

基于DloU损失与平滑约束的结构化SVM目标跟踪方法

孙子文, 袁广林, 李从利, 秦晓燕, 朱虹

引用本文

孙子文, 袁广林, 李从利, 秦晓燕, 朱虹. [基于DloU损失与平滑约束的结构化SVM目标跟踪方法](#)[J]. 计算机科学, 2024, 51(6A): 230700113-8.

SUN Ziwen, YUAN Guanglin, LI Congli, QIN Xiaoyan, ZHU Hong. [Object Tracking of Structured SVM Based on DloU Loss and Smoothness Constraints](#) [J]. Computer Science, 2024, 51(6A): 230700113-8.

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)

[基于跟踪检测时序特征融合的视频遮挡目标分割方法](#)

Occluded Video Instance Segmentation Method Based on Feature Fusion of Tracking and Detection in Time Sequence

计算机科学, 2024, 51(6A): 230600186-6. <https://doi.org/10.11896/jsjcx.230600186>

[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>

[基于自适应搜索范围调整的视觉目标跟踪](#)

Visual Object Tracking Based on Adaptive Search Range Adjustment

计算机科学, 2023, 50(11A): 221000172-6. <https://doi.org/10.11896/jsjcx.221000172>

[融合跟踪器:融合图像特征和事件特征的单目标跟踪框架](#)

Fusion Tracker:Single-object Tracking Framework Fusing Image Features and Event Features

计算机科学, 2023, 50(10): 96-103. <https://doi.org/10.11896/jsjcx.220900075>

[基于改进ByteTrack算法的红外地面多目标跟踪方法](#)

Infrared Ground Multi-object Tracking Method Based on Improved ByteTrack Algorithm

计算机科学, 2023, 50(9): 176-183. <https://doi.org/10.11896/jsjcx.220900004>

基于 DIOU 损失与平滑约束的结构化 SVM 目标跟踪方法

孙子文¹ 袁广林² 李从利¹ 秦晓燕² 朱 虹²¹ 中国人民解放军陆军炮兵防空兵学院兵器工程系 合肥 230031² 中国人民解放军陆军炮兵防空兵学院信息工程系 合肥 230031

(2215738260@qq.com)

摘要 基于结构化 SVM 的目标跟踪因其优良的性能而受到了广泛的关注,但是现有方法存在损失函数不精确和模型漂移问题。针对这两个问题,首先提出基于 DIOU 损失与平滑约束的结构化 SVM 模型。该模型采用 DIOU 函数作为损失函数,利用 t 时刻超平面法向量 w_t 与 $t-1$ 时刻超平面法向量 w_{t-1} 差值的 L_2 范数作为平滑约束。其次基于对偶坐标下降原理设计了该模型的求解算法。最后利用提出的基于 DIOU 损失与平滑约束的结构化 SVM 实现了一种多尺度目标跟踪方法。对所提出的目标跟踪方法在 OTB100 和 VOT-ST2021 数据集上进行了实验验证,实验结果表明:所提出的 Scale-DCSSVM 在 OTB 数据集上的跟踪成功率比 DeepSRDCF 高 1.1 个百分点,在 VOT-ST2021 上的 EAO 比 E. T. Track 高 1.2 个百分点。所提方法具有较优的性能。

关键词: 目标跟踪;结构化 SVM;DIOU 损失;平滑约束

中图分类号 TP391.41

Object Tracking of Structured SVM Based on DIOU Loss and Smoothness Constraints

SUN Ziwen¹, YUAN Guanglin², LI Congli¹, QIN Xiaoyan² and ZHU Hong²¹ Department of Weaponry Engineering of PLA Army Academy of Artillery and Air Defense, Hefei 230031, China² Department of Information Engineering of PLA Army Academy of Artillery and Air Defense, Hefei 230031, China

Abstract Object tracking based on structured support vector machine has been widely concerned because of its excellent performance. However, the existing methods have the problems of imprecise loss function and model drift. To solve these two problems, firstly, a structured SVM model is proposed based on DIOU loss and smoothness constraints. Secondly, DIOU function and L_2 norm of the difference between w_t and w_{t-1} are used respectively as the loss functions and the smoothness constraints in the model. Thirdly, the algorithm for the proposed model is designed with the dual coordinate descent principle. Finally, a multi-scale object tracking method is implemented via the proposed structured SVM on the basis of DIOU loss and smoothness constraints. The proposed object tracking method is experimentally validated on the OTB100 and VOT-ST2021 datasets, and the experimental results show that the tracking success rate of the Scale-DCSSVM on the OTB100 is 1.1% higher than the DeepSRDCF, and the EAO on VOT-ST2021 is 1.2% higher than the E. T. Track. The proposed object tracking method has superior performance.

Keywords Object tracking, Structured SVM, DIOU loss, Smoothness constraints

1 引言

目标跟踪是计算机视觉领域的重要研究内容,其目标是利用图像序列估计目标的状态。目标跟踪在视频监控、车辆导航、人机交互、智能交通等民用领域,以及视觉制导、目标定位和火力控制等军事领域均有重要的应用前景。近年来,虽然目标跟踪取得了较大发展,但是其仍然面临复杂背景、光照变化和目标遮挡等诸多挑战性难题,因此目标跟踪一直是计算机视觉领域的热点问题。

目标跟踪分为生成式跟踪和判别式跟踪两类。生成式跟踪的代表性方法有 IVT 跟踪^[1]、L1 跟踪^[2]和相关滤波跟踪^[3]等。判别式跟踪的代表性方法有 MIL 跟踪^[4]、TLD 跟踪^[5]、随机森林跟踪^[6]、SVM 跟踪^[7]和深度学习跟踪^[8-9]等,

其中基于 SVM 的目标跟踪由于其优异的性能而受到了广泛关注。2004 年, Avidan 首次提出基于 SVM 的目标跟踪方法^[10], 并取得了较好效果。受到结构化 SVM 在目标检测中应用的启发, 2011 年, Hare 等在 ICCV 上首次提出基于结构化 SVM 的目标跟踪方法 Struck^[7], 2015 年该文的扩展^[11]发表在顶级国际期刊《模式分析与机器智能》(PAMI)上。Struck 把目标跟踪看作结构化学习问题, 避免了传统判别式跟踪的中间分类环节, 显著提升了目标跟踪的性能。为了适应目标的变化同时又不丢失目标的时间上下文信息, 2012 年 Yao 等^[12]提出一种加权在线结构化 SVM 跟踪方法, 进一步提升了结构化 SVM 跟踪的性能。为了提高遮挡与形变目标的跟踪性能, 2013 年 Yao 等^[13]又以在线算法 pegasos 为基础, 提出一种在线结构化 SVM 用于目标跟踪。为了解决目

基金项目:安徽省自然科学基金(2008085QF325)

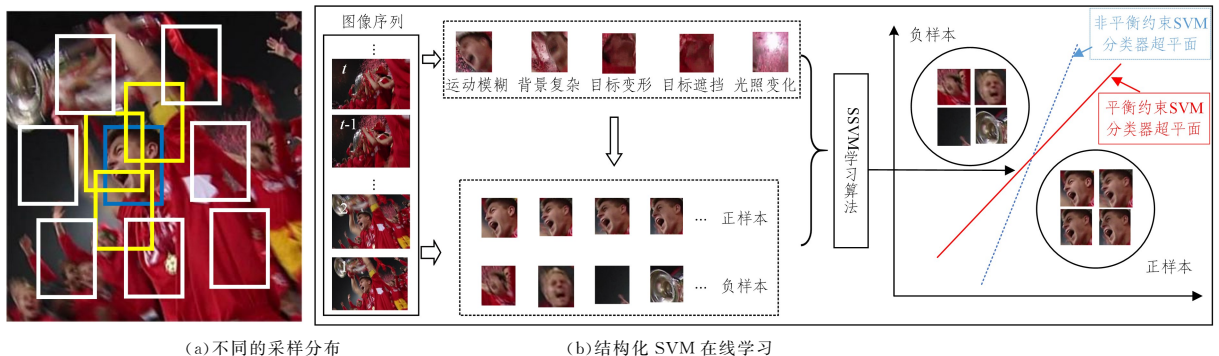
This work was supported by the Natural Science Foundation of Anhui Province, China(2008085QF325).

通信作者:袁广林(yuangl_plus@126.com)

标跟踪中存在的模型漂移问题,2012年 Bai 等^[14]提出一种在线拉普拉斯排序 SVM 跟踪。同样是为了应对模型漂移问题,2014年 Zhang 等^[15]提出了 MEEM 跟踪,该方法以在线 SVM 作为基础跟踪建立专家组跟踪目标,取得了较好的效果。2015年 Hong 等^[16]利用在线 SVM 指导反向传播特定目标的 CNN 特征到输入层,进而建立特定目标的显著图跟踪目标,该方法建立的显著图保持了目标的空间结构,增强了目标跟踪的鲁棒性。2016年 Ning 等^[17]基于对偶坐标下降原理提出一种对偶线性结构化 SVM 目标跟踪方法,该方法在保证目标跟踪鲁棒性的同时提高了目标跟踪的速度。2017年 Wang 等^[18]提出的 LMCF 跟踪利用相关滤波对基于结构化 SVM 的目标跟踪进行了加速。2018年 Ji 等^[19]采用了与 LMCF 类似的思想来加速基于目标部件的结构化 SVM 跟踪。2019年 Zuo 等^[20]提出一种基于离散傅里叶变换的交替优化方法来求解相关滤波器,并设计了多通道支持相关滤波器跟踪目标,进一步提升了基于 SVM 的目标跟踪的性能。2021年 Wang 等^[21]提出一种将 SVM 目标重检测模块融合的算法,提取目标的多特征然后将这些特征矢量融合以增强

目标的特征表达,取得不错的效果。2022年 Qian 等^[22]将现有运动特征关联方法和信号特征辅助关联方法进行改进融合,提出一种基于历史运动特征约束和 SVM 频谱分类的被动声呐目标关联跟踪方法,实验结果较优。

综上所述,基于结构化 SVM 的目标跟踪方法具有较优的跟踪性能,因而受到了广泛的关注,但是现有方法存在两个问题。一方面,现有结构化 SVM 目标跟踪利用 IoU 作为损失函数,如图 1(a)所示,当采样框(黄色框)与目标框(蓝色框)重叠时,利用 IoU 损失可以计算采样框的得分。当采样框(白色框)与目标框(蓝色框)不重叠时,采样样本的 IoU 损失没有差别。但是大部分采样框与目标框中心的距离是不同的, IoU 损失不能描述这种差别,无法体现对结构化 SVM 超平面的影响,因而影响了分类器的性能。另一方面,为了适应目标的变化,现有方法利用跟踪结果更新结构化 SVM。如图 1(b)所示,由于目标遮挡、光照变化、复杂背景、运动模糊和目标变形等,训练样本中会引入非目标信息,从而引起结构化 SVM 超平面偏移(模型漂移),进而导致跟踪失败。



(a)不同的采样分布

(b)结构化 SVM 在线学习

图 1 不同的采样分布与结构化 SVM 在线学习

Fig. 1 Different sampling distributions and structured SVM online learning

针对这两个问题,本文首先提出基于 DIoU 损失^[23]与平滑约束的结构化 SVM 模型,该模型采用 DIoU 函数作为损失函数,利用 t 时刻超平面法向量 w_t 与 $t-1$ 时刻超平面法向量 w_{t-1} 差值的 L2 范数作为平滑约束。其次,基于对偶坐标下降 (Dual Coordinate Descent, DCD) 原理^[24]设计了该模型的求解算法。最后,利用所提出的基于 DIoU 损失与平滑约束的结构化 SVM 实现了一种多尺度目标跟踪方法。采用 OTB100 数据集^[25]对所提出的目标跟踪方法进行了实验验证,并与现有的先进目标跟踪方法进行了比较。实验结果表明,本文提出目标跟踪方法在 OTB 数据集上的跟踪成功率比 DeepSRD-CF 高 1.1 个百分点,在 VOT-ST2021 上的 EAO 比 E. T. Track 高 1.2 个百分点,达到了预期的跟踪效果,与现有优秀的目标跟踪方法相比具有较好的性能。

2 结构化 SVM 和 DIoU 损失函数

2.1 结构化 SVM

受到结构化 SVM 在目标检测中应用的启发,2011年 Hare 等^[27]提出了基于结构化 SVM 的目标跟踪 Struck。Struck 把目标跟踪作为结构化学习问题,避免了传统判别式跟踪的中间分类环节,显著提升了目标跟踪的性能。在利用结构化 SVM 进行预测时,其目标是预测给定样本 $x \in \mathbf{R}^d$ 的结构化输出 $y \in \mathbf{Y}$,其中 \mathbf{Y} 可以是任意结构输出空间。在基于结

构化 SVM 的目标跟踪中, \mathbf{Y} 是矩形框空间,它的任一元素用 (x, y, w, h) 表示,其中 (x, y) 表示矩形框的中心位置, w 和 h 分别表示矩形框的宽和高。假设训练数据为 $\{(x_i, y_i)\}_{i=1}^N$, $\Phi(x, y)$ 表示矩形框 y 在图像 x 上的特征向量,则可以通过式(1)学习一个参数为 w 的分类器。

$$y^* = \arg \max_{y \in \mathbf{Y}} f(x, y, w) \quad (1)$$

其中, $f(x, y, w) = w^T \Phi(x, y)$, w 可以通过求解如下优化问题得到。

$$\begin{aligned} \min_w \frac{1}{2} \|w\|^2 + C \sum_{i=1}^N \xi_i \\ \text{s. t. } \forall i, \xi_i \geq 0 \\ \forall i, \forall y \neq y_i: \langle w, \Psi_i(y) \rangle \geq L(y_i, y) - \xi_i \end{aligned} \quad (2)$$

其中, $\Psi_i(y) = \Phi(x_i, y_i) - \Phi(x_i, y)$; C 是正则化参数; ξ_i 是松弛变量; $L(y_i, y)$ 是损失函数,表示预测输出矩形框 y 的结构误差损失。 $L(y_i, y)$ 定义为样本矩形框 y 和目标矩形框 y_i 的重合率,如式(3)所示。

$$L(y_i, y) = 1 - s(y_i, y) \quad (3)$$

其中,

$$s(y_i, y) = \frac{y_i \cap y}{y_i \cup y} \quad (4)$$

2.2 DIoU 损失函数

现有结构化 SVM 目标跟踪使用 IoU 作为损失函数, IoU

损失函数的形式如式(5)所示:

$$L_{IoU} = 1 - \frac{\mathbf{B} \cap \mathbf{B}^{gt}}{\mathbf{B} \cup \mathbf{B}^{gt}} \quad (5)$$

其中, $\mathbf{B}^{gt} = (x^{gt}, y^{gt}, w^{gt}, h^{gt})$ 是目标框的位置, $\mathbf{B} = (x, y, w, h)$ 是预测框的位置。

IoU 损失只能在采样框和目标框重叠时对样本有区分性。无论是目标跟踪还是目标检测,大部分采样框与目标框是没有重叠的,但是采样框和目标框中心的距离是不同的,即这些采样框对分类器的学习所起到的作用不同。为了解决这个问题,Zhen 等^[23]提出用于目标检测的 DIoU 损失函数,其定义如式(6)所示:

$$L_{DIoU} = 1 - IoU + \frac{\rho^2(b, b^{gt})}{c^2} \quad (6)$$

其中, b 和 b^{gt} 表示 \mathbf{B} 和 \mathbf{B}^{gt} 的中心点, $\rho(\cdot)$ 是欧几里得距离, c 是覆盖这两个位置框的最小包围框的对角线长度。

基于上述分析,可以将结构化 SVM 目标跟踪中的 IoU 损失函数改进为 DIoU 损失函数,如式(7)所示:

$$L(y_i, y) = 1 - s(y_i, y) + \frac{\rho^2(b, b^{gt})}{c^2} \quad (7)$$

3 基于 DIoU 损失与平滑约束的结构化 SVM

3.1 基于 DIoU 损失与平滑约束的结构化 SVM 模型

由引言分析可知,针对列举的问题,本文提出一种基于 DIoU 损失与平滑约束的结构化 SVM 的模型(Structured SVM Based on DIoU Loss and Smoothness Constraints, DCSSVM),该模型描述如下:

$$\min_{\mathbf{w}} \frac{1}{2} \|\mathbf{w}\|^2 + \lambda \|\mathbf{w} - \mathbf{w}_{t-1}\|^2 + C \sum_{y \neq y_i} \xi_i \quad (8)$$

$$s. t \quad \forall i, \xi_i \geq 0$$

$$\forall i, \forall y \neq y_i: \langle \mathbf{w}, \Psi_i(\mathbf{y}) \rangle \geq L(y_i, y) - \xi_i$$

其中, \mathbf{w} 为 t 时刻结构化 SVM 的法向量, \mathbf{w}_{t-1} 为 $t-1$ 时刻结构化 SVM 的法向量, λ 为平滑约束系数。 C, ξ_i 和 $\Psi_i(\mathbf{y})$ 的含义如式(2)所示; $b, b^{gt}, \rho(\cdot)$ 和 c 的含义如(6)式所示; $L(y_i, y)$ 的含义如式(7)所示。

3.2 基于 DIoU 损失与平滑约束的结构化 SVM 算法

为了求解式(8),使用拉格朗日乘子法可得到其“对偶问题”,为此引入拉格朗日乘子 α_i^y 和 β_i^y ,满足如下条件:

$$\forall i, \forall y \neq y_i: \alpha_i^y \geq 0, \beta_i^y \geq 0 \quad (9)$$

则式(8)的拉格朗日函数如下:

$$L(\mathbf{w}, \xi, \alpha, \beta) = \frac{1}{2} \|\mathbf{w}\|^2 + \lambda \|\mathbf{w} - \mathbf{w}_{t-1}\|^2 + C \sum_{i=1} \xi_i - \sum_{i, y \neq y_i} \alpha_i^y (\langle \mathbf{w}, \Psi_i(\mathbf{y}) \rangle + \xi_i - L(y_i, y)) - \sum_{i, y \neq y_i} \beta_i^y \xi_i \quad (10)$$

将拉格朗日函数 $L(\mathbf{w}, \xi, \alpha, \beta)$ 分别对 \mathbf{w} 和 ξ_i 求偏导并令其为 0 可得:

$$\mathbf{w} = \frac{\sum_{i, y \neq y_i} \alpha_i^y \Psi_i(\mathbf{y}) + 2\lambda \mathbf{w}_{t-1}}{1 + 2\lambda} \quad (11)$$

$$C - \alpha_i^y - \beta_i^y = 0 \quad (12)$$

将式(11)和式(12)代入式(10),可以将 $L(\mathbf{w}, \xi, \alpha, \beta)$ 中的 \mathbf{w}, β 和 ξ 消去,得到式(8)的对偶问题,如(13a)一式(13c)式所示:

$$\min_{\alpha \geq 0} f(\alpha) = \frac{1}{2(1+2\lambda)} \left\| \sum_{i, y \neq y_i} \alpha_i^y \Psi_i(\mathbf{y}) \right\|^2 -$$

$$\frac{\lambda \mathbf{w}_{t-1}^2 - 4\lambda \mathbf{w}_{t-1} \sum_{i, y \neq y_i} \alpha_i^y \Psi_i(\mathbf{y})}{2(1+2\lambda)} - \sum_{i, y \neq y_i} L(y_i, y) \alpha_i^y \quad (13a)$$

$$s. t \quad \forall i, \forall y \neq y_i: \alpha_i^y \geq 0 \quad (13b)$$

$$\forall i: \sum_{y \neq y_i} \alpha_i^y \leq C \quad (13c)$$

本文基于坐标优化原理^[24]设计模型式(13a)一式(13c)的求解算法。由文献[24]可知,对偶坐标下降优化算法每次利用式(14)从训练集中选择一个训练样本 k ,然后利用式(15)更新其对偶标量 α_k^y 。

$$\mathbf{y}_k^* = \arg \max_{y \in \mathbf{y}_k} L(\mathbf{y}, \mathbf{y}_k) - \mathbf{w}^T \Psi_k(\mathbf{y}) \quad (14)$$

$$\alpha_k^{y^* (new)} = \alpha_k^{y^* (old)} + \Delta \alpha_k^{y^*} \quad (15)$$

解决问题的关键是如何求得式(15)中的 $\Delta \alpha_k^{y^*}$ 。为此,首先将式(15)代入式(13a)中,然后将其转化为关于 $\Delta \alpha_k^{y^*}$ 的函数,这样得到式(16)。

$$L(\Delta \alpha_k^{y^*}) = \frac{1}{2(1+2\lambda)} (\alpha_k^{y^* (old)} + \Delta \alpha_k^{y^*})^2 \|\Psi_k(\mathbf{y}^*)\|^2 + \frac{1}{1+2\lambda} \Delta \alpha_k^{y^*} \sum_{\substack{i, i \neq k \\ y \neq y_i \\ y \neq y^*}} \alpha_i^y \Psi_i(\mathbf{y}^*) \Psi_k(\mathbf{y}^*) + \frac{2\lambda \mathbf{w}_{t-1}}{1+2\lambda} \Delta \alpha_k^{y^*} \Psi_k(\mathbf{y}) - \Delta \alpha_k^{y^*} L(\mathbf{y}_k, \mathbf{y}^*) + c \quad (16)$$

其中, c 是一个与增量 $\Delta \alpha_k^{y^*}$ 无关的常量。将(16)式对 $\Delta \alpha_k^{y^*}$ 求导并令其为 0 可得:

$$\Delta \alpha_k^{y^*} = \frac{L(\mathbf{y}_k, \mathbf{y}^*) - \mathbf{w}^T \Psi_k(\mathbf{y}^*)}{\|\Psi_k(\mathbf{y}^*)\|^2} (1+2\lambda) \quad (17)$$

根据式(13c)的约束条件,可得 $\Delta \alpha_k^{y^*}$ 的取值范围为:

$$\Delta \alpha_k^{y^*} \in [-\alpha_k^{y^* (old)}, C - \sum_y \alpha_k^y] \quad (18)$$

利用式(17)和式(18)可以得到 $\Delta \alpha_k^{y^*}$,然后结合式(11)即可得到 \mathbf{w} 的更新公式,如式(19)所示:

$$\mathbf{w}^{(new)} = \mathbf{w}^{(old)} + \frac{\Delta \alpha_k^{y^*} \Psi_k(\mathbf{y}^*)}{1+2\lambda} \quad (19)$$

在基于结构化 SVM 的目标跟踪中,随着时间的推移,结构化 SVM 中的支持向量的数量不断增加,为了保证目标跟踪的效率,需要固定支持向量的数目。为此,当结构化 SVM 中模式数超出预算时,根据(20)式选择一个支持向量并删除。本文提出的算法采用这一策略。

$$\alpha^* = \arg \min_{\alpha_i^y \in \alpha} \|\alpha_i^y \Psi_i(\mathbf{x}_t, \mathbf{y})\|^2 \quad (20)$$

综上所述,本文提出的 DCSSVM 学习算法如算法 1 所示。根据 DCSSVM 学习算法可知, t 时刻的训练样本 $\{(x_t^i, y_t^i)\}_{i=1}^N$, t 时刻的模式集 $\mathbf{P}_t = \mathbf{P}_{t-1} \cup \{(x_t^i, y_t^i)\}_{i=1}^N$, 设定初始化时的模式集是空集合。样本初始化是从样本候选区域的左上角区域开始采样,然后按照从上到下、从左到右顺序共采样 1764 个样本。候选样本的采集过程以 basketball 数据为例,尺寸大小为 $432 * 576 * 3$ 的原始图片如图 2(a)所示,图 2(b)蓝色框表示尺寸大小为 $64 * 27 * 3$ 的第一帧目标真值框,图 2(c)大红色框表示尺寸大小为 $148 * 111 * 3$ 的样本候选区域,小红色框与目标真值框的尺寸大小相同,为第一个采样样本。初始化样本采样的过程如图 2(d)一 2(f)所示。

算法 1 DCSSVM 学习算法

输入: $t-1$ 时刻 DCSSVM 的参数 \mathbf{w}_{t-1} 和模式集 \mathbf{P}_{t-1} , t 时刻的训练样

本 $\{(x_t^i, y_t^i)\}_{i=1}^N$

输出: t 时刻 DCSSVM 的参数 \mathbf{w}_t 和模式集 \mathbf{P}_t

1. $\mathbf{P}_t = \mathbf{P}_{t-1} \cup \{(x_i^j, y_i^j)\}_{i=1}^N$.
For $j=1:n_1$
2. 计算模式集 \mathbf{P}_t 中的模式数 n .
3. 令 $i=n-\lfloor(j-1) * n/n_1\rfloor$.
4. 利用式(14)从模式 \mathbf{P}_t^i 中选择选择样本 y_i^* .
5. 利用式(15)更新 y_i^* 的对偶变量 $\alpha_i^{y^*}$.
6. 利用式(19)更新 w_{t-1} .
7. 如果模式集 \mathbf{P}_t 中的支持向量数目大于预算数目, 则用式(20)更新模式集 \mathbf{P}_t .
8. 计算模式集 \mathbf{P}_t 中的模式数 n ;
For $p=1:n_2$
 令 $i=n-\lfloor(p-1) * n/n_2\rfloor$.
9. 利用式(14)从模式 \mathbf{P}_t^i 中选择选择样本 y_i^* .
10. 利用式(15)更新 y_i^* 的对偶变量 $\alpha_i^{y^*}$.
11. 利用(19)式更新 w_{t-1} .
12. End For
13. End For
14. 令 $w_t = w_{t-1}$.

说明: 算法 1 中 n_1 和 n_2 是外部循环和内循环的迭代次数, 本文分别取 5 和 10。

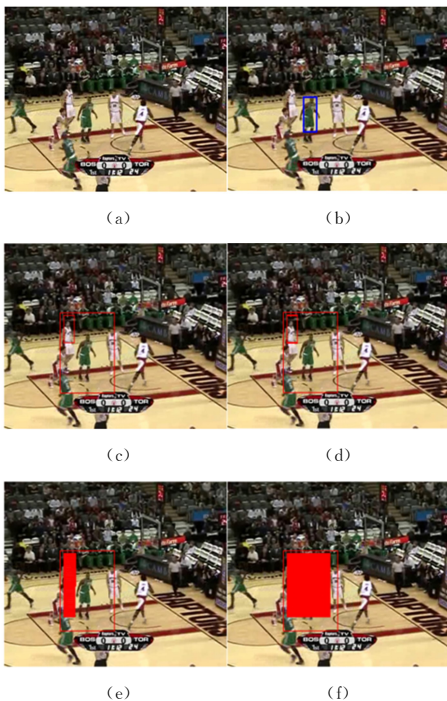


图 2 候选样本采集过程

Fig. 2 Candidate sample collection process

4 目标跟踪方法

本文使用网格搜索生成候选样本, 搜索区域的大小由跟踪目标的大小自适应确定。在学习到 DCSSVM 的超平面法向量 w 后, 利用内积运算计算候选样本的得分, 根据最大得分准则式(21)估计目标的状态。

$$y^* = \arg \max_{y \in Y} w^T \Psi_t(y) \quad (21)$$

对于目标特征, 本文选择目标的 Lab 颜色和局部秩变换 LRT 特征, LRT 特征的计算方法同文献[26]。对于多尺度目标跟踪, 本文参照文献[17]中的多尺度估计方法提出 Scale-DCSSVM, 尺度参数 $S_p = \{1, 0.995, 1.005\}$, 即在 3 种不同尺度图像上分别跟踪目标, 以最大得分作为跟踪结果。

在对目标提取 Lab 颜色和局部秩变换 LRT 特征时, 并不是对整张输入原始图片提取特征, 而是对裁剪后的搜索区域提取特征。以 basketball 数据为例, 如图 3 所示: 图 3(a) 为大小 $432 * 576 * 3$ 的原始图片; 图 3(b) 为裁剪之后大小为 $148 * 111 * 3$ 的候选区域; 图 3(c) 为裁剪后搜索区域 Lab 颜色特征图, 大小为 $148 * 111 * 3$; 图 3(d) 为裁剪后搜索区域 LRT 特征图, 大小 $148 * 111 * 16$ 。将其每一维度展开可以得到相应的特征图。

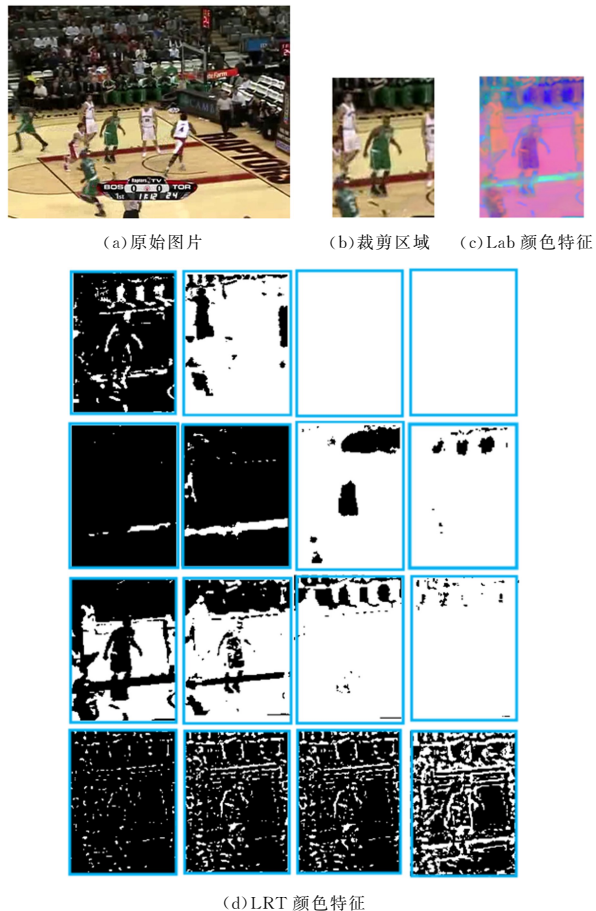


图 3 目标的 Lab 颜色和局部秩变换 LRT 特征可视化图

Fig. 3 Visualization of target's Lab color and local rank transform (LRT) feature

基于上面的分析, 本文提出的目标跟踪方法描述如算法 2 所示。

算法 2 DCSSVM 的目标跟踪方法

输入: 序列图像 $\mathbf{I} = \{I_1, I_2, \dots, I_T\}$ 和目标状态 s_1

输出: 每一帧的跟踪结果 $s_i (i=1, 2, \dots, T)$

1. 根据当前目标状态 s_1 , 采样 N 个训练样本 $\{(x_i^j, y_i^j)\}_{i=1}^N$.
2. 初始 $w_0 = 0$ 和 $P_0 = 0$.
3. 调用算法 1 初始化 $t=1$ 时刻 DCSSVM 的参数 w_1 和模式 P_1 .
For ($t=2:T$)
4. 利用滑动窗口在搜索区域中采样测试样本 $\{(x_i^j, y_i^j)\}_{i=1}^M, X_t = [x_i^j]_{i=1}^M, Y_t = [y_i^j]_{i=1}^M$.
5. 利用最大化得分准则估计目标当前状态:
$$s_t = \arg \max_{y \in Y_t} (w_t^T X_t)$$
6. 根据当前目标状态 s_t , 采样 N 个训练样本 $\{(x_i^j, y_i^j)\}_{i=1}^N$.
7. 调用算法 1 更新 DCSSVM 的 w_t .
8. End For

5 实验

5.1 实验设置

5.1.1 参数设置

本文提出的目标跟踪方法中有一些需要设置的参数,实验中这些参数固定不变。惩罚系数 C 的值为 100,支持向量的预算设为 100,训练和检测过程的搜索半径由目标区域的平方根自动确定。约束系数 λ 取 0.16。多尺度估计参数设置为 $Scale = \{1, 0.995, 1.005\}$ 。

5.1.2 数据集与评价指标

利用 OTB100^[25] 数据集和基于锚点的短期跟踪 VOT-ST2021 数据集^[16] 对本文提出的目标跟踪方法进行实验验证。其中 OTB100 数据集包含 100 个视频序列,这些视频序列具有 11 个挑战性属性,分别是光照变化(IV)、尺度变化(SV)、遮挡(OCC)、形变(DEF)、运动模糊(MB)、快速运动(FM)、平面内旋转(IPR)、平面外旋转(OPR)、目标出界(OV)、背景混杂(BC)和低分辨率(LR)。文献^[25]提出两种评价目标跟踪方法性能的指标。一种是 precision(精确率)指标,另外一种 success(成功率)指标。相关的具体计算方法参考文献^[25]。

基于短期跟踪数据集 VOT-ST2021^[16] 上的评价指标为期望平均重叠 EAO(Expected Average Overlap)、精度 Accuracy 和鲁棒性 Robustness。精度表示真值框与预测目标框之间的平均重叠的加权组合;鲁棒性表示跟踪器平均失败前的帧数百分比;EAO 是一种结合了精度和鲁棒性之后衡量整体跟踪性能的指标。具体计算方法参考文献^[27]。

5.1.3 实验环境

在 DELL XPS8930 台式机(CPU 型号为 Intel 酷睿 i7 8700K,内存为 16GB)上,使用 Matlab 和 OpenCV 实现了本文提出的跟踪方法,其中 Matlab 版本为 R2019a,OpenCV 版本为 2.4.8。

5.2 实验结果及分析

5.2.1 消融实验

为了验证本文提出的平滑约束 SSVM、DIoU 损失函数以及多尺度估计对跟踪性能的影响,本文在 OTB100 数据集上用不同版本的 DCSSVM 进行了 OPE(One pass evaluation)性能测试。本文设定非约束 SSVM 目标跟踪方法为 DCSSVM-NW,无 DIoU 损失函数目标跟踪方法为 DCSSVM-ND,两者都没有的目标跟踪方法为 DCSSVM-N2,两者都有的目标跟踪方法为 DCSSVM,有多尺度估计的目标跟踪方法为 Scale-DCSSVM。实验得到的各跟踪方法的结果如表 1 所列,表中方法的跟踪速度 FPS(Frame Per Second)是根据 OTB100 中最长的一个视频 doll(3872 帧)计算得到的。

由表 1 可以看出,Scale-DCSSVM 由于在标准方法的基础上进行了多尺度的估计,因此在 OPE 性能评估中的跟踪性能最好,但是速度较慢。DCSSVM 在 OPE 性能评估中的跟踪性能位居第二,但是速度相比 Scale-DCSSVM 提升明显。DCSSVM-ND 由于没有 DIoU 损失函数,无法对没有重叠的采样框和目标框进行区分,因此跟踪性能相较于 DCSSVM 略有下降。DCSSVM-NW 由于没有对 w 进行约束,导致跟踪方法对某一帧中的噪声较为敏感,在背景混杂、物体相似和光照变化等挑战性视频中容易发生漂移现象,因此跟踪性能相

于 DCSSVM 下降明显。DCSSVM-N2 由于既没有 DIoU 损失函数也没有对 w 进行约束,因此其跟踪性能结果排在最后。

表 1 不同版本的 DCSSVM 在 OTB100 数据集上的 OPE 性能与速度指标

Table 1 OPE performance and speed indicators of different versions of DCSSVM on OTB100 dataset

Methods	OPE		Speed/ FPS
	precision (pixels)	Success (AUC)	
DCSSVM-N2	0.766	0.541	10.36
DCSSVM-NW	0.788	0.569	10.33
DCSSVM-ND	0.801	0.584	10.35
DCSSVM(本文方法)	0.819	0.605	10.29
Scale-DCSSVM(本文方法)	0.862	0.649	5.31

5.2.2 与其他结构化 SVM 跟踪方法的比较

表 2 列出了 Scale-DCSSVM, DCSSVM, Scale-DLSSVM^[17], DLSSVM^[17], Struck^[11] 和 LMCF^[18] 这 6 种结构化 SVM 跟踪方法在 OTB100 数据集上的 OPE 性能比较结果。在跟踪速度上,本文选择一个长视频 liquor(1741 帧)进行评估。从表 2 可以看出,在 Precision 和 Success 两项指标上,本文提出的 Scale-DCSSVM 均优于其他跟踪方法。由表 2 中的比较结果可以看出,单尺度跟踪方法 DCSSVM 相较于 DLSSVM 跟踪方法在 Precision 上提高 5.3%,在 Success 上提高 6.4%。多尺度跟踪方法 Scale-DCSSVM 相较于 Scale-DLSSVM 跟踪方法在 Precision 上提高 5.6%,在 Success 上提高 8.6%。本文方法不仅在性能上有明显提升,而且目标跟踪速度几乎没有受到影响。

表 2 6 种结构化 SVM 跟踪方法在 OTB100 数据集上的 OPE 性能与速度比较

Table 2 Comparison of OPE performance and speed of six structured SVM tracking methods on OTB100 dataset

Methods	OPE		Speed/ FPS
	precision (pixels)	Success (AUC)	
Scale-DCSSVM(本文方法)	0.862	0.649	5.31
DCSSVM(本文方法)	0.819	0.605	10.29
Scale-DLSSVM ^[17]	0.806	0.563	5.39
DLSSVM ^[17]	0.766	0.541	10.36
Struck ^[11]	0.634	0.457	0.92
LMCF ^[18]	0.785	0.574	85.26

表 3 列出了 6 种结构化 SVM 跟踪方法在不同挑战性视频中 Success 的结果。从表中可以看出,本文 Scale-DCSSVM 跟踪方法在 BC(背景混杂)、FM(快速移动)、IV(光照变化)、MB(运动模糊)、DEF(目标形变)和 OCC(目标遮挡)等挑战性视频中表现较优,这一结果验证了本文方法的有效性。

表 3 6 种结构化 SVM 跟踪方法在不同属性视频中 Success 的比较

Table 3 Comparison of Success of six structured SVM tracking methods in different attribute videos

Methods	BC	FM	IV	MB	DEF	OCC
Scale-DCSSVM(本文方法)	0.642	0.647	0.633	0.626	0.663	0.609
DCSSVM(本文方法)	0.618	0.613	0.605	0.582	0.636	0.565
Scale-DLSSVM ^[17]	0.581	0.571	0.567	0.556	0.599	0.527
DLSSVM ^[17]	0.535	0.553	0.540	0.535	0.583	0.508
Struck ^[11]	0.416	0.382	0.408	0.376	0.393	0.369
LMCF ^[18]	0.606	0.590	0.602	0.546	0.623	0.542

图 4 给出了 6 种结构化 SVM 跟踪方法在背景混杂、快速移动、光照变化、运动模糊、目标形变和目标遮挡这 6 种代表性的挑战性视频中的可视化跟踪结果。其中 Basketball 视频中的主要挑战属性是快速移动、目标遮挡; Board 视频中的主要挑战属性是背景混杂; Diving 视频中的主要挑战属性是目标形变; DragonBaby 视频中的主要挑战属性是快速移动、目标形变; Matrix 视频中的主要挑战属性是运动模糊、目标遮挡、光照变化; Ironman 视频中的主要挑战属性是背景混杂、

光照变化、快速移动。由不同属性视频中的跟踪结果可以看出,相比其他跟踪方法,本文提出的跟踪方法能够有效提升在以上属性视频中的跟踪性能。但是其不足之处在于特征选择方面,本文方法采用的是传统的 LRT 和颜色特征,在特征表达上与现有基于卷积神经网络学习到的深度特征相差较多,因此在 Diving 视频中,当跟踪目标人物的肢体发生大规模形变时,本文方法并不能完全跟踪目标,这是下阶段要解决的重点问题。



图 4 6 种基于结构化 SVM 的跟踪方法在 6 个挑战性视频中的跟踪结果

Fig. 4 Tracking results of six structured SVM-based tracking methods on six challenging videos

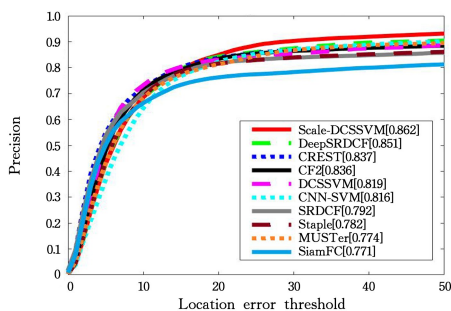
5.2.3 与其他高性能跟踪方法的比较

在 OTB100 数据集上,将本文提出的目标跟踪方法与 8 种优秀的目标跟踪方法进行比较,分别是基于深度学习

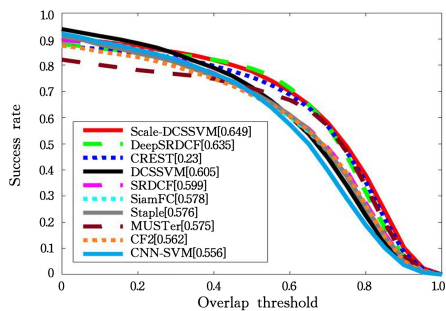
的 CNN-SVM^[28], SiamFC^[29], CREST^[30] 和 DeepSRDCF^[31], 基于相关滤波的 CF2^[32] MUSTer^[33], SRDCF^[34] 和 Staple^[35]。图 5 给出了 10 种跟踪方法在 OTB100 数据集

上的 OPE 性能曲线。从图 5 中的结果可以看出,本文提出的 Scale-DCSSVM 跟踪方法在 Precision 和 Success 两个指标上比 DeepSRDCF 分别高出 1.1% 和 1.4%,处于最高的水平。

如表 4 所列,本文多尺度方法 Scale-DCSSVM 与近两年



(a) Precision plots of OPE



(b) Success plots of OPE

图 5 10 种高性能跟踪器在 OTB100 数据集上的 OPE 性能指标曲线

Fig. 5 OPE performance metric curves of ten high-performance trackers on OTB100 dataset

表 4 本文方法与 5 种高性能跟踪方法在 VOT-ST2021 数据集上实验结果

Table 4 Experimental results of the proposed methods and five high-performance tracking methods on VOT-ST2021 dataset

	Ours	SiamRPN++ ^[36]	D3S ^[37]	TADT ^[38]	LT-Mobile ^[39]	E. T. Track ^[40]
EAO \uparrow	0.238	0.257	0.234	0.192	0.211	0.226
Accuracy \uparrow	0.451	0.472	0.442	0.403	0.407	0.414
Robustness \downarrow	0.631	0.668	0.626	0.675	0.714	0.752

结束语 本文分析了现有基于结构化 SVM 的目标跟踪存在的两个问题,即损失函数不精确和模型漂移问题。针对这两个问题,首先提出一种基于 DIoU 损失与平滑约束的结构化 SVM 模型 DCSSVM,其次基于对偶坐标优化原理设计了其求解算法,最后利用提出的基于 DIoU 损失与平滑约束的结构化 SVM 算法,实现了一种多尺度目标跟踪方法 Scale-DCSSVM。采用 OTB100 和 VOT2021 数据集对本文提出的目标跟踪方法进行实验验证和分析,本文方法能够有效应对快速运动、目标遮挡、光照变化、背景混杂等易导致模型漂移而导致跟踪失效的场景。与相关滤波方法相比,本文方法跟踪精度较高,与深度学习方法相比,本文方法具有速度方面的优势。实验结果证明了本文方法的有效性和先进性。

参考文献

[1] ROSS D A, LIM J, LINR S, et al. Incremental Learning for Robust Visual Tracking[J]. International Journal of Computer Vision, 2008, 77(1/2/3): 125-141.

[2] MEI X, LING H. Robust visual tracking using L1 minimization [C] // Proceedings of IEEE Conference on Computer Vision. Kyoto, Japan: IEEE Computer Society Press, 2009: 1436-1443.

[3] SUN J, XIANG W, TAN S K, et al. Improved Kernel Correlation Filter Tracking Algorithm[J]. Computer Engineering and Applications, 2018, 54(9): 178-182.

[4] BABENKO B, YANG M H, BELONGIE S. Robust object tracking with online multiple instance learning[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2011, 33(8): 1619-1632.

[5] KALAL Z, MIKOLAJCZYK K, MATAS J. Tracking-learning-detection[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2012, 34(7): 1409-1422.

[6] SAFFARI A, LEISTNER C, SANTNER J, et al. Online random forests[C] // Proceedings of the IEEE Conference on Computer

高性能的深度学习方法 SiamRPN++^[36], D3S^[37], TADT^[38], LT-Mobile^[39] 和 E. T. Track^[40] 共 6 种算法在 VOT-ST2021 数据集上进行比较。由表 4 可知,本文多尺度方法 Scale-DCSSVM 在 EAO 指标上略低于 SiamRPN++,但是相较于其他高性能方法来说在性能上仍然有着优势。

Vision. Kyoto, Japan: IEEE Computer Society Press, 2009: 1393-1400.

[7] HARE S, SAFFARI A, TORR P H S. Structured output tracking with kernels[C] // Proceedings of IEEE International Conference on Computer Vision. Barcelona, Spain: IEEE Computer Society Press, 2011: 263-270.

[8] WANG N, YEUNG D Y. Learning a Deep Compact Image Representation for Visual Tracking[C] // Proceedings of Neural Information Processing Systems. Curran Associates: IEEE Computer Society Press, 2013: 809-817.

[9] WEI Y Q, YANG X J. Kernel Correlation Filtering for Visual Tracking of Deep Features[J]. Computer Engineering and Applications, 2020, 56(15): 147-152.

[10] AVIDAN S. Support vector tracking[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2004, 26(8): 1064-1072.

[11] HARE S, SAFFARI A, TORR P H S. Structured output tracking with kernels[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2016, 38(10): 2096-2109.

[12] YAO R, SHI Q, SHEN C, et al. Robust Tracking with Weighted Online Structured Learning[C] // Proceedings of the European conference on Computer Vision. Berlin Heidelberg: Springer, 2012, Part III: 158-172.

[13] YAO R, SHI Q, SHEN C, et al. Part-based Visual Tracking with Online Latent Structural Learning [C] // Proceedings of Computer Vision and Pattern Recognition. Portland, USA: IEEE Computer Society Press, 2013: 2363-2370.

[14] BAI Y, TANG M. Robust Tracking via Weakly Supervised Ranking SVM[C] // Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. Providence, RI, USA: IEEE Computer Society Press, 2012: 1854-1861.

[15] ZHANG J, MA S, SCLAROFF S. Meem: Robust tracking via multiple experts using entropy minimization[C] // Proceedings of European Conference on Computer Vision. Zurich, Switzer-

- land; IEEE Computer Society Press, 2014; 188-203.
- [16] HONG S, YOU T, KWAK S, et al. Online Tracking by Learning Discriminative Saliency Map with Convolutional Neural Network[C]//Proceedings of International Conference on Machine Learning, Lille, France; IEEE Computer Society Press, 2015: 597-606.
- [17] NING J F, YANG J M, JIANG S J, et al. Object tracking via dual linear structured SVM and explicit feature map [C] // Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas; IEEE Computer Society Press, 2016; 4266-4274.
- [18] WANG M M, LIU Y, HUANG Z Y. Large Margin Object Tracking with Circulant Feature Maps [C] // Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, USA, 2017; 4800-4808.
- [19] JI Z J, FENG K, QIANY H. Part-based Visual Tracking via Structural Support Correlation Filter [J]. arXiv: 1805. 09971, 2018.
- [20] ZUO W, WU X, LIN L, et al. Learning Support Correlation Filters for Visual Tracking [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2019, 41(5): 1158-1172.
- [21] WANG C Y, WANG S Q, ZHANG L J, et al. A Multi feature DSST Target Tracking Algorithm Based on SVM Fusion [J]. Ordnance Industry Automation, 2021, 40(7): 39-45, 66.
- [22] QIAN Y N, CHEN Y W, LI G. Passive sonar target association tracking method based on historical motion feature constraints and SVM spectrum classification [J]. Journal of Electronics and Information Technology, 2023, 45(8): 2991-3001.
- [23] ZHEN Z H, WANG P, LIU W, et al. Distance-IoU Loss; Faster and Better Learning for Bounding Box Regression [C] // Proceedings of the AAAI Innovative Applications of Artificial Intelligence Conference, New York, USA, 2020; 12993-13000.
- [24] RAMANAN D. Dual coordinate solvers for large-scale structural svms [J]. arXiv: 1312. 1743v2, 2014.
- [25] WU Y, LIM J, YANG M H. Object tracking benchmark [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2015, 37(9): 1834-1848.
- [26] ZABIH R, WOODFILL J. Non-parametric Local Transforms for Computing Visual Correspondence [C] // Proceedings of the European Conference on Computer Vision, Stockholm, Sweden. IEEE Computer Society Press, 1994; 151-158.
- [27] KRISTAN M, MATAS J, LEONARDIS A, et al. The Ninth Visual Object Tracking [C] // Proceedings of the (VOT2021) Challenge Results International Conference on Computer Vision Workshops, Montreal, BC, Canada, 2021; 2711-2738.
- [28] HONG S, YOU T, KWAK S, et al. Online Tracking by Learning Discriminative Saliency Map with Convolutional Neural Network [C] // Proceedings of the International Conference on Machine Learning, Lille; IEEE Computer Society Press, 2015; 597-606.
- [29] 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, The Netherlands; IEEE Computer Society Press, 2016; 850-865.
- [30] SONG Y B, MA C, GONG L J, et al. Convolutional Residual Learning for Visual Tracking [C] // Proceedings of the IEEE International Conference on Computer Vision, Venice, Italy; IEEE Computer Society Press, 2017; 2574-2583.
- [31] DANELLJAN M, HAGER G, KHAN F S, et al. Convolutional Features for Correlation Filter Based Visual Tracking [C] // Proceedings of the IEEE International Conference on Computer Vision, Santiago; IEEE Computer Society Press, 2015; 621-629.
- [32] MA C, HUANG J B, YANG X K, et al. Hierarchical Convolutional Features for Visual Tracking [C] // Proceedings of the IEEE International Conference on Computer Vision, Santiago, Chile; IEEE Computer Society Press, 2015; 3074-3082.
- [33] HONG Z B, CHEN Z, WANG C H, et al. Multi-Store Tracker: A Cognitive Psychology Inspired Approach to Object Tracking [C] // Proceedings of the International Conference on Computer Vision and Pattern Recognition, Boston, USA; IEEE Computer Society Press, 2015; 749-758.
- [34] DANELLJAN M, GUSTAV H, KHAN F S, et al. Learning Spatially Regularized Correlation Filters for Visual Tracking [C] // Proceedings of the IEEE International Conference on Computer Vision, Santiago; IEEE Computer Society Press, 2015; 4310-4318.
- [35] BERTINETTO L, VALMADRE J, GOLODETZ S, et al. Staple; Complementary Learners for Real-Time Tracking [C] // Proceedings of the International Conference on Computer Vision and Pattern Recognition, Las Vegas, NV; IEEE Computer Society Press, 2016; 1401-1409.
- [36] LI B, WU W, WANG Q, et al. SiamRPN++: Evolution of Siamese Visual Tracking With Very Deep Networks [C] // Proceedings of the Conference on Computer Vision and Pattern Recognition, Long Beach, CA, USA, 2019; 4282-4291.
- [37] LUKEZIC A, MATAS J, KRISTAN M. D3S-A discriminative single shot segmentation tracker [C] // Proceedings of the Conference on Computer Vision and Pattern Recognition, Seattle, WA, USA, 2020; 7131-7140.
- [38] LI X, MA C, WU B Y, et al. Target-Aware Deep Tracking [C] // Proceedings of the International Conference on Computer Vision and Pattern Recognition, CA, USA, 2019; 1369-1378.
- [39] YAN B, PENG H W, WU K, et al. Lighttrack; Finding lightweight neural networks for object tracking via one-shot architecture search [C] // Proceedings of the Conference on Computer Vision and Pattern Recognition, Virtual, 2021; 15180-15189.
- [40] BLATTER P, KANAKIS M, DANELLJAN M, et al. Efficient Visual Tracking with Exemplar Transformers [C] // Proceedings of the Winter Conference on Applications of Computer Vision, Waikoloa, HI, USA, 2023; 1571-1581.



SUN Ziwen, born in 1996, Ph.D candidate. His main research interests include computer vision and object tracking.



YUAN Guanglin, born in 1976, Ph.D, associate professor. His main research interests include computer vision, object tracking and machine learning.