自动化学报
ACTA AUTOMATICA SINICA

一种新的分段式细粒度正则化的鲁棒跟踪算法

安志勇 梁顺楷 李博 赵峰 窦全胜 相忠良

Robust Visual Tracking With a Novel Segmented Fine-grained Regularization

AN Zhi-Yong, LIANG Shun-Kai, LI Bo, ZHAO Feng, DOU Quan-Sheng, XIANG Zhong-Liang

在线阅读 View online: https://doi.org/10.16383/j.aas.c220544

您可能感兴趣的其他文章

基于深度卷积特征的细粒度图像分类研究综述

A Survey on Fine-grained Image Categorization UsingDeep Convolutional Features

自动化学报. 2017, 43(8): 1306-1318 https://doi.org/10.16383/j.aas.2017.c160425

基于迁移学习的细粒度实体分类方法的研究

Fine-grained Entity Type Classification Based on Transfer Learning

自动化学报. 2020, 46(8): 1759-1766 https://doi.org/10.16383/j.aas.c190041

面向边缘计算应用的宽度孪生网络

Broad Siamese Network for Edge Computing Applications

自动化学报. 2020, 46(10): 2060-2071 https://doi.org/10.16383/j.aas.c200555

基于FR-ResNet的车辆型号精细识别研究

Fine-Grained Car Model Recognition Based on FR-ResNet

自动化学报. 2021, 47(5): 1125-1136 https://doi.org/10.16383/j.aas.c180539

基于Fg-CarNet的车辆型号精细分类研究

Fine-grained Classification of Car Models Using Fg-CarNet Convolutional Neural Network

自动化学报. 2018, 44(10): 1864-1875 https://doi.org/10.16383/j.aas.2017.c170109

基于方向场正则化的线描画生成算法

Line Drawing Generation Algorithm Based on Direction Field Regularization

自动化学报. 2021, 47(3): 685-694 https://doi.org/10.16383/j.aas.c190393

一种新的分段式细粒度正则化的鲁棒跟踪算法

安志勇 [1] 梁顺楷 [1] 李 博 [1] 赵 峰 [1] 窦全胜 [1] 相忠良 [1]

摘 要 孪生网络跟踪算法在训练阶段多数采用 L_2 正则化, 而忽略了网络架构的层次和特点, 因此跟踪的鲁棒性较差. 针对该问题, 提出一种分段式细粒度正则化跟踪 (Segmented fine-grained regularization tracking, SFGRT) 算法, 将孪生网络的正则化划分为滤波器、通道和神经元三个粒度层次. 创新性地建立了分段式细粒度正则化模型, 分段式可针对不同层次粒度组合, 利用组套索构造惩罚函数, 并通过梯度自平衡优化函数自适应地优化各惩罚函数系数, 该模型可提升网络架构的泛化能力并增强鲁棒性. 最后, 基于 VOT2019 跟踪数据库的消融实验表明, 与基线算法 SiamRPN++比较, 在鲁棒性指标上降低了 7.1% 及在平均重叠期望 (Expected average overlap, EAO) 指标上提升了 1.7%, 由于鲁棒性指标越小越好, 因此鲁棒性得到显著增强. 基于 VOT2018、VOT2019、UAV123 和 LaSOT 等主流数据库的实验也表明, 与国际前沿跟踪算法相比, 所提算法具有较好的鲁棒性和跟踪性能.

关键词 视觉跟踪, 孪生网络, 细粒度正则化, 组套索

引用格式 安志勇, 梁顺楷, 李博, 赵峰, 窦全胜, 相忠良. 一种新的分段式细粒度正则化的鲁棒跟踪算法. 自动化学报, 2023, **49**(5): 1116−1130

DOI 10.16383/j.aas.c220544

Robust Visual Tracking With a Novel Segmented Fine-grained Regularization

AN Zhi-Yong[1] LIANG Shun-Kai[1] LI Bo[1] ZHAO Feng[1] DOU Quan-Sheng[1] XIANG Zhong-Liang[1]

Abstract Most of the Siamese network tracking algorithms use L_2 regularization in the training stage, while ignoring the hierarchy and characteristic of the network architecture. As a result, such trackers have poor robustness. With this insight, we propose a segmented fine-grained regularization tracking (SFGRT) algorithm, which divides the regularization of Siamese network into three fine-grained levels, namely filter level, channel level and shape level. Then we creatively build a segmented fine-grained regularization model that constructs penalty functions based on group lasso, which combines with different levels of granularity to improve generalization ability and robustness. In addition, aiming at the imbalance of gradient magnitude of each penalty function, our approach constructs a gradient self-balancing optimization function to adaptively optimize the coefficients of each penalty function. Finally, ablation study on VOT2019 show that compared with the baseline algorithm SiamRPN++, our approach achieves relative gains of 7.1% and 1.7% in terms of robustness and expected average overlap (EAO) metrics, respectively. It means that the robustness of our tracker is significantly enhanced over baseline tracker since the smaller the robustness metrics, the better. Extensive experiments based on VOT2018, VOT2019, UAV123 and LaSOT show that the proposed algorithm has better robustness and tracking performance than related state-of-the-art methods.

Key words Visual tracking, Siamese network, fine-grained regularization, group lasso

Citation An Zhi-Yong, Liang Shun-Kai, Li Bo, Zhao Feng, Dou Quan-Sheng, Xiang Zhong-Liang. Robust visual tracking with a novel segmented fine-grained regularization. *Acta Automatica Sinica*, 2023, **49**(5): 1116−1130

目标跟踪是计算机视觉领域的前沿技术, 被广泛应用于智能监控[1]、人机交互[2] 和增强现实[3] 等领域. 该方向存在着许多挑战性属性, 如遮挡、形变、尺度变化、镜头移动以及快速运动等[4], 因此如何增强模型的泛化能力和鲁棒性, 已成为目标跟踪研究的关键.

当前, 相关滤波和孪生网络是目标跟踪最流行的两大范式. 其中, 相关滤波是一种有监督的线性回归技术, 它通过循环移动训练样本实现稠密采样, 利用了快速傅里叶变换的特性, 在线学习相关滤波器来定位连续帧中的目标对象[4]. 而孪生网络将跟踪视为相似性学习问题, 通过端到端的离线训练来学习目标图像和搜索区域之间的相似性, 孪生网络

收稿日期 2022-07-03 录用日期 2022-11-12
Manuscript received July 3, 2022; accepted November 12, 2022
国家自然科学基金 (61976125, 62176140), 山东省自然科学基金 (ZR2021MF068, ZR2021MF015, ZR2021MF107, ZR2021QF134) 资助
Supported by National Natural Science Foundation of China (61976125, 62176140) and Natural Science Foundation of Shandong Province (ZR2021MF068, ZR2021MF015, ZR2021MF107, ZR2021QF134)
本文责任编委 黄攀峰
Recommended by Associate Editor HUANG Pan-Feng
1. 山东工商学院计算机科学与技术学院 烟台 264003
1. School of Computer Science and Technology, Shandong Technology and Business University, Yantai 264003

跟踪器能够学习物体运动和外观之间的一般关系, 并可以用来定位训练中未见过的目标.

基于相关滤波的跟踪算法具有较高的跟踪精度和实时性, 受到了较多学者的关注. Tang 等[5] 采用循环矩阵来生成训练样本, 并利用离散傅里叶变换将其对角化, 减少了计算量. RPCF 算法[6] 在滤波器上引入加权约束来等效实现感兴趣区域池化, 对模型的有效参数量进行压缩, 缓解了过拟合. 随着相关滤波跟踪算法在精度上不断提升, 许多研究者尝试利用正则化方法改进其鲁棒性. 仇祝令等[7] 利用注意力机制在空域和时域上自适应约束滤波器的学习, 实现了一种基于注意力学习的正则化方法. TSCF 算法[8] 在时间和空间上平滑多通道滤波器, 使得学习到的滤波器的能量分布更均匀, 针对遮挡、平面内旋转等挑战性属性具有更好的鲁棒性. Hu 等[9] 提出了一种基于增广样本的流形正则化相关滤波跟踪方法, 实现了一个半监督目标跟踪学习框架. Xu 等[10] 在组套索正则化的基础上提出了组空间–通道正则化算法应用于相关滤波跟踪算法, 降低了通道级特征和跨通道级特征的冗余度.

另外, 有较多学者将相关滤波与粒子滤波、条件随机场以及显著性等技术相结合, 获得了较为满意的跟踪性能. MCPF 算法[11] 通过多任务相关滤波器使采样粒子聚焦在目标可能的位置, 结合粒子采样策略可以有效地解决大尺度变化问题. 黄树成等[12] 考虑到响应图中数值具有连续性, 定义响应值的求解为一个连续条件随机场的学习问题, 将相关滤波与条件随机场结合, 设计了一个端到端的深度卷积神经网络. 张伟俊等[13] 提出了三个模型, 包括像素级概率性表征模型、显著性观测模型和基于运动估计的观测模型, 对复杂背景、形变和平面内旋转具有更好的鲁棒性. 然而, 相关滤波采用的循环位移操作引入了边界效应, 由于循环位移, 训练样本中的负样本并不是真实的背景内容, 而是一个较小图像块的不断位移合成的重复, 因此模型在训练过程中看到的背景样本较少, 限制了其判别能力, 同时上述技术也无法避开边界效应.

基于深度学习及孪生网络的目标跟踪也取得了显著性进展, 成为当前的主流方法. 郭文等[14] 提出了基于深度相对度量学习的目标跟踪方法, 构建深度相对度量学习模型, 挖掘跟踪目标在大尺度图像块里的结构相对关系. SiamFC 算法[15] 以 AlexNet[16] 为基础构建了孪生网络跟踪框架, 具有较高的鲁棒性和高速性能. 此后孪生网络系列跟踪算法 (SiamRPN[17], SiamRPN++[18], SiamMask[19], SiamDW[20], SiamBAN[21−22]) 将跟踪问题视为计算模板和搜索区域互相关后的相似度, 根据分类与回归两个分支分别预测相似度和位置. SiamDW 算法发现了主干网络影响孪生网络的三个重要因素, 分别是最后一层的感受野大小、网络总步长和特征的填充, 并且提出了一个孪生网络结构的设计思路和新的网络结构. SiamMask 算法提出了对视觉目标跟踪 (Visual object tracking, VOT) 和视频目标分割 (Visual object segmentation, VOS) 的统一框架. AFST 算法[23] 提出一种基于像素上直接预测方式的高鲁棒性跟踪算法, 简化了分类任务和回归任务的复杂程度, 并消除了锚框和目标误匹配问题. SiamBAN 算法构造了边框自适应头网络, 实现了一种无锚框跟踪框架, 且通过多层次预测自适应融合了不同层次特征, 使得算法定位精确且对目标外观变化具有较强的鲁棒性. 文献 [24] 提出了一种基于 Transformer 的全新跟踪框架, 包括特征提取、类 Transformer 融合和头部预测模块, 并开发了一个基于注意力机制的特征融合网络, 替代了传统的互相关层. HiFT 算法[25] 提出了一种层次特征转换器来学习多层次特征之间的关系, 并设计了一个简洁的特征调制层, 进一步利用了孪生网络中的层次特征. 最新的孪生网络跟踪算法 AFAT[26] 设计了质量预测网络模块, 可以从时空角度对潜在的跟踪失败进行可靠的预测, 结合到 SiamRPN++中, 减少了跟踪失败的次数. 上述孪生网络跟踪算法多数采用 L_2 正则化方法, 然而, 权值往往作为特定组合存在于卷积核中, 权值之间具有局部相关性, 即具有一种自然的分组结构, 而 L_2 缺乏对于这种局部相关性的考虑, 忽视了卷积核中的分组结构, 因此模型的鲁棒性较低.

针对该问题, 本文提出将孪生网络进行细粒度分层, 划分为滤波器、通道和神经元三个粒度层次, 分别对应了卷积层的滤波器、通道和神经元的三种自然分组结构, 利用组套索构造不同粒度层次的正则项优化网络权值, 因此模型的鲁棒性比 L_2 正则化表现更优.

本文主要贡献总结如下:

1) 针对孪生网络算法在鲁棒性方面的不足, 本文将孪生网络的训练划分为滤波器、通道和神经元三个粒度层次, 利用组套索构造惩罚函数. 根据检索分析, 这是首次将该划分方式应用于孪生网络跟踪模型中.

2) 针对简单分层无法表征组间相关性, 本文提出分段式细粒度正则化. 针对不同层次的粒度组合, 可缓解神经网络的过拟合, 从而提高孪生网络模型的鲁棒性. 自适应调节各惩罚函数的梯度量级, 防止训练过程中过度学习一些粒度层次而忽视其他粒

度任务, 可进一步增强模型的鲁棒性.

3) 本文所提算法在 VOT2019[27]、VOT2018[28]、UAV123[29] 和 LaSOT[30] 4 个主流数据集上与其他国际前沿算法相比, 本文算法具有较好的鲁棒性和平均重叠期望 (Expected average overlap, EAO) 性能, 特别是在部分遮挡、光照变化、旋转、相机运动等挑战性属性下具有更好的鲁棒性和有效性.

1 相关工作

1.1 孪生网络

孪生网络跟踪算法 (SiamFC, SiamRPN, SiamRPN++, SiamMask, SiamBAN, SiamDW) 因其高精度和端到端的学习而在视觉跟踪领域引起了极大的关注. SiamFC 采用孪生网络作为特征提取器, 并首先引入互相关层以获得响应图, 该算法进行多尺度测试, 速度快但精度表现一般. 为了提升跟踪精度, SiamRPN 引入了区域提议网络 (Region proposal network, RPN) 进行边框回归替代多尺度测试, 以获得更紧凑的目标边界框, 之后的 DaSiamRPN[31]、SiamDW、SiamRPN++及 SiamBAN 等也继承了边框回归的思想. 另外, SiamBAN 采用了无锚框的边框回归, 使得算法定位精确且对目标外观变化具有较强的鲁棒性. 当前孪生网络跟踪算法的改进策略主要分为三种, 一种引入了干扰对抗模块, 如 DaSiamRPN 引入了一个干扰感知模块, 以提高模型的辨别能力. 另一种采用了更深更复杂的网络架构, 如 SiamRPN++ 和 SiamDW 以不同方式消除了零填充的不良影响, 将 ResNet[32] 等现代深度网络架构引入了孪生网络跟踪框架, 增强了模型的泛化能力. 第三种方式是引入注意力机制, 如 SA-Siam[33]、RASNet[34] 和 CGACD[35] 等跟踪器嵌入了多种注意力模块, 帮助模型区分目标与背景. 以上策略均着力于网络结构的优化, 而忽略了网络权值的优化. 实际上, 网络的泛化能力不只取决于网络结构, 网络权值也是主因. 因此, 本文以正则化为切入点, 主张设计一种适应孪生网络结构的正则化方法, 在离线训练阶段约束网络权值的学习以增强算法的鲁棒性.

1.2 跟踪算法的正则化

正则化是利用对具体任务的先验知识, 以约束和惩罚的形式控制模型学习, 增强模型泛化能力. STRCF[36] 结合了空间和时间正则化以缓解相关滤波边界效应. ARCF[37] 强制限制响应图的变化率, 抑制了异常变化, 使得跟踪器更加鲁棒和准确. AutoTrack[38] 自动且自适应地学习时空正则化项, 有助于在跟踪过程中适应各种变化. Liu 等[39] 通过对卷积神经网络学习的特征施加度量学习正则化项, 使得孪生网络更加鲁棒. DaSiamRPN 通过随机平移、尺度缩放、高斯模糊等数据增强策略提升了模型的抗干扰能力并扩充了训练样本. IRCA-Siam[40] 在训练数据中输入噪声数据作为正则化, 提高模型的泛化能力. Jia 等[41] 在孪生网络跟踪算法上应用了对抗性训练策略, 作为一种正则化增强了模型应对恶意对抗样本时的鲁棒性. 目前, 大多数孪生网络跟踪算法通过 L_2 正则化优化权值, 然而卷积层中存在一些固有的不同维度的分组结构, 例如过滤器、通道、神经元. 但是 L_2 正则化只能沿神经元维度调整网络权值, 不能恰当地约束其他更高维度的分组结构, 效率较低. 由于复杂场景中的视觉跟踪对跟踪器的鲁棒性提出了很高的要求, 因此迫切需要一种更合适的方法来替代孪生网络的 L_2 正则化.

2 本文方法

基于分段式细粒度正则化的目标跟踪算法的训练框架, 如图 1 所示. 其中特征提取网络分别接受当前帧的模板和搜索区域作为输入, 通过 Depth-wise 互相关操作得到响应映射图. 区域提议网络模块的分类分支和回归分支分别输出分类得分图和回归得分图, 结合标签信息计算出对应的损失值, 分类损失采用交叉熵损失函数, 而回归损失采用 Smooth L_1 损失函数. 分段式细粒度正则化模块把整体网络划分为 4 个段分别取网络权值, 然后通过细粒度组套索正则化模块为每个段计算 3 种细粒度组套索惩罚函数, 这些惩罚函数构成了分段式正则化矩阵. 梯度自平衡优化方法通过正则化系数矩阵约束分段式正则化矩阵, 得到最终的分段式细粒度正则项, 并通过随机梯度下降方法求解梯度自平衡优化函数, 更新下次迭代的正则化系数矩阵.

2.1 细粒度组套索正则化

本文所提出的分段式细粒度正则化跟踪 (Segmented fine-grained regularization tracking, SFGRT) 算法的损失函数由 RPN 分类分支损失函数 L^{CLS} 和加权回归分支损失函数 ρL^{REG} 构成, 记作式 (1):

$$L(\boldsymbol{y}, \boldsymbol{l}) = L^{\mathrm{CLS}} + \rho L^{\mathrm{REG}} \tag{1}$$

其中, $\boldsymbol{y}$ 和 $\boldsymbol{l}$ 分别是预测函数的结果和样本标签. 分类分支的损失函数 L^{CLS} 采用了交叉熵损失函数. 而回归分支的损失函数 L^{REG} 采用了 Smooth L_1 损失函数:

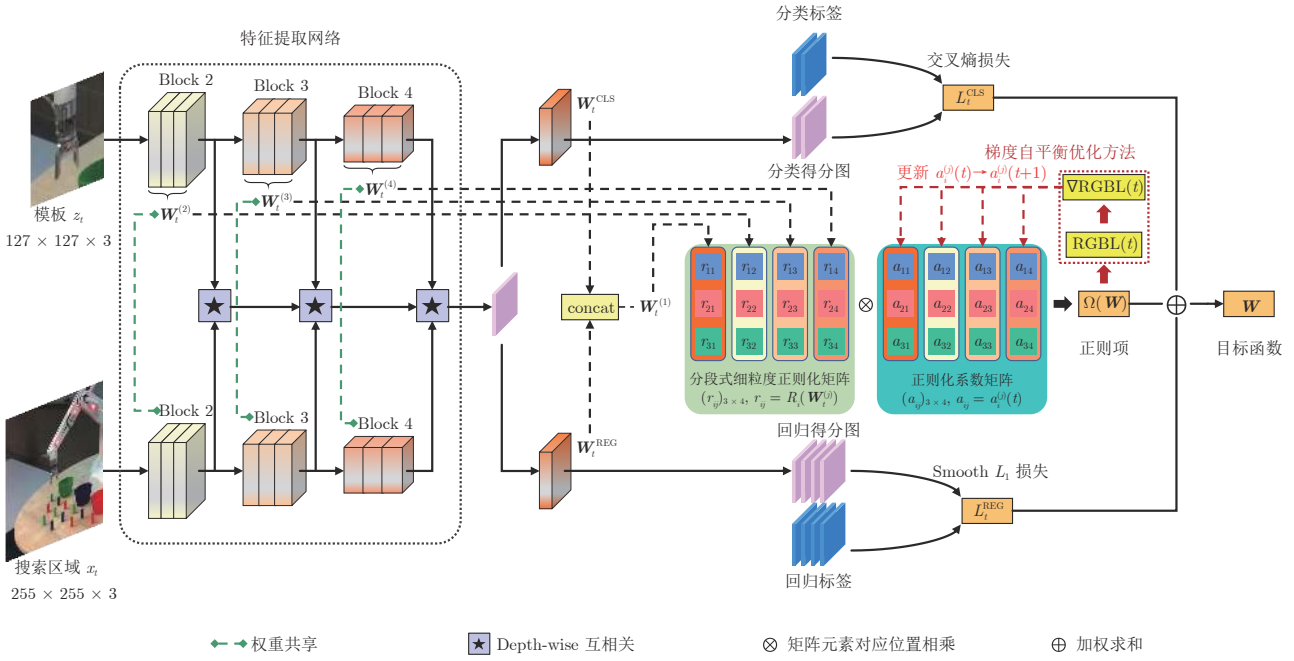


图 1　分段式细粒度正则化跟踪算法的训练框架图

Fig. 1　The training framework of the segmented fine-grained regularization tracking

$$L^{\mathrm{REG}} = \sum_{i=0}^{3} \mathrm{smooth}_{L_1}(\delta[i],\, \sigma) \tag{2}$$

$$\mathrm{smooth}_{L_1}(x,\, \sigma) = \begin{cases} 0.5\sigma^2 x^2, & |x| < \dfrac{1}{\sigma^2} \\ |x| - \dfrac{1}{2\sigma^2}, & |x| \geq \dfrac{1}{\sigma^2} \end{cases} \tag{3}$$

式 (2) 中, $\delta[i]$ 表示预测边界框和真实边界框之间的标准坐标距离.

本文将目标跟踪的分类和回归任务描述为单样本检测问题. 单样本检测的任务是学习一组参数 $\widetilde{\boldsymbol{W}}$, 以最小化预测函数 $\boldsymbol{y} = \varsigma(\varphi(\boldsymbol{z}_n,\, \boldsymbol{W}) * \varphi(\boldsymbol{x}_n,\, \boldsymbol{W}))$ 的损失函数 $L(\boldsymbol{y},\, \boldsymbol{l})$ 的期望[17], 结合结构风险最小化, 避免预测函数 ς 对训练集过拟合, 对于 N 个样本 $\boldsymbol{x}_n$、模板帧 $\boldsymbol{z}_n$ 和样本标签 $\boldsymbol{l}_n$, 优化目标为:

$$\arg\min_{\boldsymbol{W}} \sum_{n=1}^{N} L(\varsigma(\varphi(\boldsymbol{z}_n,\, \boldsymbol{W}) * \varphi(\boldsymbol{x}_n,\, \boldsymbol{W})),\, \boldsymbol{l}_n) + \lambda\Omega(\boldsymbol{W}) \tag{4}$$

其中, ς 表示 RPN 预测函数, φ 表示孪生网络特征提取函数, $*$ 表示互相关[18]. $\Omega(\boldsymbol{W})$ 是对复杂模型的惩罚. 如图 2 所示, 本文定义了细粒度正则项 $\Omega(\boldsymbol{W}) = \sum_{i=1}^{3} R_i(\boldsymbol{W})$, 分别为神经元 (Shape-wise) 正则化 R_1、通道 (Channel-wise) 正则化 R_2 和滤波器 (Filter-wise) 正则化 R_3. λ 是正则化率. 本文采用组套索函数 $R(\boldsymbol{W}) = \sum_{g=1}^{G} \|\boldsymbol{w}_g\|_2$ 得到 G 组参数 $\boldsymbol{w}_g$ [42].

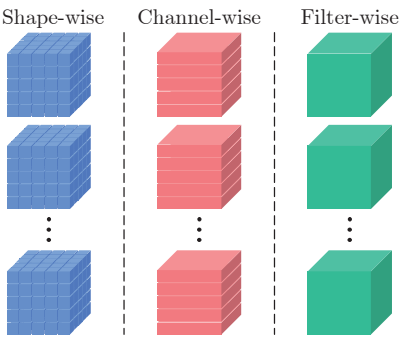


图 2　细粒度组套索示意图

Fig. 2　Fine-grained group lasso

若把权值参数 $\boldsymbol{W}$ 按卷积层维度展开, 可以得到一个长度为 L 的序列, 序列中的第 l 个元素是一个 4D 张量 $\boldsymbol{W}^l \in \mathbf{R}^{N_l \times C_l \times H_l \times B_l}$, N_l、C_l、H_l 和 B_l 分别是网络第 l 层的滤波器数、通道数、特征图高度和特征图宽度. 第 l 层的 Shape-wise 组套索惩罚函数为:

$$R_1(\boldsymbol{W}^l) = \sum_{n_l=1}^{N_l} \sum_{c_l=1}^{C_l} \sum_{h_l=1}^{H_l} \sum_{b_l=1}^{B_l} \left\| \boldsymbol{w}^l_{n_l,\, c_l,\, h_l,\, b_l} \right\|_2 \tag{5}$$

$\|\cdot\|_2$ 是 L_2 范数, $\|\boldsymbol{w}\|_2 = \sqrt{\sum_{i=1}^{|\boldsymbol{w}|} (w_i)^2}$, $|\boldsymbol{w}|$ 是参数 $\boldsymbol{w}$ 的元素 w_i 的个数. 类似地, 本文把 Channel-wise 组套索和 Filter-wise 组套索的惩罚函数分别表示为:

$$R_2(\boldsymbol{W}^l)=\sum_{c_l=1}^{C_l}\left\|\boldsymbol{w}^l_{:,c_l,:,:}\right\|_2 \tag{6}$$

$$R_3(\boldsymbol{W}^l)=\sum_{n_l=1}^{N_l}\left\|\boldsymbol{w}^l_{n_l,:,:,:}\right\|_2 \tag{7}$$

式 (6) 的 $\boldsymbol{w}^l_{:,c_l,:,:}$ 表示由卷积神经网络第 l 层所有滤波器第 c_l 通道构成的权值张量. 式 (7) 的 $\boldsymbol{w}^l_{n_l,:,:,:}$ 表示由卷积神经网络第 l 层第 n_l 滤波器构成的权值张量.

细粒度组套索惩罚函数是基于滤波器、通道和神经元的特征表示与图像本身具有一致性这一先验知识的, 这种特征分组方式在一定程度上符合图像的视觉感知. 而且, 该特征分组策略的类似应用[43−47]在一定程度上表明了其有效性.

2.2 分段式细粒度正则化

卷积神经网络的一个特性是: 底层卷积层和顶层卷积层学习的特征存在差异. 底层卷积层学习到局部图像特征如边缘、纹理、色度等, 而顶层卷积层学习到全局语义特征如鼻子形状、耳朵轮廓等. 自然地, 卷积层构成的残差块之间存在的层次结构也会造成这种差异的产生. Wen 等[47] 实验表明正则化对深层卷积层的惩罚应大于浅层卷积层. 高维空间包含了神经网络高层中的语义信息[48], 语义层次会随着卷积层的深度加深而升高, 高语义层次的特征信息主要由高维卷积层结构所包含. 为了进一步探究 3 种细粒度组套索正则化与网络层次的关系, 本文进行了多次微调实验, 每次只在一个网络部分 (Block2, Block3, Block4, RPN) 上加入一种细粒度组套索正则化 (Shape-wise, Channel-wise, Filter-wise), 在相同学习率策略下微调 5 个轮次, 最后在 VOT2019 上与基线方法 (无细粒度组套索正则化微调 5 轮) 进行对比, 指标采用平均重叠期望 (EAO).

图 3 实验结果表明不同残差块组合与网络层数有关, 其中本文的 Block2、Block3、Block4、RPN 所处的网络深度逐渐加深, 且 RPN 相当于一个深层网络结构. 如图 3 所示, 与基线方法进行比较, Shape-wise 正则化在 RPN 上出现下降, 而在 Block2 上提升显著, 因此说明 Shape-wise 正则化适合浅层约束. Channel-wise 正则化在 Block3、Block4、RPN 上出现明显提升, 且 Block3 和 Block4 的提升较为显著, 因此 Channel-wise 正则化适合中间层的约束. 而 Filter-wise 正则化在 Block2 和 Block3 上不如基线方法, 而在 Block4、RPN 上有少量提升, 因此 Filter-wise 适合深层网络约束. 由此本文设计了一种细粒度组套索的分段式正则化模型, 以卷积层为单位作为正则化的对象进行二次细分, 将优化目标拓展为式 (8):

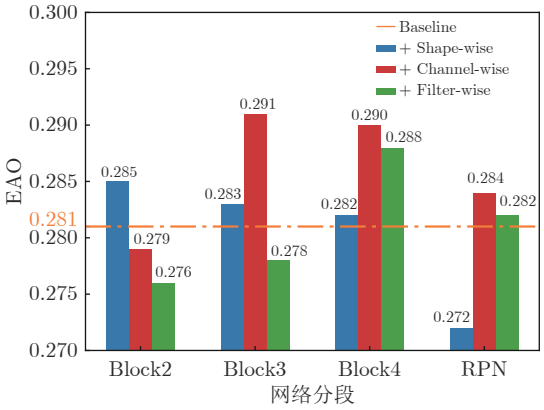


图 3 细粒度组套索正则化在各网络分段的效果对比

Fig. 3 Comparison of the effects of fine-grained group lasso regularization in each network segment

$$\arg\min_{\boldsymbol{W}}\sum_{n=1}^{N}L(\varsigma(\varphi(\boldsymbol{z}_n,\boldsymbol{W})*\varphi(\boldsymbol{x}_n,\boldsymbol{W})),\boldsymbol{l}_n)+\lambda\sum_{p=1}^{P}\sum_{i=1}^{3}a_i^{(p)}R_i(\boldsymbol{W}^{(p)}) \tag{8}$$

$\{\boldsymbol{W}^{(1)},\boldsymbol{W}^{(2)},\cdots,\boldsymbol{W}^{(P)}\}=\boldsymbol{W}$, 本文把参数 $\boldsymbol{W}$ 划分为 P 个子集 $\boldsymbol{W}^{(p)}$, 通过初始化正则化系数 $a_i^{(p)}$ 为 1 或 0, 使得每个子集可以有不同的正则化组合.

图 4 为分段式细粒度正则化示意图, 本文根据上述分析, 分别将 Shape-wise 正则化置于浅层 Block, Channel-wise 正则化置于中层 Block, 而 Filter-wise 正则化置于深层 Block 和 RPN. 而且, 为了充分学习不同层间的相关关系, 对各 Block 的正则化进行叠加组合. 具体地, 本文将 ResNet-50 按前向传播顺序划分为 4 段, 分别是第二残差块、第三残差块、第四残差块和 RPN 头网络, 在这 4 个段中选取卷积核尺寸为 3×3、步长为 1 的无偏置项卷积层共 26 个. 如图 4 所示, 起始立方体是第一残差块, 因其不参与参数更新, 因此不需要考虑对其进行正则化. 其余立方体从左到右分别是第二残差块 $\boldsymbol{W}^{(2)}$、第三残差块 $\boldsymbol{W}^{(3)}$、第四残差块 $\boldsymbol{W}^{(4)}$、头网络 (RPN 分类分支和回归分支) $\boldsymbol{W}^{(5)}=[\boldsymbol{W}^{\mathrm{CLS}},\boldsymbol{W}^{\mathrm{REG}}]$. 划分完成后, 在 4 个段中应用不同粒度的细粒度组套索惩罚函数组合, 第一残差块只使用 Shape-wise 正则化, 第二残差块把 Shape-wise 正则化和 Channel-wise 正则化结合使用, 第三残差块同时使用 3 种粒度的组套索正则化, 头网络部分结合了 Channel-wise 正则化和 Filter-wise 正则化. 若目

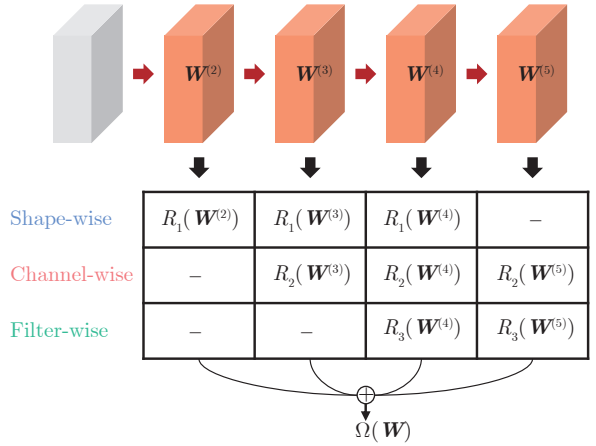


图 4　分段式细粒度正则化模型示意图

Fig. 4　Segmented fine-grained regularization model

标卷积层序列 S 划分为 P 段, 第 p 段用集合 S_p 表示, 则第 p 段的 Shape-wise 组套索惩罚函数可以用式 (9) 表示:

$$R_1(\boldsymbol{W}^{(p)})=\sum_{l\in S_p}\sum_{n_l=1}^{N_l}\sum_{c_l=1}^{C_l}\sum_{h_l=1}^{H_l}\sum_{b_l=1}^{B_l}\left\|\boldsymbol{w}^l_{n_l,c_l,h_l,b_l}\right\|_2 \tag{9}$$

第 p 段的 Channel-wise 组套索和 Filter-wise 组套索的惩罚函数分别用式 (10) 和式 (11) 表示:

$$R_2(\boldsymbol{W}^{(p)})=\sum_{l\in S_p}\sum_{c_l=1}^{C_l}\left\|\boldsymbol{w}^l_{:,c_l,:,:}\right\|_2 \tag{10}$$

$$R_3(\boldsymbol{W}^{(p)})=\sum_{l\in S_p}\sum_{n_l=1}^{N_l}\left\|\boldsymbol{w}^l_{n_l,:,:,:}\right\|_2 \tag{11}$$

然而, 分段式细粒度正则化导致多个惩罚函数的梯度量级不平衡, 训练会导致过度学习一些粒度层次而忽视其他粒度任务, 同时降低训练速度. 针对该问题, 本文采用梯度自平衡优化策略[49], 实现各细粒度组套索惩罚函数的梯度量级的稳定动态平衡, 加快训练速度, 梯度自平衡方法如图 5 所示. 在第 t 次迭代训练时, 梯度自平衡优化方法的任务是学习各惩罚函数的系数 $a_i^{(p)}(t)$, 最小化第 p 段各组套索惩罚函数 $R_i(\boldsymbol{W}_t^{(p)})$ 的梯度张量的 1-范数 $G_i^{(p)}(t)$ 的数学期望 $\mathrm{E}(G^{(p)}(t))$. 因此本文定义梯度自平衡优化函数 $\mathrm{RGBL}(t)$, 形式如式 (12) 所示:

$$\mathrm{RGBL}(t)=\sum_{p=1}^{P}\sum_{i=1}^{3}\left|G_i^{(p)}(t)-\mathrm{E}(G^{(p)}(t))\right| \tag{12}$$

其中, $G_i^{(p)}(t)$ 是对梯度张量的量化, 计算公式如下:

$$G_i^{(p)}(t)=\left\|\nabla_{\boldsymbol{W}}a_i^{(p)}(t)R_i(\boldsymbol{W}_t^{(p)})\right\|_1 \tag{13}$$

式 (13) 中的 ∇ 是梯度符号, 而 $\|\cdot\|_1$ 表示 1-范数的计算公式, $\|\boldsymbol{w}\|_1=\sum_{i=1}^{|\boldsymbol{w}|}|w_i|$, $|\boldsymbol{w}|$ 是张量 $\boldsymbol{w}$ 的元素 w_i 的个数. 此处 1-范数实际表示了张量的绝对值之和, 用于量化各惩罚函数回传梯度张量间的量级大小.

为了验证分段式细粒度正则化和细粒度组套索的性能差异, 本文采用类激活热力图 (Class activation map, CAM) 进行可视化, 选取 VOT2019 中 hand 序列的第 47、107、266 帧. 如图 6 所示, 第一列是跟踪框对比, 实线框表示真实框, 线状虚线框表示分段式细粒度正则化, 点状虚线框表示细粒度组套索. 第二、三列分别是分段式细粒度正则化的 CAM、细粒度组套索的 CAM. 可以看出, 仅采用细粒度组套索算法的跟踪框尺度大小和中心点位置出现了失准, 具体表现为当跟踪目标发生运动和形变等变化时, 预测框发生了中心点偏移和尺度放大问题. 在 CAM 中发现仅采用细粒度组套索的算法存在图像像素敏感区域偏小的问题, 只关注到图像中目标的局部特征, 如序列中手 (目标) 的关节等突出部位, 而对于全局特征不敏感. 这是因为细粒度组套索的简单分层方式不足以表征特征的组间相关性, 跟踪器难以把握图像各特征间的相关联系, 全局特征利用的不充分, 导致跟踪容易失准. 加入分

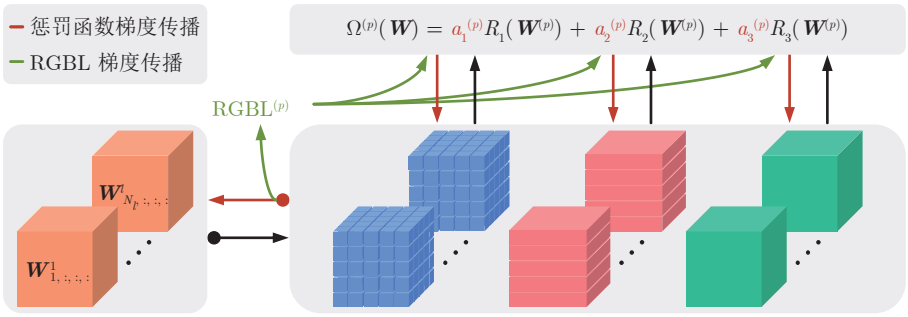


图 5　梯度自平衡优化方法

Fig. 5　The gradient self-balancing optimization approach

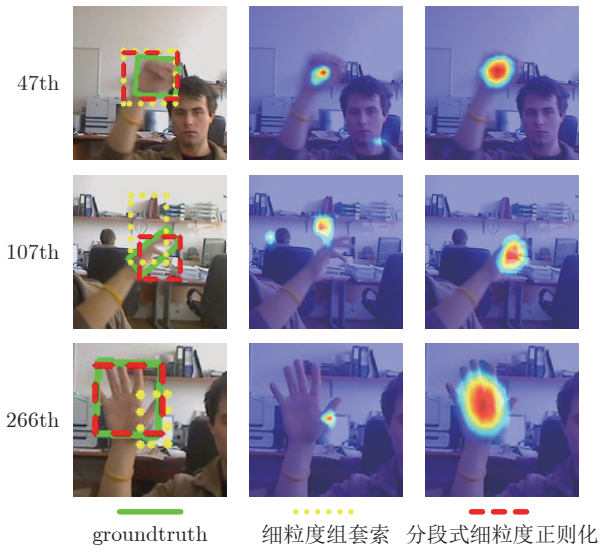


图 6 分段式细粒度正则化和细粒度组套索的跟踪效果对比

Fig. 6 Comparison of tracking results between segmented fine-grained regularization and fine-grained group lasso

段式细粒度正则化方法后, 因为在粒度层次和卷积层层次上都扩展了特征分组, 观察到模型利用的图像特征能够充分覆盖整个跟踪目标, 跟踪器对快速运动 (47th)、复杂背景 (107th)、形变 (266th) 等干扰不敏感. 从 CAM 中可以看出, 跟踪器能够正确感知目标的全局特征, 跟踪过程中未发生中心点偏移和尺度放大的问题, 反映算法的鲁棒性得到提升.

3 实验结果与分析

3.1 实验设置

实验软硬件平台. 本文实验的硬件环境为 Intel i7-10 700 CPU@2.90 GHz、64 GB 的内存和 Nvidia GeForce RTX 3 090 显卡, 操作系统为 Ubuntu 16.04 (64bit) 的标准个人电脑, 软件环境为 Pytorch 1.8.0.

实验细节及参数设置. 本文选取了 6 个公开数据集训练模型, 包括 COCO[50]、ImageNet DET[51]、ImageNet VID[51]、YouTube BoundingBoxes[52]、GOT10k[53]、LaSOT. 迭代训练过程总共有 20 个轮次, 网络参数 $\boldsymbol{W}$ 的参数更新方式采用次梯度方法 SGD-L1(naive). 本文采用了动态学习率的训练方式, 初始学习率设置为 0.001, 前 5 轮采用预热策略使学习率提升至 0.005, 之后的 15 轮采用余弦退火策略使学习率降至 0.00005. 本文使用了在 ImageNet[51] 预训练的 ResNet-50 模型对残差网络部分进行参数初始化, 并在前 10 轮冻结该部分参数的更新. 正则化系数 $a_i^{(p)}$ 的参数更新方式采用了随机梯度下降. 正则化率 λ 设置为 0.00001, 梯度自平衡方法的学习率设置为 0.00005.

数据集. 本文在 4 个广泛使用的视觉跟踪基准数据集上进行了评估, 分别为 VOT2018、VOT2019、UAV123 和 LaSOT. 其中, VOT2018 包含了 60 个视频序列, 所有的视频都被标记为旋转边界框, 这给视觉跟踪的任务带来了巨大的挑战. VOT2019 使用更具挑战性的视频来取代一些简单的视频, 而视频的总数保持不变. UAV123 提供了高分辨率的跟踪数据集, 包括 123 个完全标注的航拍视频序列, 视频总帧数超过 110 000 帧. 而最近的 LaSOT 是一个高质量的大规模数据集, 包含 1 400 个训练视频和 280 个测试视频, 其平均视频长度甚至达到了 2 512 帧.

比较方法. 为了进一步验证本文 SFGRT 算法的鲁棒性与有效性, 选取了目前最前沿的跟踪算法与本文算法进行全面的对比实验. 这些算法包括 SiamBAN、SiamCAR[54]、CLNet[55]、ROAM++[56]、HiFT、ATOM[57]、SiamMask、SiamRPN++、GFS-DCF[10]、SiamDW、SPLT[58]、C-RPN[59]、MemDTC[60]、LADCF[61]、DaSiamRPN、SiamRPN、ECO[62]、SPM[63]. 由于部分算法未给出全部数据集的测试结果, 因此本文在每个数据集的实验分析中只比较有结果的算法.

评价指标. 本文使用各个基准所提供的对应评价指标来评估所提出的算法和其他先进跟踪方法的性能. EAO、鲁棒性 (Robustness) 和准确性 (Accuracy) 被用来衡量模型在 VOT2018 和 VOT2019 上的性能. 其中, EAO 是一个综合性量化指标, 为了减少重新初始化引入的随机误差, VOT 从各测试序列截取短序列, 进行一次不重新初始化的跟踪, 计算所有短序列有效帧的平均重叠率即可得到 EAO, 该指标同时表现了算法鲁棒性和准确性. 而鲁棒性是一个逆向指标, 它表示了目标在跟踪过程中丢失或失败的次数, 该指标的数值越小, 反映跟踪器越鲁棒. 准确性表示真实框和算法预测的边界框之间的平均重叠率. 而 UAV123 和 LaSOT 采用了一次通过评估方式 (One pass evaluation, OPE), 提供了两个评价指标来比较所有算法, 分别是成功率图和精度图. 成功率图说明了在阈值 [0, 1] 范围内的成功帧的比率, 其中成功帧指那些重叠率超过给定阈值的帧, 曲线下面积被用来对所有的跟踪器进行排名. 精度图描述了在一系列阈值下每个跟踪器的平均距离精度 (Diatance precision, DP) 的得

分, DP 被定义为预测位置在给定阈值内的帧的百分比.

3.2　消融实验

为了验证本文方法的有效性, 以 SiamRPN++为基线模型, 分别叠加细粒度组套索及分段式细粒度正则化进行了实验. 表 1 为 VOT2019 的消融实验结果, 采用 Accuracy、Robustness 和 EAO 三个评价指标进行性能评估. 其中基线算法 SiamRPN++采用了 L_2 正则化, 使用 ResNet-50 作为骨干网络.

表 1　VOT2019 上的消融实验

Table 1　Ablation study on VOT2019

	基线算法	+细粒度组套索	+分段式细粒度正则化
EAO ↑	0.287	0.293	0.304
Accuracy ↑	0.595	0.600	0.586
Robustness ↓	0.467	0.456	0.396

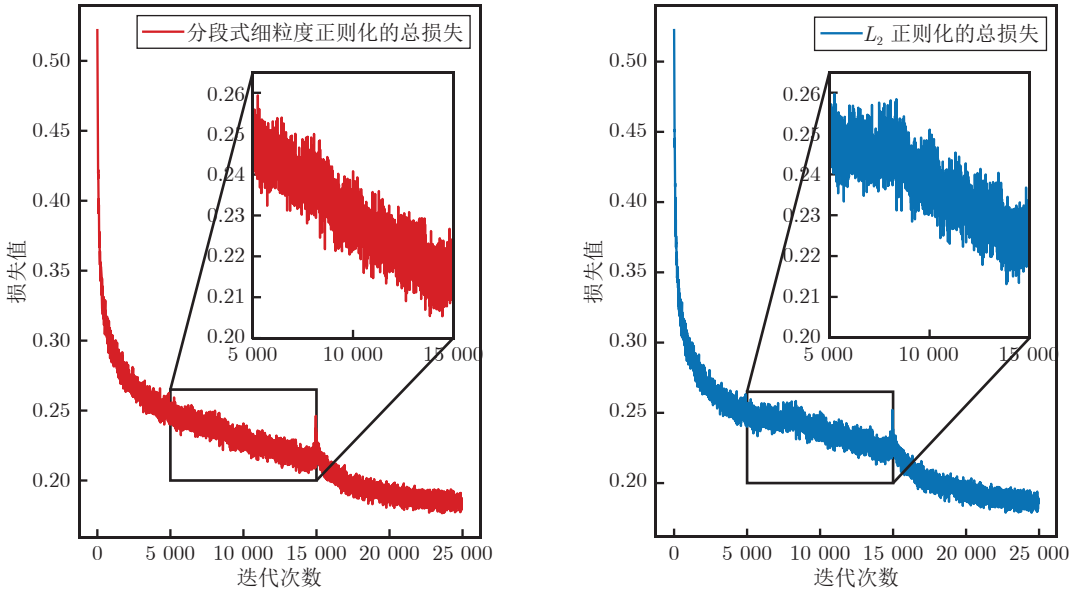
如表 1 所示, 增加了细粒度组套索的算法相比 SiamRPN++基线算法, 在 EAO 和 Accuracy 上均有所提升, 分别提高了 0.6% 和 0.5%; 在 Robustness 上降低了 1.1%, 鲁棒性得到增强. 而在基线算法上增加了分段式细粒度正则化后综合评价 EAO 达到了 0.304, Robustness 达到了 0.396 且比基线算法降低了 7.1%, 可见分段式细粒度正则化能够显著增强算法的鲁棒性.

另外, 分析表 1 发现, 分段式细粒度正则化方法的 Accuracy 指标略微下降了 0.9%. 主要原因是正则化方法优化后的跟踪框倾向于较宽松地包围目标, 为快速适应目标的变化留出余裕, 相比起紧凑型的跟踪框, 丢失和重启次数会明显减少, 而 EAO 指标的提升也反映了该策略的综合评价更高.

由于 SiamRPN++采用了 L_2 正则化, 没有考虑到网络的层次结构, 因此跟踪性能不稳定. 细粒度组套索在一定程度上能够提升跟踪性能, 但这种简单划分不能够有效体现网络层次结构, 因此性能提升有限. 而分段式细粒度正则化, 从滤波器、通道和神经元三个层次组合分组, 可针对不同层次粒度组合和优化, 增强了模型的泛化能力, 实验结果表明分段式细粒度正则化可显著增强跟踪算法的鲁棒性.

3.3　收敛性分析

为了探究本文提出的分段式细粒度正则化在训练阶段发挥的作用, 本文还对比了分段式细粒度正则化和 L_2 正则化方法的训练损失曲线图. 如图 7 所示, 图 7(a) 是分段式细粒度正则化的训练损失曲线, 图 7(b) 是 L_2 正则化方法的训练损失曲线. 从图 7(b) 的局部放大图中可以看出, 在训练中期 (5 000 ~ 15 000 次迭代) L_2 正则化方法的总损失值出现了区间震荡, 无法顺利下降导致损失曲线不平滑, 导致收敛速度减慢. 从图 7(a) 中可以看出, 分



(a) 分段式细粒度正则化的损失曲线

(a) Loss curve of segmented fine-grain regularization

(b) L_2 正则化的损失曲线

(b) Loss curve of L_2 regularization

图 7　训练损失曲线对比

Fig. 7　Comparison of training loss curve

段式细粒度正则化方法的损失值下降更稳定, 训练损失曲线更平滑, 从局部放大图中可以看出, 在 15000 次迭代后本文方法的损失值小于 L_2 正则化的损失值. 整体上 L_2 正则化和分段式细粒度正则化的收敛性基本一致, 局部上分段式细粒度正则化收敛更加平稳.

3.4 与 SOTA 方法的对比

实验 1. VOT2018 数据集上的实验

表 2 为 VOT2018 上的不同算法性能比较, 和基线算法 SiamRPN++相比, 本文算法在 EAO 指标上提升了 0.8%, 在逆向指标 Robustness 上降低了 3.7%, 反映了算法的综合评价和鲁棒性评价得到增强. 在 Robustness 指标上, 本文算法优于其他基于锚框的孪生网络跟踪算法 (SiamRPN, SiamRPN++, SiamMASK). 与目前最先进的无锚框跟踪算法 SiamBAN 相比, Robustness 仅差 1.9%. 类似地, 相关滤波算法 GFS-DCF 亦在组套索基础上进行正则化方法的拓展, 与其相比, 本文算法的 EAO 和 Accuracy 分别高 2.5% 和 7.8%, 但 Robustness 没有 GFS-DCF 好, 原因是相关滤波跟踪算法采用了模板更新策略, 可快速适应目标的外观变化, 而本文算法没有更新策略.

表 2 在 VOT2018 上与 SOTA 算法的比较

Table 2 Comparison with SOTA trackers on VOT2018

算法	出版	EAO ↑	Accuracy ↑	Robustness ↓
SiamRPN	CVPR2018	0.383	0.586	0.276
SiamRPN++	CVPR2019	0.414	0.600	0.234
SiamMask	CVPR2019	0.380	0.609	0.276
LADCF	ITIP2019	0.389	0.503	0.159
ATOM	CVPR2019	0.401	0.590	0.204
GFS-DCF	ICCV2019	0.397	0.511	0.143
SiamBAN	CVPR2020	0.452	0.597	0.178
SFGRT (Ours)	—	0.422	0.589	0.197

实验 2. VOT2019 数据集上的实验

表 3 为 VOT2019 上的不同算法性能比较, 本文算法的 Robustness 指标得分达到了 0.396, 在该性能指标上均优于 SiamDW、SiamMASK、Roam++ 等算法, 并与目前最先进的无锚框跟踪算法 SiamBAN 持平, 反映了本文所提 SFGRT 算法的鲁棒性达到了先进水平. 与 ATOM 算法相比, Accuracy 指标低 1.7%, 原因是 ATOM 引入了目标检测方向最新的 IoU-Net 框架, 优化了边界框回归, 使得预测框在尺度上更接近真实框. 最后与基线算法 SiamRPN++进行比较, 在综合指标 EAO 上提升了 1.7%, 在逆向指标 Robustness 上显著降低了 7.1%. 本文算法在不改动网络结构的条件下, 在离线训练阶段设计适应孪生网络的正则化约束基线模型学习, 即可显著地增强鲁棒性, 实验表明了网络的泛化能力不只取决于网络结构, 网络权值亦是主因.

表 3 在 VOT2019 上与 SOTA 算法的比较

Table 3 Comparison with SOTA trackers on VOT2019

算法	出版	EAO ↑	Accuracy ↑	Robustness ↓
SPM	CVPR2019	0.275	0.577	0.507
SiamRPN++	CVPR2019	0.287	0.595	0.467
SiamMask	CVPR2019	0.287	0.594	0.461
SiamDW	CVPR2019	0.299	0.600	0.467
MemDTC	PAMI2019	0.228	0.485	0.587
ATOM	CVPR2019	0.292	0.603	0.411
Roam++	CVPR2020	0.281	0.561	0.438
SiamBAN	CVPR2020	0.327	0.602	0.396
SFGRT (Ours)	—	0.304	0.586	0.396

实验 3. UAV123 数据集上的实验

图 8 比较了本文提出的 SFGRT 算法与目前最前沿的目标跟踪算法在 UAV123 基准下的精度图和成功率图, 本文算法在精度和成功率上分别领先基线算法 2.4% 和 1.2%. 相比于最新的无锚框孪生网络跟踪算法 SiamCAR, 本文算法在精度上领先 1.5%, 且在成功率上与其接近持平.

实验 4. LaSOT 数据集上的实验

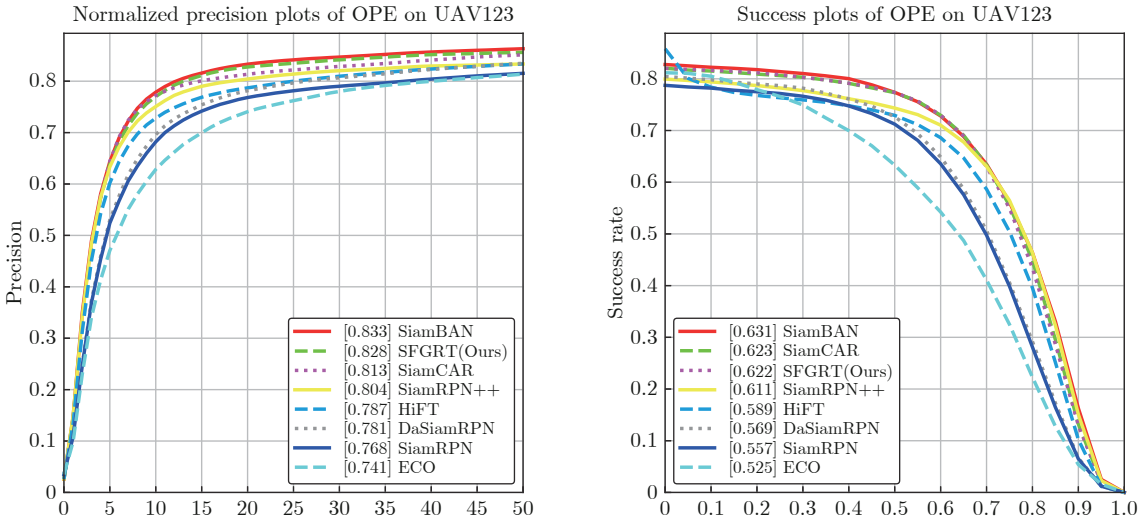
图 9 是 SFGRT 算法与目前最前沿的跟踪算法在 LaSOT 基准下的归一化精度图和成功率图的性能比较结果. 可以看出, 本文算法在该基准下达到了先进水平. 且与 SiamRPN++算法相比, 本文所提出的 SFGRT 算法分别在归一化精度和成功率上提升了 2.0% 和 0.3%.

3.5 讨论

本文设计了 4 个实验进行分析和讨论, 实验 1 分析了各 SOTA 算法在 UAV123 上不同挑战性属性下的性能. 实验 2 分析了各算法在 LaSOT 上不同挑战性属性下的性能. 实验 3 选取了 VOT2019 部分视频序列, 分析了本文算法 SFGRT 和基线算法 SiamRPN++的差异. 实验 4 对比分析了各算法的模型大小和速度.

实验 1. UAV123 不同挑战属性影响下的性能分析

UAV123 为无人机航拍视频数据集, 本文选择了光照变化 (Illumination variation, IV)、部分遮挡 (Partial occlusion, POC)、超出视野 (Out-of-

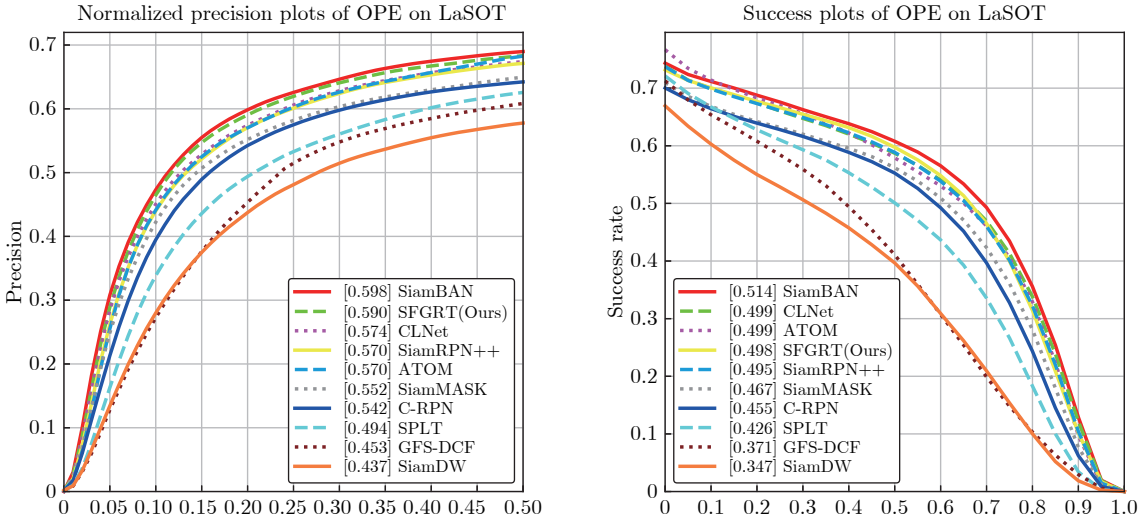


(a) 归一化精度图对比　(b) 成功率图对比

(a) Comparison of precision plots　(b) Comparison of success plots

图 8　UAV123 基准上的性能对比

Fig. 8　Comparison of tracking results on UAV123



(a) 归一化精度图对比　(b) 成功率图对比

(a) Comparison of normalized precision plots　(b) Comparison of success plots

图 9　LaSOT 基准上的性能对比

Fig. 9　Comparison of tracking results on LaSOT

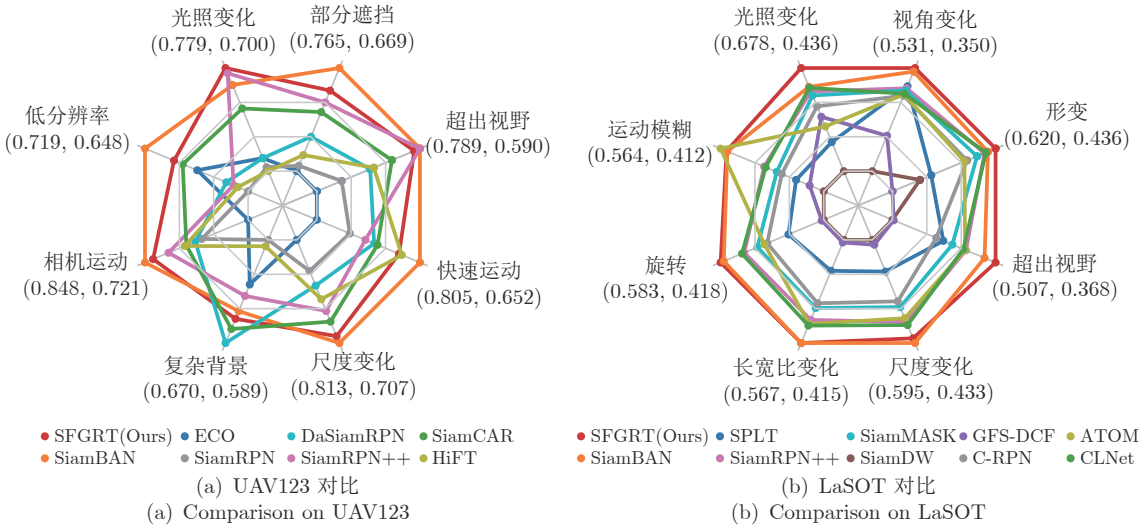
view, OV)、快速运动 (Fast motion, FM)、复杂背景 (Background clutter, BC)、低分辨率 (Low resolution, LR)、相机运动 (Camera motion, CM)、尺度变化 (Scale variation, SV) 8 项视频属性, 表 4 对 SiamBAN、HiFT、SiamCAR 等 8 个 SOTA 跟踪器的精度 (Precision) 指标进行了统计.

从表 4 中可以看出本文所提 SFGRT 算法在光照变化、部分遮挡、低分辨率、相机运动、尺度变化和超出视野这六项挑战属性下的精度指标均达到了先进水平. 尤其在光照变化属性下, 本文算法取得了最佳性能. 和基线算法 SiamRPN++比较发现, SFGRT 算法仅在超出视野性能出现少量下滑 (1.1%), 而其余七项的性能均实现了提升, 尤其体现在快速运动和低分辨率这两项属性上, 分别实现了 5.0% 和 4.1% 的显著提升. 图 10(a) 是参与测试的八个算法的蛛网图对比.

表 4 在 UAV123 基准上与 SOTA 算法在 8 个挑战性属性下的精度对比

Table 4 Comparison of precision with SOTA trackers on 8 challenging attributes on UAV123

Attribute	ECO	SiamRPN	DaSiamRPN	SiamRPN++	SiamCAR	SiamBAN	HiFT	SFGRT
	CVPR2017	CVPR2018	ECCV2018	CVPR2019	CVPR2020	CVPR2020	ICCV2021	—
POC	0.669	0.674	0.701	0.733	0.724	0.765	0.684	0.744
IV	0.710	0.703	0.710	0.775	0.748	0.766	0.700	0.779
CM	0.721	0.778	0.786	0.819	0.797	0.848	0.799	0.838
FM	0.652	0.701	0.737	0.724	0.742	0.805	0.778	0.774
SV	0.707	0.739	0.754	0.780	0.791	0.813	0.768	0.806
BC	0.624	0.589	0.670	0.633	0.659	0.645	0.594	0.651
OV	0.590	0.638	0.693	0.789	0.735	0.789	0.700	0.778
LR	0.683	0.648	0.663	0.658	0.693	0.719	0.655	0.699
Overall	0.741	0.768	0.781	0.804	0.813	0.833	0.787	0.828



(a) UAV123 对比 (a) Comparison on UAV123

(b) LaSOT 对比 (b) Comparison on LaSOT

图 10 UAV123 和 LaSOT 基准上的多挑战属性下精度对比

Fig. 10 Comparison of precision under different challenging attributes on UAV123 and LaSOT benchmarks

实验 2. LaSOT 不同挑战属性影响下的性能分析

LaSOT 数据集包含多种现实挑战属性, 本文选择了 IV、OV、SV、旋转 (Rotation, ROT)、运动模糊 (Motion blur, MB)、形变 (Deformation, DEF)、视角变化 (Viewpoint change, VC)、长宽比变化 (Aspect ratio change, ARC) 共 8 项属性, 表 5 对 CLNet、ATOM、GFS-DCF 等 10 个性能 SOTA 的跟踪器在不同属性下的归一化精度指标进行了统计.

由表 5 所示, 可见本文算法在 8 项挑战属性下的归一化精度指标均取得了最优或次优. 在运动模糊下稍逊于 ATOM, 在尺度变化下略低于 SiamBAN. 和基线算法 SiamRPN++ 进行对比, 本文算法在所有挑战属性下均取得了提升, 特别在光照变化和运动模糊两项属性下提升显著, 分别提升了 4.5% 和 4.7%. 图 10(b) 是参与测试的 10 个算法在 LaSOT 基准下的性能对比蛛网图, 可以看出本文所提 SFGRT 算法的蛛网图面积最大且形状近似正八边形, 反映了本文算法在这 8 项挑战属性下无短板, 算法具有最佳且最全面的鲁棒性.

实验 3. VOT2019 数据集部分序列跟踪结果可视化比较

图 11 为本文方法与基线算法 SiamRPN++在 dinosaur、gymnastics2、wheel、soccer1 4 个视频序列的对比结果, 所选视频序列存在相机运动、快速运动、旋转、遮挡、复杂背景等多种挑战属性. dinosaur 和 gymnastics2 两个序列都存在旋转干扰, 如 dinosaur 中, 受到旋转干扰, SiamRPN++跟踪框先出现了严重的中心点偏移和尺度放大, 导致在后续帧中跟踪失败, 但本文方法能够成功跟踪目标,

表 5　在 LaSOT 基准上与 SOTA 算法在 8 个挑战性属性下的归一化精度对比

Table 5　Comparison of norm precision with SOTA trackers on 8 challenging attributes on LaSOT

Attribute	SPLT	C-RPN	SiamDW	SiamMask	SiamRPN++	GFS-DCF	ATOM	SiamBAN	CLNet	SFGRT
	ICCV2019	CVPR2019	CVPR2019	CVPR2019	CVPR2019	ICCV2019	CVPR2019	CVPR2020	ICCV2021	—
DEF	0.520	0.578	0.500	0.593	0.604	0.436	0.574	0.609	0.606	0.620
VC	0.505	0.491	0.350	0.499	0.502	0.427	0.493	0.526	0.494	0.531
IV	0.524	0.603	0.436	0.625	0.633	0.581	0.560	0.642	0.640	0.678
MB	0.465	0.486	0.412	0.493	0.510	0.443	0.564	0.556	0.508	0.557
ROT	0.488	0.520	0.418	0.534	0.552	0.425	0.524	0.579	0.555	0.583
ARC	0.473	0.518	0.415	0.524	0.539	0.423	0.544	0.567	0.546	0.567
SV	0.496	0.540	0.433	0.548	0.568	0.447	0.563	0.595	0.572	0.589
OV	0.447	0.438	0.368	0.458	0.474	0.372	0.473	0.495	0.471	0.507
Overall	0.494	0.542	0.437	0.552	0.570	0.453	0.570	0.598	0.574	0.590

图 11　VOT2019 测试集部分视频序列的跟踪结果

Fig. 11　Tracking results for some video sequences on VOT2019

因为基于分段式细粒度正则化的孪生网络模型对目标外观变化的适应能力更强. 在 wheel 序列中出现了明显的相机运动和快速运动, 尽管目标始终出现在镜头中央, 基线算法的预测框仍出现了错误的尺度估计. 在 soccer1 视频序列中存在严重的背景干扰和遮挡现象, 基线模型缺乏泛化能力, 容易受到背景和遮挡干扰, 尺度估计不准确. 相比之下, 本文算法的跟踪框的尺度更加接近真实框, 框内的前景比例更高, 即重叠率更高. 因此, SFGRT 算法在上述场景中有较好的稳定性和准确性.

实验 4. 算法实时性分析

不同目标跟踪算法的平均处理帧速率如表 6 所示, 包括基线算法 SiamRPN++、SiamMask、SiamBAN 和本文的 SFGRT 算法. SiamMask 算法的模型容量最小, 且平均处理帧速率最高. 其次是无锚框跟踪算法 SiamBAN. 本文算法只在网络权值上进行正则化约束优化, 因此模型大小与基线算法相等, 都是 431.2 MB, 且跟踪速度也基本一致, 二者数值之差在测量误差范围之内.

表 6　不同跟踪算法的模型大小和平均帧速率对比

Table 6　Comparison of model size and average framerate for different trackers

算法	出版	模型大小 (MB)	帧速率 (FPS)
SiamRPN++	CVPR2019	431.2	80.20
SiamMask	CVPR2019	86.1	106.43
SiamBAN	CVPR2020	430.9	81.76
SFGRT (Ours)	—	431.2	79.99

4　结论

本文旨在设计适应孪生网络跟踪框架的正则化, 提出了一种面向孪生网络的分段式细粒度正则化目标跟踪算法. 构造不同粒度的组套索惩罚函数应用于孪生网络, 分段式正则化策略解决了简单分层无法表征组间相关性学习问题. 另外, 本文发现在目标函数中多个惩罚函数会引起梯度量级不平衡的问题, 针对该问题采用梯度自平衡策略可进一步提升训练速度和模型的鲁棒性. 实验结果表明, 本文所提出的方法有效提高了跟踪器的鲁棒性, 基于 VOT2019、VOT2018、UAV123 和 LaSOT 4 个主流数据库, 与当前相关的主流 SOTA 方法进行比较, 本文所提 SFGRT 算法优于其他算法, 在鲁棒性上优势明显, 特别是在部分遮挡、尺度变化、光照、旋转和相机移动等挑战属性下具有良好的跟踪性能.

References

1 Xing D T, Evangeliou N, Tsoukalas A, Tzes A. Siamese transformer pyramid networks for real-time UAV tracking. In: Proceedings of IEEE/CVF Winter Conference on Applications of Computer Vision (WACV). Waikoloa, USA: IEEE, 2022. 1898–1907

2 Fang L P, Liang N X, Kang W X, Wang Z Y, Feng D D. Real-time hand posture recognition using hand geometric features and fisher vector. *Signal Processing: Image Communication*, 2020, **82**: Article No. 115729

3 Ballester I, Fontán A, Civera J, Strobl K H, Triebel R. DOT: Dynamic object tracking for visual SLAM. In: Proceedings of IEEE International Conference on Robotics and Automation (ICRA). Xi′an, China: IEEE, 2021. 11705–11711

4 Wu Y, Lim J, Yang M H. Object tracking benchmark. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2015, **37**(9): 1834–1848

5 Tang M, Yu B, Zhang F, Wang J Q. High-speed tracking with multi-kernel correlation filters. In: Proceedings of IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City, USA: IEEE, 2018. 4874–4883

6 Sun Y X, Sun C, Wang D, He Y, Lu H C. ROI pooled correlation filters for visual tracking. In: Proceedings of IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Long Beach, USA: IEEE, 2019. 5783–5791

7 Qiu Zhu-Ling, Zha Yu-Fei, Wu Min, Wang Qing. Learning attentional regularized correlation filter for visual tracking. *Acta Electronica Sinica*, 2020, **48**(9): 1762–1768
(仇祝令, 查宇飞, 吴敏, 王青. 基于注意力学习的正则化相关滤波跟踪算法. 电子学报, 2020, **48**(9): 1762–1768)

8 Zhu Jian-Zhang, Wang Dong, Lu Hu-Chuan. Learning temporal-spatial consistency correlation filter for visual tracking. *Scientia Sinica Informationis*, 2020, **50**(1): 128–150
(朱建章, 王栋, 卢湖川. 学习时空一致性相关滤波的视觉跟踪. 中国科学: 信息科学, 2020, **50**(1): 128–150)

9 Hu H W, Ma B, Shen J B, Shao L. Manifold regularized correlation object tracking. *IEEE Transactions on Neural Networks and Learning Systems*, 2018, **29**(5): 1786–1795

10 Xu T Y, Feng Z H, Wu X J, Kittler J. Joint group feature selection and discriminative filter learning for robust visual object tracking. In: Proceedings of IEEE/CVF International Conference on Computer Vision (ICCV). Seoul, South Korea: IEEE, 2019. 7949–7959

11 Zhang T Z, Xu C S, Yang M H. Multi-task correlation particle filter for robust object tracking. In: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Honolulu, USA: IEEE, 2017. 4819–4827

12 Huang Shu-Cheng, Zhang Yu, Zhang Tian-Zhu, Xu Chang-Sheng, Wang Zhi. Improved deep correlation filters via conditional random field. *Journal of Software*, 2019, **30**(4): 927–940
(黄树成, 张瑜, 张天柱, 徐常胜, 王直. 基于条件随机场的深度相关滤波目标跟踪算法. 软件学报, 2019, **30**(4): 927–940)

13 Zhang Wei-Jun, Zhong Sheng, Xu Wen-Hui, Wu Ying. Correlation filter based visual tracking integrating saliency and motion cues. *Acta Automatica Sinica*, 2021, **47**(7): 1572–1588
(张伟俊, 钟胜, 徐文辉, Wu Ying. 融合显著性与运动信息的相关滤波跟踪算法. 自动化学报, 2021, **47**(7): 1572–1588)

14 Guo Wen, You Si-Si, Gao Jun-Yu, Yang Xiao-Shan, Zhang Tian-Zhu, Xu Chang-Sheng. Deep relative metric learning for visual tracking. *Scientia Sinica Informationis*, 2018, **48**(1): 60–78
(郭文, 游思思, 高君宇, 杨小汕, 张天柱, 徐常胜. 深度相对度量学习的视觉跟踪. 中国科学: 信息科学, 2018, **48**(1): 60–78)

15 Bertinetto L, Valmadre J, Henriques J F, Vedaldi A, Torr P H S. Fully-convolutional Siamese networks for object tracking. In: Proceedings of the 14th European Conference on Computer Vision. Amsterdam, The Netherlands: Springer, 2016. 850–865

16 Krizhevsky A, Sutskever I, Hinton G E. ImageNet classification with deep convolutional neural networks. In: Proceedings of the 25th International Conference on Neural Information Processing Systems. Lake Tahoe, USA: ACM, 2012. 1097–1105

17 Li B, Yan J J, Wu W, Zhu Z, Hu X L. High performance visual tracking with Siamese region proposal network. In: Proceedings of IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City, USA: IEEE, 2018. 8971–8980

18 Li B, Wu W, Wang Q, Zhang F Y, Xing J L, Yan J J. SiamRPN++: Evolution of Siamese visual tracking with very deep networks. In: Proceedings of IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Long Beach, USA: IEEE, 2019. 4282–4291

19 Wang Q, Zhang L, Bertinetto L, Hu W M, Torr P H S. Fast online object tracking and segmentation: A unifying approach. In: Proceedings of IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Long Beach, USA: IEEE, 2019. 1328–1338

20 Zhang Z P, Peng H W. Deeper and wider Siamese networks for real-time visual tracking. In: Proceedings of IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Long Beach, USA: IEEE, 2019. 4591–4600

21 Chen Z D, Zhong B N, Li G R, Zhang S P, Ji R R. Siamese box adaptive network for visual tracking. In: Proceedings of IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Seattle, USA: IEEE, 2020. 6667–6676

22 Chen Z D, Zhong B N, Li G R, Zhang S P, Ji R R, Tang Z J, et al. SiamBAN: Target-aware tracking with Siamese box adaptive network. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, DOI: 10.1109/TPAMI.2022.3195759

23 Tan Jian-Hao, Zheng Ying-Shuai, Wang Yao-Nan, Ma Xiao-Ping. AFST: Anchor-free fully convolutional Siamese tracker with searching center point. *Acta Automatica Sinica*, 2021, **47**(4): 801–812
(谭建豪, 郑英帅, 王耀南, 马小萍. 基于中心点搜索的无锚框全卷积孪生跟踪器. 自动化学报, 2021, **47**(4): 801–812)

24 Chen X, Yan B, Zhu J W, Wang D, Yang X Y, Lu H C. Transformer tracking. In: Proceedings of IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Nashville, USA: IEEE, 2021. 8122–8131

25 Cao Z, Fu C H, Ye J J, Li B W, Li Y M. HiFT: Hierarchical feature transformer for aerial tracking. In: Proceedings of IEEE/CVF International Conference on Computer Vision (ICCV). Montreal, Canada: IEEE, 2021. 15437–15446

26 Xu T Y, Feng Z H, Wu X J, Kittler J. AFAT: Adaptive failure-aware tracker for robust visual object tracking. arXiv preprint arXiv: 2005.13708, 2020.

27 Kristan M, Matas J, Leonardis A, Felsberg M, Pflugfelder R, Kämäräinen J K, et al. The seventh visual object tracking VOT2019 challenge results. In: Proceedings of IEEE/CVF International Conference on Computer Vision Workshop (ICCVW). Seoul, South Korea: IEEE, 2019. 2206–2241

28 Kristan M, Leonardis A, Matas J, Felsberg M, Pflugfelder R, Zajc L Č, et al. The sixth visual object tracking VOT2018 challenge results. In: Proceedings of the 14th European Conference on Computer Vision Workshops. Munich, Germany: Springer, 2018. 3–53

29 Mueller M, Smith N, Ghanem B. A benchmark and simulator for UAV tracking. In: Proceedings of the 14th European Conference on Computer Vision. Amsterdam, The Netherlands: Spring-

er, 2016. 445−461

30 Fan H, Lin L T, Yang F, Chu P, Deng G, Yu S J, et al. LaSOT: A high-quality benchmark for large-scale single object tracking. In: Proceedings of IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Long Beach, USA: IEEE, 2019. 5369−5378

31 Zhu Z, Wang Q, Li B, Wu W, Yan J J, Hu W M. Distractor-aware Siamese networks for visual object tracking. In: Proceedings of the 15th European Conference on Computer Vision. Munich, Germany: Springer, 2018. 103−119

32 He K M, Zhang X Y, Ren S Q, Sun J. Deep residual learning for image recognition. In: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Las Vegas, USA: IEEE, 2016. 770−778

33 He A F, Luo C, Tian X M, Zeng W J. A twofold Siamese network for real-time object tracking. In: Proceedings of IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City, USA: IEEE, 2018. 4834−4843

34 Wang Q, Teng Z, Xing J L, Gao J, Hu W M, Maybank S. Learning attentions: Residual attentional Siamese network for high performance online visual tracking. In: Proceedings of IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City, USA: IEEE, 2018. 4854−4863

35 Du F, Liu P, Zhao W, Tang X L. Correlation-guided attention for corner detection based visual tracking. In: Proceedings of IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Seattle, USA: IEEE, 2020. 6835−6844

36 Li F, Tian C, Zuo W M, Zhang L, Yang M H. Learning spatial-temporal regularized correlation filters for visual tracking. In: Proceedings of IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City, USA: IEEE, 2018. 4904−4913

37 Huang Z Y, Fu C H, Li Y M, Lin F L, Lu P. Learning aberrance repressed correlation filters for real-time UAV tracking. In: Proceedings of IEEE/CVF International Conference on Computer Vision (ICCV). Seoul, South Korea: IEEE, 2019. 2891−2900

38 Li Y M, Fu C H, Ding F Q, Huang Z Y, Lu G. AutoTrack: Towards high-performance visual tracking for UAV with automatic spatio-temporal regularization. In: Proceedings of IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Seattle, USA: IEEE, 2020. 11920−11929

39 Liu X N, Zhou Y, Zhao J Q, Yao R, Liu B, Zheng Y. Siamese convolutional neural networks for remote sensing scene classification. *IEEE Geoscience and Remote Sensing Letters*, 2019, **16**(8): 1200−1204

40 Fiaz M, Mahmood A, Baek K Y, Farooq S S, Jung S K. Improving object tracking by added noise and channel attention. *Sensors*, 2020, **20**(13): Article No. 3780

41 Jia S, Ma C, Song Y B, Yang X K. Robust tracking against adversarial attacks. In: Proceedings of the 16th European Conference on Computer Vision. Glasgow, UK: Springer, 2020. 69−84

42 Yuan M, Lin Y. Model selection and estimation in regression with grouped variables. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 2006, **68**(1): 49−67

43 Nie F P, Huang H, Cai X, Ding C. Efficient and robust feature selection via joint ℓ_2, 1 -norms minimization. In: Proceedings of the 23rd International Conference on Neural Information Processing Systems. Vancouver, Canada: ACM, 2010. 1813−1821

44 Bach F, Jenatton R, Mairal J, Obozinski G. Structured sparsity through convex optimization. *Statistical Science*, 2012, **27**(4): 450−468

45 Yoon J, Hwang S J. Combined group and exclusive sparsity for deep neural networks. In: Proceedings of the 34th International Conference on Machine Learning. Sydney, Australia: JMLR.org, 2017. 3958−3966

46 Hu Y H, Li C, Meng K W, Qin J, Yang X Q. Group sparse optimization via $L_{p,q}$ regularization. *The Journal of Machine Learning Research*, 2017, **18**(1): 960−1011

47 Wen W, Wu C P, Wang Y D, Chen Y R, Li H. Learning structured sparsity in deep neural networks. In: Proceedings of the 30th International Conference on Neural Information Processing Systems. Barcelona, Spain: ACM, 2016. 2082−2090

48 Szegedy C, Zaremba W, Sutskever I, Bruna J, Erhan D, Goodfellow I, et al. Intriguing properties of neural networks. arXiv preprint arXiv: 1312.6199, 2013.

49 Chen Z, Badrinarayanan V, Lee C Y, Rabinovich A. GradNorm: Gradient normalization for adaptive loss balancing in deep multitask networks. In: Proceedings of the 35th International Conference on Machine Learning. Stockholm, Sweden: PMLR, 2018. 793−802

50 Lin T Y, Maire M, Belongie S, Hays J, Perona P, Ramanan D, et al. Microsoft COCO: Common objects in context. In: Proceedings of the 13th European Conference on Computer Vision. Zurich, Switzerland: Springer, 2014. 740−755

51 Russakovsky O, Deng J, Su H, Krause J, Satheesh S, Ma S A, et al. ImageNet large scale visual recognition challenge. *International Journal of Computer Vision*, 2015, **115**(3): 211−252

52 Real E, Shlens J, Mazzocchi S, Pan X, Vanhoucke V. YouTube-BoundingBoxes: A large high-precision human-annotated data set for object detection in video. In: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Honolulu, USA: IEEE, 2017. 7464−7473

53 Huang L H, Zhao X, Huang K Q. GOT-10k: A large high-diversity benchmark for generic object tracking in the wild. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2021, **43**(5): 1562−1577

54 Guo D Y, Wang J, Cui Y, Wang Z H, Chen S Y. SiamCAR: Siamese fully convolutional classification and regression for visual tracking. In: Proceedings of IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Seattle, USA: IEEE, 2020. 6268−6276

55 Dong X P, Shen J B, Shao L, Porikli F. CLNet: A compact latent network for fast adjusting Siamese trackers. In: Proceedings of the 16th European Conference on Computer Vision. Glasgow, UK: Springer, 2020. 378−395

56 Yang T Y, Xu P F, Hu R B, Chai H, Chan A B. ROAM: Recurrently optimizing tracking model. In: Proceedings of IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Seattle, USA: IEEE, 2020. 6717−6726

57 Danelljan M, Bhat G, Khan F S, Felsberg M. ATOM: Accurate tracking by overlap maximization. In: Proceedings of IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Long Beach, USA: IEEE, 2019. 4660−4669

58 Yan B, Zhao H J, Wang D, Lu H C, Yang X Y. 'Skimming-perusal' tracking: A framework for real-time and robust long-term tracking. In: Proceedings of IEEE/CVF International Conference on Computer Vision (ICCV). Seoul, South Korea: IEEE, 2019. 2385−2393

59 Fan H, Ling H B. Siamese cascaded region proposal networks for real-time visual tracking. In: Proceedings of IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Long Beach, USA: IEEE, 2019. 7952−7961

60 Yang T Y, Chan A B. Visual tracking via dynamic memory networks. *IEEE Transactions on Pattern Analysis and Machine In-*

telligence, 2021, **43**(1): 360−374

61 Xu T Y, Feng Z H, Wu X J, Kittler J. Learning adaptive discriminative correlation filters via temporal consistency preserving spatial feature selection for robust visual object tracking. *IEEE Transactions on Image Processing*, 2019, **28**(11): 5596−5609

62 Danelljan M, Bhat G, Khan F S, Felsberg M. ECO: Efficient convolution operators for tracking. In: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Honolulu, USA: IEEE, 2017. 6931−6939

63 Wang G T, Luo C, Xiong Z W, Zeng W J. SPM-Tracker: Series-parallel matching for real-time visual object tracking. In: Proceedings of IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Long Beach, USA: IEEE, 2019. 3643−3652

安志勇 山东工商学院计算机科学与技术学院副教授. 2008 年获得西安电子科技大学计算机系统结构专业博士学位. 主要研究方向为计算机视觉, 目标跟踪.
E-mail: azytyut@163.com
(**AN Zhi-Yong** Associate professor at the School of Computer Science and Technology, Shandong Technology and Business University. He received his Ph.D. degree in computer system architecture from Xidian University in 2008. His research interest covers computer vision and object tracking.)

梁顺楷 山东工商学院计算机科学与技术学院硕士研究生. 2019 年获得广东工业大学物联网工程专业学士学位. 主要研究方向为计算机视觉, 目标跟踪.
E-mail: keith1063@163.com
(**LIANG Shun-Kai** Master student at the School of Computer Science and Technology, Shandong Technology and Business University. He received his bachelor degree in internet of things from Guangdong University of Technology in 2019. His research interest covers computer vision and object tracking.)

李 博 山东工商学院计算机科学与技术学院副教授. 2013 年获得东北大学计算机系统结构专业博士学位. 主要研究方向为人工智能, 机器学习. 本文通信作者.
E-mail: libokkkkk@sdtbu.edu.cn
(**LI Bo** Associate professor at the School of Computer Science and Technology, Shandong Technology and Business University. He received his Ph.D. degree in computer system architecture from Northeastern University in 2013. His research interest covers artificial intelligence and machine learning. Corresponding author of this paper.)

赵 峰 山东工商学院计算机科学与技术学院教授. 2008 年获得西安电子科技大学计算机应用技术专业博士学位. 主要研究方向为人工智能, 机器学习, 医学图像分析和金融大数据分析. E-mail: zhaofeng1016@126.com
(**ZHAO Feng** Professor at the School of Computer Science and Technology, Shandong Technology and Business University. He received his Ph.D. degree in computer application technology from Xidian University in 2008. His research interest covers artificial intelligence, machine learning, medical image analysis, and financial big data analysis.)

窦全胜 山东工商学院计算机科学与技术学院教授. 2005 年获得吉林大学计算机应用技术专业博士学位. 主要研究方向为计算智能, 数据挖掘, 知识工程和知识处理.
E-mail: douqsh@sdtbu.edu.cn
(**DOU Quan-Sheng** Professor at the School of Computer Science and Technology, Shandong Technology and Business University. He received his Ph.D. degree in computer application technology from Jilin University in 2005. His research interest covers computational intelligence, data mining, knowledge engineering, and knowledge management.)

相忠良 山东工商学院计算机科学与技术学院讲师. 2015 年获得韩国东西大学信息技术专业博士学位. 主要研究方向为机器学习, 贝叶斯网络学习.
E-mail: zlxiang@sdtbu.edu.cn
(**XIANG Zhong-Liang** Lecturer at the School of Computer Science and Technology, Shandong Technology and Business University. He received his Ph.D. degree in information technology from Dongseo University in 2015. His research interest covers machine learning and Bayesian network learning.)