

# 重构特征联合的多域卷积神经网络跟踪算法

杨大伟<sup>1,2</sup>, 巩欣飞<sup>1\*</sup>, 毛琳<sup>1,2</sup>, 张汝波<sup>1,2</sup>

<sup>1</sup>大连民族大学机电工程学院, 辽宁 大连 116600;

<sup>2</sup>大连民族大学国家民委智能感知与先进控制重点实验室, 辽宁 大连 116600

**摘要** 针对 MDNet 跟踪算法网络模型中存在的特征稳健性差以及目标背景信息丢失导致跟踪失败的问题, 提出一种基于重构特征联合的多域卷积神经网络视觉跟踪算法。基于末端卷积层提取的目标高级特征, 使用反卷积操作上采样, 获得了包含目标背景信息重构特征, 再通过联合目标高级特征和背景信息重构特征的方式增强特征的稳健性, 达到了有效区分目标和背景的目的, 适用于解决跟踪过程中出现的目标遮挡、形变、光照变化等问题。将本文算法分别在 OTB50 和 VOT2015 跟踪测试集上进行测试, 与 MDNet 算法相比, 跟踪精度提升 1.53%, 跟踪成功率提升 2.03%。

**关键词** 机器视觉; 目标跟踪; 卷积神经网络; 重构特征; 特征联合; 反卷积

中图分类号 TP391.41

文献标识码 A

doi: 10.3788/LOP56.191501

## Multi-Domain Convolutional Neural Network Tracking Algorithm Based on Reconstructed Feature Combination

Yang Dawei<sup>1,2</sup>, Gong Xinfei<sup>1\*</sup>, Mao Lin<sup>1,2</sup>, Zhang Rubo<sup>1,2</sup>

<sup>1</sup>College of Mechanical and Electronic Engineering, Dalian Minzu University, Dalian, Liaoning 116600, China;

<sup>2</sup>Key Laboratory of Intelligent Perception and Advanced Control State Ethnic Affairs Commission, Dalian Minzu University, Dalian, Liaoning 116600, China

**Abstract** The tracking algorithm always receives an inaccurate object position because of the poor robustness of the features in the multi-domain network tracking (MDNet) algorithm network model and the loss of the target background information. In this study, we propose a multi-domain convolutional neural network visual tracking algorithm based on the combined reconstructed features. This algorithm performs the deconvolution upsampling operation on an advanced object feature to obtain reconstructed features containing the background information. This advanced object feature is extracted using the end convolutional layer and is combined with the reconstructed feature, which can enhance the robustness of the feature and effectively distinguish an object from the background, thereby improving the object tracking accuracy in situations such as object occlusion, illumination change, and object deformation. The proposed algorithm is tested using the OTB50 and VOT2015 tracking test sets. When compared with the MDNet algorithm, the tracking accuracy and tracking success rate of the proposed algorithm are improved by 1.53% and 2.03%, respectively.

**Key words** machine vision; target tracking; convolutional neural networks; reconstructed feature; feature combination; deconvolution

**OCIS codes** 100.4999; 150.1135; 150.0155

## 1 引言

当前视觉跟踪算法普遍存在的难点是解决跟踪

目标光照变化、遮挡、姿态变化、运动模糊和背景混杂等导致的跟踪失效问题。为解决这些难题, 国内外众多优秀学者提出了各种优秀算法, 当前视觉跟

收稿日期: 2019-03-22; 修回日期: 2019-03-27; 录用日期: 2019-04-19

基金项目: 国家自然科学基金(61673084)、辽宁省自然科学基金(20170540192)

\* E-mail: chengshux@foxmail.com

踪算法可以分为生成式和判别式两类。

基于生成方法的视觉跟踪通过对历史帧中的跟踪目标进行建模来得到模板,再利用模板在后续帧中搜寻与目标最为匹配的跟踪区域。基于生成类的跟踪方法由于只对跟踪目标的信息进行分析,未保留背景信息,所以在复杂环境背景下存在稳健性差、跟踪精度不足等问题。其中比较经典的生成类算法有文献[1]提出的 ASMS 算法,该算法提出了一种基于均值漂移的尺寸自适应算法,通过采用背景比率加权(BRW)和尺寸估计能很好地、自适应地跟踪目标,但该算法对目标遮挡和变形情况的适应性较差,存在稳健性较差的问题。

判别类算法在历史帧中结合跟踪目标和背景的信息训练一个分类器,在后续帧中通过分类器将最优目标区域与背景分离开来。该类方法能有效地利用视频图像中的大部分信息,稳健性较好。文献[2]提出的 Struck 算法是早期的经典判别类算法,该算法采用了一种自适应的基于在线学习的结构化输出支持向量机(SVM)预测分类器,通过引入输出空间直接输出跟踪结果,避开跟踪算法的分类环节;Struck 算法的实时性比较好,但对目标全遮挡、半遮挡的问题难以解决,稳健性较差。文献[3]提到的 MOSSE 算法首次使用了相关滤波的方法,该方法通过频域快速傅里叶变换大大加快了运算速度,具有很好的实时性效果,但该算法只采用单通道灰度图进行跟踪,最终导致算法中目标特征表达信息不足。文献[4]提出了 CSK 算法,该算法针对目标框重叠和特征冗余的问题,采用循环矩阵的方法进行密集采样,保持了很好的实时性,但因只采用了灰度特征,目标信息丢失,导致算法稳健性较差。文献[5]针对 CSK 算法灰度特征信息不足的问题,采用多通道方向梯度直方图(HOG)特征核函数融合的方法,提出了 KCF/DCF 算法,该算法对光照变化、目标遮挡有着很好的适应效果,但对目标尺度变化

的适应效果较差。目前主流的判别类跟踪方法主要分为提取卷积网络(CNN)特征和其他特征相结合的相关滤波算法和基于深度卷积神经网络分类预测的算法。其中基于深度特征的相关滤波算法有:1)文献[6]在 KCF/DCF 算法的基础上提出的 C-COT 算法,该算法通过增加深度卷积神经网络来提取 CNN 特征,由于其特征稳健性差,对目标遮挡、形变情况的适应效果较差;2)文献[7]在 C-COT 算法的基础上提出的 ECO 算法,该算法通过采用联合 CNN、HOG、CN 特征的方法来提高特征表达的有效融合,是目前综合表现最为出色的算法之一。基于深度卷积神经网络分类预测方法的判别类算法有:文献[8]基于 VGG-net 网络提出的 MDNet 算法,该算法是基于卷积神经网络分类预测特性的一种离线学习和在线更新相结合的多域卷积神经网络跟踪算法,该算法在 OTB<sup>[9]</sup>和 VOT<sup>[10]</sup>测试集上都表现得极为出色,是目前基于深度学习的跟踪算法中综合表现最好的算法之一。

判别类算法能随时通过模板来更新,具有良好的自适应性,但其存在的问题主要是特征提取不充分,背景信息易丢失,导致算法稳健性差。本文针对判别类算法中普遍存在的特征稳健性不足的问题,在 MDNet 的基础上,提出一种基于重构特征联合的多域卷积神经网络视觉跟踪算法(RCNet)。

## 2 RCNet 算法

### 2.1 MDNet 算法分析

MDNet 算法提出了一个由共享层和全连接层组成的网络,并且把每一个视频序列视为一个特定域<sup>[11]</sup>,该特定域为辨别目标和背景的二元分类器。图 1 为 MDNet 网络结构简图。在 MDNet 算法中,把视觉跟踪过程中目标的外观特征分为通用特征和特定特征。通用特征是与域无关的属于目标的共同属性,是目标在跟踪过程中光照变化、运动模糊、尺

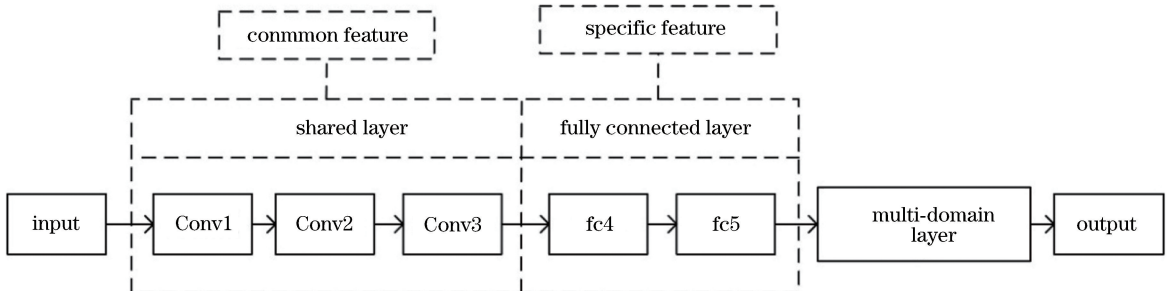


图 1 MDNet 网络结构简图

Fig. 1 MDNet network structure diagram

度变化等的有用特征,在离线训练阶段可利用大量标注的多域视频图像对网络共享层预训练来学习。特定特征是包含特定域信息的特征,是不同目标在跟踪过程中特定信息的表达,可通过跟踪过程中的初始帧来学习,同时需在后续跟踪过程中在线更新。

由于 MDNet 的网络模型在离线训练过程中学

习得到的共享层权重在跟踪过程中保持不变,共享层权重直接决定网络模型对特征的提取能力,因此本文依据文献[12]提出的卷积层特征可视化研究方法对 MDNet 网络共享层中提取的深度特征进行研究,其中 MDNet 跟踪网络的 Conv3 层中提取的特征如图 2 所示。

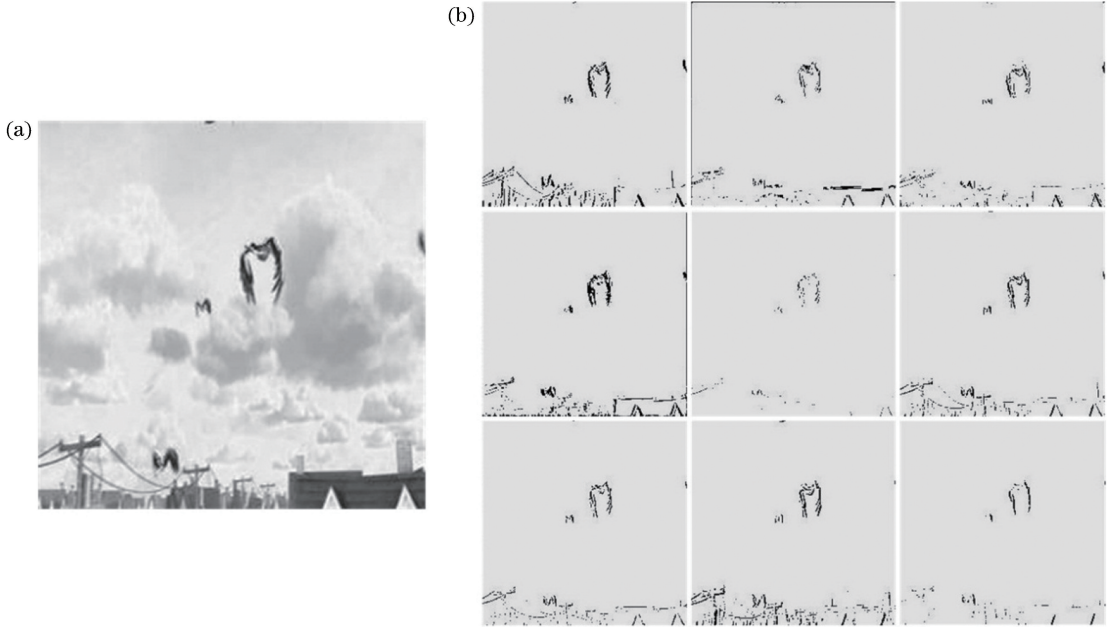


图 2 特征可视化结果(以部分特征为例)。(a)输入图像;(b)第三卷积层特征

Fig. 2 Feature visualization results (taking some features as examples). (a) Input image; (b) features of Conv3 layer

由图 2 可知,Conv3 层提取的特征大部分包含跟踪目标特征的轮廓、纹理等细节信息,但未包含目标的重要背景信息。在 MDNet 算法中,全连接层和特定域层是具备分辨目标和背景的分类器。分类器输入的通用特征中目标信息和背景信息不平衡,使得分类器对目标所处背景的敏感性降低,不能适应环境的变化,分类能力降低。这是导致跟踪算法失效的最大因素。

## 2.2 重构特征

为了增加 Conv3 层中提取特征的背景信息量,在 Conv3 层中提取目标特征的基础上采用上采样操作,重构 Conv3 层中的输入特征,扩大背景信息在特征信息量中的信息熵。文献[13]提出的反卷积神经网络可以实现上采样,实现从特征图到输入图片的反卷积过程。反卷积计算与卷积计算实现过程相同,不同的是 padding 的大小。设图片的输入尺寸为  $h$ ,卷积核尺寸为  $k$ ,步长为  $l$ ,padding 为  $q$ ,输入特征图的尺寸为  $s$ ,卷积的计算公式为

$$s = (h + 2q - k) / l + 1. \quad (1)$$

反卷积是将特征图重构到原输入尺寸,计算公

式为

$$h = l(s - 1) + k - 2q, \quad (2)$$

反卷积实现如图 3 所示。

为了验证从目标特征反卷积上层特征并重建背景信息特征的想法,本文在 MDNet 的 Conv3 层后增加反卷积层重构特征,并进行特征可视化研究。图 4(b)为 Conv3 层提取的特征可视化结果,该结果中只包含目标大量的轮廓、纹理等细节信息,缺少目标所处的背景信息。图 4(c)为通过反卷积操作上采样得到的重构特征,重构了图 4(b)中缺少的背景信息特征。

## 2.3 特征联合

为了增强特征的表达能力,采用特征联合的方法融合原有提取的卷积特征和反卷积得到的重构特征。在 MDNet 网络中,设图像  $x$  在 Conv3 卷积的函数模型为  $H$ ,全连接层的函数模型为  $F$ ,函数输出为  $Y$ ,则网络的计算模型为

$$Y = F[H(x)]. \quad (3)$$

特征联合是分别对提取的不同特征进行加权叠加,将原有通道扩充一倍,把不同的特征同时作为网络下一层的输入,设在 Conv3 后面增加反卷积层,反

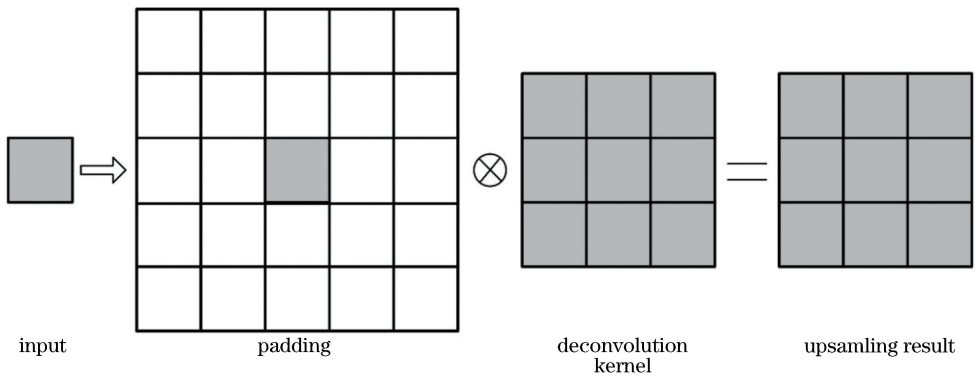


图3 反卷积实施过程

Fig. 3 Deconvolution implementation process

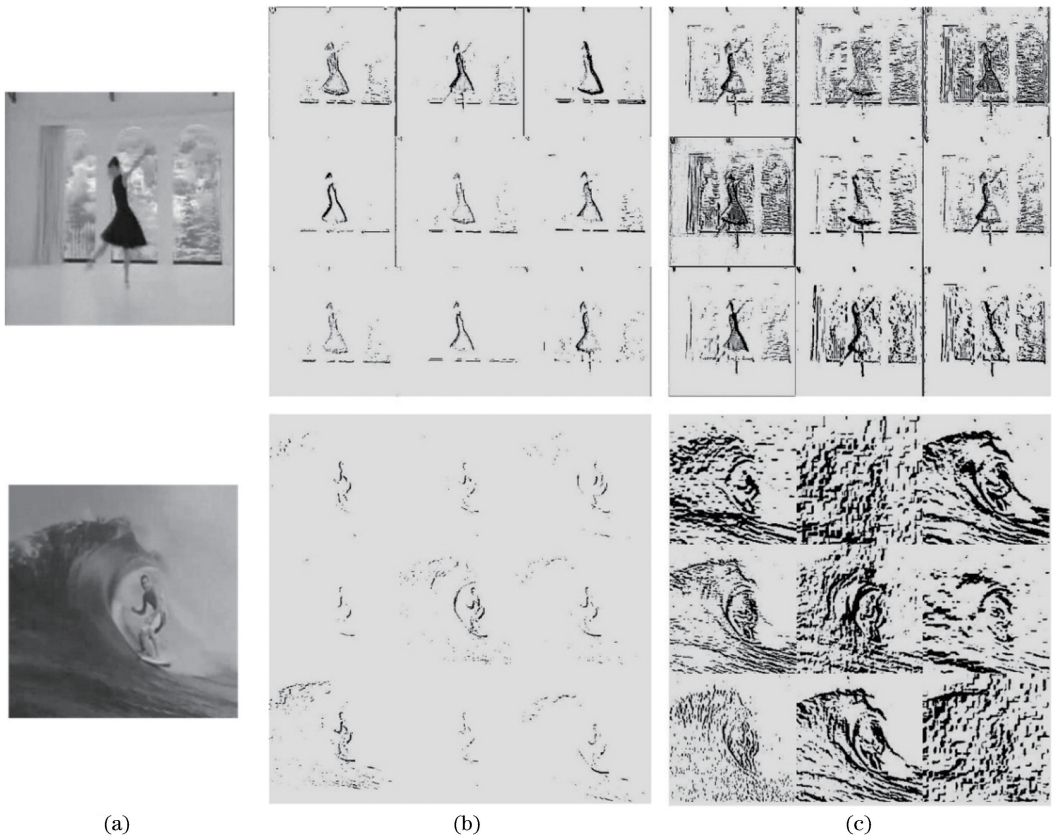


图4 重构特征可视化结果(以部分特征为例)。(a)输入图片;(b)第三卷积层特征;(c)重构特征

Fig. 4 Visualization results of reconstructed features (taking some features as examples). (a) Input image; (b) features of Conv3 layer; (c) reconstructed features

卷积层的函数模型为  $G$ , 加权系数为  $q$ , 原有卷积层的加权系数为  $p$ , 再对提取特征进行特征联合操作, RCNet 反卷积计算模型为

$$Y = F\{p \cdot H(x) + q \cdot G[H(x)]\}, \quad (4)$$

特征联合实现过程如图 5 所示。

#### 2.4 RCNet 结构研究

通过反卷积的上采样操作可以对卷积层末端提取的缺少背景信息的目标高级特征进行特征重构, 使其包含大量的背景信息。本文在 MDNet 跟踪算法的

基础上, 提出了一种新颖的基于特征联合的网络结构, 该方法首先通过增加隐藏层中的卷积层(Conv4、Conv5)作为反卷积操作来达到背景信息重构的目的。为了防止重构特征中背景信息过多溢出, 再通过特征联合的方式联合未改进前的卷积层(Conv3)末端提取的目标特征和重构特征, 使输入全连接层的特征中包含的目标和背景信息平衡, 以此来增强全连接的输入, 达到增强特定域学习特定特征的能力。本文提出的重构特征联合网络(RCNet)结构如图 6 所示。

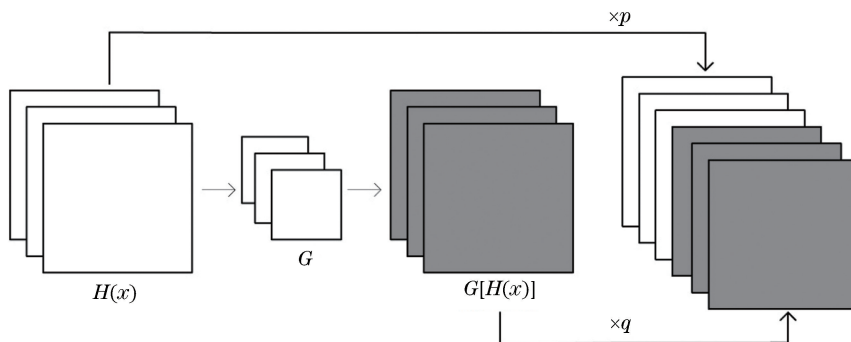


图 5 特征联合

Fig. 5 Feature combination

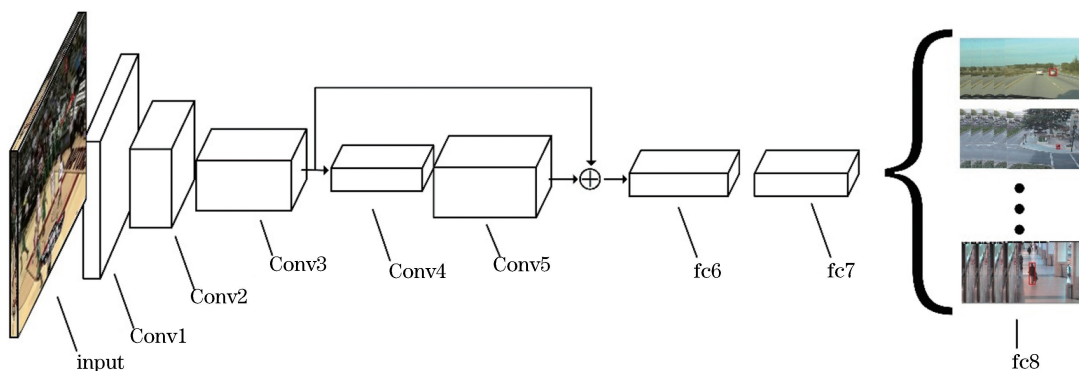


图 6 RCNet 网络结构( $\oplus$ 表示特征联合)

Fig. 6 RCNet network structure ( $\oplus$  indicates feature combination)

为了验证重构特征联合的合理性,取 OTB 跟踪测试集中的图片作为输入,对跟踪网络中卷积层

Conv3、Conv5 以及联合以后的 feature map 进行可视化,得到如图 7 所示结果。

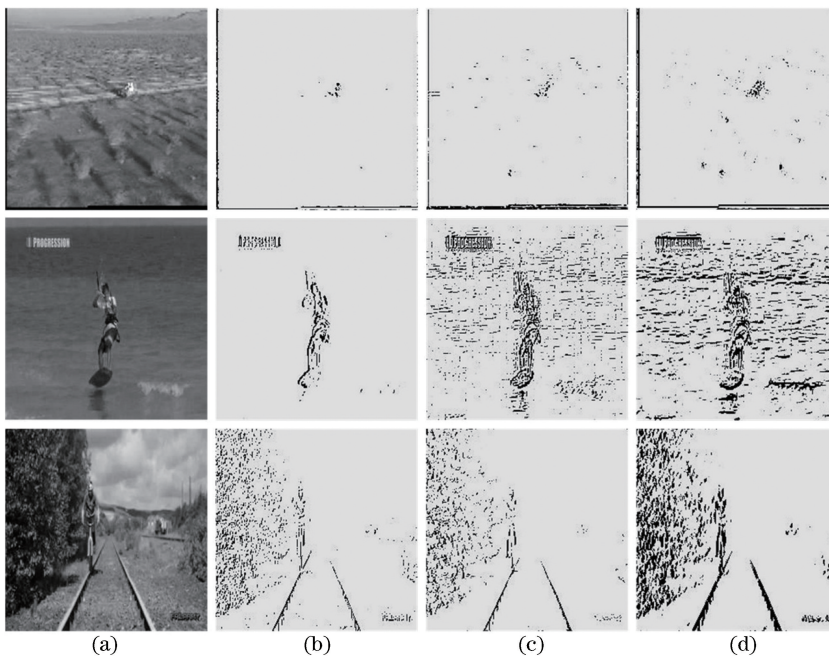


图 7 特征联合分析。(a)输入图片;(b)第三卷积层特征;(c)第五卷积层特征;(d)联合特征

Fig. 7 Feature combination analysis. (a) Input image; (b) features of Conv3 layer; (c) features of Conv5 layer; (d) combination features

通过对重构特征联合的可视化操作,可以观测到 Conv3 层提取到的特征大都包含大量细节信息的目标轮廓特征。Conv5 层提取到的重构特征包含目标所处环境的背景信息,在进行特征联合后,融合特征相当于在轮廓特征的基础上丰富了特征的信息,既保持了目标特征的细节信息,又保留了目标背景信息的特征,增加了特征的有效性。可以看出,卷积特征联合后,特征得到非常有效的增强,验证了本文提出的特征联合增强的设想,保证了实验结果的合理性。

### 3 算法实施细则

RCNet 跟踪算法的整个过程分为离线训练和在线跟踪过程,设网络中各层的卷积核权重为  $w$ ,其中  $w_1 \sim w_5$  在网络离线训练阶段学习, $w_6 \sim w_8$  在每个新的训练序列中随机获取。本文的网络模型采用 MDNet 离线训练数据,采用 VOT2013 和 VOT2014 跟踪数据集作为训练集。

离线训练:

输入:训练集中的图片

1) 加载 VGG 网络模型。

2) 初始化网络模型,将网络模型各层修改为本文的结构,对各层卷积核  $\{w_1(96 \times 7 \times 7), w_2(256 \times 5 \times 5), w_3(512 \times 3 \times 3), w_4(512 \times 3 \times 3), w_5(512 \times 3 \times 3), w_6(1024 \times 3 \times 3), w_7(1024 \times 1024), w_8\}$  随机初始化。

3) 从全部的  $n$  个训练集中随机选取某一帧图片,组成一个 mini-batch(小批次)。

4) 从一个 mini-batch 中,取第  $i$  个训练数据 ( $i=1, 2, 3, \dots, n$ ),将该帧中目标的 groundtruth 通过随机采样的方法在目标周围得到与 groundtruth 尺寸相同的  $m$  个样本框,计算样本框与 groundtruth 的重叠率。重叠率大于设定阈值  $\lambda$  时,判定样本框为正样本,标签为 1,小于等于阈值  $\lambda$  时,判定标签为 0。

5) 把  $m$  个带标签的样本框作为网络的输入 ( $m \times 3 \times 107 \times 107$ ),根据网络的输出对  $w_1 \sim w_7$  进行更新。

6) 一个 mini-batch 训练结束后重复第 3)~5)步。

7) 达到设定的训练次数后,停止训练。

输出:预训练得到的网络结构和权重参数。

在线跟踪过程分为跟踪过程和更新过程。跟踪过程是在离线训练的基础上,基于跟踪的第一帧图像对网络模型的权重参数  $w_6 \sim w_8$  进行学习;更新

过程是在网络输出得分小于阈值或在规定的几帧内对权重参数  $w_6 \sim w_8$  进行更新。在线跟踪过程中,离线训练得到的权重参数  $w_1 \sim w_5$  保持不变。

在线跟踪:

输入:测试集中的图片以及第一帧的 groundtruth。

1) 在目标的第一帧位置随机采样得到与 groundtruth 尺寸相同的  $k$  个样本框,计算重叠率,并将其与设定阈值  $\gamma$  进行比较。重叠率大于设定阈值时判定为正样本,小于时判定为负样本。

2) 随机初始化  $w_8$ ,以带标签的  $k$  个样本作为输入,对权重参数  $w_6 \sim w_8$  进行更新。

3) 在上一帧目标周围采样,并将采样结果作为输入,网络输出各个输入的得分,若最大得分大于得分阈值  $\eta$ ,则判定为成功帧,作为网络更新的数据。

4) 若跟踪结果得分小于得分阈值  $\eta$ ,以成功帧的样本对网络中的  $w_6 \sim w_8$  进行更新。

5) 每隔一定的帧时,同样以成功帧的样本对网络中的  $w_6 \sim w_8$  进行更新。

6) 重复第 3)~5)步,直到跟踪结束。

输出:网络预测的目标跟踪结果。

### 4 仿真分析

本文使用在视觉跟踪领域运用广泛且常被用来比较跟踪算法性能的 OTB50<sup>[9]</sup> 基准测试集,以及在 VOT 竞赛中广泛使用的 VOT2015<sup>[10]</sup> 测试集进行验证。基于该测试集,比较了本文算法与当前各种表现优异的算法,验证了本文算法的优异性能。本文算法的实验平台设置如下:GPU 为 GeForce GTX1060 6G,CPU 为 4 核 3.6 GHz 的 intel core i3-810,内存为 8 GB,配置 MatConvNet 的 MATLAB2014 软件。

#### 4.1 OTB50 测试结果

OTB 测试集把所有跟踪的影响因素分为遮挡、姿态变形等 11 种,并对算法的每种影响因素进行了统计。为了验证本文所提算法的优异性能,本实验采用了一些当前比较优秀的算法作为对照组在实验平台进行了实验。这些优秀算法包括相关滤波类算法 CSK<sup>[4]</sup>、KCF<sup>[5]</sup>、C-COT<sup>[6]</sup>、ECO<sup>[7]</sup> 算法,基于深度学习分类的 MDNet<sup>[8]</sup>、CNN-SVM<sup>[14]</sup>、TCNN<sup>[15]</sup> 算法,以及经典判别类算法 Struck<sup>[2]</sup>、TLD<sup>[16]</sup> 算法。

在 OTB50 的测试集,平均跟踪精度(Precision plots)和平均跟踪成功率(Success plots)分别由中心位置误差和边框重叠率计算得出。如图 8 所示,

基于特征融合的 RCNet 算法在跟踪精度和跟踪成功率上都表现得最为优秀,其中:在跟踪精度方面,RCNet(0.932)较 MDNet(0.918)提升 1.53%;在跟

踪成功率方面,RCNet(0.702)较 MDNet(0.688)提升 2.03%。图 9 为各算法在 OTB50 上的跟踪成功率属性图。

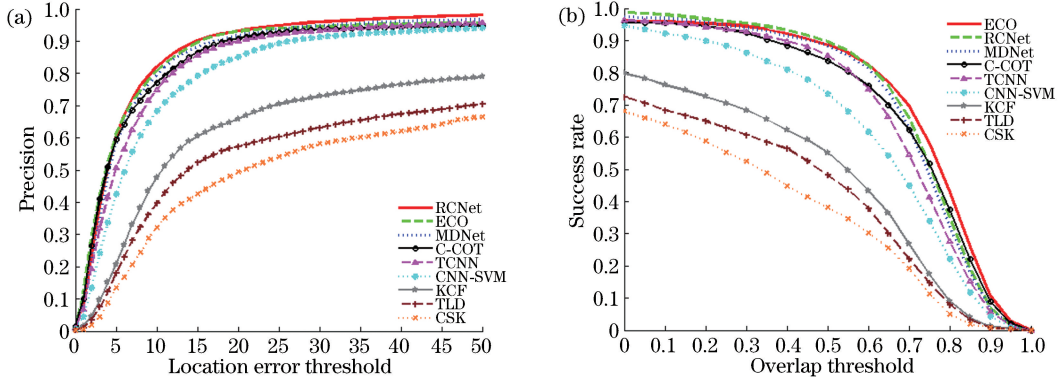


图 8 OTB50 实验结果。(a)跟踪精度得分;(b)跟踪成功得分属性图

Fig. 8 OTB50 experimental results. (a) Tracking precision score; (b) tracking success score

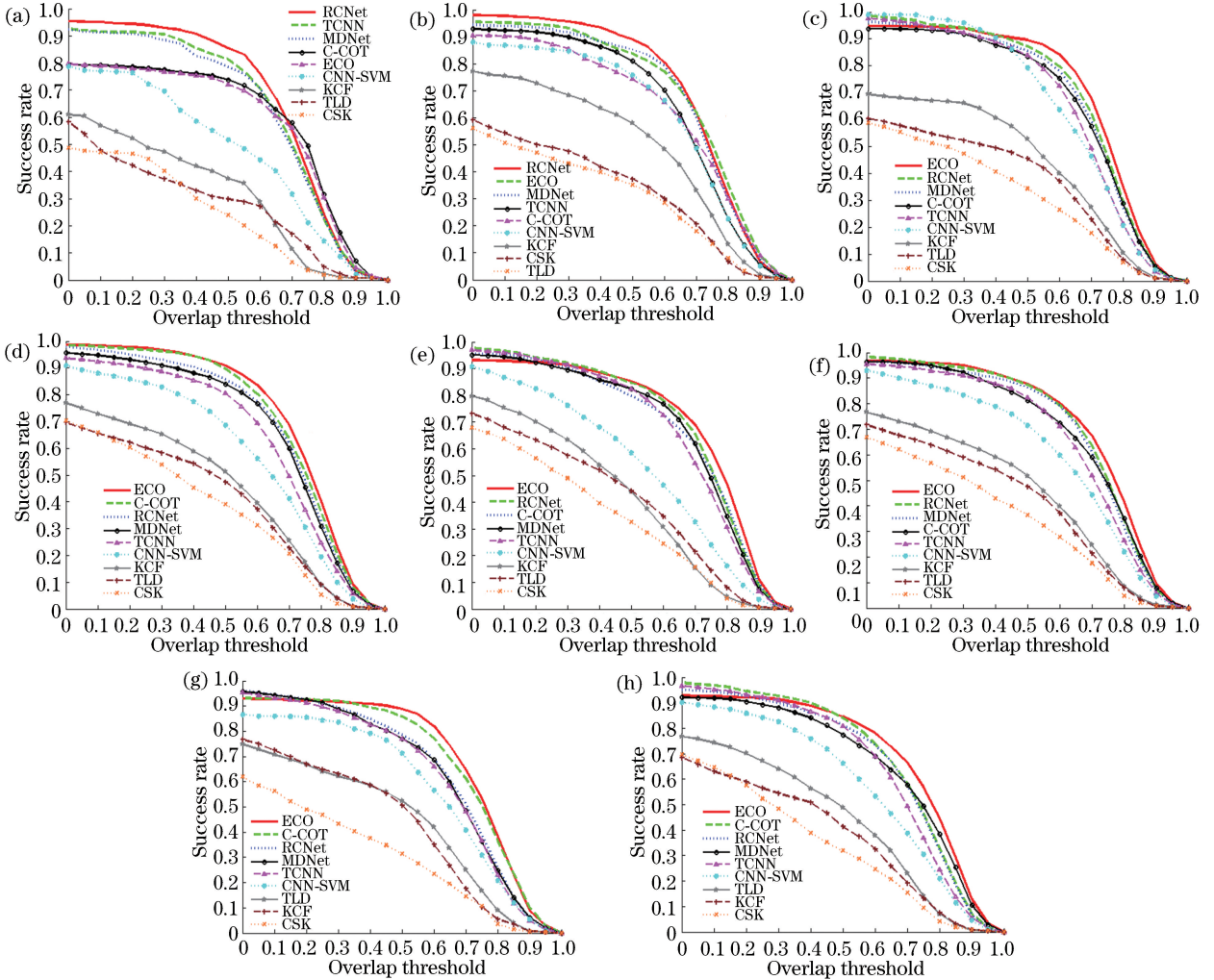


图 9 OTB50 跟踪成功率属性图。(a)低分辨率;(b)背景混杂;(c)形变;(d)遮挡;  
(e)尺度变化;(f)目标旋转;(g)运动模糊;(h)光照变化

Fig. 9 Maps of tracking success rate attributes. (a) Low resolution; (b) background clutter; (c) deformation; (d) occlusion; (e) scale variation; (f) out-of-plane rotation; (g) motion blur; (h) illumination variation

基于深度学习分类预测的跟踪算法的优点是跟踪成功率和跟踪精度较高,缺点是实时性差,表1为本文在GPU为GeForce GTX1060 6G和CUDA-8.0的实验平台上测试所得的各类深度学习类跟踪算法在OTB50上的平均跟踪速度。与MDNet算法相比,本算法牺牲了一定的跟踪速度,极大地提升了算法的跟踪能力。

#### 4.2 VOT2015 测试结果

为了进一步比较改进后的RCNet与MDNet算法的性能差异,用RCNet算法与MDNet算法在

表1 各类深度学习跟踪算法在OTB50上的平均跟踪速度  
Table 1 Average tracking speed of various deep learning tracking algorithms based on OTB50

Algorithm	CNN-SVM	MDNet	TCNN	RCNet
Tracking speed / (frame · s <sup>-1</sup> )	1.45	1.21	2.14	1.06

VOT2015测试集上进行了多帧测试。将MDNet和RCNet算法对同一测试集的跟踪效果进行对比,挑选了一些MDNet算法跟踪失效的测试集,并利用RCNet算法对其进行跟踪,跟踪效果得到改善,如图10所示。

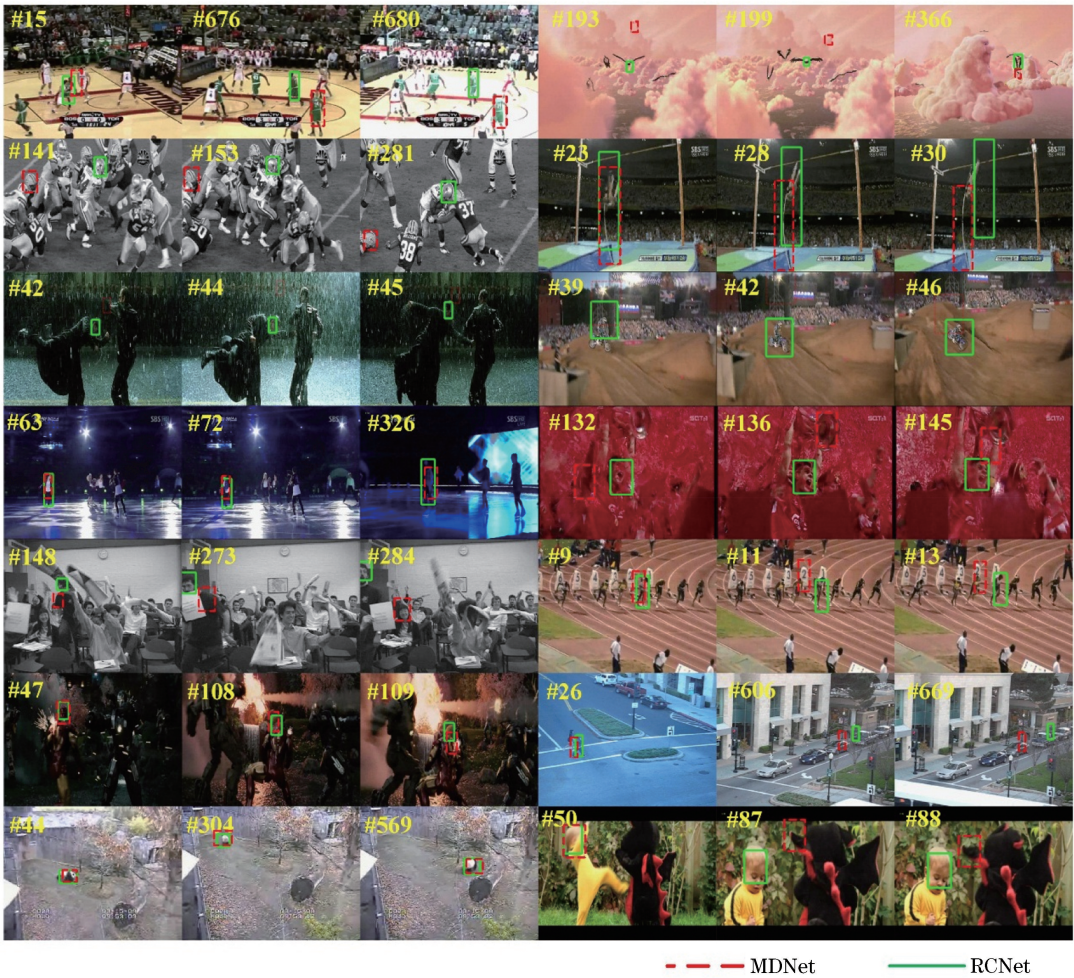


图10 VOT2015测试结果  
Fig. 10 VOT2015 testing results

## 5 结 论

提出一种基于重构特征联合的多域卷积神经网络视觉跟踪算法(RCNet),通过重构目标的背景信息,联合目标的细节特征,增加全连接层输入的有效输入,增强网络特定域的预测分类能力,大幅提升了算法性能。与MDNet算法相比,本文算法可以很好地解决视觉跟踪中背景混杂、目标遮挡、目标快速

运动、低分辨率图片等各种跟踪问题。但本算法的实时性不强,还不能满足实际应用的需要,这将是下一步研究的关键。为了验证了算法的优异性能,将RCNet与众多优秀算法在OTB和VOT标准测试集上进行了比较,表现较为出色。

#### 参 考 文 献

- [1] Vojir T, Noskova J, Matas J. Robust scale-adaptive



- mean-shift for tracking [J]. *Pattern Recognition Letters*, 2014, 49: 250-258.
- [2] 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.
- [3] Ge B Y, Zuo X Z, Hu Y J. Long-term object tracking based on feature fusion [J]. *Acta Optica Sinica*, 2018, 38(11): 1115002.  
葛宝义, 左宪章, 胡永江. 基于特征融合的长时目标跟踪算法[J]. *光学学报*, 2018, 38(11): 1115002.
- [4] Henriques J F, Caseiro R, Martins P, *et al.* Exploiting the circulant structure of tracking-by-detection with kernels[M] // Fitzgibbon A, Lazebnik S, Perona P, *et al.* *Computer vision—ECCV 2012. Lecture notes in computer science*. Berlin, Heidelberg: Springer, 2012, 7575: 702-715.
- [5] Zhou H Y, Yang Y, Wang S Y. Multiple object tracking algorithm based on kernel correlation filter [J]. *Laser & Optoelectronics Progress*, 2018, 55(9): 091502.  
周海英, 杨阳, 王守义. 基于核相关滤波器的多目标跟踪算法[J]. *激光与光电子学进展*, 2018, 55(9): 091502.
- [6] Danelljan M, Robinson A, Khan F S, *et al.* Beyond correlation filters: learning continuous convolution operators for visual tracking[M] // Leibe B, Matas J, Sebe N, *et al.* *Computer vision-ECCV 2016. Lecture notes in computer science*. Cham: Springer, 2016, 9909: 472-488.
- [7] Danelljan M, Bhat G, Khan F S, *et al.* ECO: efficient convolution operators for tracking[C] // 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), July 21-26, 2017, Honolulu, HI. New York: IEEE, 2017: 6638-6646.
- [8] Nam H, Han B. Learning multi-domain convolutional neural networks for visual tracking[C] // 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), June 27-30, 2016, Las Vegas, NV, USA. New York: IEEE, 2016: 4293-4302.
- [9] 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. New York: IEEE, 2013: 2411-2418.
- [10] Kristan M, Matas J, Leonardis A, *et al.* The visual object tracking VOT2015 challenge results[C] // 2015 IEEE International Conference on Computer Vision Workshop (ICCVW), December 7-13, 2015, Santiago, Chile. New York: IEEE, 2015: 564-586.
- [11] Lin S Z, Zheng Y, Lu X F, *et al.* Adaptive tracking algorithm for aerial small targets based on multi-domain convolutional neural networks and autoregression model[J]. *Acta Optica Sinica*, 2017, 37(12): 1215006.  
蔺素珍, 郑瑶, 禄晓飞, 等. 基于多域卷积神经网络与自回归模型的空中小目标自适应跟踪方法[J]. *光学学报*, 2017, 37(12): 1215006.
- [12] Zeiler M D, Fergus R. Visualizing and understanding convolutional networks [M] // Fleet D, Pajdla T, Schiele B, *et al.* *Computer vision-ECCV 2014. Lecture notes in computer science*. Cham: Springer, 2014, 8689: 818-833.
- [13] Shelhamer E, Long J, Darrell T. Fully convolutional networks for semantic segmentation [J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2017, 39(4): 640-651.
- [14] Hong S, You T, Kwak S, *et al.* Online tracking by learning discriminative saliency map with convolutional neural network[C] // Proceedings of the 32nd International Conference on International Conference on Machine Learning, July 6-11, 2015, Lille, France. Massachusetts: JMLR. org, 2015, 37: 597-606.
- [15] Nam H, Baek M, Han B. Modeling and propagating CNNs in a tree structure for visual tracking[J/OL]. (2016-08-25)[2019-02-29]. <https://arxiv.org/abs/1608.07242>.
- [16] Kalal Z, Mikolajczyk K, Matas J. Tracking-learning-detection[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2012, 34(7): 1409-1422.