

基于前景优化的视觉目标跟踪算法

谢青松¹, 刘晓庆², 安志勇¹, 李博¹

(1. 山东工商学院计算机科学与技术学院, 山东烟台 264000; 2. 山东工商学院信息与电子工程学院, 山东烟台 264000)

摘要: 将目标分割技术引入跟踪领域是当前的研究热点. 目前, 基于分割的跟踪算法往往根据分割结果计算最小外接矩形, 以此作为跟踪框, 但复杂的目标运动使得跟踪框内包含较多背景, 从而导致精度下降. 针对该问题, 本文提出了一种基于前景优化的视觉目标跟踪算法, 将跟踪框的尺度和角度优化统一于前景优化框架中. 首先评估跟踪框内的前景比例, 若小于设定阈值, 则对跟踪框分别进行尺度和角度优化; 在尺度优化模块中, 结合回归框计算跟踪框的条件概率, 根据条件概率的结果分情形进行尺度优化; 角度优化模块中, 针对跟踪框设定多个偏移角度, 利用前景IoU (Intersection over Union) 极大策略选择最优跟踪框角度. 结果证明, 将本文方法应用于SiamMask算法, 精度在VOT2016, VOT2018和VOT2019数据集分别提升约3.2%, 3.7%和3.6%, 而EAO分别提升约1.8%, 1.9%和1.6%. 另外, 本文的方法针对基于分割的跟踪算法具有一定的普适性.

关键词: 目标分割; 目标跟踪; 前景优化; 尺度优化; 角度优化

中图分类号: TP391.4

文献标识码: A

文章编号: 0372-2112(2022)07-1558-09

电子学报 URL: <http://www.ejournal.org.cn>

DOI: 10.12263/DZXB.20210641

Visual Object Tracking Algorithm Based on Foreground Optimization

XIE Qing-song¹, LIU Xiao-qing², AN Zhi-yong¹, LI Bo¹

(1. School of Computer Science and Technology, Shandong Technology and Business University, Yantai, Shandong 264000, China;

2. School of Information and Electronic Engineering, Shandong Technology and Business University, Yantai, Shandong 264000, China)

Abstract: The introduction of object segmentation technology into the tracking field is a current research hotspot. At present, the tracking algorithm based on segmentation often calculates the minimum bounding rectangle as the bounding box according to the segmentation result. However, the complex target movement makes the bounding box contain more background, which leads to a decrease in accuracy. In response to the problem, this paper proposes a visual object tracking algorithm based on foreground optimization, which unifies the optimization of the scale and angle in the bounding box into the foreground optimization frame. First, the foreground ratio in the bounding box is evaluated. If it is less than the set threshold, the scale and angle of the bounding box are optimized; in the scale optimization module, the conditional probability of the bounding box is calculated in combination with the regression box, and the scale is optimized according to the results of the conditional probability; in the angle optimization module, many deviation angles are set for the bounding box, and the optimal bounding box angle is chosen by the foreground IoU (Intersection over Union) maximum strategy. The proposed method is applied to the SiamMask algorithm. Results show that the accuracy is improved by about 3.2%, 3.7% and 3.6% in the VOT2016, VOT2018 and VOT2019 data sets, respectively, while EAO is increased by about 1.8%, 1.9% and 1.6%, respectively. Moreover, our method has a certain universality for segmentation-based tracking algorithms.

Key words: object segmentation; object tracking; foreground optimization; scale optimization; angle optimization

1 引言

目标跟踪是根据初始帧目标计算其在后续帧的位置, 该技术在智能视频监控^[1,2]、人机交互^[3]、虚拟现

实^[4]和视觉导航^[5]等领域得到广泛应用, 成为计算机视觉领域的重要分支^[6-8]. 目前的跟踪方法主要分为生成式和判别式两大类. 早期方法主要以生成式为主, 如光流法^[9,10], 粒子滤波^[11], MeanShift^[12], CamShift^[13], 但是

收稿日期: 2021-05-19; 修回日期: 2021-10-07; 责任编辑: 李勇锋

基金项目: 国家自然科学基金(No.62072285); 山东省自然科学基金(No.ZR202102230438, No.ZR202102180144, No.ZR202103070119, No.ZR2020MA030); 山东省重点研发计划(软科学)(No.2020RKB01017); 山东工商学院校级教学改革项目(No.11688202023)

该类方法只依赖目标信息进行跟踪,而目标变化存在多样性和复杂性,因此很容易跟丢目标.与生成式不同,判别式方法通过训练分类器来区分目标和背景,在精度上存在较大优势,目前主要以相关滤波和深度学习为主.相关滤波中,Bolme等^[14]提出了最小平方和滤波器算法,但该方法存在不稳定性,同时缺乏目标尺度估计和多通道等问题.Danelljan等^[15]提出了DSST跟踪算法,利用岭回归算法对相关滤波进行改进,同时增加了尺度滤波器估计目标尺度,然而该方法仅在尺度窗口增加和降低的趋势估计上容易产生漂移,导致小尺度的震荡,检测性能有待提高.目前,更多的是基于深度学习的方法^[16,17],DLT算法^[18]将深度学习方法应用到目标跟踪中.它针对在线训练时正样本缺失的问题,提出了一种离线训练与在线调整相结合的方法.

然而,目前的跟踪方法大都使用轴对齐的水平框跟踪目标,相比水平框,旋转框能够更紧凑的包围目标,并且可提供目标的运动方向,方向信息可以进一步解决如动作分类等诸多计算机视觉问题.虽然旋转框存在较大的优势,但对其角度和尺度的准确估计却是一个难题,相关文献^[19,20]对此进行了研究,然而这些算法在跟踪速度和精度方面都有局限性.孪生网络中的SiamMask^[21]算法将目标分割^[22,23]引入了跟踪领域兼顾了这些问题,该算法能够预测目标的分割掩码,并且实时适应最小外接矩形作为跟踪框.尽管该网络在跟踪性能上取得了很大进展,但是当非刚性目标运动时,实时生成的最小外接矩形的准确度较低,框内包含较多背景.而本文认为,这些背景主要是由于跟踪框尺度和角度的不准确引入的.因此本文将SiamMask作为基准算法,提出了一种基于前景优化的视觉目标跟踪算法,设计了尺度优化和角度优化两个模块,来提升跟踪性能.本文主要贡献总结如下:

(1)针对分割导致的跟踪框尺度不准确,提出了尺度优化策略.该策略不仅保证了目标运动时尺度的稳定性,并且在丢失目标信息的前提下,较大的提升了跟踪框内的前景比例.

(2)针对分割导致的跟踪框角度不准确,提出了角度优化策略.优化后的跟踪框可适应目标不同程度的旋转且与目标角度保持一致,进一步提升了跟踪框内的前景比例.

(3)在VOT2016^[24]、VOT2018^[25]和VOT2019^[26]三个旋转框标记的数据集上进行大量实验证明,本文提出的前景优化策略具有较高的精度,在一定程度上改进了分割与跟踪结合存在的缺陷.

2 相关工作

近年来,基于深度学习的孪生网络系列跟踪算

法^[27-33]取得了优异的成绩.其中,SiamFC^[27]是早期孪生网络的代表,受到广泛关注,运行速度可达到58 fps,然而该网络在跟踪时需要进行多尺度测试,跟踪精度很不理想.SiamRPN^[28]在此基础上借鉴了Faster-RCNN^[34]的区域建议网络,使用边框回归替代多尺度测试,得到最大响应的回归框,准确度显著优于SiamFC.Li等提出SiamRPN++^[29]解决了孪生网络中的平移不变性问题,并提出了一种新的模型来执行分层和深度聚合,提高精度的同时降低了模型的尺度规模.

SiamMask^[21]算法与他们不同,该算法与分割技术相结合,跟踪框可旋转.但是由于SiamMask直接对分割掩码拟合最小外接矩形作为跟踪框,在目标发生运动时跟踪框精度往往较低.之后,Chen等提出SiamMask-E^[35]算法优化了SiamMask的跟踪框拟合过程,该算法首先采用最小二乘法对分割掩码的轮廓进行椭圆拟合,再进一步得到椭圆的最小外接矩形.该方式在一定程度上提升了跟踪框的准确性,但是在目标运动幅度较大时,该方法的跟踪框稳定性较差,精度并不高.

因此,本文在第三节针对SiamMask跟踪框存在的问题,在尺度和角度上分别进行了优化,来提升跟踪框内的前景比例,抑制背景的干扰.

3 基于前景优化的视觉目标跟踪算法

本文的整体流程如图1所示,共分为三个部分.第一部分为SiamMask网络工作, f_θ 为孪生网络,分支网络中包含回归网络 r_θ 、分割网络 h_θ 、分类网络 c_θ .其中,分类损失使用交叉熵损失,记为 L_{cls} ;回归损失采用归一化坐标的 $smooth_{L1}$ 损失,记为 L_{reg} ;分割部分的损失函数记为 L_{mask} ,具体计算在3.1节.第二部分为尺度优化模块,在优化之前先对SiamMask输出的跟踪框内前景比例进行计算,从而判断是否需要优化,不需要直接输出原跟踪框,否则进行尺度优化,包含两种优化情形($M \cap R'$, AS/MS).第三部分为角度优化模块,该模块针对跟踪框的原始角度设定了多个偏移,依次与前景计算IoU(Intersection over Union),选择最高的进行输出.具体实施细节在3.3和3.4节.

3.1 损失函数

特征提取时,模板帧 z 和搜索帧 x 两个分支经过相同的卷积神经网络 f_θ ,分别生成特征图 $f_\theta(z)$ 和 $f_\theta(x)$,两个特征图进行深度互相关确定目标位置:

$$g_\theta(z, x) = f_\theta(z) * f_\theta(x) \quad (1)$$

其中 $*$ 为深度互相关操作, $g_\theta(z, x)$ 为 $f_\theta(z)$ 和 $f_\theta(x)$ 互相关后生成的响应图,响应图中每个空间元素记为一个RoW(Response of a candidate Window).

在三个分支中,分类损失^[28]为交叉熵损失,记为

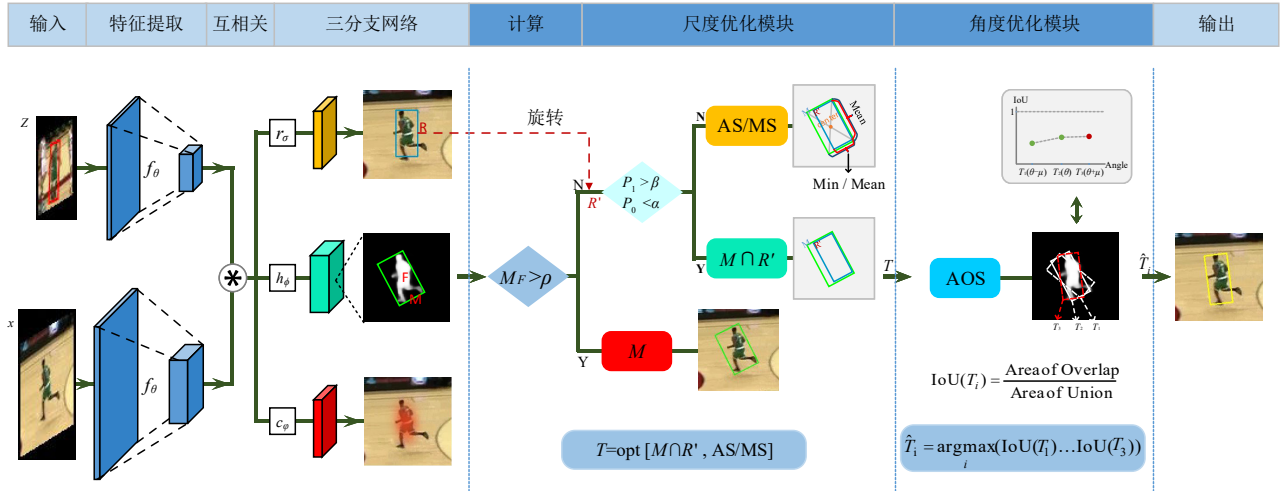


图1 基于前景优化的视觉目标跟踪算法整体流程框架. 图中的 R 为回归分支输出的回归框, R' 为旋转后的回归框, F 为目标分割掩码, M 为掩码的最小外接矩形, T 为尺度优化后输出的跟踪框, \hat{T}_i 为角度优化后输出的最终跟踪框,AS(Adaptive Strategy)为自适应策略,MS(Mean Strategy)为均值策略,AOS(Angle Offset Strategy)为角度偏移策略.

L_{cls} . 回归分支中,由神经网络 r_σ 进行回归,损失函数采用归一化坐标的 smooth_{L_1} 损失,因此回归损失 $^{[28]}L_{reg}$ 为

$$L_{reg} = \sum_{i=0}^3 \text{smooth}_{L_1}(\delta[i], \sigma) \quad (2)$$

$\delta[0] \sim \delta[3]$ 分别表示为锚框与真实框回归后得到的平移量($\delta[0], \delta[1]$)和尺度缩放($\delta[2], \delta[3]$). 分割网络中,由神经网络 h_ϕ 对响应图 g 中的每个RoW预测成 $w \times h$ 的二元掩码,损失函数 $^{[21]}$ 为

$$L_{mask}(\theta, \phi) = \sum_n \left(\frac{1+y_n}{2wh} \sum_{ij} \log(1 + e^{-c_n^{ij} m_n^{ij}}) \right) \quad (3)$$

每个RoW标记为一个二元标签 y_n ,如果RoW的一个锚框与真实框的IoU值 ≥ 0.6 ,则 $y_n=1$;否则 $y_n=-1$. m_n 表示第 n 个RoW预测的二元掩码, c_n^{ij} 为 m_n 中位置 (i, j) 处所对应的二元标签,前景值为1,背景值为-1. 因此,本文主干网络损失函数为

$$L_{loss} = \lambda_1 L_{mask} + \lambda_2 L_{cls} + \lambda_3 L_{reg} \quad (4)$$

损失函数 L_{loss} 中,分别使用参数 $\lambda_1, \lambda_2, \lambda_3$ 对三个分支进行加权. 网络训练完成后,在跟踪阶段会输出当前帧的回归框,目标的分割掩码及其最小外接矩形作为跟踪框. 在3.3节和3.4节分别对该跟踪框进行尺度和角度的优化.

3.2 前景比例计算

前景即为目标的分割掩码部分,表示为 F ;跟踪框即为分割掩码的最小外接矩形,表示为 M ,旋转角度记为 θ . M 的顶点坐标表示为

$$M = [(x_A, y_A), (x_B, y_B), (x_C, y_C), (x_D, y_D)] \quad (5)$$

最小外接矩形即采用面积最小的旋转矩形对前景进行包围,因此前景属于最小外接矩形内. 在跟踪框 M 内,将前景 F 所占的比例记为 M_F :

$$M_F = \frac{F_area}{M_area} \quad (6)$$

F_area 和 M_area 分别为前景 F 的面积和跟踪框 M 的面积. M_area 可通过勾股定理得到,前景是一个二值矩阵,记为 $A_{FPI} \in \{0, 1\}$,矩阵面积为

$$F_area = A_{FPI} \left[\sum_{x=1}^X \sum_{y=1}^Y H(x, y) \right] \quad (7)$$

X, Y 为矩阵的行和列, $H(x, y)$ 为矩阵中坐标 (x, y) 处的值. 本文发现当 $M_F > \rho$ 时,跟踪框 M 内前景所占的比例较高,因此直接输出跟踪框 M 不需要进行优化. 否则,对跟踪框的尺度和角度分别进行优化.

3.3 尺度优化策略

尺度优化策略是一种基于前景的紧凑跟踪框优化方法,即通过优化跟踪框的尺度使它更加紧凑的包围目标,避免引入过多背景. 经过观察,边框回归得到的回归框在目标发生运动时,尺度更加稳定. 因此在本节中,本文结合了回归框来进行尺度优化. 首先,需要对回归框进行旋转预处理,成为与跟踪框 M 一致的角度:

$$\begin{aligned} x_i &= (x_j - x_0) \times \cos(\theta) - (y_j - y_0) \times \sin(\theta) + x_0 \\ y_i &= (x_j - x_0) \times \sin(\theta) + (y_j - y_0) \times \cos(\theta) + y_0 \end{aligned} \quad (8)$$

该操作将回归框 R 围绕中心点 (x_0, y_0) 旋转了 θ 度,旋转后记为 R' . (x_j, y_j) 为 R 的顶点坐标,对应的 (x_i, y_i) 为 R' 的顶点坐标,记为:

$$R' = [(x_a, y_a), (x_b, y_b), (x_c, y_c), (x_d, y_d)] \quad (9)$$

此时,两者旋转角度一致,通过条件概率计算两者的相似性与差异性,根据结果分为两种优化情形. 在这里,将目标搜索区域 $\{(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)\}$ 记为样本空间 S . 当某个样本点出现在矩形 M 内即为跟踪框 A 发生,出现在矩形 R' 内即为跟踪框 B 发生.

首先,通过条件概率公式计算在跟踪框 A 发生的条件下,跟踪框 B 也发生的概率,记为 $P1$:

$$P1 = P(B|A) = \frac{P(A \cap B)}{P(A)} \quad (10)$$

其次,计算在跟踪框 B 发生的条件下,跟踪框 A 不发生的概率,记为 $P0$:

$$P0 = P(\bar{A}|B) = \frac{P(\bar{A} \cap B)}{P(B)} \quad (11)$$

其中, $P(A)$ 、 $P(B)$ 、 $P(A \cap B)$ 、 $P(\bar{A} \cap B)$ 分别为跟踪框 A 发生的概率、跟踪框 B 发生的概率、跟踪框 A 和 B 同时发生的概率、跟踪框 B 发生但 A 不发生的概率,任意概率计算公式为:

$$P(X) = \frac{X \text{ 的样本数目}}{S \text{ 的样本数目}} \quad (12)$$

条件概率 $P1$ 反映了 M 和 R' 两框在位置和尺度的相似性, $P0$ 反映了两者的差异性. 当 $P1$ 越大同时 $P0$ 越小时,表明两者相似性越大,说明跟踪框 M 的准确性较高;相反,当 $P1$ 越小或者 $P0$ 越大时,此时跟踪框 M 的可靠性较低. 定义阈值 α 、 β 分别评估条件概率 $P0$ 和 $P1$,将尺度优化分为两种情形.

情形 1: 当同时满足 $P1 > \beta$ 和 $P0 < \alpha$ 时,此时 M 的可靠性较高,尺度优化为:

$$T = M \cap R' \quad (13)$$

此时 M 的准确性较高,框内的背景像素较少,目标通常为刚性物体或者运动幅度不大的非刚性物体,取两者交集部分作为尺度优化后的跟踪框 T ,跟踪框 T 相较 M 向目标的中心位置处进行了缩小,过滤掉了目标四周的干扰背景,因此跟踪框内前景比例得到了提高.

情形 2: 当 $P1 \leq \beta$ 或者 $P0 \geq \alpha$ 时,该情形的两框差异较大,若采用式(13)进行优化,可能会发生局部标记或者偏移标记的情况,从而丢失目标信息. 因此针对情形 2,本文提出了两种优化策略:均值策略(Mean Strategy, MS)和自适应策略(Adaptive Strategy, AS),在消融实验中本文会对两种策略的性能进行比较分析.

(1) 均值策略(Mean Strategy, MS)

均值策略即不考虑目标的运动方向,对两框在高度和宽度方向均采用平均策略,即:

$$\begin{cases} T_w = 0.5 \times (R'_w + M_w) \\ T_h = 0.5 \times (R'_h + M_h) \end{cases} \quad (14)$$

其中, R'_w 和 R'_h 分别为 R' 的宽高, M_w 和 M_h 分别为 M 的宽高(在任意矩形中,宽 \leq 高). 为了避免计算的复杂以及主观的倾向性,均值策略直接对 M 和 R' 两个框的宽度和高度分别取均值,作为优化后跟踪框 T 的宽和高,记为 T_w 和 T_h ,中心坐标和旋转角度不变.

(2) 自适应策略(Adaptive Strategy, AS)

经过分析,情形 2 中跟踪框 M 包含的干扰背景较多,这些背景一般是由目标的剧烈运动引入,这种目标以人物等非刚体为主,因为四肢的变化很容易引入背景. 当人物在水平方向运动时,由于四肢的摆动在目标宽度方向引入背景的可能性极大;而在垂直方向运动时,在目标的高度方向引入背景的可能性较小.

由于目标运动方向的不同,引入背景的可能性也不一样. 然而均值策略并没有考虑到该问题,因此本文通过判断目标的运动方向提出了自适应策略,根据目标中心位置变化判断垂直方向和水平方向是否有快速移动状态,然后分别设置不同的优化方式.

首先,根据 SiamMask 算法计算目标中心位置,将前一帧的目标中心记为 (x_1, y_1) ,当前帧中心记为 (x_2, y_2) . 因此,通过目标中心的移动路径与 x 轴形成的角度来判断目标运动方向,角度计算如下:

$$\begin{cases} \delta = \arctan\left(\frac{|y_2 - y_1|}{|x_2 - x_1|}\right) \times \frac{180}{\pi}, & |y_2 - y_1| > \varepsilon \\ \delta = 0^\circ, & |y_2 - y_1| \leq \varepsilon \end{cases} \quad (15)$$

$|y_2 - y_1| > \varepsilon$ 代表了目标在 y 轴上的移动距离大于阈值 ε ,因此当角度 $\delta \geq \gamma^\circ$ 时,判断为目标在垂直方向快速运动. 而如果在 y 轴上移动距离小于阈值,则无需计算角度,将 δ 记为 0° . 由于水平方向引入背景的可能性更大,所以垂直方向快速运动还需要限制 M 和 R' 两框的宽度差 w 小于阈值 ε ,即 $|M_w - R'_w| = w < \varepsilon$. 反之,判断为目标在水平方向的快速运动. 因此,自适应策略的优化公式为:

$$\begin{cases} T_w = \begin{cases} 0.5 \times (R'_w + M_w), & \delta \geq \gamma^\circ \text{ and } w < \varepsilon \\ \min(R'_w, M_w), & \text{else} \end{cases} \\ T_h = 0.5 \times (R'_h + M_h) \end{cases} \quad (16)$$

在该策略中,当目标移动与 x 轴形成的角度 $\delta \geq \gamma^\circ$ 且 w 小于阈值 ε ,即满足目标在高度上快速运动,并且水平方向变化不大,此时跟踪框在高度和宽度引入背景的可能性相同,从而对跟踪框 M 和回归框 R' 两者的宽度和高度分别取均值作为新跟踪框的宽和高,宽记为 T_w ,高记为 T_h ,中心坐标和旋转角度不变.

若角度 $\delta < \gamma^\circ$ 或者宽度差 $w \geq \varepsilon$,即说明了目标在水平方向发生显著运动. 此时,跟踪框在水平方向的边往往会被运动的四肢拉伸,导致了跟踪框 M 的宽度与目标宽度差距较大,而高度与目标不会产生太大的偏差. 因此,在不丢失目标信息并且保证更高前景比例的基础上,对稳定性较差的宽度设置两者最小值,记为 T_w ,而高度设置两者均值,记为 T_h ,中心坐标和旋转角度不变.

总之,均值策略 MS 相对保守,直接对两框宽高取

均值避免了计算的复杂性,而自适应策略 AS 则根据目标快速移动方向选择不同的优化方式,得到的跟踪框更准确.

最终,跟踪器在情形 1 和 2 中选择最符合当前帧情形的策略进行优化,然后输出优化后的跟踪框 T , 坐标记为:

$$T=[(x_1, y_1), (x_2, y_2), (x_3, y_3), (x_4, y_4)] \quad (17)$$

本节的尺度优化是一种基于前景的紧凑跟踪框优化方法,即通过优化跟踪框的尺度提升其前景比例. 该策略在目标剧烈运动时依然能够保证跟踪框的准确性与稳定性,消融实验表明,该策略提升了跟踪的精度以及整体性能. 在 3.4 节将尺度优化后的跟踪框 T 进行角度优化.

3.4 角度优化策略

本节在尺度优化基础上,对角度存在缺陷进行改进,进一步提升了跟踪框内前景所占的比例. 具体为:(1)对尺度优化后的跟踪框 T 角度设定多个偏移,包含顺时针和逆时针.(2)将各角度的旋转框处理成二值矩阵,与前景矩阵进行点乘.(3)根据结果计算两的交集与并集比,即 IoU. 最终,选择 IoU 值最高的旋转框输出.

跟踪框 T 的原始角度为 θ , 对其顺时针和逆时针分别设置 n 个偏移阈值,间隔为 μ . 即生成了不同角度的旋转框集合 $\{T_1(\theta-n\mu), \dots, T_n(\theta-\mu), T_{n+1}(\theta), T_{n+2}(\theta+\mu), \dots, T_{2n+1}(\theta+n\mu)\}$, 由 $T_i (i=1, 2, \dots, n, \dots, 2n+1)$ 表示任意一个角度的旋转框.

前景 F 为二值矩阵 \mathbf{A}_{FPI} , 因此需要将旋转框 T_i 部分处理为二值矩阵才能得到两者的交集,记为 $\mathbf{A}_{T_i \text{PI}} \in \{0, 1\}$. 交集矩阵为

$$\mathbf{A}_{\text{IPI}} = \mathbf{A}_{\text{FPI}} \otimes \mathbf{A}_{T_i \text{PI}} \quad (18)$$

其中, \otimes 为矩阵中对应元素点乘, \mathbf{A}_{IPI} 即为两者的交集矩阵. 矩阵面积计算方式如式(7)得到,记为 I_{area} . 因此,两者的 IoU 为

$$\text{IoU}(T_i) = \frac{I_{\text{area}}}{T_i_{\text{area}} + F_{\text{area}} - I_{\text{area}}} \quad (19)$$

T_i_{area} 为 T_i 面积与 T 相等, F_{area} 为前景面积. $\text{IoU}(T_i)$ 即为 T_i 与前景的 IoU, 跟踪器输出令 IoU 最大的旋转框 T_i . 因此,角度偏移策略表示为:

$$\hat{T}_i = \arg \max_i (\text{IoU}(T_1), \text{IoU}(T_2) \cdots \text{IoU}(T_n) \cdots \text{IoU}(T_{2n+1})) \quad (20)$$

角度优化策略通过调整跟踪框的角度来避免角度不准确带来的损失. IoU 常被用在目标检测中衡量预测框与真实框的差异,本节采用 IoU 策略准确的评估了前景与不同角度旋转框之间的差异,因此当 IoU 最大时,该角度相对其他角度更准确. 消融实验表明,本文的角

度优化策略增强了跟踪的鲁棒性以及整体性能.

4 实验与分析

本文的实验是在单目标跟踪的测试平台 VOT (Visual Object Tracking) 数据集上进行的,分别是 VOT2016、VOT2018 和 VOT2019,均是旋转框标记. 评价指标包含精度 A, 鲁棒性 R, EAO (Expected Average Overlap). 精度采用预测框与真实框的 IoU 值进行评估,值越大精度越高;鲁棒性根据跟踪失败次数来计算(当前帧精度为零时判定为跟踪失败),失败的次数越多鲁棒性越差;EAO 是评估跟踪性能的综合指标,被认为是 VOT 数据集中最重要的评价参数.

4.1 实验配置

实验环境:64 位的 Windows 10 操作系统, CPU: Intel(R) Xeon(R) Gold 5218, GPU: GeForce RTX 2080 Ti. Python 3.7 实验, Matlab R2018b 绘图. 实验参数:3.1 节中,设置损失函数的权重参数 $\lambda_1=32, \lambda_2=\lambda_3=1$. 3.3 节中,2016 数据集设定阈值 $\beta=0.85$;2018 和 2019 数据集设定阈值 $\beta=0.8$;在三个数据集中设定共同参数: $\rho=0.9, \alpha=0.2, \varepsilon=8, \gamma=50$. 3.4 节中, $n=1, \mu=20^\circ$.

4.2 消融实验

为了测试每个模块的有效性,本文对 3.3 节的尺度优化和 3.4 节的角度优化两模块进行了消融实验,结果如表 1 所示. 这里, SO (Scale Optimization) 为本文尺度优化模块, SO (MS) 为均值策略优化尺度, SO (AS) 为自适应策略优化尺度. AO (Angle Optimization) 为本文角度优化模块.

首先,表 1 分析了 SiamMask+SO (MS) 和 SiamMask+SO (AS) 两种策略的性能,实验表明两种策略在精度和 EAO 上均优于基准算法 SiamMask,特别是 SiamMask+SO (AS) 优于 SiamMask+SO (MS) 的结果. 同时,表 1 分析了 SiamMask+SO (MS)+AO 和 SiamMask+SO (AS)+AO 的结果,显然 SiamMask+SO (AS)+AO 算法在精度和 EAO 上具有更高的性能. 因此,自适应策略 (AS) 比均值策略 (MS) 的准确度更高,整体性能更佳. 在三个数据集上, SiamMask+SO (AS)+AO 算法相比 SiamMask 算法精度上分别提升约 3.2%、3.7%、3.6%, EAO 分别提升了 1.8%、1.9% 和 1.6%. 另外,本文提出的 SiamMask+SO (AS)+AO 在三个数据集的精度方面均优于 SiamMask_E 算法,同时 EAO 指标在 VOT2016 和 VOT2019 上均优于 SiamMask_E 算法.

其次,本文对比了使用自适应策略的 SiamMask+SO (AS) 以及 SiamMask+SO (AS)+AO 的结果,在 SiamMask+SO (AS) 基础上进行角度优化 AO 后,在 VOT2016、VOT2018 和 VOT2019 数据集上, EAO 分别提升 1.5%、0.8%、0.1%,同时鲁棒性在三个数据集上得

到普遍得到增强,因此角度优化 AO 具有一定的有效性.

表 1 三个数据集下消融实验结果(粗体为最佳性能)

	VOT2016			VOT2018			VOT2019			平均速度(fps)
	A ↑	R ↓	EAO ↑	A ↑	R ↓	EAO ↑	A ↑	R ↓	EAO ↑	
SiamMask	0.620	0.214	0.436	0.597	0.248	0.406	0.596	0.467	0.283	88
SiamMask_E	0.645	0.210	0.452	0.627	0.248	0.427	0.625	0.482	0.298	—
SiamMask+SO(MS)	0.644	0.228	0.437	0.623	0.267	0.410	0.622	0.487	0.294	88
SiamMask+SO(AS)	0.654	0.233	0.439	0.634	0.276	0.417	0.633	0.502	0.298	88
SiamMask+SO(MS)+AO	0.642	0.219	0.443	0.622	0.258	0.415	0.619	0.482	0.294	75
SiamMask+SO(AS)+AO	0.652	0.224	0.454	0.634	0.267	0.425	0.632	0.497	0.299	75

另外,表 1 分析发现本文的优化策略会降低鲁棒性的性能,主要原因是优化后的跟踪框更加紧凑的包围目标,某些情况下跟踪框变小会导致目标丢失,使得重启次数增加,从而整体鲁棒性相比基准算法变差. 在运行速度方面,表 1 统计了三个数据集上的平均速度, SiamMask 算法速度为 88 fps, SiamMask+SO(MS) 和 SiamMask+SO(AS) 也均为 88 fps, 即尺度优化策略对速度没有影响;但增加角度优化后的 SiamMask+SO(MS)+AO 和 SiamMask+SO(AS)+AO 的速度为 75 fps, 由于角度优化需要计算多个角度与前景的 IoU 值,速度有所降低,但仍然能够保持跟踪的实时性.

4.3 与国际前沿算法的比较

在尺度优化中本文采用了自适应策略(AS),因此图 2~图 4 中标注的 Ours 均为 SiamMask+SO(AS)+AO 的跟踪结果. 本文在 VOT2016, VOT2018, VOT2019 三个数据集中共与 12 个国际前沿算法进行了整体性能的比较. 分别是 SiamMask^[21], Siam-RPN^[28], SiamRPN++^[29], DaSiamRPN^[36], SPM^[37], Siam R-CNN^[38], SiamDW^[39], TADT^[40], ASRCF^[41], Meta-Tracker^[42], UpdateNet^[43], ROAM++^[44]. Siam-Mask 的结果数据是在本文的机器测试得到的,其余各算法的结果文件均来自文献[24~26]或者原作者的项目网页中.

图 2 为六个跟踪器在 VOT2016 的 EAO 得分图,图 3 为 VOT2018 得分图. 左侧为跟踪器 EAO 得分排序,如图所示,本文的 EAO 得分最高. 在 VOT2018 数据集中,采用本文优化策略后的 SiamMask 算法超过了 SiamRPN++ 和 Siam R-CNN 两个前沿的算法.

4.4 讨论

本文在相机移动、丢失、光照、运动、遮挡、尺度变化六个干扰属性下分别进行了测试. 结果表明,本文算法在相机移动、运动和尺度变化三个属性下,相比 SiamMask 算法精度得到很大的提升,并且远高于其他国际前沿算法. 在其他属性下,也依然可以保持较高的精度与稳定的性能.

图 4 为部分帧的跟踪效果,四个目标均为非刚性

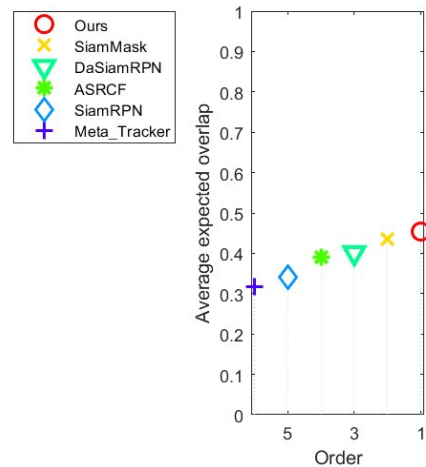


图 2 VOT2016 EAO 得分图

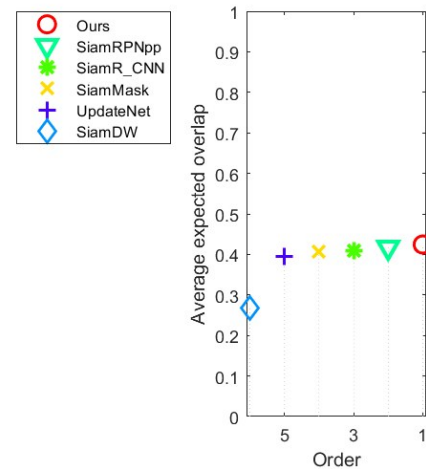


图 3 VOT2018 EAO 得分图

物体,都发生了不同程度的运动、旋转和尺度变化. 如图所示, SiamMask 网络的跟踪框偏大并且旋转角度与目标不一致,引入了过多的干扰背景. 相比之下,本文跟踪框的尺度和角度都更加接近真实框,框内的前景比例更高. 因此,本文的优化策略在目标运动,尺度变化,旋转等复杂场景时有较好的稳定性和准确性.

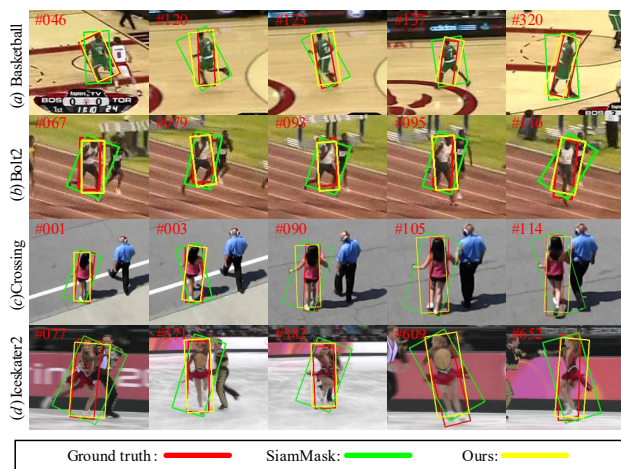


图4 部分视频跟踪效果

5 结论

针对基于分割的跟踪方法得到的跟踪框不准确的问题,本文提出了一种基于前景优化的视觉目标跟踪算法.设计了尺度和角度两个优化模块,来分别提升跟踪框内的前景比例,抑制背景干扰.本文在一定程度上融合和优化了目标分割与跟踪技术,在VOT2016, VOT2018和VOT2019三个数据集上,与SiamMask算法相比,精度平均提升约3.5%,EAO平均提升约1.76%,与其他国际前沿的算法相比具有较好的跟踪性能,特别是在相机移动、尺度变化和运动变化等复杂场景下都保持较高的精度.然而本文算法的鲁棒性会有所下降,针对基于分割的跟踪算法鲁棒性优化将是我们的进一步研究方向.

参考文献

- [1] MATHUR G, SOMWANSHI D, BUNDELE M. Intelligent video surveillance based on object tracking[C]//Proceedings of the 2018 3rd International Conference and Workshops on Recent Advances and Innovations in Engineering. Jaipur, India: IEEE Press, 2018: 1-6.
- [2] GAJJAR V, KHANDHEDIYA Y, GURNANI A. Human detection and tracking for video surveillance: A cognitive science approach[C]//Proceedings of the IEEE International Conference on Computer Vision Workshops. Venice, Italy: IEEE Press, 2017: 2805-2809.
- [3] CHEN B, SAHDEV R, TSOTSOS J, et al. Person following robot using selected online ada-boosting with stereo camera[C]//Proceedings of the 2017 14th Conference on Computer and Robot Vision. Edmonton, Canada: IEEE Press, 2017: 48-55.
- [4] 张爱梅. 目标跟踪瞄准系统的虚拟现实研究[D]. 西安: 西安电子科技大学, 2004.
- [5] ZHANG Ai-mei. Virtual Reality Research of Target Tracking and Aiming System[D]. Xi'an: Xidian University, 2004. (in Chinese)
- [5] AGARWAL N, CHIANG C, SHARMA A. A study on computer vision techniques for self-driving cars[C]//Proceedings of the International Conference on Frontier Computing. Singapore: Springer, 2019: 629-634.
- [6] 孟球, 杨旭. 目标跟踪算法综述[J]. 自动化学报, 2019, 45(7): 1244-1260.
- MENG Lu, YANG Xu. A survey of object tracking algorithms[J]. Acta Automatica Sinica, 2019, 45(7): 1244-1260. (in Chinese)
- [7] WU Y, LIM J, YANG M H. Online object tracking: A benchmark[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Portland, USA: IEEE Press, 2013: 2411-2418.
- [8] 李玺, 查宇飞, 张天柱, 等. 深度学习的目标跟踪算法综述[J]. 中国图象图形学报, 2019, 24(12): 2057-2080.
- LI Xi, ZHA Yu-fei, ZHANG Tian-zhu, et al. Survey of visual object tracking algorithms based on deep learning[J]. Journal of Image and Graphics, 2019, 24(12): 2057-2080. (in Chinese)
- [9] SUN D Q, ROTH S, BLACK M J. Secrets of optical flow estimation and their principles[C]//Proceedings of the 2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. San Francisco, CA: IEEE, 2010: 2432-2439.
- [10] XU L, JIA J Y, MATSUSHITA Y. Motion detail preserving optical flow estimation[C]//Proceedings of the 2010 IEEE Conference on Computer Vision and Pattern Recognition. San Francisco, USA: IEEE, 2010: 1293-1300.
- [11] NUMMIARO K, KOLLER-MEIER E, VAN GOOL L. An adaptive color-based particle filter[J]. Image and Vision Computing, 2003, 21(1): 99-110.
- [12] DU K, JU Y F, JIN Y L, et al. Object tracking based on improved MeanShift and SIFT[C]//Proceedings of the 2nd International Conference on Consumer Electronics, Communications and Networks. Yichang, China: IEEE, 2012: 2716-2719.
- [13] EXNER D, BRUNS E, KURZ D, et al. Fast and robust CAMShift tracking[C]//Proceedings of the 2010 IEEE Conference on Computer Vision and Pattern Recognition. San Francisco, USA: IEEE, 2010: 9-16.
- [14] BOLME D, BEVERIDGE J, DRAPER B, LUI Y. Visual object tracking using adaptive correlation filters[C]//Proceedings of the 2010 IEEE Computer Society Conference

- on Computer Vision and Pattern Recognition. San Francisco, USA: IEEE, 2010: 2544-2550.
- [15] DANELLJAN M, et al. Discriminative scale space tracking[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2017, 39(8): 1561-1575.
- [16] 丁新尧, 张鑫. 基于显著性特征的选择性目标跟踪算法[J]. 电子学报, 2020, 48(1): 118-123.
DING Xin-yao, ZHANG Xin. Visual tracking with salient features and selective mechanism[J]. Acta Electronica Sinica, 2020, 48(1): 118-123. (in Chinese)
- [17] 王治丹, 蒋建国, 齐美彬, 等. 基于最大池图匹配的形变目标跟踪方法[J]. 电子学报, 2017, 45(3): 704-711.
WANG Zhi-dan, JIANG Jian-guo, QI Mei-bin, et al. Deformable object tracking based on max-pooling graph matching[J]. Acta Electronica Sinica, 2017, 45(3): 704-711. (in Chinese)
- [18] WANG N Y, YEUNG D Y. Learning a deep compact image representation for visual tracking[C]//Proceedings of the 26th International Conference on Neural Information Processing Systems. Lake Tahoe, Nevada: Curran Associates Inc, 2013: 809-817.
- [19] ROUT L, MISHRA D, et al. Rotation adaptive visual object tracking with motion consistency[C]//Proceedings of the 2018 IEEE Winter Conference on Applications of Computer Vision(WACV). Lake Tahoe, USA: IEEE, 2018: 1047-1055.
- [20] HUA Y, ALAHARI K, SCHMID C. Online object tracking with proposal selection[C]//Proceedings of the IEEE International Conference on Computer Vision. Santiago, Chile: IEEE, 2015: 3092-3100.
- [21] WANG Q, ZHANG L, BERTINETTO L, et al. Fast online object tracking and segmentation: A unifying approach[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Long Beach, USA: IEEE Press, 2019: 1328-1338.
- [22] PINHEIRO P O, COLLOBERT R, DOLLAR P. Learning to segment object candidates[C]//Proceedings of the Neural Information Processing Systems. Montreal, Canada: ACM, 2015: 1990-1998.
- [23] PINHEIRO P O, LIN T, COLLOBERT R, et al. Learning to refine object segments[C]//Proceedings of the European Conference on Computer Vision. Amsterdam: Springer, 2016: 75-91.
- [24] KRISTAN M, LEONARDIS A, MATAS J, et al. The visual object tracking vot2016 challenge results[C]//Proceedings of the European Conference on Computer Vision. Amsterdam: Springer, 2016: 777-823.
- [25] 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. Munich, Germany: Springer, 2018: 3-53.
- [26] KRISTAN M, LEONARDIS A, MATAS J, et al. The seventh visual object tracking vot2019 challenge results[C]//Proceedings of the IEEE International Conference on Computer Vision Workshop. Seoul, Korea: IEEE, 2019: 2206-2241.
- [27] BERTINETTO L, VALMADRE J, HENRIQUES J F, et al. Fully-convolutional Siamese networks for object tracking[C]//Proceedings of the European Conference on Computer Vision. Amsterdam: Springer, 2016: 850-865.
- [28] 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. Salt Lake, USA: IEEE Press, 2018: 8971-8980.
- [29] LI B, WU W, WANG Q, et al. SiamRPN++: Evolution of Siamese visual tracking with very deep networks[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Long Beach, USA: IEEE Press, 2018: 4277-4286.
- [30] 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. Seattle, USA: IEEE, 2020: 6268-6276.
- [31] CHEN Z, ZHONG B, LI G, et al. Siamese box adaptive network for visual tracking[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Seattle, USA: IEEE, 2020: 6667-6676.
- [32] 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. Salt Lake City, USA: IEEE, 2018: 4834-4843.
- [33] DU Y, LIU P, ZHAO W, et al. Correlation-guided attention for corner detection based visual tracking[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Seattle, USA: IEEE, 2020: 6835-6844.
- [34] REN S, HE K, GIRSHICK R, et al. Faster R-CNN: Towards real-time object detection with region proposal networks[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2017, 39(6): 1137-1149.

- [35] CHEN B, TSOTSOS JOHN K. Fast visual object tracking using ellipse fitting for rotated bounding boxes[C]//2019 IEEE/CVF International Conference on Computer Vision Workshop (ICCVW). Seoul: IEEE, 2019: 2281-2289.
- [36] ZHU Z, WANG Q, LI B, et al. Distractor-aware Siamese networks for visual object tracking[C]//Proceedings of the IEEE European Conference on Computer Vision. Munich, Germany: Springer, 2018:103-119.
- [37] WANG G, LUO C, XIONG Z, et al. SPM-tracker: Series-parallel matching for real-time visual object tracking[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Long Beach, USA: IEEE, 2019: 3638-3647.
- [38] VOIGTLAENDER P, LUITENET J, et al. Siam R-CNN: Visual tracking by re-detection[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Seattle, USA: IEEE, 2020: 6577-6587.
- [39] ZHANG Z, PENG H. Deeper and wider Siamese networks for real-time visual tracking[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Long Beach, USA: IEEE, 2019: 4586-4595.
- [40] LI X, MA C, WU B, et al. Target-aware deep tracking [C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Long Beach, USA: IEEE, 2019:1369-1378.
- [41] DAI K, WANG D, LU H, et al. Visual tracking via adaptive spatially-regularized correlation filters[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Long Beach, USA: IEEE, 2019:4665-4674.
- [42] PARK E, BERG A C. Meta-tracker: Fast and robust online adaptation for visual object trackers[C]//Proceedings of the IEEE European Conference on Computer Vision. Munich, Germany: Springer, Cham, 2018:587-604.
- [43] ZHANG L, GONZALEZ-GARCIA A, WEIJER J, et al. Learning the model update for Siamese trackers[C]//Proceedings of the IEEE International Conference on Computer Vision. Seoul, Korea: IEEE, 2019:4010-4019.
- [44] YANG T, XU P, HU R, et al. ROAM: Recurrently optimizing tracking model[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Seattle, USA: IEEE, 2020: 6717-6726.

作者简介



谢青松 男,1965年1月生于山东青岛,现为山东工商学院计算机科学与技术学院教授、硕士生导师,主要研究方向为智能算法、图像识别、目标跟踪。

E-mail: qs_xie@163.com



刘晓庆 女,1996年5月生于山东青岛,现为山东工商学院信息与电子工程学院硕士研究生,主要研究方向为目标跟踪。

E-mail: 1322699199@qq.com



安志勇(通讯作者) 男,1975年10月生于山西怀仁,工学博士,现为山东工商学院计算机科学与技术学院副教授,研究方向为计算机视觉、目标跟踪等。

E-mail: azytyut@163.com



李博 男,1980年3月生于辽宁锦州,东北大学信息科学与工程学院计算机系统结构博士,研究领域为大数据与机器学习。