基于孪生神经网络在线判别特征的视觉跟踪算法

仇祝令, 查宇飞*, 朱鹏, 吴敏
空军工程大学航空工程学院, 陕西 西安 710038

摘要 基于孪生神经网络的跟踪算法是利用离线训练的网络提取目标的特征并进行匹配,从而实现跟踪。在离线训练过程中,网络学到的是相似目标的通用特征,因此当有相似目标干扰时,用这种通用特征表达特定目标将会导致跟踪性能下降,甚至丢失目标。为提高对相似目标的判别能力,通过在线更新网络参数,使网络能够在通用特征的基础上,进一步学到当前目标的特定特征,这样不仅能有效地区分目标与背景,还能消除相似目标的干扰。实验在 OTB50 和 OTB100 数据库上进行,结果表明该算法可以提高对网络提取特征的判别力,实现对目标的稳健性跟踪。

关键词 机器视觉; 视觉追踪; 离线训练; 在线更新
中图分类号 TN919.82 **文献标识码** A **doi**: 10.3788/AOS201939.0915003

Visual Tracking Algorithm Based on Online Feature Discrimination with Siamese Network

Qiu Zhuling, Zha Yufei*, Zhu Peng, Wu Min
Aeronautics Engineering College, Air Force Engineering University, Xi'an, Shaanxi 710038, *China*

Abstract Tracking algorithms with Siamese network use the offline training network to extract features from the target for matching and tracking. In the offline training process, the network learns the common features of similar goals. In the case of interference from similar targets, using common features to express specific targets will lead to degradation of tracking performance and even loss of targets. To improve the feature discriminative ability for similar targets, we update the parameters of network online, and make the network further learn the specific characteristics of the current target based on the common features. The proposed method can not only effectively distinguish the target and background, but also eliminate interference from similar targets. We conduct a large number of experiments on the OTB50 and OTB100 databases. The results show that the proposed algorithm can improve the discriminative ability to features extracted by the network and achieve robust tracking of the target.

Key words machine vision; visual tracking; offline training; online update
OCIS codes 150.1135; 100.4996; 100.4999

1 引 言

作为计算机视觉研究领域的重要课题,目标跟踪具有广泛的应用前景,在过去几十年一直备受关注,并取得长足的发展。卷积神经网络因学习获得的深度特征具有强大的目标表示能力,逐渐取代传统的手工特征,在图像识别和目标检测等任务中取得突破性进展,近几年被引入到目标跟踪任务中,用于提高跟踪的稳健性和准确性[1-2]。

最近,基于孪生神经网络的目标跟踪方法在一些流行的跟踪数据库和竞赛中展现了优秀的性能。这些方法通过一个 Y 型网络[3],基于超大的视频目标检测(VID)[3]数据库和离线训练网络,学习目标的通用特征表示,这种表示同时具有度量图像间相似程度的作用,在后续的图像帧中通过判别与初始帧目标最相似的区域,从而稳健地估计目标的状态。但是,在跟踪过程中,因网络参数巨大和训练数据缺乏,无法在线更新网络以适应目标在时域上的变化,在目标周围出现相似目标时,跟踪性能下降,甚至会丢失目标。

导致上述现象的一个主要原因在于网络所学到的特征对时域变化目标判别力不够。相似性匹配算法大都通过相似性匹配训练网络,利用最后一层卷积层来提取目标的语义信息,使相似目标具有类似

收稿日期: 2019-03-13; **修回日期**: 2019-04-06; **录用日期**: 2019-05-06
基金项目: 国家自然科学基金(61773397,61703423,61701524)
* **E-mail**: 2530858535@qq.com

的语义信息。然而此类算法只采用离线训练网络，并没有实现网络的在线更新。因此，网络只学到了相似目标的通用特征，当目标周围出现相似的干扰项时，离线训练所学到的通用特征无法将其进行区分，即无法实现对特定目标的表达。

因此，人们尝试利用在线训练网络来提升目标跟踪的准确性。与基于端到端学习型相关滤波器(CFNet)[4]直接对在跟踪过程中网络提取的特征进行多项式拟合来表示目标不同，Guo 等[5]提出动态孪生网络(Dsiam)，通过在线学习目标外观变化和背景抑制，提高跟踪性能。Wang 等[6]通过 U 形网络(RASNet)保留目标更多细节信息。双重孪生网络的视觉物体跟踪方法[7]是利用图像识别模型来弥补相似模型的不足，从而提高所学特征的判别力。但是，这些方法都是利用训练好的深度特征来描述语义目标，而训练样本中并不包含跟踪目标，导致特征对当前跟踪目标的表达力不够，特别是当背景中含有相似语义目标时，跟踪器将无法鉴别，导致跟踪失败。为提高网络对特定目标的判别能力，本文直接利用跟踪结果对深度特征进行增量学习，在跟踪过程中直接增加训练步骤，以获得表示当前特定目标判别力更强的深度特征，从而能够区分跟踪目标和相似背景。

2　本文方法

基于全卷积孪生网络的目标跟踪(SiamFC)[3]算法采用大量离线数据训练网络参数，学习相似目标之间的通用特征。在线跟踪中，采用此种通用特征来实现对目标的表示，其损失函数为

$$L=\mathrm{argmin}\frac{1}{N}\sum_{i=1}^{N}\{L[v(f(x_i),f(z_i)),y_i]\},\tag{1}$$

式中：N 为正样本个数；v 为由目标模板 x 和搜索区域 z 得到的响应，即 $v=f(x_i)*f(z_i)$，$f(x_i)$ 为目标模板 x 经最后一层卷积层得到的特征，$f(z_i)$ 为搜索区域 z 经最后一层卷积层得到的特征，$*$ 为卷积；y_i 为样本所对应的标签，$y_i\in\{+1,-1\}$。

通过离线训练，SiamFC 算法学到了相似目标之间的共性。但是，目标跟踪任务是对当前特定目标进行跟踪，单独的通用特征只能反映当前目标的共性部分，对当前目标的特定部分有所忽视。

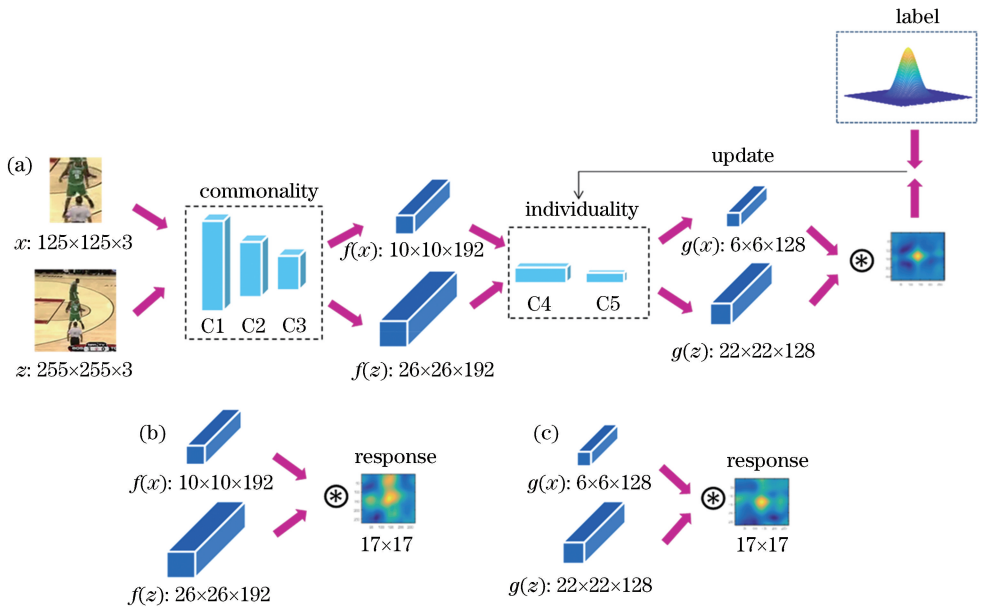

图 1　网络在线更新流程图。(a)网络训练模型；(b)通用特征响应；(c)特定特征响应

Fig. 1　Flow chart of network online update. (a) Network training model; (b) response of common features; (c) response of special features

SiamFC 算法通过离线训练目标模型，当目标周围出现干扰选项时无法区分。本文算法以 SiamFC 算法为基础算法，结合离线训练的通用特征和在线更新的特定特征，以提高对特定目标的表达能力。离线训练是通过大量数据训练网络参数，学习相似目标之间的通用特征；在线更新则是根据当前目标训练网络，提取当前目标的特定特征，进而更新目标模型，提高目标跟踪的准确度。图 1 为在线更新网络结构图。其中，图 1(a)表示网络训练模型，x 通过首帧目标框确定，输入大小为 125×125×3；z 为不同帧数所保存的正样本，大小为 255×255×3。将 x 和 z 输入网络中，通过在目标模板与

搜索区域的网络 C1 层、C2 层和 C3 层中提取相似目标之间的通用特征得到 $f(x)$和 $f(z)$。利用卷积层 C4 层和 C5 层提取相似目标的特定特征得到 $g(x)$和 $g(z)$，将所获得的特定特征进行卷积得到模板在搜索区域上的响应图，利用正样本来提取深度层次特征，基于得到的响应与样本标签可实现对网络参数的更新。图 1(b)表示对相似目标的通用特征进行相关卷积得到的响应图，图 1(c)表示对特定特征进行相关卷积得到的响应图。

2.1　通用特征学习

本文网络为拥有五层卷积层的深度卷积图像分类神经网络(AlexNet)[8]，在离线训练过程中，采用与 SiamFC 算法类似的思路，利用 VID 数据库实现对网络参数的训练，学到相似目标间的通用特征。损失函数 L 为逻辑回归函数。在跟踪过程中，保持前三层卷积层参数不变，用来提取目标的共性特征 $f(x)$和 $f(z)$。

$$v_c = R[f(x), f(z)], \tag{2}$$

式中：$f(x)$为目标模板 x 经过前三层卷积层得到的通用特征；$f(z)$为搜索区域 z 经过前三层卷积层得到的通用特征；v_c 为目标模板与搜索区域经过 $R(\cdot)$相关得到的响应。

由图 1 可知，通用特征可以提取目标模板 x 与搜索区域 z 中相似的部分，实现跟踪目标与背景的初步区分。从图 1(b)可知：根据对通用特征进行相关卷积运算得到的响应图可以区分目标与背景，但不能有效区分有相似干扰时的特定目标。

2.2　特定特征学习

离线训练学习的是相似目标间的通用特征，但通用特征会对特定目标和相似干扰项同时产生较大的响应，无法进行区分。本文以离线训练好的网络为基础，根据样本在线训练卷积层 C4、C5 的参数，通过离线训练和在线更新，可以同时表达目标的共性与个性。

输入首帧的目标模板 x，算法的搜索区域 z。先将 x 和 z 输入网络中，经过固定参数的前三层卷积层提取通用特征 $f(x)$和 $f(z)$，将通用特征 $f(x)$和 $f(z)$经后两层卷积层学习当前目标的个性，得到特定特征 $g(x)$和 $g(z)$。

$$v_p = R[g(x), g(z)], \tag{3}$$

式中，v_p 为根据目标模板与搜索区域得到的响应值。将响应值与标签同时代入损失函数 L 中，实现对后两层卷积层参数的更新。

在线跟踪中，在通用特征的基础上利用正样本来学习特定目标的特定特征，进行相关卷积得到其响应图，实现对特定目标的有效区分。其损失函数为

$$L = \operatorname{argmin} \frac{1}{N} \sum_{i=1}^{N} \{L(v_p, y_i)\}。 \tag{4}$$

由图 1(c)可知，经过最后一层卷积层得到的特征可以实现目标模板与背景的区分，目标模板只对搜索区域中属于目标的部分进行响应，所得到的响应图只包含特定目标，从而实现对目标的区分。

2.3　网络在线更新

在跟踪的初始阶段，根据首帧目标框绝对准确的特点，在首帧样本中提取正负样本，可以确定训练样本的准确性。但是在跟踪阶段，仅采用首帧样本远远满足不了更新的要求，因为首帧样本数量太少，并且没有目标和背景的变化。而高分样本是通过设定阈值，在跟踪精确度的帧中提取目标样本并保存，这样可以有效提高样本的数目，并对样本的准确性也有一定的保证。

考虑到跟踪的效率，每帧更新在很大程度上会对跟踪速度产生较大的影响。为减小在线更新对跟踪速度的影响，本文主要采用首帧更新、间隔更新和失败更新三种方式。

1) 首帧更新。在首帧利用首帧模板 x_f 和搜索区域 z_f 实现对网络参数 w_f 的更新，可使网络初步学习当前目标的个性，对后续更快地学习目标个性具有较大作用。

$$w_f = U(x_f, z_f), \tag{5}$$

式中，w_f 为首帧更新后的网络参数，$U(\cdot)$是更新网络操作。

2) 间隔更新。利用高分样本并采用间隔若干帧的方式来更新网络参数：

$$w_i = \frac{1}{N} \sum_{i=1}^{N} U(x_f, z_i), \text{ if } T = M \times K \quad (K = 1, 2, 3, \cdots), \tag{6}$$

式中：w_i 为间隔更新后的网络参数；z_i 为第 i 个样本的搜索区域；T 为当前帧数；M 为更新间隔；K 为更新次数。

3) 失败更新。当跟踪失败时，说明当前网络对目标的表达能力不够，无法将目标与背景进行有效区分，在此时利用高分样本进行更新，可以提高网络对目标的表达力。

$$w_a = \frac{1}{N} \sum_{i=1}^{N} U(x_f, z_i), \text{ if } v_p < Q, \tag{7}$$

式中：w_a 为失败更新后的网络参数；Q 为判断跟踪

失败的阈值。通过在线更新网络参数，在浅层网络提取的通用特征基础上，精调深层网络以适应目标变化，进一步消除相似目标的干扰。

图 2 为本文算法与 SiamFC 算法的对比图，选取 basketball、girl2、liquor 3 个具有相似干扰的序列。第二行是 SiamFC 算法对三个序列的目标响应图，当目标周围出现相似干扰项时，SiamFC 算法中所学到的特征无法很好地区分目标与干扰项。第三行是本文算法的目标响应图，从中可以看出通过增加网络的在线更新，利用通用特征与特定特征之间的互补，能够有效提高网络对特定目标的判别力，从而可以很好地区分目标与背景。

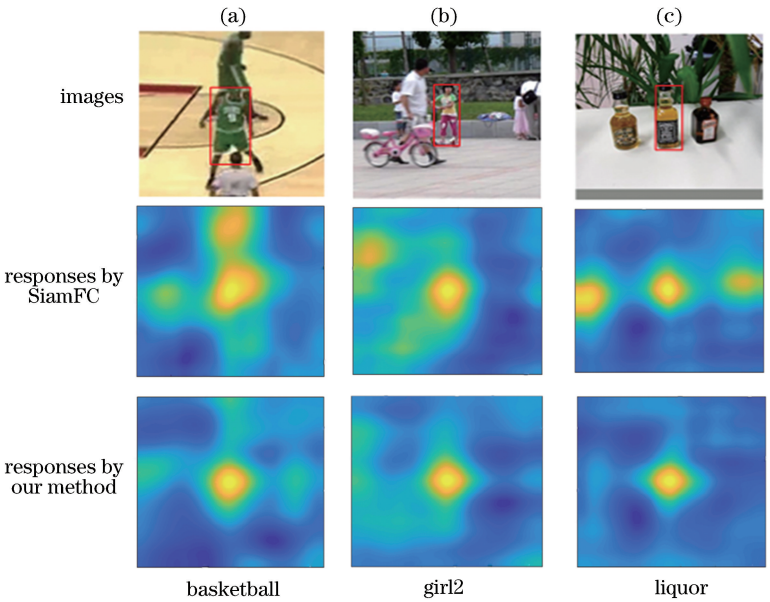

图 2　相似干扰序列中跟踪目标响应图。(a)篮球；(b)女孩；(c)酒

Fig. 2　Tracking target responses in similar interference sequence. (a) Basketball; (b) girl2; (c) liquor

3　实验分析

实验的硬件环境为 multi-core CPU、NVIDIA 1080TI GPU，软件环境为 MATLAB 2017a。离线训练的数据是根据用于检测任务的视频序列 ILSVRC2015[9] 进行改造得到的专门用于训练孪生网络的数据。训练网络为拥有五层卷积层的 AlexNet，学习率 $\eta=0.001$。在第一帧初始化时，采用较多的迭代次数(150 次)训练模型，使其能够收敛；在跟踪过程中，由于目标变化较小，采用较少的迭代次数(5～10 次)就能够使模型收敛；离线训练过程中，实现对五层卷积层参数的更新，在线更新过程中，固定前三层卷积层的参数，实现对后两层卷积层参数的更新。测试视频来自于当前目标跟踪领域最常用的 OTB50[10] 和 OTB100[11] 数据库。

3.1　参数设置标准

为证明增加模型更新对算法的提升，在 OTB100 数据库中进行实验，并分别就不同的更新条件(首帧更新、间隔更新、失败更新和整体性能)对算法进行说明。其中，首帧更新是采用首帧的目标模板和搜索区域对网络参数进行训练；间隔更新是每间隔 30 帧利用保存的样本对网络参数进行更新；失败更新是当出现跟踪失败的情况时对网络参数进行更新；整体性能是同时采用首帧更新、间隔更新和失败更新后得到的实验结果。经过多次实验得到失败更新中的阈值 $Q=0.3$。

图 3 中，基准算法为 SiamFC 算法，OSFC_F 是在 SiamFC 算法的基础上增加首帧更新，OSFC_FI 是在 SiamFC 算法的基础上增加首帧更新和间隔更新策略，OSFC_FF 是在 SiamFC 算法的基础上增加首帧更新和失败更新，OSFC_OP 是在 SiamFC 算法的基础上增加首帧更新、间隔更新和失败更新。从图中可知，只采用首帧更新的实验结果得分较低，精确率和成功率只能达到 0.669 和 0.491。当增加间隔更新后，通过 30 帧间隔对网络参数进行更新，实验结果有了很大的提升，精确率和成功率分别达到了 0.810 和 0.587，说明使用时间信息更新网络参数可以使网络模型更加适应目标的时域变化。在首帧更新的基础上增加失败更新，利用之前保存的正样本对网络模型进行更新，提高网络模型对正负样本的判别能力，精确率和成功率分别达到了 0.795 和 0.579。当同时使用首帧更新、间隔更新和失败更

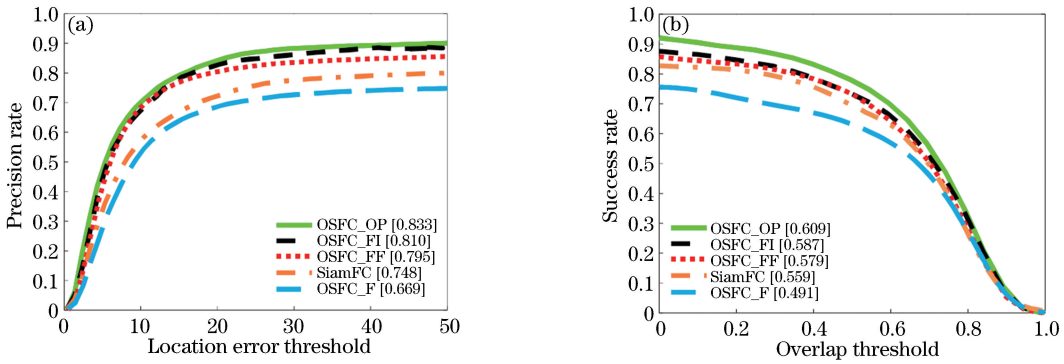

图 3　不同更新策略下的精确率与成功率对比图。(a)精确率图;(b)成功率图

Fig. 3　Comparison of precision and success rates under different update strategies. (a) Plot of precision rate; (b) plot of success rate

新时,算法的精确度和成功率最高,分别为 0.833 和 0.609,在 SiamFC 算法的基础上分别提高了 8.5% 和 5%。在线更新方式的引进使得跟踪性能得到提升,但是更新的方法需要采用梯度下降法对新收集到的样本进行迭代训练和网络模型在线更新,这将会导致更新时间与解析解相比所需要的时间更长,使得跟踪速度有所下降。

3.2　OTB 数据库结果分析

3.2.1　总体性能分析

选取 OTB50 和 OTB100 数据库作为测试数据库,分别与其他算法进行比较与分析。

1) OTB50 基准数据库。选取基于判别式相关滤波网络的视觉跟踪(DCFNet)[12]、基于认知心理学的目标跟踪(MUSTer)[13]、基于完全卷积网络的视觉跟踪(FCNT)[14]、基于卷积神经网络学习判别显著图的在线跟踪(CNN-SVM)[15]、孪生网络实例跟踪研究(SINT)[16]、基于学习型空间正则化相关滤波的视觉跟踪(SRDCF)[17]、多专家跟踪(MEEM)[18]、SiamFC 算法与本文算法进行对比。其中,DCFNet、SINT、FCNT 是基于孪生网络的算法,SiamFC3s 是在 3 尺度条件下 SiamFC 算法公布的实验结果,SiamFC 是根据公布的代码在 5 尺度条件下得到的实验结果,同时也是本文的基准算法。由图 4 可知,本文算法在 SiamFC 的基础上实现了精确度为 11%,成功率为 7.1%的提高。

2) OTB100 基准数据库。OTB100 数据库是在 OTB50 数据库上改进的数据库,为证明本文算法的有效性,在 OTB100 数据库上也进行了对比实验,对比算法有基于深度学习型空间正则化相关滤波的视觉跟踪(DeepSRDCF)[19]、对冲深度跟踪(HDT)[20]、基于分层卷积特征的视觉跟踪(HCF)[21]、DCFNet、CNN-SVM、SRDCF、MEEM、SiamFC 算法。由图 5 表明,本文算法的精确度为 0.833,成功率为 0.609,在 SiamFC 算法的基础上分别提高了 8.5%和 5%。

3.2.2　属性分析

为全面评估跟踪算法在不同难点属性上的性能,图6给出了各个跟踪器在各个难点属性下精确

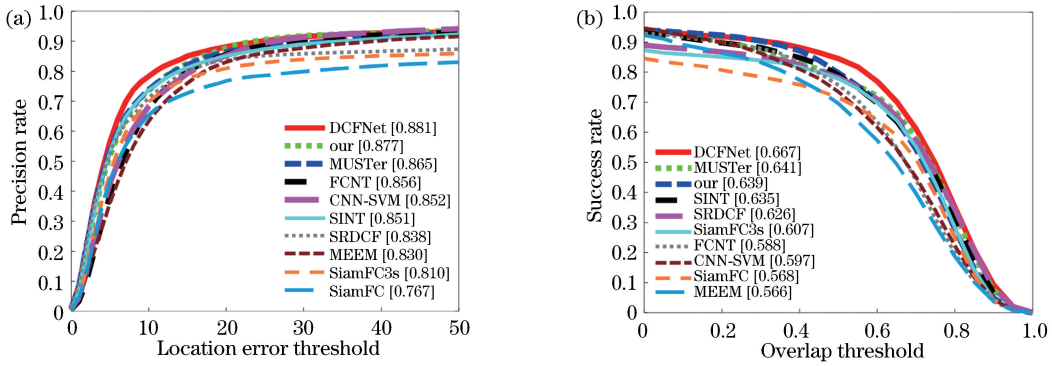

图 4　不同算法在 OTB50 视频库中的精确率与成功率对比图。(a)精确率图;(b)成功率图

Fig. 4　Comparison of precision and success rates of different algorithms in the OTB50 video library. (a) Plot of precision rate; (b) plot of success rate

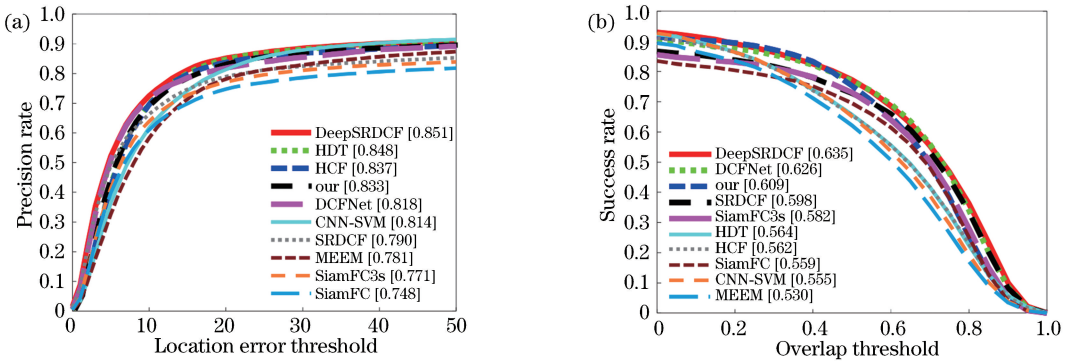

图 5　不同算法在 OTB100 视频库中的精确率与成功率对比图。(a)精确率图;(b)成功率图

Fig. 5　Comparison of precision and success rates of different algorithms in the OTB100 video library. (a) Plot of precision rate; (b) plot of success rate

率的分析曲线。通过分析曲线可以看出,在 8 个难点属性里,本文算法在平面旋转、低分辨率、运动模糊、遮挡、超出视野、平面外旋转、光照变化、尺度变化条件下取得较好的成绩,分别达到了 0.829,0.957,0.790,0.834,0.798,0.860,0.830,0.898,并且相比于 SiamFC 算法有了较大的提高。由图 6 可知,通过在线模型更新,提高模型的特征表达能力,本文算法在大部分环境下的稳健性均有所提高。

3.2.3　定性分析

离线训练学习相似目标之间的共性特征,在线更新学习当前目标的个性特征,通过结合目标的共

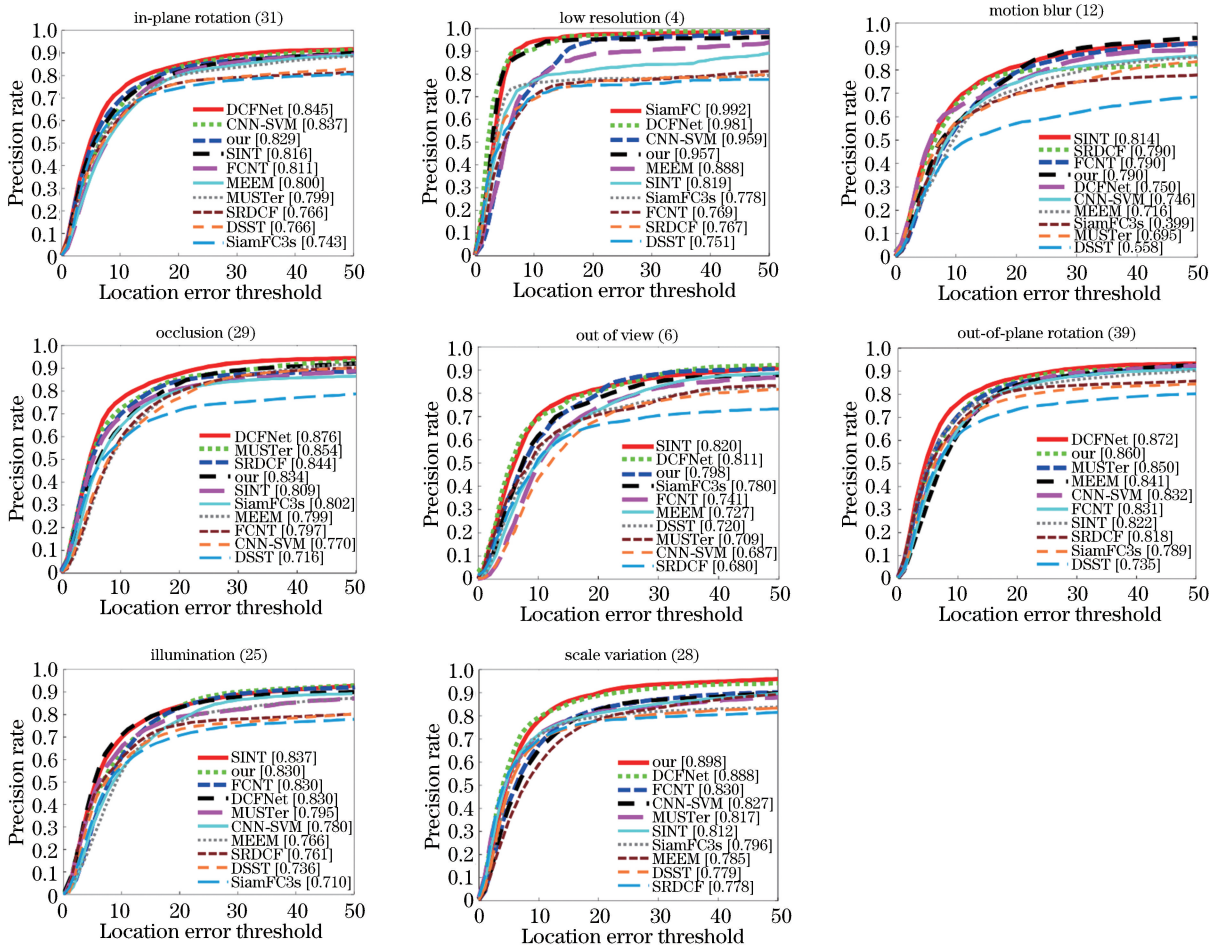

图 6　不同难度属性下的各算法跟踪结果的精确率曲线图

Fig. 6　Precision rate graph of each algorithm under different difficulty attributes

性特征和个性特征，可以提高对当前目标的判别能力。为了证明这种判定能力，从 OTB100 数据库中选取 6 个具有各种跟踪难点的视频序列，与 DCFNet、HCF、CNN-SVM、SiamFC3s、SiamFC 进行对比验证。

由图 6 可知，通过与各算法进行对比，发现本文算法对各跟踪难点都比较稳健，特别是在处理相似目标的问题方面。为了可视化显示跟踪效果，本文结合部分测试视频的难点属性和部分视频的部分跟踪结果进行针对性分析，如图 7 所示。

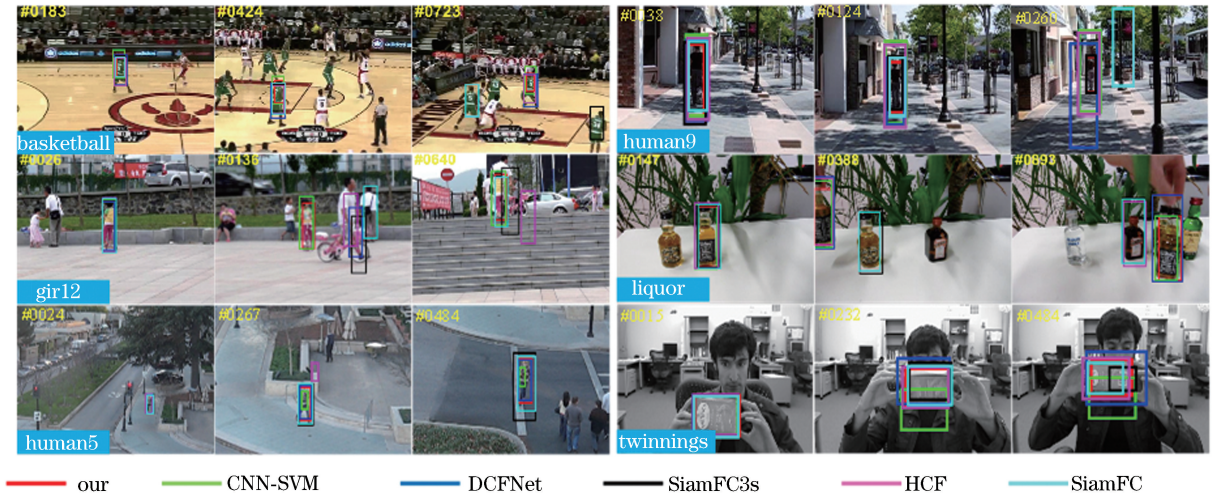

图 7　各算法在不同难度属性的视频中的实际追踪效果图

Fig. 7　Actual tracking effect of each algorithm for vidios with different difficulty attributes

对于 basketball 测试视频的跟踪，其主要难点在于相似目标的干扰。参与测评的跟踪器在第 723 帧开始产生分化，此时目标周围有较为密集的相似干扰，并且有部分相似的背景对目标有了部分遮盖。CNN-SVM 和 SiamFC3s 都直接跟丢了目标，其他跟踪器都有不同程度地偏离目标或者是有跟踪效果变差的情况出现。而本文算法由于模型更新方式的改进，在面对较多相似的干扰目标时依旧能达到优异的效果。

对于 girl2 测试视频的跟踪，其主要难点在于背景中的人物对于目标的遮挡和相似干扰。由于相似的遮挡出现，除了本算法和 CNN-SVM 之外的算法都同时跟丢了目标，CNN-SVM 的目标框大小并不能精准地定位目标。在后续帧中除 HCF 和 DCFNet 之外的其他算法都找回了目标，但是跟踪效果如第 640 帧中所显示，效果都受到了不同程度的影响。但是本文算法跟踪效果一直较好，在较大遮挡的情况下依旧能够有非常出色的跟踪效果。

对于 liquor 测试视频的跟踪，其主要难点是场景中其他背景对目标产生的相似干扰以及目标本身的快速移动，由于难度属性的叠加，本测试视频对于跟踪器的要求相对较高，在第 388 帧中由于目标发生快速移动并且超出了边框，SiamFC 算法跟踪器直接跟丢目标，并且各个算法都出现了跟踪效果变差的情况。在第 893 帧中，背景中出现相似干扰，只有本文算法和 CNN-SVM 还有较为优异的跟踪效果，其他的算法都产生了跟丢或是效果变差的现象。说明本文算法对于快速移动也有很好的跟踪效果。

对于 twinnings 测试视频的跟踪，其主要难点是场景中复杂的背景对目标跟踪的较大干扰。在第 15 帧中所有的跟踪算法都有较好的效果，但是在第 232 帧中，由于目标开始移动，此时背景对于目标的干扰开始加强，各算法的跟踪效果开始分化，DCFNet 和 CNN-SVM 的跟踪效果开始变差，并开始脱离目标。在第 484 帧中，除了本文算法，其他算法的跟踪效果都有不同程度地变差，并且其他算法的目标框都发生偏移或是尺度与目标不符合的情况。说明本文算法在复杂背景下也能有很好的跟踪效果。

4　结　　论

目前，基于相似性匹配算法仅使用离线训练网络参数来学习相似目标之间的一般相似性。在在线跟踪过程中，由于跟踪目标是特定目标，因此一般特征无法实现对当前特定目标的表达，会受到周围相似目标的干扰。基于这些问题，增加基于相似性匹配算法的网络在线更新，利用跟踪结果对网络的最后两层进行增量学习，网络基于通用特征学习当前

特定目标的特定特征，并将通用特征与特定特征相结合，实现对特定目标的表达。

参 考 文 献

[1] Li S S, Zhao G P, Wang J Y. Distractor-aware object tracking based on multi-feature fusion and scale-adaption[J]. Acta Optica Sinica, 2017, 37(5): 0515005.
李双双, 赵高鹏, 王建宇. 基于特征融合和尺度自适应的干扰感知目标跟踪[J]. 光学学报, 2017, 37(5): 0515005.

[2] Li Z D, Zhong Y, Chen M, *et al*. Fast face image retrieval based on depth feature[J]. Acta Optica Sinica, 2018, 38(10): 1010004.
李振东, 钟勇, 陈蔓, 等. 基于深度特征的快速人脸图像检索方法[J]. 光学学报, 2018, 38(10): 1010004.

[3] Bertinetto L, Valmadre J, Henriques J F, *et al*. Fully-convolutional Siamese networks for object tracking[M]//Hua G, Jégou H. Computer vision-ECCV 2016 workshops. Lecture notes in computer science. Cham: Springer, 2016, 9914: 850-865.

[4] Valmadre J, Bertinetto L, Henriques J, *et al*. End-to-end representation learning for correlation filter based tracking[C]//2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), July 21-26, 2017, Honolulu, HI. New York: IEEE, 2017: 2805-2813.

[5] Guo Q, Feng W, Zhou C, *et al*. Learning dynamic Siamese network for visual object tracking[C]//2017 IEEE International Conference on Computer Vision (ICCV), October 22-29, 2017, Venice. New York: IEEE, 2017: 1763-1771.

[6] Wang Q, Teng Z, Xing J L, *et al*. Learning attentions: residual attentional Siamese network for high performance online visual tracking[C]//2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, June 18-23, 2018, Salt Lake City, UT. New York: IEEE, 2018: 4854-4863.

[7] He A F, Luo C, Tian X M, *et al*. A twofold Siamese network for real-time object tracking[C]//2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, June 18-23, 2018, Salt Lake City, UT, USA. New York: IEEE, 2018: 4834-4843.

[8] Krizhevsky A, Sutskever I, Hinton G E. ImageNet classification with deep convolutional neural networks[J]. Communications of the ACM, 2017, 60(6): 84-90.

[9] 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. New York: IEEE, 2016: 770-778.

[10] Wu Y, Lim J, Yang M H. Online object tracking: a benchmark[C]//2013 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), June 23-28, 2013, Portland, OR, USA. New York: IEEE, 2013: 2411-2418.

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

[12] Wang Q, Gao J, Xing J, *et al*. DCFNet: discriminant correlation filters network for visual tracking[J/OL]. (2017-04-13)[2019-03-15]. https://arxiv.org/abs/1704.04057.

[13] Hong Z B, Chen Z, Wang C H, *et al*. MUlti-store tracker (MUSTer): a cognitive psychology inspired approach to object tracking[C]//2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), June 7-12, 2015, Boston, MA, USA. New York: IEEE, 2015: 749-758.

[14] 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. New York: IEEE, 2015: 3119-3127.

[15] Girshick R, Donahue J, Darrell T, *et al*. Rich feature hierarchies for accurate object detection and semantic segmentation[C]//2014 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), June 23-28, 2014, Columbus, OH, USA. New York: IEEE, 2014: 580-587.

[16] Tao R, Gavves E, Smeulders A W M. Siamese instance search for tracking[C]//2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), June 27-30, 2016, Las Vegas, NV, USA. New York: IEEE, 2016: 1420-1429.

[17] Danelljan M, Hager G, Khan F S, *et al*. Learning spatially regularized correlation filters for visual tracking[C]//2015 IEEE International Conference on Computer Vision (ICCV), December 7-13, 2015, Santiago, Chile. New York: IEEE, 2015: 4310-4318.

[18] Zhang J M, Ma S G, Sclaroff S. MEEM: robust tracking via multiple experts using entropy minimization[M]//Fleet D, Pajdla T, Schiele B, *et al*. Computer vision-ECCV 2014. Lecture notes in computer science. Cham: Springer, 2014, 8694: 188-203.

[19] Danelljan M, Hager G, Khan F S, *et al*. Convolutional features for correlation filter based visual tracking [C]//2015 IEEE International Conference on Computer Vision Workshop (ICCVW), December 7-13, 2015, Santiago, Chile. New York: IEEE, 2015: 58-66.

[20] Qi Y K, Zhang S P, Qin L, *et al*. Hedged deep tracking [C]//2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), June 27-30, 2016, Las Vegas, NV, USA. New York: IEEE, 2016: 4303-4311.

[21] 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. New York: IEEE, 2015: 3074-3082.