

# 相关滤波融合卷积残差学习的目标跟踪算法

杨亚光\*\*, 尚振宏\*

昆明理工大学信息工程与自动化学院, 云南 昆明 650500

**摘要** 针对复杂场景中传统单一手工特征表达能力不足, 以及模型更新过程中由于误差累积导致模型退化问题, 提出了基于相关滤波融合卷积残差学习的目标跟踪算法。将融合了多特征的相关滤波算法定义为神经网络中的一层, 将特征提取、响应图生成、模型更新整合到端到端的神经网络中进行模型训练; 为解决在线更新过程中模型退化问题, 引入残差学习方式引导模型更新。在基准数据集 OTB-2013 和 OTB-2015 上的实验结果表明, 本文算法能够有效应对复杂场景中运动模糊、形变和光照等变化, 具备较高跟踪精度与鲁棒性。

**关键词** 图像处理; 目标跟踪; 相关滤波; 端到端学习; 残差学习

中图分类号 TP391.4

文献标志码 A

doi: 10.3788/LOP57.121012

## Object Tracking Algorithm Based on Correlation Filtering and Convolution Residuals Learning

Yang Yaguang\*\*, Shang Zhenhong\*

Faculty of Information Engineering and Automation, Kunming University of Science and Technology, Kunming, Yunnan 650500, China

**Abstract** Aiming at the problem of insufficient expression ability of traditional single manual feature and model degradation caused by error accumulation in the process of model updating in complex scenes, Based on this, the object tracking algorithm based on correlation filtering and convolution residual learning is proposed. The multi-feature correlation filtering algorithm is defined as a layer in the neural network, and the feature extraction, response graph generation, and model update are integrated into the end-to-end neural network for model training. In order to reduce the degradation of model during online updating, the residual learning mode is introduced to guide model updating. The proposed method is validated on the benchmark datasets OTB-2013 and OTB-2015. The experimental results show that the proposed algorithm can effectively deal with motion blur, deformation, and illumination in the complex scene, and has high tracking accuracy and robustness.

**Key words** image processing; object tracking; correlation filter; end-to-end learning; residuals learning

**OCIS codes** 100.2960; 100.4999

## 1 引言

运动目标跟踪属于计算机视觉中一个重要研究方向, 有着广泛应用, 如视频监控, 人机交互, 无人驾驶等。计算机在面对复杂多变的外部环境变化和运动目标形变情况下, 往往会造成目标漂移和跟丢, 因此, 该研究方向仍面临巨大挑战和进步空间。现有的目标跟踪方法如下。

### 1.1 相关滤波目标跟踪方法

相关滤波(CF)最初用在信号处理领域, 描述信号间的相关性。Bolme 等<sup>[1]</sup>最先将 CF 引入目标跟踪中来。之后, Henriques 等<sup>[2]</sup>使用方向梯度直方图(HOG)特征, 并加入核函数, 提出了核相关滤波(KCF)算法。Danelljan 等<sup>[3]</sup>提出了基于颜色特征的自适应维度下降(CN)算法。以上算法仅使用单个特征, 在背景遮挡和形变等复杂场景下鲁棒性较

收稿日期: 2019-09-02; 修回日期: 2019-10-09; 录用日期: 2019-11-02

基金项目: 国家自然科学基金(61462052)

\* E-mail: shangzhenhong@126.com; \*\* E-mail: 1107298031@qq.com

差。针对这一问题,后来的研究者创造性地将多种特征融合在一起。例如,比例自适应核相关滤波(SAMF)算法<sup>[4]</sup>和 staple 算法<sup>[5]</sup>,在 KCF 算法 HOG 特征基础上,加上了颜色特征。与单一特征目标跟踪算法相比,多特征融合算法具有更高的跟踪精度和准确率。然而现有相关滤波目标跟踪方法存在两个不足:1)特征提取和相关滤波器训练独立进行,无法从端到端的训练中受益;2)相关滤波器更新时大多采用移动平均方法,模型自适应性和稳定性较差。

## 1.2 深度学习目标跟踪方法

深度学习追踪(DLT)<sup>[6]</sup>是第一个把深度模型运用在单目标跟踪任务上的跟踪算法。Wang 等<sup>[7]</sup>提出的算法分析了不同层卷积神经网络特征的差异。随后,Krizhevsky 等<sup>[8]</sup>的研究显示出深度特征结合相关滤波器的巨大优势,但该算法没有对尺度变化进行处理<sup>[9]</sup>。本文使用 ImageNet<sup>[10]</sup>数据集上预训练的 VGG-16<sup>[11]</sup>网络作为图像特征提取网络。为防止因噪声更新造成的模型退化,本文借鉴了 CREST<sup>[12]</sup>算法中的残差学习思想,利用残差学习<sup>[13]</sup>来弥补预测响应图和真值响应图之间的偏差

并指导模型更新,不同的是,CREST 侧重于训练时利用残差促进模型的拟合,本文引入残差学习主要用于弥补跟踪过程中基本卷积层可能造成的损失。

本文在广泛使用的目标跟踪数据集 OTB-2013<sup>[14]</sup>和 OTB-2015<sup>[15]</sup>上进行了实验,结果表明,在复杂多变的场景下,相比相关滤波算法和深度学习算法,本文提出的基于相关滤波融合卷积残差学习的目标跟踪方法有更高跟踪精度和准确率。

## 2 本文方法

本文算法网络框架图如图 1 所示,输入第  $T$  帧图像,提取以第  $T-1$  帧预测的目标对象为中心的候选块;将候选块输入模型特征提取模块中,特征提取网络选用 VGG-16 具有最佳表征跟踪目标信息能力的中间层 conv4-3 来提取图像特征;接着基本卷积层输出图像最大响应映射。理想情况下,基本卷积层输出和真实值相等,然而,实际情况二者存在一定偏差;此时空间残差层和时间残差层捕获基本卷积层输出和真实高斯响应差异;合并基本卷积层与时空残差层输出的响应映射为最终响应图。

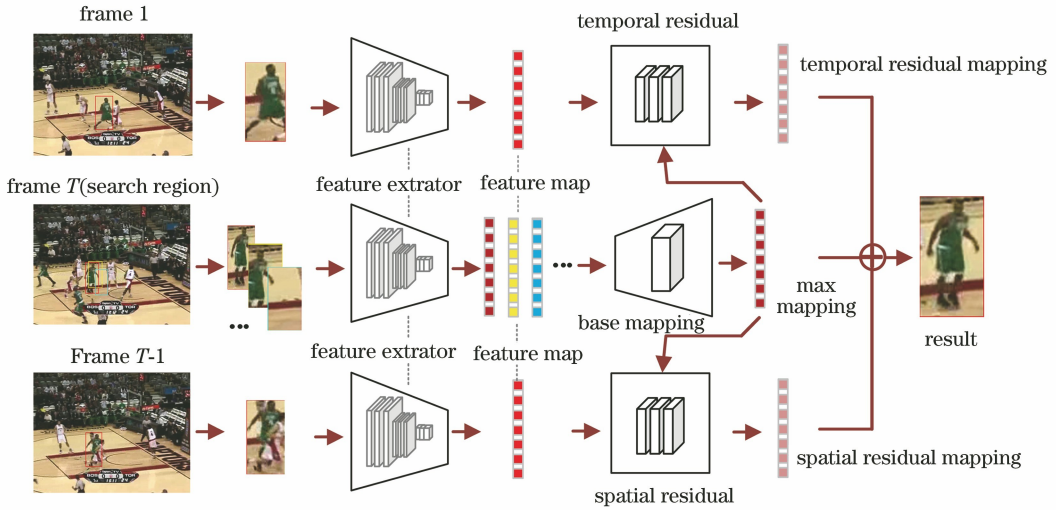


图 1 本文算法网络框架图

Fig. 1 Network frame diagram of proposed algorithm

### 2.1 基本卷积层

本文借鉴了 Bertinetto 等<sup>[5]</sup>的多特征融合思路,将对运动模糊和光照具有鲁棒性的 HOG 特征和对形变具有鲁性的颜色特征进行融合,优势互补从而使鲁棒性更好。具体形式如下:

$$f(x) = \gamma_{\text{templ}} f_{\text{templ}}(x) + \gamma_{\text{hist}} f_{\text{hist}}(x), \quad (1)$$

$$f_{\text{templ}}(x; \mathbf{h}) = \sum_{u \in \tau} \mathbf{h}[u]^T \varphi_x[u], \quad (2)$$

$$f_{\text{hist}}(x; \boldsymbol{\beta}) = g(\psi_x; \boldsymbol{\beta}), \quad (3)$$

$$g(\psi_x; \boldsymbol{\beta}) = \boldsymbol{\beta}^T \left( \frac{1}{H} \sum_{u \in H} \psi[u] \right), \quad (4)$$

$$l(x, p, \boldsymbol{\theta}) = d \{ p, \arg \max_{q \in S} f [T(x, q); \boldsymbol{\theta}] \}. \quad (5)$$

(1)式表示 HOG 特征  $f_{\text{templ}}$  和颜色特征  $f_{\text{hist}}$  的融合,权重系数分别为  $r_{\text{templ}}$  和  $r_{\text{hist}}$ 。(2)~(4)式分别对应 HOG 特征  $f_{\text{templ}}$  得分和颜色特征  $f_{\text{hist}}$  得分的

具体计算公式。其中 HOG 特征得分(2)式,是一个  $K$  通道特征图像  $\varphi_x$  的线性函数,其中  $\tau \rightarrow \mathbf{R}^K$ ,表示  $\tau$  是  $K$  通道特征图像,从  $x$  中获得并定义在有限网格上, $\tau \subset \mathbf{Z}^2$ ,式中  $\mathbf{h}$  表示权重向量, $u$  表示图像  $x$  中的一个像素位置。其中颜色特征得分[(3)式和(4)式]是从一个  $M$  通道特征图像  $\varphi_x$  计算得到,其中  $H \rightarrow \mathbf{R}^M$ ,表示  $H$  是  $M$  通道特征图像,从  $x$  中获得并定义在有限网格上, $H \subset \mathbf{Z}^2$ ,式中  $\boldsymbol{\beta}$  表示权重向量。(5)式表示融合后的损失函数,其中参数  $\boldsymbol{\theta} = (\mathbf{h}, \boldsymbol{\beta})$ ,系数  $r_{\text{templ}}$  和  $r_{\text{hist}}$  可以看作隐含在  $\mathbf{h}$  和  $\boldsymbol{\beta}$  中, $d(p, q)$  表示当选择的矩形框是  $q$  而正确矩形框是  $p$  时的损失值。

研究人员 Song 等<sup>[12]</sup>将单一特征判别相关滤波(DCFs)算法形式化为基本卷积层,受此启发,本文将多种特征融合后形式化为基本卷积层,利用神经网络反向传播算法训练单层卷积层来等价替换传统相关滤波算法训练方式。对于输入样本  $\mathbf{X}$ ,对应的输出高斯响应为  $\mathbf{Y}$ ,那么训练相关滤波器  $\mathbf{W}$  可以转化为求解以下最小化问题。

$$\mathbf{W}^* = \operatorname{argmin} \|\mathbf{W} * \mathbf{X} - \mathbf{Y}\|^2 + \lambda \|\mathbf{W}\|^2, \quad (6)$$

式中: $\lambda$  表示正则化参数。损失函数表示为

$$L(\mathbf{W}) = \frac{1}{N} \sum_i^{[N]} L_w(\mathbf{X}^{(i)}) + \lambda r(\mathbf{W}), \quad (7)$$

式中: $N$  为样本数量; $L_w(\mathbf{X}^{(i)}) (i \in N)$  为第  $i$  个样本的损失函数; $r(\mathbf{W})$  为权重衰减。当  $N = 1$  时,将  $r(\mathbf{W})$  形式化为 L2 范数。则(7)式表示为

$$L(\mathbf{W}) = L_w(\mathbf{X}) + \lambda \|\mathbf{W}\|^2. \quad (8)$$

当  $L_w(\mathbf{X}) = \|\mathbf{F}(\mathbf{X}) - \mathbf{Y}\|^2$ ,  $\mathbf{F}(\mathbf{X})$  为神经网络输

出, $\mathbf{Y}$  表示真值,此时  $L_w(\mathbf{X})$  等价于  $\mathbf{F}(\mathbf{X})$  和  $\mathbf{Y}$  之间的 L2 损失。损失函数[(8)式]等价于相关滤波目标函数[(6)式]。

### 2.2 卷积残差学习

图 2 为基本卷积层和残差学习层的结构,如果将输入第  $K$  帧的理想响应映射表示为  $H(K)$ ,从基层网络的实际输出表示为  $F_B(K)$ ,那么,残差学习  $F_R(K)$  可表示为

$$F_R(K) = H(K) - F_B(K). \quad (9)$$

因此,加入残差学习后,帧  $K$  最终的实际响应结果  $F(K)$  可表示为

$$F(K) = F_B(K) + F_R(K) = F_B(K, \{\mathbf{W}_B\}) + F_R(K, \{\mathbf{W}_R\}). \quad (10)$$

时间残差用来捕捉空间残差无效时的差异,它的网络结构与空间残差结构相似。时间残差输入从包含初始对象外观的第一帧中提取。若  $K_t$  表示输入  $K$  的第  $t$  帧,则时空残差可以表示为

$$F(K_t) = F_B(K_t) + F_{SR}(K_t) + F_{TR}(K_t), \quad (11)$$

式中: $F_B(K_t)$  是基本卷积层响应值; $F_{SR}(K_t)$  是空间残差层响应值; $F_{TR}(K_t)$  是时间残差层响应值。当目标对象外观经历较小变化时,基本卷积层输出与真实值差异很小,残差层对最终响应结果影响较弱;当目标对象外观经历较大变化时,比如快速运动导致背景模糊,基本卷积层很难区分目标与背景,此时残差层弥补基本卷积层输出与真值软标签(前一帧预测得到的最大响应目标作为真值软标签)之间的差异。

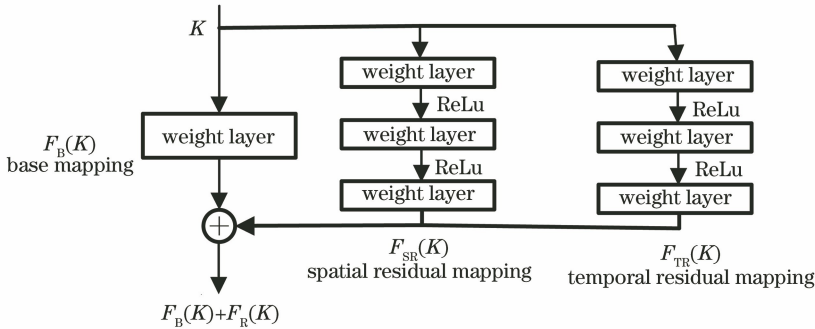


图 2 基本卷积层和时空残差层

Fig. 2 Basic convolution layer and temporal-spatial residual layer

### 2.3 本文算法流程

本文算法流程如下。

1) 模型初始化。本文网络含有参数的层有 1 层基本卷积层,3 层时间残差层和 3 层空间残差层,更新对象为各层权值参数  $\mathbf{W}$  和偏置参数  $b$ 。首先

基础卷积层和残差层参数根据零均值高斯分布随机初始化;然后给定带有目标位置的第一帧图像,以目标中心位置为中心提取训练块;将训练块传入本文网络中的特征提取模块(预训练的 VGG-16 网络)进行特征提取。迭代训练至损失函数  $J(\mathbf{W}, b)$  [等

价于(8)式]小于设定阈值 0.02 时停止,此时模型参数初始化完成。更新参数使用的前向传播和反向传播公式分别为

$$\begin{cases} Z^i = W^i A^{i-1} + b^i \\ A^i = g^i(Z^i) \end{cases}, \quad (12)$$

$$\begin{cases} W_i^l = W_i^l - \alpha \frac{\partial}{\partial W_i^l} J(W, b) \\ b_i^l = b_i^l - \alpha \frac{\partial}{\partial b_i^l} J(W, b) \end{cases}, \quad (13)$$

式中: $l=(0,1,2)$ 分别表示基本卷积层、时间残差层和空间残差层。前向传播时,网络第  $i=(1,2,3,\dots)$ 层目标函数值为  $Z^i$ ,通过激活函数  $g^i(\cdot)$ (本文选用 ReLu 函数)得到当前层输出  $A^i$ 。反向传播时,当前  $W_i^l$  减去损失函数对  $W_i^l$  偏导数与学习率  $\alpha$  的乘积为更新后权值,当前  $b_i^l$  减去损失函数对  $b_i^l$  与学习率  $\alpha$  的乘积为更新后偏置值。

2) 在线检测。当新帧出现时,根据前一帧预测的目标中心位置,提取与训练块同样大小的搜索块,将搜索块输入模型中生成响应映射。以最大响应值 [(11)式]作为定位目标。

3) 尺度估计。根据响应图数值估计最终目标尺度。在获取目标中心位置后,提取不同尺度搜索块,将搜索块发送到模型中生成响应映射,根据生成响应映射范围,更新目标对象在  $t$  帧处的宽和高,可表示为

$$(\omega_t, h_t) = \alpha(\omega_t^*, h_t^*) + (1 - \beta)(\omega_{t-1}, h_{t-1}), \quad (14)$$

式中: $\omega_t$  和  $h_t$  分别代表第  $t$  帧的宽和高; $\omega_t^*$  和  $h_t^*$  表示最大响应映射搜索块的宽和高;权重  $\alpha$  表示目标尺度平滑更新因子。

4) 模型更新。每帧搜索块中响应值最大块标准化作为训练块,通过在线检测不断生成训练数据,将训练数据输入到本文网络中更新模型参数,每  $N$  帧进行一次模型更新,更新方法同(12)式和(13)式。

## 3 实验

### 3.1 实验参数设置、数据集和评价指标

实验选用的硬件配置: Intel Corei7-7800X 3.50 GHz CPU, GTX Titan X GPU, 内存为 16 GB 的台式电脑。

实验细节: 本文的深度神经网络选用 Matconvnet<sup>[16]</sup>; 特征提取网络选用 VGG-16, 为降低计算复杂度,通过主成分分析(PCA)降维将特征

通道减少到 64 个。在线检测和尺度估计是 5 倍搜索域; 本文利用二维高斯函数生成回归目标图, 其峰值为 1.0; 平滑尺度因子  $\alpha$  设置为 0.6。在训练阶段, 应用 adam 优化器迭代计算更新系数, 并将学习率设置为  $5e-8$ , 直到损失函数 [(8)式] 降到 0.02 停止训练。模型更新阶段每 2 帧更新模型, 进行 2 次迭代计算, 学习率设置为  $2e-9$ 。

基准数据集: 本文采用 OTB-2013 和 OTB-2015 基准数据集, 它们分别包括 50 段和 100 段标注完整的视频序列, 根据目标跟踪所面临的挑战, 这些视频序列含有 11 种属性, 其中包括: 光照变化 (IV)、尺度变化 (SV)、运动模糊 (MB)、遮挡 (OCC)、形变 (DEF)、快速运动 (FM)、平面外旋转 (OPR)、出视野 (OV)、背景模糊 (BC)、平面内旋转 (IPR)、低分辨率 (LR)。

评价指标: 本文使用一次性通过评估 (OPE) 的重叠精度 (OP)、距离精度 (DP)、中心位置误差 (CLE) 作为评价标准<sup>[17]</sup>。其中, CLE 为预测位置与真值位置的欧氏距离; OP 表示重合率大于某一阈值帧占总帧的百分比, 阈值设为 0.5, 其中重合率为跟踪算法定位目标框与真值框交集; DP 表示追踪算法定位中心点与真值中心点之间的距离小于给定阈值帧占总帧的百分比, 阈值设为 20 个像素。

### 3.2 消融实验

本文方法包括基本卷积卷积层和时空残差层, 如上述第 2 节中的形式化定义, 基本卷积层是将融合后相关滤波定义为卷积神经网络中的一层卷积层, 这里作为基础网络, 残差层负责弥补基本卷积层与真值软标签之间的差异。图 3 为将基本卷积层和残差层剥离后, 分别在数据集 OTB-2015 上通过一次性通过评估方式得到的距离精度曲线图和成功率曲线图。实验表明, 随着时空残差层的融合, 基础网络性能不断提高。

### 3.3 定量评价

为验证本文基于相关滤波融合卷积残差学习的跟踪算法在形变、运动模糊、光照都具有鲁棒性, 特别选择了 9 种优异的目标跟踪算法在 OTB-2013 和 OTB-2015 基准数据集上进行对比, 这些算法包括仅适用单一 HOG 特征的 KCF 算法, 加入尺度变换的判别尺度空间目标跟踪 (DSST 算法)<sup>[18]</sup>, 基于自适应结构化方法的 Struck 算法<sup>[19]</sup> 和基于结构化支持向量机 (SVM) 的 Scale\_DLSSVM 算法<sup>[20]</sup>, 还包括融合多种特征的 Staple 算法、SAMF 算法, 以及加入了卷积特征的 CF2 算法<sup>[21]</sup>、DeepSRDCF 算

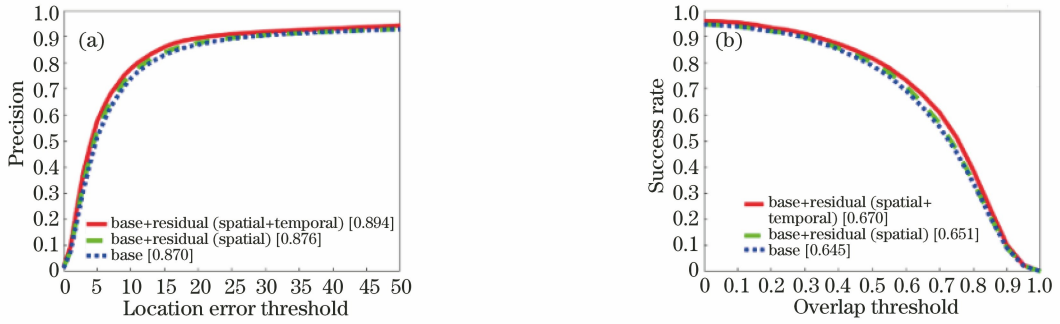


图3 在 OTB2015 数据集上与基础网络的对比。(a)距离精度;(b)成功率

Fig. 3 Comparison experiment on OTB2015 datasets with baseline. (a) Precision; (b) success rate

法<sup>[22]</sup>和端到端学习的 SiamFC 算法<sup>[23]</sup>。利用 DP 和 OP 指标在 OTB-2013 和 OTB-2015 视频序列上进行了一遍评估。评估结果如图 4 和图 5 所示。

可以看出在 OTB-2013 数据集上本文方法的 DP 为 89.2%，OP 为 67.1%，在 OTB-2015 数据集上本文方法 ours 的 DP 为 89.4%，OP 为 67.0%，都超过了 9 个对比方法。在 OTB-2015 数据集上，同样采用端到端学习，本文方法比 SiamFC 在 DP 和 OP 上分别提升了 12.3% 和 8.8%，证明了基于相关

滤波的多特征融合更具鲁棒性。相比同样采用多特征融合的 Staple(融合了 HOG 特征和颜色特征)和 SAMF(融合灰度特征、HOG 特征和颜色特征)，本文方法在 DP 上分别提升了 11.0% 和 14.3%，在 OP 上分别提升了 8.9% 和 11.7%，表明了深度学习方法更能有效利用待跟踪目标的语义信息，提升跟踪精度和准确率。在 OTB-2015 数据集上，本文算法在面临 11 类困难场景的挑战下，同样表现出了良好的跟踪性能。如表 1 和表 2 所示。

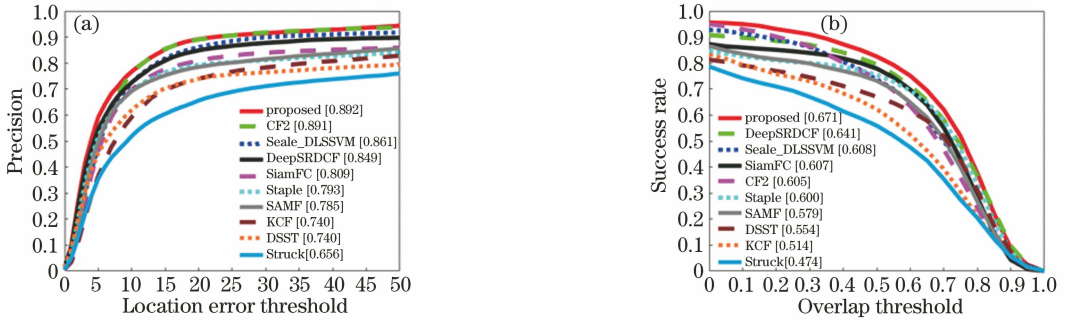


图4 10 种算法在 OTB-2013 的距离精度曲线图和成功率曲线图。(a) 距离精度;(b)成功率

Fig. 4 Distance precision plots and overlap success plots of ten algorithms in OTB-2013. (a) Precision; (b) success rate

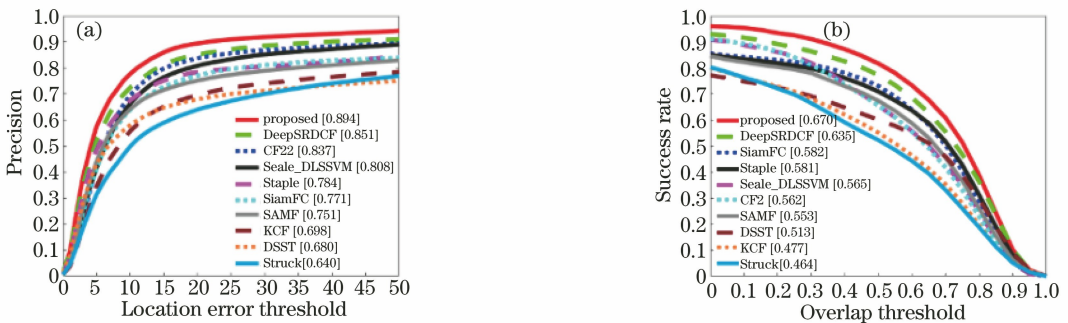


图5 10 种算法在 OTB-2015 的距离精度曲线图和成功率曲线图。(a)距离精度;(b)成功率

Fig. 5 Distance precision plots and overlap success plots of ten algorithms in OTB-2015.

(a) Precision; (b) success rate

表1 本文方法和对比方法在不同场景属性下的 DP 值

Table 1 DP values of proposed method and compared methods under different scene attributes

Item	Proposed	CF2	Scale_DLSSVM	DeepSRDCF	SiamFC	Staple	SAMF	KCF	DSST	Struck
IV	<b>0.862</b>	<u>0.816</u>	0.790	0.786	0.741	0.787	0.708	0.724	0.715	0.558
SV	<b>0.872</b>	0.798	0.758	<u>0.817</u>	0.738	0.723	0.701	0.635	0.633	0.595
OCC	<b>0.893</b>	0.765	0.789	<u>0.822</u>	0.726	0.721	0.722	0.632	0.590	0.528
DEF	<b>0.849</b>	<u>0.790</u>	0.748	0.779	0.693	0.743	0.680	0.619	0.533	0.527
MB	<b>0.883</b>	0.804	0.740	<u>0.823</u>	0.705	0.707	0.655	0.601	0.567	0.580
FM	<b>0.871</b>	<u>0.815</u>	0.727	0.814	0.743	0.697	0.654	0.621	0.552	0.606
IPR	<b>0.860</b>	<u>0.854</u>	0.820	0.818	0.742	0.770	0.721	0.701	0.691	0.629
OPR	<b>0.886</b>	<u>0.807</u>	0.802	0.835	0.756	0.738	0.739	0.677	0.644	0.587
OV	<b>0.864</b>	0.677	0.704	<u>0.781</u>	0.669	0.661	0.628	0.501	0.481	0.472
BC	<b>0.866</b>	<u>0.843</u>	0.793	0.841	0.690	0.766	0.689	0.713	0.704	0.552
LR	<b>0.872</b>	0.831	0.791	0.708	<u>0.847</u>	0.631	0.685	0.560	0.567	0.671

Note: In this table, the number marked with black is the first, and the number underlined is the second.

表2 本文方法和对比方法在不同场景属性下的 OP 值

Table 2 OP values of proposed method and compared methods under different scene attributes

Item	Proposed	CF2	Scale_DLSSVM	DeepSRDCF	SiamFC	Staple	SAMF	KCF	DSST	Struck
IV	<b>0.677</b>	0.541	0.572	<u>0.624</u>	0.574	0.596	0.530	0.482	0.556	0.428
SV	<b>0.652</b>	0.485	0.500	<u>0.607</u>	0.556	0.522	0.492	0.395	0.466	0.403
OCC	<b>0.672</b>	0.526	0.551	<u>0.603</u>	0.547	0.545	0.538	0.445	0.449	0.393
DEF	<b>0.615</b>	0.530	0.522	<u>0.567</u>	0.510	0.552	0.505	0.438	0.415	0.387
MB	<b>0.701</b>	0.585	0.591	<u>0.642</u>	0.550	0.546	0.525	0.459	0.469	0.459
FM	<b>0.673</b>	0.570	0.550	<u>0.628</u>	0.568	0.537	0.507	0.459	0.447	0.462
IPR	<b>0.622</b>	0.559	0.564	<u>0.589</u>	0.557	0.552	0.519	0.469	0.502	0.448
OPR	<b>0.648</b>	0.534	0.547	<u>0.607</u>	0.558	0.534	0.536	0.453	0.470	0.423
OV	<b>0.639</b>	0.474	0.498	<u>0.553</u>	0.506	0.481	0.480	0.393	0.386	0.365
BC	<b>0.645</b>	0.585	0.560	<u>0.627</u>	0.523	0.574	0.525	0.498	0.523	0.429
LR	<b>0.597</b>	0.439	0.433	0.475	<u>0.592</u>	0.418	0.430	0.307	0.383	0.363

Note: In this table, the number marked with black is the first, and the number underlined is the second.

由表1和表2可以看出,本文方法在11类困难场景下 DP 值和 OP 值都是第一名,表明在复杂场景下,本文方法能准确定位目标并有效跟踪,具有较强鲁棒性,能够适应各类场景跟踪任务。

### 3.4 定性评价

为了更直观地表现出本文方法在各类视频场景中的鲁棒性,挑选四类相近但不同的算法在8段具有挑战的视频序列上进行可视化跟踪并与本文方法进行对比,四组对比实验结果如图6所示。

图6中选择的8个具有挑战的视频序列依次是: jumping\_1, football\_1, carDark\_1, skiing\_1, dudek\_1, sylvester\_1, matrix\_1, ironman\_1。第一组对比实验,选择了仅使用单一特征的相关滤波算法 KCF 和 DSST 与本文方法进行对比,可以看出,在快速运动和遮挡的视频序列中本文方法更加鲁棒。第二组对比实验,选择了使用卷积特征且端到端学习的 SiamFC 算法与本文方法进行对比,结果

表明,本文方法将形状变化鲁棒的颜色直方图特征和对运动模糊和亮度变化鲁棒的 HOG 特征结合起来更能应对光照变化和形变带来的挑战。第三组对比实验,选择了多特征融合相关滤波算法 Staple 和 SAMF 与本文方法进行对比,在两段长视频序列上两种算法前期都表现出不错的性能,后期由于目标出视野导致跟踪漂移和跟丢,本文方法采用时空残差学习的方式,能够及时更新跟踪模型从而有效维持模型在目标运动中的预测性能。第四组对比实验,选择的 CF2 算法和 DeepSRDCF 算法都采用了卷积特征,在快速运动、光照变化、出视野等复杂场景下,本文方法仍能做到有效跟踪。

## 4 结论

本文提出的目标跟踪方法将融合了颜色直方图特征和 HOG 特征的相关滤波器作为卷积神经网络中的基本卷积层,从而将特征提取、响应图生成、模



图 6 7 种算法在 8 个视频序列的跟踪结果比较

Fig. 6 A visualization of the tracking results of seven algorithms on eight sequences

型更新整合到端到端的神经网络中进行模型训练和预测。同时,引入残差学习的方式应对跟踪过程中因目标外观的变化导致的模型失效。实验结果表明,本文方法能够有效应对复杂场景中的运动模糊、形变和光照等变化,具备较高的跟踪精度与鲁棒性。但由于深度学习的训练过程需要一定的时间,因此本文方法的实时性还存在不足,针对此问题,接下来将重点对训练方式进行改进,提高模型收敛速度,同时提升跟踪速度<sup>[24-25]</sup>。

### 参 考 文 献

- [1] Bolme D, Beveridge J R, Draper B A, et al. Visual object tracking using adaptive correlation filters[C]// 2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, June 13-18, 2010. San Francisco, CA, USA. IEEE, 2010: 2544-2550.
- [2] Henriques J F, Caseiro R, Martins P, et al. High-speed tracking with kernelized correlation filters[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2015, 37(3): 583-596.
- [3] Danelljan M, Khan F S, Felsberg M, et al. Adaptive color attributes for real-time visual tracking [C] // 2014 IEEE Conference on Computer Vision and Pattern Recognition, June 23-28, 2014. Columbus, OH, USA. IEEE, 2014: 1090-1097.
- [4] Li Y, Zhu J K. A scale adaptive kernel correlation filter tracker with feature integration[M]// Computer Vision - ECCV 2014 Workshops. Cham: Springer International Publishing, 2015: 254-265.
- [5] Bertinetto L, Valmadre J, Golodetz S, et al. Staple: complementary learners for real-time tracking [C] // 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), June 27-30, 2016. Las Vegas, NV, USA. IEEE, 2016: 1401-1409.
- [6] Wang N, Yeung D Y. Learning a deep compact image representation for visual tracking [C] // Proceedings of Advances in Neural Information Processing Systems, 2013: 809-817.
- [7] Wang L J, Ouyang W L, Wang X G, et al. Visual tracking with fully convolutional networks[C]// 2015 IEEE International Conference on Computer Vision (ICCV), December 7-13, 2015. Santiago, Chile. IEEE, 2015: 3119-3127.
- [8] Krizhevsky A, Hinton G. Learning multiple layers of features from tiny images [R]. Technical Report, University of Toronto, 2009.
- [9] Yang J F, Zhang J P. Long time target tracking based on kernel correlation filtering [J]. Laser & Optoelectronics Progress, 2019, 56(2): 021502. 杨剑锋, 张建鹏. 基于核相关滤波的长时间目标跟踪 [J]. 激光与光电子学进展, 2019, 56(2): 021502.
- [10] Deng J, Dong W, Socher R, et al. ImageNet: a large-scale hierarchical image database [C] // 2009 IEEE Conference on Computer Vision and Pattern Recognition, June 20-25, 2009. Miami, FL. IEEE, 2009: 248-255.
- [11] Jing W, An J F, Jiang G, et al. Standoff fully-polarimetric holographic-radar-imaging experiments in terahertz regime for concealed target detection [J]. Chinese Journal of Lasers, 2019, 46(6): 0614032.

- 经文, 安健飞, 江舸, 等. 针对隐藏目标检测的太赫兹波站开式全极化全息雷达成像实验[J]. 中国激光, 2019, 46(6): 0614032.
- [12] Song Y B, Ma C, Gong L J, et al. CREST: convolutional residual learning for visual tracking[C]//2017 IEEE International Conference on Computer Vision (ICCV), October 22-29, 2017. Venice. IEEE, 2017: 2555-2564.
- [13] He K M, Zhang X Y, Ren S Q, et al. Deep residual learning for image recognition [C] // 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), June 27-30, 2016. Las Vegas, NV, USA. IEEE, 2016: 770-778.
- [14] Wu Y, Lim J, Yang M H. Online object tracking: a benchmark[C]//2013 IEEE Conference on Computer Vision and Pattern Recognition, June 23-28, 2013. Portland, OR, USA. IEEE, 2013: 2411-2418.
- [15] Wu Y, Lim J, Yang M H. Object tracking benchmark [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2015, 37(9): 1834-1848.
- [16] Vedaldi A, Lenc K. MatConvNet - convolutional neural networks for MATLAB [EB/OL]. 2014: arXiv:1412.4564[cs.CV].
- [17] Everingham M, van Gool L, Williams C K I, et al. The pascal visual object classes (VOC) challenge[J]. International Journal of Computer Vision, 2010, 88(2): 303-338.
- [18] Danelljan M, Hager G, Khan F S, et al. Discriminative scale space tracking [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2017, 39(8): 1561-1575.
- [19] Hare S, Golodetz S, Saffari A, et al. Struck: structured output tracking with kernels [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2016, 38(10): 2096-2109.
- [20] Pan Z F, Zhu Y L. Kernelized correlation filters object tracking method with multi-scale estimation [J]. Laser & Optoelectronics Progress, 2016, 53(10): 101501.  
潘振福, 朱永利. 多尺度估计的核相关滤波器目标跟踪方法[J]. 激光与光电子学进展, 2016, 53(10): 101501.
- [21] Vedaldi A, Zisserman A. Efficient additive kernels via explicit feature maps [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2012, 34(3): 480-492.
- [22] Ma C, Huang J B, Yang X K, et al. Hierarchical convolutional features for visual tracking [C] // 2015 IEEE International Conference on Computer Vision (ICCV), December 7-13, 2015. Santiago, Chile. IEEE, 2015: 3074-3082.
- [23] Danelljan M, Hager G, Shahbaz Khan F, et al. Convolutional features for correlation filter based visual tracking [C] // Proceedings of the IEEE International Conference on Computer Vision Workshops. 2015: 58-66.
- [24] Xiong C Z, Che M Q, Wang R L, et al. Adaptive model update via fusing peak-to-sidelobe ratio and mean frame difference for visual tracking [J]. Acta Photonica Sinica, 2018, 47(9): 0910001.  
熊昌镇, 车满强, 王润玲, 等. 融合峰旁比和帧差均值自适应模型更新的视觉跟踪[J]. 光子学报, 2018, 47(9): 0910001.
- [25] He X D, Zhou S Z. Fast scale adaptive kernel correlation filtering algorithm for target tracking [J]. Laser & Optoelectronics Progress, 2018, 55(12): 121501.  
何雪东, 周盛宗. 快速尺度自适应核相关滤波目标跟踪算法[J]. 激光与光电子学进展, 2018, 55(12): 121501.