

基于深度空间正则化的相关滤波跟踪算法

蒲磊¹,冯新喜²,侯志强³,查宇飞⁴,余旺盛²

(1. 空军工程大学研究生院,陕西西安 710077; 2. 空军工程大学信息与导航学院,陕西西安 710077;
3. 西安邮电大学计算机学院,陕西西安 710121; 4. 西北工业大学计算机学院,陕西西安 710072)

摘要: 近年来,结合深度特征的相关滤波算法由于较高的跟踪精度在视觉跟踪领域受到了广泛的关注. 对训练样本的周期性假设一方面提高了计算效率,但是也引入了边界效应,限制了算法性能的进一步提升. 通过对深度特征表达能力的深入挖掘,本文提出了一种新的跟踪框架. 由于深层特征具有良好的语义信息,选取 VGG 网络第五层卷积特征提取目标的空间可靠区域,将该区域信息用于对样本进行裁剪并引入目标函数,建立空间约束模型,接着采用 ADMM 算法进行迭代求解. 为了进一步提高算法的长时跟踪能力,提出一种简单有效的遮挡判断方法. 实验结果表明,所提出的算法在跟踪精度和成功率上优于大多数先进的算法.

关键词: 视觉跟踪; 空间正则化; 深度特征; 相关滤波; 模型更新

中图分类号: TP391.4 **文献标识码:** A **文章编号:** 0372-2112 (2020)10-2025-08

电子学报 URL: <http://www.ejournal.org.cn> **DOI:** 10.3969/j.issn.0372-2112.2020.10.021

Correlation Filter Tracking Based on Deep Spatial Regularization

PU Lei¹, FENG Xin-xi², HOU Zhi-qiang³, ZHA Yu-fei⁴, YU Wang-sheng²

(1. Graduate College, Air Force Engineering University, Xi'an, Shaanxi 710077, China;

2. Institute of Information and Navigation, Air Force Engineering University, Xi'an, Shaanxi 710077, China;

3. School of Computer Science and Technology, Xi'an University of Posts and Telecommunications, Xi'an, Shaanxi 710121, China;

4. School of Computer Science, Northwestern Polytechnical University, Xi'an, Shaanxi 710072, China)

Abstract: In recent years, the correlation filter based algorithm combined with deep features has received extensive attention. The period assumption of the training samples improves the computational efficiency, but also introduces the boundary effect, which limits the further improvement of the tracking performance. By exploring the deep feature representation ability, a new tracking framework is proposed. Since the deep features have good semantic information, the fifth layer convolution feature of VGG network is used to extract the spatially reliable region of the target, and the region information is introduced into the objective function to establish a spatial constraint model. Then iteratively solved by ADMM algorithm. In order to further improve the long-time tracking ability, a simple and effective method of occlusion detection is proposed. Experimental results show that the proposed algorithm outperforms most advanced algorithms in tracking accuracy and success rate.

Key words: visual tracking; spatial regularization; deep features; correlation filter; model update

1 引言

近年来,计算机视觉受到广泛的关注与研究,在图像分类^[1]、目标检测^[2]、图像分割^[3]等多个领域都取得了长足的发展. 而视觉跟踪作为计算机视觉的一个子问题,也是当前的热点研究方向之一^[4-7]. 在给定初始帧目标状态信息的前提下,视觉跟踪的任务是实现后续帧目标状态信息(位置、尺度等)的准确估计. 尽管经过了几十年的发展,实现鲁棒稳定的跟踪依然是一

个很困难的问题. 因为现实场景中跟踪任务总是受到各种各样的挑战,比如遮挡、背景杂波、快速运动等复杂场景,训练样本有限以及跟踪算法对实时性的需求. 因此,如何从非常有限的训练样本中构建一个用于区分目标和背景的鲁棒外观模型至关重要.

自2010年以来,基于相关滤波的跟踪算法获得广泛的关注并取得了很多的研究成果,几乎主导了视觉跟踪领域. 相关滤波算法通过循环样本和高斯标签对模型进行训练,并利用快速傅里叶变换加速学习和检

测. Bolme 等人^[8]提出了 MOSSE 算法,在灰度特征上实现了极高的跟踪速度. Henriques 等人在此基础上,提出了 CSK^[9],引入循环矩阵用于对训练样本稠密采样进行近似,同时也可以利用快速傅里叶变换进行求解.为了提升特征的表达能力, Danelljan 和 Henriques 先后将多通道颜色特征和 Hog 特征引入进来,提出了 CN^[10]和 KCF^[11]算法,进一步提高了跟踪精度.与 MOSSE 相比, KCF 算法在 OTB2013 上利用 HOG 特征和高斯核在保证较高速度的前提下,获得了更高的跟踪性能.然而对样本的周期性假设,使得训练样本普遍不是真实样本,产生了边界效应.严重降低了模型的判别力,进一步限制了算法性能的进一步提升.

在近几年的研究中,许多算法成功将深度特征的强大表达能力和相关滤波的快速计算进行结合,在大量的数据集上均取得了很大的性能提升.基于深度特征的相关滤波方法主要就是将传统的手工特征替换为在 ImageNet^[12]上预训练模型的深度特征(VGG^[13], ResNet^[14]等),来作为目标的外观模型. Ma 等人^[15]较早地将多层深度特征和相关滤波算法进行结合,并对深度特征不同层的表达能力进行了研究,提出了一种由粗到精的融合定位方法.不同于 HCF 采用固定权值对多层特征进行融合估计, Qi 等人^[16]考虑到环境的复杂变化,采用自适应权重的方式取得了很好的跟踪效果.在特征表达上,这些方法主要是将传统的 hog 特征替换为深度特征,再用于对滤波器进行训练. Valmadre 等人^[17]从整体的思路出发,将相关滤波器嵌入到神经网络中,作为和卷积层类似的一个模块,实现了端到端的训练.虽然这些方法实现了精度和鲁棒性的提升,但是仍然受到循环矩阵周期性假设产生的边界效应的影响,同时循环卷积要求滤波器和搜索区域大小一致,这就进一步限制了该算法的检测范围.

为了缓解这一问题,诸如 CFLB^[18]、SRDCF^[19]、BACF^[20]、STRCF^[21]等方法相继被提出,通过压制模型边界的响应,提高模型对目标和背景的判别能力,从而一定程度地缓解边界效应.在 SRDCF 中, Danelljan 等人采用逆高斯约束来代替原来的 L2 正则项,强制性地对边界响应进行抑制.另一方面, CFLB 以及后续的改进版本 BACF 从样本的角度出发,在更大的搜索区域内裁剪小范围固定尺寸的目标区域,增加了真实样本的数量,并且也扩大了滤波器的检测范围,这些策略都一定程度上缓解了边界效应.

为了充分挖掘深度特征的语义表达能力,本文提出了一种基于深度空间正则化的相关滤波算法.将从预训练模型提取的特征一方面用于建立目标的外观模型,另一方面也用于提取目标的空间分布区域,将该区域图用二值矩阵进行表示并嵌入到目标函数中,相比

于固定的矩形框,该方法可以更好地减少目标区域内背景的干扰,一方面减轻了边界效应,一方面也提升了模型的判别力.由于提出的目标函数没有闭合解,本文通过 ADMM^[22]进行快速求解,同时在大量的数据集上取得了很好的效果.

2 本文算法

本文提出了基于深度空间正则化的相关滤波跟踪算法.首先提取 VGG 网络的第四层特征用于建立目标的外观模型,并用于滤波器训练,接着利用 VGG 网络的第五层特征提取目标的空间分布,并转换为二值矩阵,用于对模型进行约束,最后采用 ADMM 进行快速求解.

2.1 多通道相关滤波算法

在详细介绍本文算法之前,首先对传统的相关滤波算法进行简单地介绍.相关滤波算法以判别式学习为核心,并引入循环矩阵实现对样本密集采样的近似,接着通过快速傅里叶变换在频域内进行求解,取得了较好的跟踪性能和较高的计算效率,引起了人们的广泛关注.相关滤波算法可以表示为一个岭回归问题,定义如下损失函数:

$$L(\mathbf{h}) = \min_{\mathbf{h}} \|\mathbf{f} \otimes \mathbf{h} - \mathbf{g}\|^2 + \lambda \|\mathbf{h}\|^2 \quad (1)$$

其中, $\mathbf{f} = \{\mathbf{f}_d\}_{d=1,D}$ 表示提取的多通道特征, D 为通道数, $\mathbf{h} = \{\mathbf{h}_d\}_{d=1,D}$ 代表与特征对应的滤波器, \otimes 表示对特征和滤波器进行循环相关运算, \mathbf{g} 代表输出标签,通常采用二维高斯函数表示, $\mathbf{f}_d, \mathbf{h}_d, \mathbf{g} \in \mathbb{R}^{c_s \times c_s}$. λ 是正则化系数,用于防止模型过拟合,提高泛化性能.为了加速运算,一般将式(1)变换到傅里叶域进行求解:

$$L(\mathbf{h}) = \min_{\mathbf{h}} \left\| \sum_{d=1}^D \text{diag}(\hat{\mathbf{f}}_d) \bar{\mathbf{h}}_d - \hat{\mathbf{g}} \right\|^2 + \lambda \sum_{d=1}^D \|\hat{\mathbf{h}}_d\|^2 \quad (2)$$

其中, $\hat{\mathbf{f}}_d$ 表示对 \mathbf{f}_d 进行傅里叶变换,并且将变换后的矩阵拉伸成一个列向量, $\hat{\mathbf{f}}_d \in \mathbb{R}^{c_s \times c_s}$. $\text{diag}(\hat{\mathbf{f}}_d)$ 表示一个 $D \times D$ 的对称矩阵. $\bar{\mathbf{h}}_d$ 表示对 $\hat{\mathbf{h}}_d$ 求复共轭操作,对单个通道的滤波器可以计算如下:

$$\hat{\mathbf{h}}_d = (\text{diag}(\hat{\mathbf{f}}_d) \bar{\hat{\mathbf{g}}}) \odot^{-1} \left(\sum_{d=1}^D \text{diag}(\hat{\mathbf{f}}_d) \bar{\hat{\mathbf{f}}}_d + \lambda \right) \quad (3)$$

式中, \odot^{-1} 代表点除操作.上述方法的主要问题在于训练样本受到边界效应的影响,除了初始样本,其他样本都是通过循环移位产生的虚假样本.由于缺乏真实的负样本,使得滤波器的判别能力受到一定的限制,同时由于样本数较少,容易发生过拟合,使得泛化性能降低.

2.2 深度空间正则化相关滤波模型

为了使得滤波器可以学到更多关于目标本身的信息,并尽可能的降低背景信息的影响,考虑到高层深度特征对目标的良好语义表达,本文提出了基于深度空间正则化的相关滤波模型.引入目标空间可靠分布区

域作为掩膜,直接作用于滤波器模板,将目标以外的区域置零,提出的目标函数如下所示:

$$L(\mathbf{h}) = \min_{\mathbf{h}} \|\mathbf{f} \otimes (\mathbf{p} \odot \mathbf{h}) - \mathbf{g}\|^2 + \lambda \|\mathbf{h}\|^2 \quad (4)$$

其中 \mathbf{p} 为空间可靠性二值矩阵,目标区域为 1,背景区域为 0, $\mathbf{p} \in \mathbb{R}^{c_w \times c_h}$, \odot 代表点乘操作. 这样处理后该函数便不再有闭合解,只有通过 ADMM 方法进行迭代求解.

首先我们引入对偶变量 \mathbf{h}_c :

$$\mathbf{h}_c - \mathbf{p} \odot \mathbf{h} = \mathbf{0} \quad (5)$$

建立如下增广拉格朗日方程:

$$\begin{aligned} L(\hat{\mathbf{h}}_c, \mathbf{h}, \hat{\mathbf{I}} | \mathbf{p}) = & \|\text{diag}(\hat{\mathbf{f}}) \hat{\mathbf{h}}_c - \hat{\mathbf{g}}\|^2 + \frac{\lambda}{2} \|\mathbf{h}_p\|^2 \\ & + [\hat{\mathbf{I}}^H (\hat{\mathbf{h}}_c - \hat{\mathbf{h}}_p) + \overline{\hat{\mathbf{I}}^H (\hat{\mathbf{h}}_c - \hat{\mathbf{h}}_p)}] \\ & + \mu \|\hat{\mathbf{h}}_c - \hat{\mathbf{h}}_p\|^2 \end{aligned} \quad (6)$$

其中 $\hat{\mathbf{I}}$ 表示拉格朗日乘子, $\mu > 0$, 采用 $\mathbf{h}_p \equiv (\mathbf{p} \odot \mathbf{h})$ 进行运算. 通过 ADMM 迭代求解可以将问题分为以下的两个子问题:

$$\hat{\mathbf{h}}_c^{i+1} = \underset{\hat{\mathbf{h}}_c}{\text{argmin}} L(\hat{\mathbf{h}}_c, \mathbf{h}^i, \hat{\mathbf{I}}^i | \mathbf{p}) \quad (7)$$

$$\mathbf{h}^{i+1} = \underset{\mathbf{h}}{\text{argmin}} L(\hat{\mathbf{h}}_c^{i+1}, \mathbf{h}, \hat{\mathbf{I}}^i | \mathbf{p}) \quad (8)$$

拉格朗日乘子的更新方式为:

$$\hat{\mathbf{I}}^{i+1} = \hat{\mathbf{I}}^i + \mu (\hat{\mathbf{h}}_c^{i+1} - \hat{\mathbf{h}}^{i+1}) \quad (9)$$

可以各自得到如下闭合解:

$$\hat{\mathbf{h}}_c^{i+1} = (\hat{\mathbf{f}} \odot \hat{\mathbf{g}} + (\mu \hat{\mathbf{h}}_c^i - \hat{\mathbf{I}}^i)) \odot^{-1} (\hat{\mathbf{f}} \odot \hat{\mathbf{f}} + \mu^i) \quad (10)$$

$$\mathbf{h}^{i+1} = \mathbf{p} \odot \mathbf{F}^{-1} [\hat{\mathbf{I}}^i + \mu^i \hat{\mathbf{h}}_c^{i+1}] / \left(\frac{\lambda}{2D} + \mu^i \right) \quad (11)$$

其中, \mathbf{F}^{-1} 表示傅里叶反变换, μ 的更新方式为 $\mu^{i+1} = \beta \mu^i$.

2.3 提取目标空间可靠分布区域

在 Kiani Galoogahi 和 Danelljan 等人的早期工作中,都曾考虑采用空间可靠区域对滤波器进行约束学习,其中 Galoogahi 等人提出的 CFLB 在扩大搜索区域的前提下,采用一个较小的矩形区域对滤波器进行约束,以此来减轻边界效应的影响. 但是目标经常会进行旋转变形,使得矩形的约束很难对目标区域进行很好的近似,导致引入大量的背景信息,降低了滤波器的判别性能. 而 Danelljan 等人提出的 SRDCF 采用一个逆高斯分布的空间矩阵来约束滤波器的响应值大小,该空间图根据滤波器大小先验确定,在跟踪过程中保持不变,难以适应目标在跟踪过程中的各种变化.

为了更好地适应目标的变化,同时进一步缓解边界效应,本文充分发掘深度特征对目标的良好表示和定位能力,利用高层特征提取目标的空间范围. 接着将其用于对滤波器训练进行空间约束,使滤波器可以学到更多目标区域的特征.

本文中我们提取 VGG 网络的第四层特征用于定

位,提取第五层特征用于提取空间区域了,特征可以表示为 $\mathbf{f} \in \mathbb{R}^{c_w \times c_h \times D}$. 从通道的方向, \mathbf{f} 可以看做由 $c_w \times c_h$ 个 D 维向量组成,将每个向量称为一个深度描述子^[23],用 \mathbf{x} 表示. 由于单个图像难以获得稳定的目标表征,本文采用连续的多个目标图像进行协同处理,以获得他们之间的内在相关性. 这里我们通过 PCA 算法^[24]对 N 个连续图像的深度特征描述子 $(\mathbf{X}^1, \dots, \mathbf{X}^N)$ 进行相关性分析,其中 $\mathbf{X} = \{\mathbf{x}_{i,j} \in \mathbb{R}^D\}, i \in 1, \dots, c_h, j \in 1, \dots, c_w$. 首先得到所有描述子的均值向量:

$$\bar{\mathbf{x}} = \frac{1}{K} \sum_n \sum_{i,j} \mathbf{x}_{(i,j)}^n \quad (12)$$

其中 $K = c_h \times c_w \times N$. 接着计算协方差矩阵:

$$\text{Cov}(\mathbf{x}) = \frac{1}{K} \sum_n \sum_{i,j} (\mathbf{x}_{(i,j)}^n - \bar{\mathbf{x}}) (\mathbf{x}_{(i,j)}^n - \bar{\mathbf{x}})^T \quad (13)$$

计算协方差矩阵的特征值和特征向量,按照特征值得大小 $\lambda_1 \geq \dots \geq \lambda_d \geq 0$ 得到特征向量 ξ_1, \dots, ξ_d .

由于第一主成分具有最大方差,我们将特征向量 ξ_1 作为主投影方向. 在一幅图像中位置 (i, j) 的深度特征描述子,计算它的主分量 p^1 为

$$p_{(i,j)}^1 = \xi_1^T (\mathbf{x}_{(i,j)} - \bar{\mathbf{x}}) \quad (14)$$

依据空间位置,所有来自同一图像的 $p_{(i,j)}^1$ 被组合为一个维度为 $c_h \times c_w$ 的二维矩阵. 我们将这个矩阵称为指示矩阵:

$$\mathbf{P}^1 = \begin{bmatrix} p_{(1,1)}^1 & p_{(1,2)}^1 & \dots & p_{(1,c_w)}^1 \\ p_{(2,1)}^1 & p_{(2,2)}^1 & \dots & p_{(2,c_w)}^1 \\ \vdots & \vdots & \ddots & \vdots \\ p_{(c_h,1)}^1 & p_{(c_h,2)}^1 & \dots & p_{(c_h,c_w)}^1 \end{bmatrix} \quad (15)$$

\mathbf{P}^1 矩阵中的正值代表深度描述子之间的正相关,负值代表负相关. 其中值的绝对值越大,代表相关性越高. 由于 ξ_1 是在多个图像上获得的,正相关区域可以表示出目标在这些图像中的位置. 将零作为区分背景和目标的阈值,矩阵中大于零的区域表示目标区域,小于零的区域表示背景区域. 另外,如果矩阵中没有正值,表示该图像中没有目标,可以用来对目标的遮挡丢失进行检测判断,并作为模型更新的依据. 可以获得 \mathbf{P}^1 矩阵中为正值最大连接区域作为目标区域.

由于从第五层特征得到的指示矩阵的维度和滤波器维度不一致,我们采用双线性插值进行上采样,在插值后的指示矩阵上将为正值的区域设为 1,负值为 0,并获得最大联通区域以去掉部分杂波点的影响,这样就得到了我们需要的空间可靠性区域矩阵 \mathbf{p} :

$$p_{i,j} = \begin{cases} 1, & \text{if } P_{i,j}^1 > 0 \\ 0, & \text{otherwise} \end{cases} \quad (16)$$

2.4 模型更新策略

在跟踪过程中,模型更新是很关键的一步. 尤其是

在相关滤波跟踪框架下,良好的模型更新策略将提高算法对复杂场景的处理能力,以防止出现跟踪漂移甚至目标丢失. 现有的方法大多采用每帧更新或者固定间隔帧更新策略^[25],在目标出现长时遮挡的时候难以应对. 在本文中,我们通过观察空间可靠性矩阵的响应情况,提出了一种遮挡判断方法. 如果出现目标遮挡或者丢失,指示矩阵大部分均为负值.

设定更新阈值 ψ ,当正值数量 $N_p > \psi$ 时,则通过下式进行滤波器模型更新以及目标统计模型的更新,目标统计模型的更新:

$$\mathbf{h}_i = (1 - \alpha)\mathbf{h}_{i-1} + \alpha\mathbf{h} \quad (17)$$

式中, \mathbf{h}_i 表示当前帧更新后的滤波器模板, \mathbf{h}_{i-1} 表示更新前的模板, \mathbf{h} 为当前帧提取的目标直方图, α 为更新率.

当正值数量 $N_p \leq \psi$ 时,减小更新率,目标统计模型的更新:

$$\mathbf{h}_i = \left(1 - \frac{\alpha}{4}\right)\mathbf{h}_{i-1} + \frac{\alpha}{4}\mathbf{h} \quad (18)$$

2.5 算法流程

本文跟踪算法的主要算法流程如图 1 所示. 我们采用在 ImageNet 数据集上训练的 VGGNet-19 作为特征提取器. 首先移除全连接层,并使用 Cov4-5 层和 Cov5-5 层的输出作为深层特征.

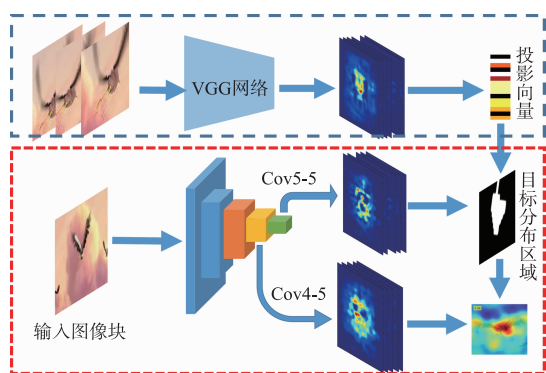


图1 算法流程

图 1 中蓝色虚线框代表初始化部分,通过提取初始连续的多帧图像进行协同处理,得到关于当前目标表示的主成分投影向量. 红色框表示滤波器训练过程,将当前帧图片输入 VGG 网络中,提取 Cov5-5 层特征和预先计算的投影向量进行计算得到相关性指示矩阵 \mathbf{p} ,将该矩阵进行上采样后用于约束滤波器的响应范围,使得滤波器系数主要集中在目标区域. 检测过程和传统的相关滤波算法一致.

3 仿真实验

为了充分验证本文算法的有效性,在多个数据集上对算法性能进行了对比评估. 首先采用 MATLAB 和

C++ 混合编程对本文算法进行了实现,并利用 MatConvNet 工具箱^[26]实现对预训练神经网络的前向传播. 为了进一步提高效率,在 Intel Xeon 2.4GHz 的处理器上对本文算法进行测试,并利用 GPU (GTX 1080Ti/64GB RAM)进行加速.

测试参数设置如下:将 VGG19 预训练模型作为特征提取器,去掉全连接层,将 Cov4-5 卷积层的输出作为目标特征表达, Cov5-5 层的特征用于进行相关性运算并提取空间可靠区域. 为了消除边界不连续性,我们通过汉明窗对每个卷积层的提取特征通道进行边界平滑. 初始化的时候提取连续的 5 帧图像计算主成分向量. 为了增加检测范围,目标搜索框大小是目标大小的 4 倍. 对于式(4)中的正则化参数 $\lambda = 10^{-4}$,尺度估计采用和文献[27]相同的参数设置. 增广拉格朗日系数为 u 为 5,更新阈值 ψ 设为 10. 我们在 OTB2013^[28] 和 OTB2015^[29] 数据集上将本文算法与其他 11 种主流的跟踪算法进行比较,这些方法包括:

- (1) 基于空间正则化的跟踪器,如 SRDCF^[19], STRCF^[21], SRDCFdecon^[30];
- (2) 基于 CNN 的跟踪器,如 HCF^[15], HDT^[16], SiamFC^[31], DCFNet^[32] 和 DeepSRDCF^[33];
- (3) 基于 CF 的跟踪器,如 KCF^[11], DSST^[27] 和 Staple^[34].

这些算法均使用默认参数,本文采用跟踪精度、成功率以及跟踪速度对算法性能进行评估.

3.1 定性分析

定性分析主要将本文算法和 STRCF、KCF、DCFNet、SRDCF、SiamFC、HCF、DeepSRDCF 和 HDT 在五个具有挑战性的图像视频上进行对比,包括 bird1, CarScale, MotorRolling, clifBar, Soccer. 如图 2 所示,本文算法能够更好地应对这些复杂场景.

(1) 目标遮挡:以“bird1”和“Soccer”为例,目标在跟踪过程中受到背景遮挡,在“bird1”中 144 帧以后目标完全被云雾遮挡,在此期间得到的指示矩阵,普遍为负值,本文算法根据指示矩阵的变化情况调整模型更新率,使得当目标重新出现能够准确将其捕获,而其它算法由于模型破坏严重,无法重新检测到目标;在“Soccer”第 100 帧到 200 帧左右中,目标也一直受到背景一定程度的遮挡,本文算法同样可以进行有效的搜寻和跟踪.

(2) 尺度变化:在“MotorRolling”、“CarScale”、“clifBar”中都存在不同程度地尺度变化. 本文算法由于对滤波器的抑制考虑到了尺度的影响,使得算法对尺度变化有着更好的适应性,可以获得更加准确的定位精度,在此基础上通过尺度金字塔的方式可以获得更加稳定的跟踪结果.



图2 9种算法的部分跟踪结果对比

(3) **快速运动**:目标外观发生变化会导致当前帧信息与模板匹配程度降低,增加跟踪难度,特别是对于相关滤波而言,边界效应会导致边缘的样本割裂,快速运动的目标难以得到较好的响应.例如“MotorRolling”和“clifBar”,目标的快速运动和形变增加了跟踪难度,本文算法由于对滤波器模板的响应范围进行了约束,使得算法可以获得更加的检测区域,对快速运动有着较好的应对能力.

(4) **目标形变**:以“MotorRolling”为例,目标在跟踪过程中不断发生形态变化或者空间里的旋转,导致模板难以一直保持较高的匹配,部分算法会出现跟踪漂

移的现象.传统的基于 HOG 特征的 KCF、SRDCF、STRCF 等算法均无法稳定的跟踪目标,而采用深度特征的 HCF 和本文算法可以实现很好的跟踪表现,进一步证明了深度特征具有良好的语义表达能力.

3.2 整体性能

跟踪基准 (OTB) 数据集包含两组:OTB2013, 51 个序列和 OTB2015, 具有 100 个序列.

我们采用跟踪精度、成功率、跟踪速度三个指标在 OTB 数据集上对算法的性能进行评估.图 3 和图 4 显示了跟踪精度和成功率的曲线图.总体而言,该算法表现良好.通过与基于空间正则化的方法比较,本文算法显

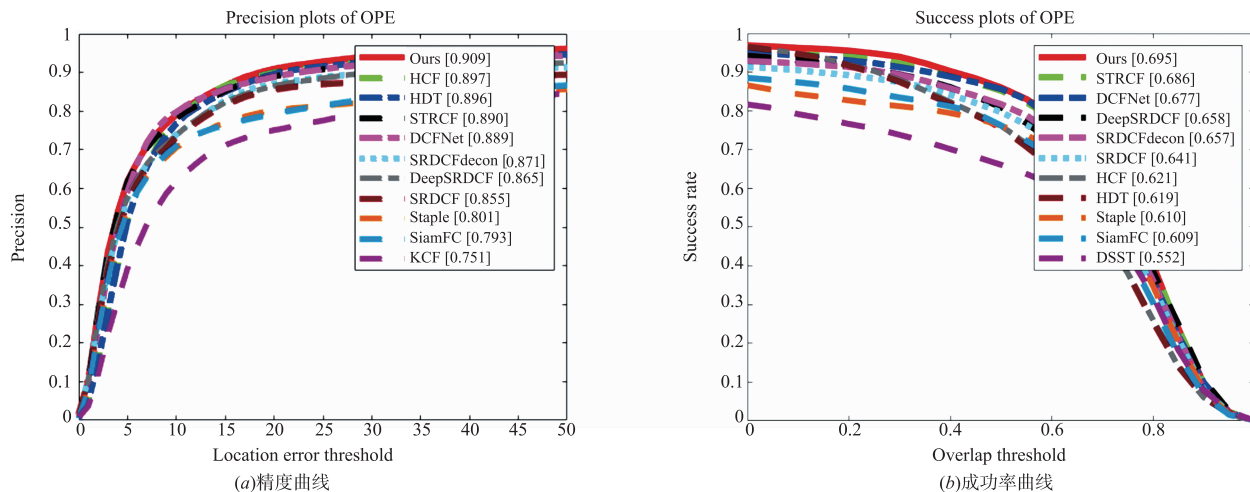


图3 OTB2013数据集上的精度和成功率曲线

著优于传统方法,在 OTB2013 和 OTB2015 上均优于 SRDCF、STRCF 等方法,在 OTB2013 上的跟踪精度和成功率分别为 0.909 和 0.695,在 OTB2015 上达到了 0.875 和 0.662,相比于同样基于深度特征和空间正则化的 DeepSRDCF 算法,在 OTB2015 上精度提升了

1.6%,成功率上提升了 1.9%。与基于 CNN 的方法比较,本文方法优于基于多层深度特征融合的 HCF、HDT 算法,与基于 Siamese 网络的 DCFNet 和 SiamFC 相比,同样具有明显的性能优势。

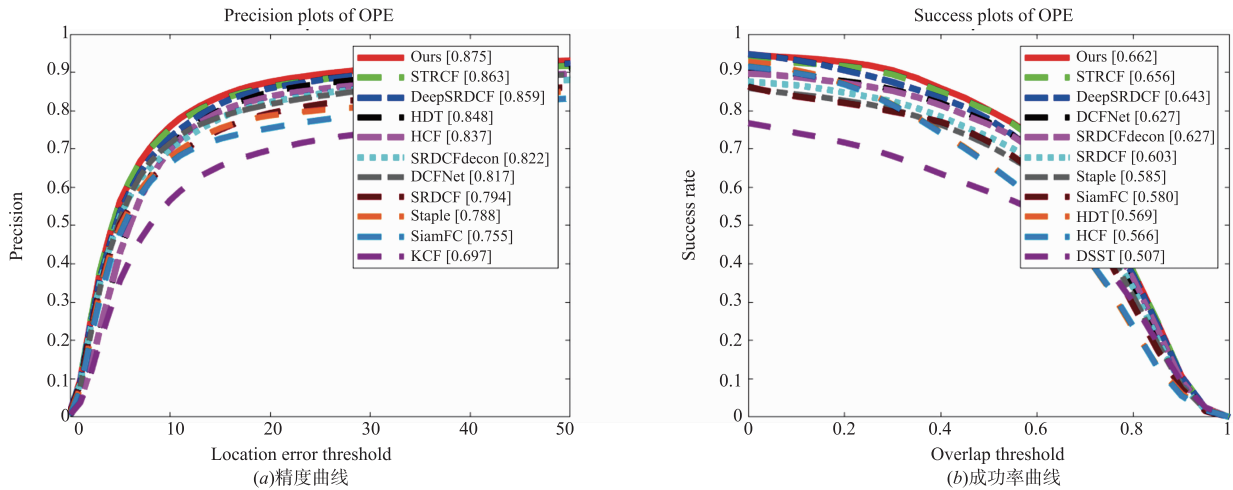


图4 OTB2015数据集上的精度和成功率曲线

表 1 显示了几种基于深度学习的跟踪算法在 OTB2015 上的性能和速度对比,在跟踪速度方面,SiamFC 表现最好(83.7 FPS),其次是 DCFNet(78.4FPS),而

本文算法的运行速度为 9.5FPS,虽然速度稍慢,但是在跟踪精度和成功率上均远高于上述两种算法.同时和 DeepSRDCF 相比,本文在速度上依然有着很大的优势.

表 1 深度学习算法跟踪性能对比

	Ours	HCF	HDT	SiamFC	DCFNet	DeepSRDCF
成功率	0.662	0.566	0.610	0.580	0.627	0.643
跟踪精度	0.875	0.837	0.848	0.755	0.817	0.859
跟踪速度(FPS)	9.5	10.2	2.7	83.7	78.4	0.2

3.3 特征表示的对比分析

为了进一步验证本文算法的有效性,分别选择 HOG 特征和 CN 特征对算法进行实现,并和本文算法进行对比,如图 5 示.如果将本文算法的特征表示换为传统的手工特征,算法的跟踪性能将会有较大幅度的

下降.主要原因在于传统手工特征对目标的语义表达能力极其有限,使得获得的可靠性矩阵难以很好的挖掘目标的区域信息.但是深度特征存在运算复杂度较高的问题,难以满足实时性的需求,后续将考虑采用更加轻量化的深度网络,同时在算法中对传统特征和深

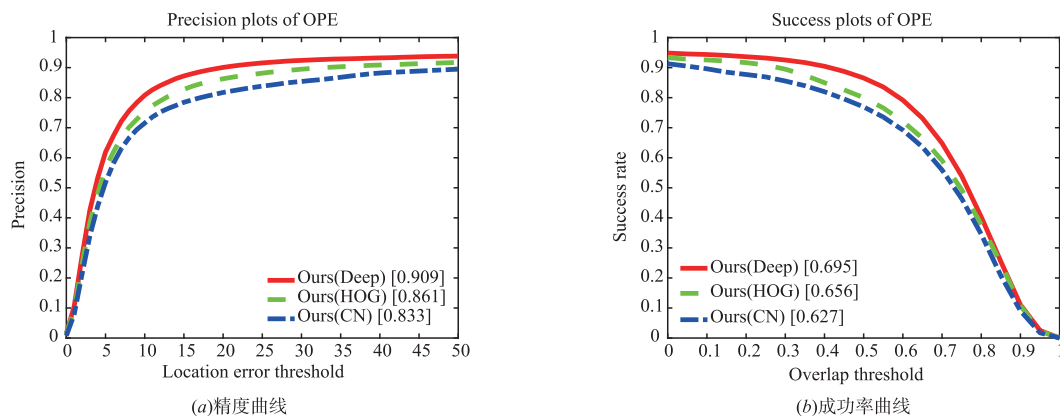


图5 不同特征表示在OTB2015上的对比分析

度特征进行合理的融合,在跟踪性能和跟踪速度上寻求一定的平衡.

4 结束语

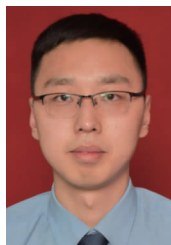
本文提出了一种新的跟踪框架,将深度特征和相关滤波器更紧密地结合起来,一方面用于对特征进行表达,另一方面提取目标的可靠区域,用于对滤波器进行空间约束,较好地提升了模型的判别能力.接着通过对指示矩阵进行分析,提出了一种简单有效的遮挡判断准则,提高了算法的长时跟踪性能.实验结果表明,基于深度空间正则化的鲁棒视觉跟踪方法,能很好地应对快速运动、遮挡、尺度变化、目标形变等多种复杂场景.目前的研究主要集中在对损失函数的改进上来缓解边界效应,在今后的工作中,将考虑时间一致性和通道特征之间的关系,以提高特征对特定目标的判别能力.

参考文献

- [1] RAWAT W, WANG Z. Deep convolutional neural networks for image classification: A comprehensive review [J]. *Neural Computation*, 2017, 29(9): 2352 – 2449. doi: 10.1162/neco_a_00990.
- [2] GIRSHICK R, DONAHUE J, DARRELL T, et al. Rich feature hierarchies for accurate object detection and semantic segmentation [A]. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition [C]*. USA: IEEE, 2014. 580 – 587. doi: 10.1109/CVPR. 2014. 81.
- [3] LONG J, SHELHAMER E, DARRELL T. Fully convolutional networks for semantic segmentation [A]. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition [C]*. USA: IEEE, 2015. 3431 – 3440. doi: 10.1109/TPAMI. 2016. 2572683.
- [4] SMEULDERS A W M, CHU D M, CUCCHIARA R, et al. Visual tracking: an experimental survey [J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2013, 36(7): 1442 – 1468. doi: 10.1109/TPAMI. 2013. 230.
- [5] WANG Naiyan, SHI Jianping, YEUNG Dityan, et al. Understanding and diagnosing visual tracking systems [A]. *IEEE International Conference on Computer Vision [C]*. Santiago, Chile: IEEE, 2015. 3101 – 3109. doi: 1109/ICCV. 2015. 355.
- [6] 余旺盛, 李卫华, 侯志强. 分层结构直方图及其应用 [J]. *电子学报*, 2017, 45(11): 2617 – 2624.
YU Wang-sheng, LI Wei-hua, HOU Zhi-qiang. Hierarchical structure histogram and its application [J]. *Acta Electronica Sinica*, 2017, 45(11): 2617 – 2624. (in Chinese)
- [7] 查宇飞, 吴敏, 库涛, 等. 基于位置敏感模型的深度跟踪算法研究 [J]. *电子学报*, 2019, 47(10): 2076 – 2082.
ZHA Yufei, WU Min, KU Tao, et al. Deep tracking algorithm research based on location-sensitive model [J]. *Acta Electronica Sinica*, 2019, 47(10): 2076 – 2082. (in Chinese)
- [8] BOLME D S, BEVERIDGE J R, DRAPER B A, et al. Visual object tracking using adaptive correlation filters [A]. *IEEE Conference on Computer Vision and Pattern Recognition [C]*. San Francisco, CA, US: IEEE, 2010. 2544 – 2550. doi: 10.1109/CVPR. 2010. 5539960.
- [9] HENRIQUES J F, CASEIRO R, MARTINS P, et al. Exploiting the circulant structure of tracking-by-detection with kernels [A]. *Proceedings of European Conference on Computer Vision [C]*. Florence, Italy, 2012. 702 – 715. doi: 10.1007/978-3-642-33765-9_50.
- [10] DANELLJAN M, SHAHBAZ K F, FELSBURG M, et al. Adaptive color attributes for real-time visual tracking [A]. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition [C]*. USA: IEEE, 2014. 1090 – 1097. doi: 10.1109/CVPR. 2014. 143.
- [11] HENRIQUES J F, RUI C, MARTINS P, et al. High-speed tracking with kernelized correlation filters [J]. *IEEE Transactions on Pattern Analysis & Machine Intelligence*, 2015, 37(3): 583 – 596. doi: 10.1109/tpami. 2014. 2345390.
- [12] RUSSAKOVSKY O, DENG J, SU H, et al. Imagenet large scale visual recognition challenge [J]. *International Journal of Computer Vision*, 2015, 115(3): 211 – 252. doi: 10.1007/s11263-015-0816-y.
- [13] SIMONYAN K, ZISSERMAN A. Very deep convolutional networks for large-scale image recognition [A]. *International Conference on Learning Representations [C]*. ICLR, 2015. 1 – 14.
- [14] HE K, ZHANG X, REN S, et al. Deep residual learning for image recognition [A]. *IEEE Conference on Computer Vision and Pattern Recognition [C]*. USA: IEEE, 2016. 770 – 778. doi: 10.1109/CVPR. 2016. 90.
- [15] MA Chao, HUANG Jiabin, YANG Xiaokang, et al. Hierarchical convolutional features for visual tracking [A]. *IEEE International Conference on Computer Vision [C]*. Santiago, Chile: IEEE, 2015. 3074 – 3082. doi: 10.1109/ICCV. 2015. 352.
- [16] QI Yuankai, ZHANG Shenging, QIN Lei, et al. Hedged deep tracking [A]. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition [C]*. USA: IEEE, 2016. 4303 – 4311. doi: 10.1109/CVPR. 2016. 466.
- [17] Valmadre J, Bertinetto L, Henriques, João F, et al. End-to-end representation learning for Correlation Filter based tracking [A]. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition [C]*. USA: IEEE, 2017. 5000 – 5008. doi: 10.1109/CVPR. 2017. 531.

- [18] GALOOGAHI H K, SIM T, LUCEY S. Correlation filters with limited boundaries[A]. IEEE Conference on Computer Vision and Pattern Recognition [C]. Boston, MA, USA: IEEE, 2015. 4630 – 4638. doi:10.1109/CVPR. 2015. 7299094.
- [19] DANELLJAN M, HAGER G, SHAHBAZ KHAN F, et al. Learning spatially regularized correlation filters for visual tracking[A]. Proceedings of the IEEE International Conference on Computer Vision[C]. USA: IEEE, 2015. 4310 – 4318. doi:10.1109/ICCV. 2015. 490.
- [20] Galoogahi H K, Fagg A, Lucey S. Learning background-aware correlation filters for visual tracking [A]. ICCV [C]. USA: IEEE, 2017. 1135 – 1143.
- [21] LI Feng, TIAN Cheng, ZUO Wangmeng, et al. Learning spatial-temporal regularized correlation filters for visual tracking[A]. Computer Vision and Pattern Recognition [C]. Salt Lake City, USA, 2018. 4904 – 4913.
- [22] Boyd S. Distributed optimization and statistical learning via the alternating direction method of multipliers[J]. FNT in Machine Learning, 2011, 3 (1): 1 – 122. doi: 10.1561/22000000016.
- [23] Wei X-S, Luo J-H, Wu J, Zhou Z-H. Selective convolutional descriptor aggregation for fine-grained image retrieval[J]. IEEE TIP, 2017, 26(6): 2868 – 2881.
- [24] Pearson K. On lines and planes of closest fit to systems of points in space [J]. Philosophical Magazine, 1901, 2 (11): 559 – 572.
- [25] DANELLJAN M, ROBINSON A, KHAN F S, et al. Beyond correlation filters: Learning continuous convolution operators for visual tracking[A]. European Conference on Computer Vision[C]. Cham: Springer, 2016. 472 – 488. doi:10.1007/978-3-319-46454-1_29.
- [26] VEDALDI A, LENC K. Matconvnet: Convolutional neural networks for matlab[A]. Proceedings of the 23rd ACM International Conference on Multimedia[C]. USA: ACM, 2015. 689 – 692. doi:10.1145/2733373. 2807412.
- [27] DANELLJAN M, HÄGER G, KHAN F, et al. Accurate scale estimation for robust visual tracking[A]. Proceedings of British Machine Vision Conference[C]. Nottingham, UK, 2014. 65. 1 – 65. 11. doi:10.5244/C. 28. 65.
- [28] WU Yi, LIM Jongwoo, YANG Minghsuan. Online object tracking: a benchmark[A]. IEEE Conference on Computer Vision and Pattern Recognition [C]. Portland, OR, USA: IEEE, 2013. 2411 – 2418. doi:10.1109/CVPR. 2013. 312.
- [29] WU Y, LIM J, YANG M H. Object tracking benchmark [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2015, 37 (9): 1834 – 1848. doi: 10.1109/TPAMI. 2014. 2388226.
- [30] DANELLJAN M, HAGER G, SHAHBAZ KHAN F, et al. Adaptive decontamination of the training set: A unified formulation for discriminative visual tracking [A]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition[C]. USA: IEEE, 2016. 1430 – 1438.
- [31] BERTINETTO L, VALMADRE J, HENRIQUES J F, et al. Fully-convolutional siamese networks for object tracking[A]. European Conference on Computer Vision[C]. Cham: Springer, 2016. 850 – 865.
- [32] Wang Q, et al. DCFNet: Discriminant correlation filters network for visual tracking [A]. Computer Vision and Pattern Recognition [C]. DBLP, 2017. arXiv: abs/1704.04057.
- [33] DANELLJAN M, HAGER G, SHAHBAZ KHAN F, et al. Convolutional features for correlation filter based visual tracking[A]. Proceedings of the IEEE International Conference on Computer Vision Workshops[C]. USA: IEEE, 2015. 58 – 66. doi:10.1109/ICCVW. 2015. 84.
- [34] Bertinetto L, et al. Staple: Complementary learners for real-time tracking[A]. 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR) [C]. Las Vegas, NV, USA: IEEE, 2016. 1401 – 1409. doi:10.1109/CVPR. 2016. 156.

作者简介



蒲 磊 (通信作者) 男, 1991 年 9 月出生, 四川遂宁人。分别于 2014 年和 2017 年获空军工程大学信息与导航学院信息与通信工程专业学士学位和硕士学位, 现于空军工程大学研究生院攻读博士学位, 主要研究方向为计算机视觉与目标跟踪。
E-mail: warmstoner@163.com



冯新喜 男, 1964 年 10 月出生, 陕西富平人。1991 年获西北工业大学博士学位, 现为空军工程大学信息与导航学院教授、博士研究生导师, 主要研究领域为信息融合、信号处理、目标跟踪等。
E-mail: fengxinxi2005@aliyun.com