZHANG Z, PENG H. Deeper and wider siamese networks for real-time visual tracking[C] // Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2019:4591-4600.
- [23] WANG Q, ZHANG L, BERTINETTO L, et al. Fast Online Object Tracking and Segmentation: A Unifying Approach [C] // 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2019:1328-1338.
- [24] CHEN Z, ZHONG B, LI G, et al. Siamese box adaptive network for visual tracking[C] // Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2020:6668-6677.
- [25] 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.
- [26] VOIGTLAENDER P, LUITEN J, TORR P H S, et al. Siam r-cnn: Visual tracking by re-detection [C] // Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2020:6578-6588.



**WANG Chao**, born in 1983, Ph.D, lecturer, master supervisor, is a member of China Computer Federation. His main research interests include video/image processing and computer vision.