

DOI:10.19651/j.cnki.emt.2108710

# 融合语义特征网络的孪生网络目标跟踪算法

付谱平<sup>1,2,3</sup> 叶俊<sup>1,2,3</sup>

(1.昆明理工大学信息工程与自动化学院 昆明 650500; 2.昆明理工大学云南省人工智能重点实验室 昆明 650500;  
3.昆明理工大学云南省计算机技术应用重点实验室 昆明 650500)

**摘要:**针对基于孪生网络的反向传播滤波器跟踪算法 CFNet 在遭遇相似物干扰或背景信息与前景目标相似的情况下容易导致模型漂移跟踪效果下降的情况,提出一种融合语义特征网络的孪生网络目标跟踪算法。在图像处理中,通过深度卷积神经网络的深层网络可以提取到丰富的语义信息,这些语义信息在目标发生相似物干扰、运动模糊、目标严重变形等情景时,对目标进行辨识是非常有用的。提出的算法在 CFNet 的原有网络结构上,增加一个语义特征网络,与 CFNet 的外观特征网络形成互补,两个特征网络的训练是独立的以保持两种特征的异质性,在得到各自的响应图后,通过计算这两个响应图的置信度来进行融合,提高了算法的判别能力。实验表明,与其他常用的 5 个算法相比,本文算法达到了最优,能够有效的跟踪目标。

**关键词:**孪生网络;反向传播滤波器;语义特征网络;外观特征网络;目标跟踪

**中图分类号:** TN911.73 **文献标识码:** A **国家标准学科分类代码:** 520.604

## Siamese network target tracking algorithm fused with semantic feature network

Fu Puping<sup>1,2,3</sup> Ye Jun<sup>1,2,3</sup>

(1. Faculty of Information Engineering and Automation, Kunming University of Science and Technology, Kunming 650500, China;  
2. Yunnan Key Laboratory of Artificial Intelligence, Kunming University of Science and Technology, Kunming 650500, China;  
3. Yunnan Key Laboratory of Computer Technology Application, Kunming University of Science and Technology, Kunming 650500, China)

**Abstract:** Aiming at the situation that CFNet, the back propagation filter tracking algorithm based on the siamese network, is likely to cause the model drift tracking effect to decrease when it encounters the interference of similar objects or the background information is similar to the foreground target, a siamese network target tracking algorithm fused with semantic feature network is proposed. In image processing, through the deep network of deep convolutional neural network, rich semantic information can be extracted. These semantic information can cause similar interference, motion blur, severe target deformation, etc. In situations, it is very useful to identify the target. In the proposed algorithm, a semantic feature network is added to the original network structure of CFNet, which is complementary to the appearance feature network of CFNet. The training of the two feature networks is independent to maintain the heterogeneity of the two features and obtain their respective response maps. Later, the fusion is performed by calculating the confidence of the two response graphs, which improves the discriminative ability of algorithm. Tests show that, the algorithm in this paper achieves the optimum and can track the target effectively compared with other 5 commonly used algorithms.

**Keywords:** siamese network; backpropagation filter; semantic feature network; appearance feature network; target tracking

## 0 引言

在计算机视觉领域中有诸如图像增强、图像滤波与降噪、三维重建等等课题,而目标跟踪也是其中的一个研究领

域。在发展几十年后,目标跟踪算法主要形成了生成式模型算法和判别式模型算法<sup>[1-2]</sup>。生成式模型算法相当于目标图像不断的在搜索图像上进行穷尽搜索的模型算法,找到相似度最高的位置作为预测的目标跟踪的区域。深度卷

收稿日期:2021-12-28

积神经网络<sup>[5]</sup>是计算机视觉应用中学习图像表示的有力工具,正是有了深度卷积神经网络的快速发展,才使基于孪生网络的跟踪算法得以出现并取得了不错的成绩。孪生网络跟踪算法可以通过离线训练来学习目标的一般性特征。但是因为训练数据有限且网络参数非常多,使其在跟踪过程中在线更新网络参数时,难以赶上目标在时域上的变化<sup>[4]</sup>。当目标周围出现相似目标进行相似性判别时,会出现无法判别的情况而导致跟踪失败。

因此,为了解决上述问题,许多相关研究者开始尝试利用在线训练网络来提高跟踪的准确性和稳健性。SO-DLT<sup>[6]</sup>(transferring rich feature hierarchies for robust visual tracking)根据其特有的跟踪器在线上轻量级调整,提高了跟踪性能。MDNet<sup>[6]</sup>(learning multi-domain convolutional neural networks for visual tracking)通过他们特有的基于MDNet的在线跟踪框架进行在线更新,提高了跟踪精度。CFNet<sup>[7]</sup>(end-to-end representation learning for correlation filter based tracking)通过其特有的CFNet框架来进行端到端的学习,提高了跟踪的稳健性。DSiam<sup>[8]</sup>(learning dynamic siamese network for visual object tracking)通过学习目标外观变化、减轻背景杂波的影响,提高了跟踪精度。然而,一方面这些方法的一部分是通过离线训练的深度模型来进行目标跟踪,深度模型包含的特定对象信息非常少,当目标背景中出现与目标高度相似的背景时,跟踪器的判断力将会减弱。一方面这些方法中的一部分的在线更新的成本非常高。

为使网络能够更加有效的区分目标和背景,使跟踪过程中更专注于目标本身,并提高对相似目标的判别能力,受文献<sup>[9-11]</sup>启发,本文在原CFNet网络结构基础上,增加一个语义特征网络,组成一个相当于双重孪生网络的网络框架,并分别训练它们以保持其异质性,来提高对相似目标的判别能力。

## 1 CFNet 跟踪算法介绍

CFNet是一种特殊的孪生网络跟踪算法,它的其中一个分支增加了一个相关滤波CF层,并且这两个分支对输入的图像都使用相同的卷积特征变换,在经过卷积特征变换后,用训练图像学习一个线性模板,然后应用互相关操作来搜索测试图像。CFNet算法网络框架如图1所示,图中的五角星表示互相关操作。

CFNet是在全卷积孪生网络的目标跟踪算法SiamFC<sup>[12]</sup>的基础上,将相关滤波器与深度特征进行深度的整合,在SiamFC的网络结构上嵌入了CF层,将卷积神经网络CNN与CF的在线学习效率的优势结合起来,实现端到端的训练。实现这种训练的核心就是可以把CF作为这个网络中的CNN层,并且这个CNN层是可以进行微分操作的,这样就可以通过CF网络层反向的将误差在CNN层中进行传播。

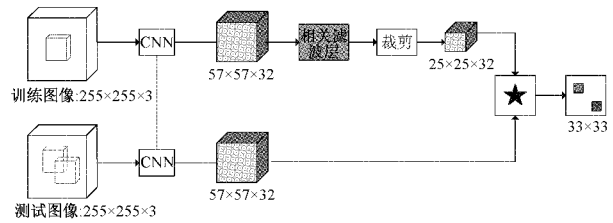


图1 CFNet算法网络框架

CFNet是在SiamFC网络结构中的特征映射和互相关算子之间增加了一个CF层,这种变化可以形式化表示为:

$$h_{\rho,s,b}(x',z') = sw(f_{\rho}(x') * f_{\rho}(z')) + b \quad (1)$$

CF块 $w = w(x)$ 是从训练特征映射 $x = f_{\rho}(x')$ 中计算出一个标准的相关滤波分类模板 $w$ ,它的求解过程是通过求解傅里叶域中的岭回归问题来解决的。它的效果相当于创建一个对平移具有鲁棒性的模板。引入标度 $s$ 和偏倚度 $b$ ,使得分数范围更适合逻辑回归。式中: $x'$ 和 $z'$ 为该网络通过从视频集中随机抽取的数百万对样本示例 $(x',z')$ 的集合,\*表示互相关操作, $\rho$ 为可学习参数, $h_{\rho,s,b}(x',z')$ 为两支分支输出进行相关计算该点的响应值。

CFNet也作为判别式目标跟踪模型,训练判别器来区分正样本、负样本,通过最小化的损失函数得到最优模型如式(2)所示,其逻辑损失函数如式(3)所示。

$$\operatorname{argmin}_{\rho} \sum_i l(h_{\rho,s,b}(x'_i, z'_i), c_i) \quad (2)$$

$$l(y, v) = \log(1 + \exp(-yv)) \quad (3)$$

式中: $c_i \in \{+1, -1\}$ 表示 $(x'_i, z'_i)$ 集合中的每对样本示例在空间映射中的样本标签, $y \in \{+1, -1\}$ 为实际标签, $v$ 表示得分映射中的实际得分。

此外,在获得最优模板 $w$ 过程中,需要在频域上进行有效求解,求解方程组如下:

$$\begin{cases} \hat{k} = \frac{1}{n}(\hat{x}^* \circ \hat{x}) + \lambda \mathbf{I} \\ \hat{a} = \frac{1}{n}\hat{k} - 1 \circ \hat{y} \\ \hat{w} = \hat{a} \circ \hat{x} \end{cases} \quad (4)$$

式中: $\hat{x} = Fx$ 为 $x$ 在傅里叶域上的变换, $x^*$ 为 $x$ 的复共轭, $\hat{y}$ 为 $y$ 在傅里叶域上的变换, $\circ$ 表示元素乘法, $\lambda$ 为正数, $\mathbf{I}$ 为单位矩阵, $\hat{k}$ 为定义循环线性核矩阵的信号, $\hat{a}$ 为拉格朗日乘子符号, $\hat{w}$ 为 $w$ 在傅里叶域进行计算。

CFNet算法模型通过损失函数经过相关滤波网络层反向梯度传播,使卷积神经网络中的学习参数 $\rho$ 能够得到训练,得到适宜良好的跟踪器,上述公式的详细推导反向梯度传递的具体过程详如文献<sup>[7]</sup>所示。

## 2 本文算法

与SiamFC跟踪算法一样,CFNet跟踪算法训练的孪

神经网络对于有语义信息的背景,尤其是与目标是相同类别的物体,往往容易跟踪失败。因此为了提高 CFNet 跟踪算法对相似目标的判别能力,对 CFNet 原有网络进行补充,加入一个语义特征网络来用于增强 CFNet 对于相似目标的判别能力。语义特征是由深度卷积神经网络提取的图像的高级信息,高层次的语义特征信息在目标外观发生变化时,更具有辨识能力。

2.1 本文算法网络框架

本文所提出的网络结构如图 2 所示,图 2 中的五角星表示互相关操作,“+”号表示响应图融合操作。输入的图片是由上一帧图像和当前帧图像组成。图 2 中的  $Z^c$  表示具有上下文信息的目标图像, $Z$  表示只包含跟踪目标信息

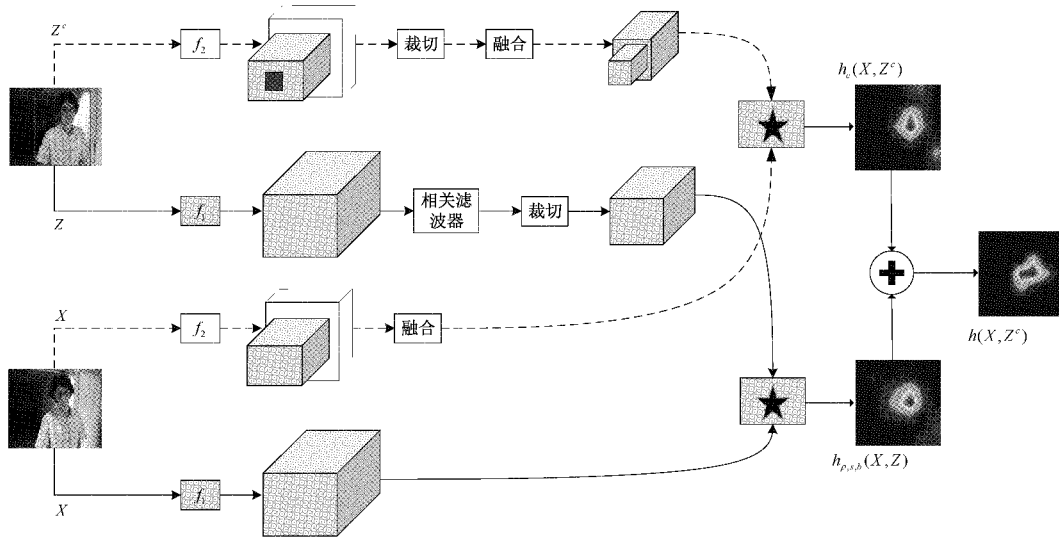


图 2 本文算法网络框架图

2.2 外观特征网络部分

由  $Z$  和  $X$  所在分支网络(实线部分)组成的外观特征网络部分,就是 CFNet 网络结构,提取的特征由  $f_1$  这个卷积神经网络实现,其响应图的计算公式参照式(1),其公式可以形式化表示如下:

$$h_{\rho,s,b}(X, Z) = stw(f_{1,\rho}(X)) * f_{1,\rho}(Z) + b \quad (5)$$

其训练和测试的过程与第一章所述的 CFNet 跟踪算法一致。

2.3 语义特征网络部分

由  $Z^c$  和  $X$  所在分支网络(虚线部分)组成的语义特征网络部分,采用的深度卷积网络是一个预训练好的网络  $f_2$ , 通过  $f_2$  来提取深度特征,提取的语义特征信息层次越深其表征能力更强,因此本文算法选择  $f_2$  输出的最后两层卷积特征 Conv-4 和 Conv-5。

随着  $f_2$  网络层数的增加,其提取的特征图的空间分辨率会下降,为了让提取的语义特征能够更好的适应后续的互相关操作,需要对提取到的分层特征进行融合,图 2 中插入了一个融合模块,通过采用双线性插值的方法来扩大 Conv-4 和 Conv-5 的特征图的分辨率,使它们的尺寸大小

一致,实现融合。搜索区域  $X$  提取的 Conv-4 和 Conv-5 的卷积特征在融合后,其特征向量设为  $g(f_2(X))$ 。双线性插值原理公式如下:

$$P'_i = \sum_j \epsilon_{ij} P_j \quad (6)$$

式中:  $P_j$  表示原特征图,  $P'_i$  表示在经过插值后放大的特征图。 $\epsilon_{ij}$  为插值系数。

$Z^c$  提取的特征包含了整个图像的上下文信息,需要对其提取的卷积特征进行裁剪,将只包含跟踪目标信息的卷积特征  $f_2(Z)$  从  $f_2(Z^c)$  中裁剪出来,然后再通过融合模块进行融合,并通过双线性插值法扩大其分辨率。其响应图的计算公式如下:

$$h_c(X, Z^c) = g(f_2(X)) * g(f_2(Z)) \quad (7)$$

在式(7)中,  $*$  表示互相关操作。它的训练过程与 CFNet 跟踪算法在不添加相关滤波 CF 层退化 SiamFC 跟踪算法后的训练过程一致,其损失函数与式(3)一致,目标图像特征图  $g(f_2(Z))$  会在搜索区域图像特征图  $g(f_2(X))$  上进行穷尽搜索,计算它们的相似度,最后映射成一个相似度得分图,得分图表示如下:

$$L(y, v) = \frac{1}{|D|} \sum_{u \in D} l(y[u], v[u]) \quad (8)$$

在式(8)中,  $D$  表示最后的得分映射,  $u$  表示得分映射中的每一个位置,  $y$  和  $v$  的含义与式(3)中的含义一样, 其训练的卷积参数  $\theta_c$  通过随机梯度下降法 SGD 来进行多次迭代, 表示如下:

$$\operatorname{argmin}_{\theta_c} \frac{1}{N} \sum_{i=1}^N L(y_i, h_c(Z_i^c, X_i; \theta_c)) \quad (9)$$

在式(9)中,  $N$  为训练样本的总对数,  $y_i$  为实际的响应图标签,  $(Z_i^c, X_i)$  为训练的图像对, 通过 SGD 进行多次迭代训练, 来得到最优的卷积参数  $\theta_c$ 。

#### 2.4 响应图加权融合

通过图 2 所示的网络框架, 输入的图像经过本文算法网络框架后得到了两个响应图, 它们的尺度是一样的, 需要对这两个响应图进行置信度检测, 通过检测到的置信度大小来调整这两个响应图的权重占比, 来进行最终的响应图融合, 融合后的响应图  $h(X, Z^c)$  可以预测跟踪目标的位置, 其融合公式如下:

$$h(X, Z^c) = \lambda h_{\rho, s, b}(X, Z) + (1 - \lambda) h_c(X, Z^c) \quad (10)$$

式(10)中,  $\lambda$  为加权系数。  $\lambda$  计算公式如下:

$$\lambda = \frac{APME\{h_{\rho, s, b}\}}{APME\{h_{\rho, s, b}\} + APME\{h_c\}} \quad (11)$$

$$APME = \frac{F_{\max}^2}{\operatorname{mean}(\sum_{row, col} F_{row, col} \cdot \epsilon(F_{row, col} - (F_{\max} + F_{\min})/2))} \quad (12)$$

在式(11)(12)中,  $F_{\max}$ 、 $F_{\min}$  分别表示响应图中的最大值和最小值,  $F_{row, col}$  表示所响应图中第  $row$  行第  $col$  列的响应值。  $\epsilon(\cdot)$  为单位阶跃函数, 主要目的是为了那些响应值非常小的点不影响对跟踪目标的选取, 将其对跟踪目标的选取的影响降为 0。  $APME$  为置信图平均峰值中值能量,  $APME$  既可以反应所检测的响应图的置信图中的波动程度, 也可以反应跟踪模板的置信度。  $APME$  的值越大, 跟踪目标的可信度越高, 则在响应图融合中占比的权重也应当更大。  $APME\{h\}$  为响应图  $h$  的置信度。

### 3 实验结果与分析

#### 3.1 实验环境

本文算法的实验环境如表 1 所示。

表 1 实验环境

名称	指标
操作系统	Ubuntu18.04
处理器	AMD Ryzen7 1700 Eight-Core 3.20 GHz
图形处理器	NVIDIA GeForce 2080Ti
内存	16 G
Python/Tensorflow	3.7/2.4
Cuda/CUDNN	10.2/7.5

#### 3.2 训练集和测试集

本文算法使用的训练集为 ILSVRS-2015<sup>[13]</sup>, 训练集包含了近 4 000 个视频序列和 30 多种目标, 其中被标注的帧数超过 100 万帧。在训练过程中, 每次训练都需要在训练集中随机抽取图像对, 其中一个图像需要裁剪出中心包含跟踪目标图像  $Z$  的  $Z^c$ , 另一个图像需要裁剪出的中心是目标被真实标注过矩形框的  $X$  块。

本文算法使用的测试集为 OTB100<sup>[14]</sup>, 该测试集包含 98 个视频, 100 个测试场景, 包含了背景杂乱、运动模糊、快速运动等多种挑战。

#### 3.3 网络结构

本文算法的网络结构所采用的深度卷积网络的核心 CNN 都是 VggNet-19<sup>[15]</sup>, VggNet-19 相对 AlexNet<sup>[16]</sup> 而言, 它提取的深层特征具有更强的语义信息, 具有更强的表征能力。对于本文所述的外观特征网络部分,  $f_1$  的网络结构与 CFNet 的网络结构是一致的, 它是 VggNet-19 网络去掉了网络中计算量较大的全连接层。对于本文所述的语义特征网络部分,  $f_2$  网络结构是在 ImageNet 上进行了预训练的 VggNet-19 网络。  $f_1$  和  $f_2$  的维度都是相同的。

#### 3.4 评估指标

本文算法采用成功率和跟踪精度这两个指标来对算法进行评估。所谓的成功率就是通过计算在跟踪过程中预测的目标跟踪区域与真实的目标跟踪区域之间的重叠率, 并设定一个阈值, 如果所计算的重叠率大于这个阈值, 则认为当前帧是有效帧, 计算所有的有效帧数与总帧数的百分比, 即为目标跟踪的成功率。所谓的跟踪精度就是通过计算在跟踪过程中预测的目标跟踪矩形框的中心位置与真实目标跟踪矩形框的中心位置的欧式距离, 并设定一个阈值, 如果所计算的欧式距离小于这个阈值, 则认为当前帧是有效帧, 计算所有的有效帧数与总帧数的百分比, 即为目标跟踪的跟踪精度。

#### 3.5 消融实验

为了更直观的看到语义特征网络对本文算法的有益效果, 本文分别将只包含语义特征网络部分的跟踪算法、只具有外观特征网络的跟踪算法即原 CFNet 跟踪算法以及同时具有语义特征网络部分和外观特征网络部分的本文算法进行对比, 如表 2 所示, 充分说明具有语义特征网络和外观特征网络的本文算法提高了目标的跟踪效果, 形成了互补效果。

表 2 加入语义特征网络的 OTB 评价分析

外观特征 网络部分	语义特征 网络部分	OTB100	
		跟踪精度	成功率
✓		0.793	0.752
	✓	0.692	0.748
✓	✓	0.847	0.779

3.6 定量分析

为了进一步的证明本文算法的有效性,本文算法不仅将与 SiamFC、CFNet 这两种基于孪生网络结构的跟踪算法进行比较,还将与基于相关滤波的判别类模型的跟踪算法 KCF<sup>[17]</sup>、SAMF<sup>[18]</sup>、DSST<sup>[19]</sup>算法进行比较,使对比实验更加完善和更具说服力,并且采用 OPE(一次通过)的评估方式,如图 3 所示,本文算法不论是跟踪精度和成功率都达到了最佳。

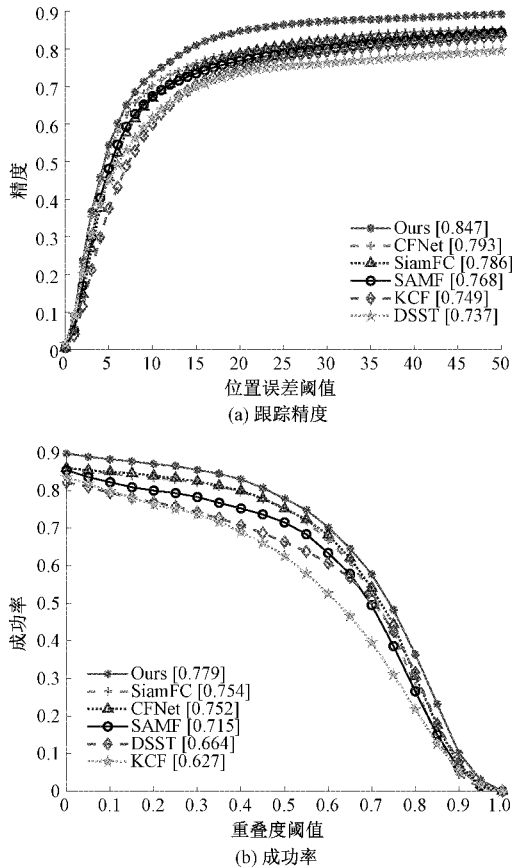


图 3 不同跟踪算法的跟踪精度和成功率曲线图

更加具体的详细数据如表 3 所示,受限于实验环境和硬件的不同,不同算法的跟踪结果可能有所不同,仅用于参考和比较。此外,本文算法的跟踪速度性能由于添加了一个语义特征网络进行训练,所以速度相比 SiamFC 和 CFNet 有所下降,但速度依然达到了 59 fps,非常适用于在对实时性有一定要求但要求不是非常高的部分视频监控领域。

表 3 不同跟踪算法的对比结果

算法	平均跟踪精度	平均成功率	跟踪速度性能/fps
Ours	0.847	0.779	59
SiamFC	0.786	0.754	77
CFNet	0.793	0.752	88
KCF	0.749	0.627	241
SAMF	0.768	0.715	20
DSST	0.737	0.664	42

为了说明本文算法的稳健性,表 4 和 5 分别给出了本文算法与其他 5 种算法在 9 种具有挑战性的干扰因素的平均跟踪精度和平均跟踪成功率。对结果排名第一的数值进行加粗,排名第二的数值使用下划线进行标记。可以看出本文算法在不同的干扰因素下,不论是在平均跟踪精度还是平均跟踪成功率上,大都是处于第一的位置。

3.7 定性分析

为了进一步分析本文算法的跟踪性能,通过在 OTB100 选取一些具有挑战性的视频序列来进行定性分析。如图 4 所示,(a)(b)(c)(d)(e)都存在相似物干扰或者背景与前景目标相似的场景,背景具有语义信息,本文算法通过语义特征网络可以提取特征的高级语义信息,与外观特征网络提取的特征信息形成了良好的互补,能够成功的跟踪目标。此外,这些视频序列还存在旋转、快速运动、运动模糊等这些干扰因素,并从图 4 中的定性结果来看,

表 4 不同跟踪算法在 OTB100 上的平均跟踪精度

算法	平面外旋转	平面内旋转	快速运动	运动模糊	光照变化	形变	遮挡	尺度变化	背景杂乱
Ours	<b>0.830</b>	<b>0.795</b>	<b>0.797</b>	<b>0.723</b>	<b>0.793</b>	<b>0.805</b>	<b>0.873</b>	<b>0.805</b>	<b>0.806</b>
SiamFC	0.761	0.735	<u>0.723</u>	<u>0.671</u>	0.709	0.696	0.766	<u>0.773</u>	0.719
CFNet	<u>0.771</u>	0.754	0.691	0.660	0.728	0.739	0.736	0.767	<u>0.796</u>
KCF	0.733	0.716	0.602	0.605	0.728	0.763	0.776	0.666	0.740
SAMF	0.750	0.704	0.608	0.584	0.682	<u>0.792</u>	<u>0.822</u>	0.717	0.659
DSST	0.739	<u>0.760</u>	0.513	0.507	<u>0.730</u>	0.675	0.709	0.705	0.678

在许多其它跟踪算法的跟踪效果不好或者跟踪失败的情况下,本文算法依然跟踪成功,这充分说明在 CFNet 跟踪

算法基础上,添加一个语义特征网络,能够与外观特征网络形成很好的互补,增强目标跟踪的跟踪性能。

表5 不同跟踪算法在 OTB100 上的平均跟踪成功率

算法	平面外旋转	平面内旋转	快速运动	运动模糊	光照变化	形变	遮挡	尺度变化	背景杂乱
Ours	<b>0.739</b>	<u>0.707</u>	<b>0.768</b>	<b>0.737</b>	<b>0.743</b>	<u>0.778</u>	<b>0.799</b>	<u>0.729</u>	<u>0.742</u>
SiamFC	<u>0.721</u>	<b>0.715</b>	<u>0.699</u>	<u>0.636</u>	0.670	0.644	0.733	<b>0.741</b>	0.681
CFNet	0.713	0.698	0.673	0.635	0.668	0.697	0.686	0.714	<b>0.752</b>
KCF	0.605	0.602	0.557	0.555	0.581	0.601	0.638	0.468	0.657
SAMF	0.680	0.642	0.593	0.583	0.641	<b>0.787</b>	<u>0.766</u>	0.633	0.638
DSST	0.641	0.668	0.503	0.493	<u>0.681</u>	0.648	0.644	0.468	0.609

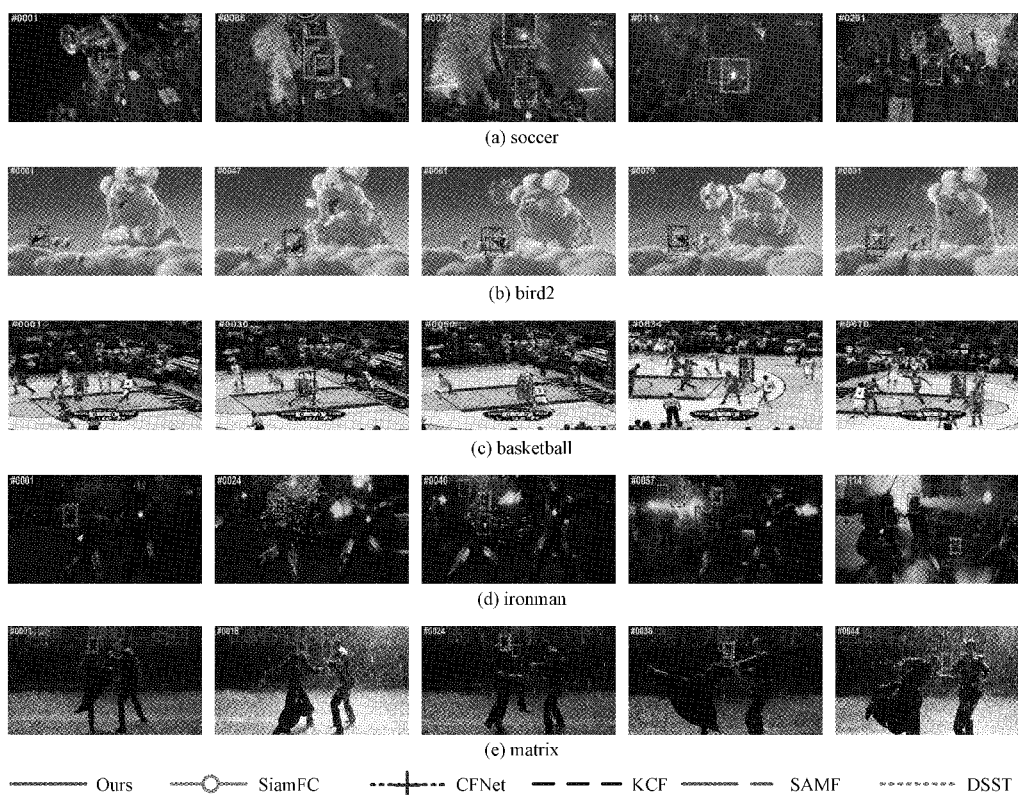


图4 六种跟踪算法的跟踪效果图

## 4 结 论

本文算法在基于孪生网络的反向传播滤波器跟踪算法 CFNet 的基础上,提出了一个融合语义特征网络的孪生网络目标跟踪算法,通过添加一个语义特征网络来提取目标的高级语义信息,与外观特征网络提取的目标特征进行互补,根据各自得到的响应图的置信度来进行融合,有力的提高了在遇到相似目标干扰或者背景信息与前景相似的情况下的判别能力,并通过在 OTB100 基准数据集上的实验结果验证了本文算法在各种干扰因素下对目标跟踪的鲁棒性,在一定程度上提高了目标跟踪的跟踪效果。接下来的工作是继续研究如何在保证目标跟踪的精确度情况下进一步提高跟踪的实时性。

## 参考文献

- [1] 卢湖川,李佩霞,王栋. 目标跟踪算法综述[J]. 模式识别与人工智能, 2018, 31(1): 61-76.
- [2] 陈云芳,吴懿,张伟. 基于孪生网络结构的目标跟踪算法综述[J]. 计算机工程与应用, 2020, 56(6): 10-18.
- [3] KHAN A, SOHAIL A, ZAHOORA U, et al. A survey of the recent architectures of deep convolutional neural networks [J]. Artificial Intelligence Review, 2019: 1-87.
- [4] 陈志旺,张忠新,宋娟,等. 基于目标感知特征筛选的孪生网络跟踪算法[J]. 光学学报, 2020, 40(9): 110-126.

- [5] WANG N, LI S, GUPTA A, et al. Transferring rich feature hierarchies for robust visual tracking [J]. Computer Science, 2015, DOI: arXiv: 1501.04587, 2015.
- [6] NAM H, HAN B. Learning multi-domain convolutional neural networks for visual tracking[J]. IEEE, 2016, DOI: 10.1109/CVPR.2016.465.
- [7] 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), IEEE, 2017; 2805-2813.
- [8] GUO Q, WEI F, ZHOU C, et al. Learning dynamic Siamese network for visual object tracking[C]. 2017 IEEE International Conference on Computer Vision(ICCV), IEEE, 2017; 1763-1771.
- [9] 杨梅, 贾旭, 殷浩东, 等. 基于联合注意力孪生网络目标跟踪算法[J]. 仪器仪表学报, 2021, 42(1): 127-136.
- [10] 易宇凡, 瞿少成, 陶林. 基于重检测和目标遮挡判定的相关滤波跟踪实现 [J]. 电子测量技术, 2020, 43(7): 93-96.
- [11] 周士杰, 彭元喜, 彭学锋. 基于双重孪生网络与相关滤波器的目标跟踪算法[C]. 第二十二届计算机工程与工艺年会暨第八届微处理器技术论坛论文集, 2018; 256-266.
- [12] BERTINETTO L, VALMADRE J, HENRIQUES J F, et al. Fully-convolutional siamese networks for object tracking [C]. Proc of the European Conference on Computer Vision, 2016; 850-865.
- [13] 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.
- [14] WU Y, LIM J, YANG M H. Object tracking benchmark [J]. IEEE Transactions on Pattern Analysis Machine Intelligence, 2015, 37 (9): 1834-1848.
- [15] SIMONYAN K, ANDREW Z. Very deep convolutional networks for large-scale image recognition[J]. Computer Science, 2014, DOI: arXiv: 1409.1556, 2014.
- [16] ALEX K, SUTSKEVER I, GEOFFREY E H. Imagenet classification with deep convolutional neural networks [J]. Advances in Neural Information Processing Systems, 2012, 141(5): 1097-1105.
- [17] HENRIQUES J F, CASEIRO R, MARTINS P, et al. High-speed tracking with kernclized correlation filters[J]. IEEE Transactions on Pattern Analysis Machine Intelligence, 2015, 37(3): 583-596.
- [18] YANG L, ZHU J. A scale adaptive kernel correlation filter tracker with feature integration[C]. Proc of the European Conference on Computer Vision, 2014; 254-265.
- [19] DENELIJAN M, KHAN F S, FELSBURG M, et al. Discriminative scale space tracking [J]. IEEE Transactions on Pattern Analysis Machine Intelligence, 2017, 39(8): 1561-1575.

### 作者简介

付谱平(通信作者), 硕士研究生, 主要研究方向为计算机视觉、目标跟踪。

E-mail: 1294502579@qq.com

叶俊, 硕士研究生, 主要研究方向为计算机视觉、图像处理。

E-mail: 1548622388@qq.com