计算机系统应用 ISSN 1003-3254, CODEN CSAOBN
Computer Systems & Applications,2020,29(1):184−189 [doi: 10.15888/j.cnki.csa.007242]

E-mail: csa@iscas.ac.cn
http://www.c-s-a.org.cn
Tel: +86-10-62661041

基于细节提取的运动目标追踪算法①

李　科[1], 蔡坚勇[1,2,3,4], 张明伟[1], 卢依宏[1], 曾远强[1]

[1](福建师范大学 光电与信息工程学院, 福州 350007)
[2](福建师范大学 医学光电科学与技术教育部重点实验室, 福州 350007)
[3](福建师范大学 福建省光子技术重点实验室, 福州 350007)
[4](福建师范大学 福建省光电传感应用工程技术研究中心, 福州 350007)
通讯作者: 蔡坚勇, E-mail: cjy@fjnu.edu.cn

摘　要: 目前运动目标追踪任务中干扰具有很大的欺骗性, 目标追踪算法容易被带有陷阱的数据集所欺骗. 为了提升追踪算法在追踪数据集上的效果, 本文提出基于 SiamFC 孪生网络上改进的 DPP-SiamFC 追踪算法, 该算法在原网络基础上引入 DPP (Detail-Perserving Pooling) 池化层和残差网络, 有效的保留目标的细节特征. 本文并在 VOT2017 追踪数据集上验证网络性能, 实验结果达到了网络性能提升的效果.

关键词: DPP 池化层; DPP-SiamFC; 残差网络; 多重任务; 细节特征

引用格式: 李科,蔡坚勇,张明伟,卢依宏,曾远强.基于细节提取的运动目标追踪算法.计算机系统应用,2020,29(1):184–189. http://www.c-s-a.org.cn/1003-3254/7242.html

Moving Target Tracking Algorithm Based on Detail Extraction

LI Ke[1], CAI Jian-Yong[1,2,3,4], ZHANG Ming-Wei[1], LU Yi-Hong[1], ZENG Yuan-Qiang[1]

[1](College of Photonic and Electronic Engineering, Fujian Normal University, Fuzhou 350007, China)
[2](Key Laboratory of Optoelectronic Science and Technology for Medicine (Ministry of Education), Fujian Normal University, Fuzhou 350007, China)
[3](Fujian Provincial Key Laboratory of Photonics Technology, Fujian Normal University, Fuzhou 350007, China)
[4](Fujian Provincial Engineering Technology Research Center of Photoelectric Sensing Application, Fujian Normal University, Fuzhou 350007, China)

Abstract: At present, the interference in the moving target tracking task is very deceptive, and the target tracking algorithm is easily deceived by the data set with traps. In order to improve the tracking algorithm's effect on tracking dataset, this study proposes an improved DPP-SiamFC tracking algorithm based on SiamFC twinning network. This algorithm introduces DPP (Detail-Perserving Pooling) pooling layer and residual network based on the original network, effectively retaining the details of the target. This study also verifies network performance on the VOT2017 tracking dataset, the experimental results have achieved the goal of improving network performance.

Key words: DPP pooling layer; DPP-SiamFC; residual network; multitasking; details of the target

近些年, 由于深度学习的火热, 在视频中的运动目标追踪中出现了很多新方法. 就追踪任务而言, 可分为 MOT (Multiple Object Tracking) 和 VOT (Visual Object Tracking) [1–3]. MOT 主要是同时追踪多个目标, 对抗干扰能力要求不高, VOT 则是在干扰条件下持续追踪单个目标. 基于监督学习算法的主流目标追踪方法的可分为, 生成法和判别法两种. 两种方法都是通过数据集训练模型, 达到预测结果的目的. 不同的是生成法先求

① 基金项目: 福建省自然科学基金 (2017J01744)
Foundation item: Natural Science Foundation of Fujian Province (2017J01744)
收稿时间: 2019-06-19; 修改时间: 2019-07-16; 采用时间: 2019-07-19; csa 在线出版时间: 2019-12-27

出联合概率$p(x,y)$, 再通过$p(y|x)=p(x,y)/p(x)$得到条件概率; 判别法则是直接学习条件概率. 两种方法得到的条件概率均可转换为目标框中的像素得分. 然而对于追踪任务而言判别法效果优于生成法[4], 判别法开山之作 SiamFC 的出现, 使得追踪任务取得很大的进展, 但是它仍然无法处理多重干扰数据集. 本文提出的 DPP-SiamFC 神经网络 (Detial-Preserving Pooling Fully-Convolutional Siamese networks) 是对 SiamFC 网络的改进, 可在旋转、快速移动、变形、遮挡和相似性干扰等数据集上取得更好的效果. 本文采用的验证数据集是 VOT2017 (包含的种类有 bag、ball、basketball、birds 等 40 多种类)[5–8].

图 1 中 SiamFC 由对称的两个神经网络架构组成, Z 代表标注的图片, X 为候选图片. 在 X 上计算候选区域和预测区域重叠面积的得分, 从而计算出预测精确度, φ 通常是若干卷积层和池化层 (经典 Alexnet 采用 5 层卷积层), 网络通过 φ 函数得到 128 个通道的特征图, 并将两个特征图通过深度卷积进行融合定位视频中目标位置.

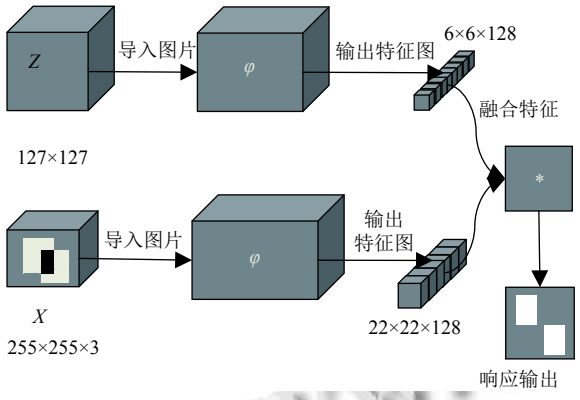

图 1 SiamFC 网络基础架构

1 视频中运动目标追踪

由于 CNN 网络在目标检测领域取得良好的效果, 能有效记忆目标的特征, 所以 CNN 网络也被引入追踪网络算法中[9]. SiamFC 网络在 VOT2016 (Visual Object Tracking) 竞赛中获得良好的比赛成绩, 相较 KCF (Kernelized correlation filter) 有较大的提升, SiamFC 在 ILSVRC2015 数据集上进行训练, 训练两条分支的网络参数权重 ω 和偏置 b[5,7,8,10–16]. 在得到稳定的网络模型后, 可进行在线追踪的任务. 进行追踪任务时, SiamFC 只需要读入初始标定的目标, 即可持续在未标定视频中连续追踪特定目标, 给出预测的目标位置框, 并计算与 GroundTruth 集合的重叠面积, 从而得到预测精确度.

图 1 中两个孪生的 φ 在实际网络中可用 5 层卷积神经网络代替, 其中 Conv1 和 Conv2 卷积层之后有 Pool1 和 Pool2 池化层. 两个池化层目的是减少网络参数的个数, 但同时也会失去目标的一些细节特征. 上述情况在 VOT2017 数据集上表现尤为明显[17,18]. 因而对于追踪方法来说, 一定的细节保留是必要的. DPP 池化层能保留目标物的一些细节特征, 对于追踪方法中的一些细节判别和寻找提供一定的帮助. 因而我们在每层网络都引入 DPP 池化层同时又在 Conv1 和 Conv3 层之后添加到融合层的残差网络. 本文的残差网络解决网络深度增加引起的梯度消失问题, DPP 池化层主要解决特征提取时的细节丢失问题[9].

1.1 残差网络

DPP-SiamFC 网络不仅在 SiamFC 网络上每层引入 DPP 池化层, 还引入 Conv1 和 Conv3 的池化层之后到融合层的残差网络. 残差网络能很大程度将输入的特征引入输出, 而并不带来很多网络开销. 在网络达到一定深度以后能很好帮助前馈网络, 同时降低错误率. SiamFC 的 Conv1–Conv5 层是类似于 AlexNet 的神经网络. 定义$f(x)$为输入值, $g(f(x))$为输入经过 CNN 网络卷积池化的函数, 则加入残差网络进行融合的表达式如式 (1) 所示:

$$h(x)=\alpha f(x)+\beta g(f(x)) \tag{1}$$

式 (1) 所示的残差网络将一部分输入特征直接引入网络输出, 使得网络的梯度下降的更快, α和β为调节参数.

1.2 DPP (Detail-Preserving Pooling) 细节保留池化

DPP 细节保留池化是应用于目标检测的 CNN 网络 Conv 卷积层之后的池化层, 目的是改善原来 CNN 检测网络的池化层对目标细节特征的丢失. 目标检测比较常用的 Avg-Pooling 和 Max-Pooling 分别利用池化区域的平均值和最大值来代替原来的像素点, 而在目标追踪领域常用的是 Max-Pooling. 随着网络层数以及数据集难度的增加, Max-Pooling 和 Avg-Pooling 丢失目标特征的弊端将逐渐展现出来. DPP 池化的结构如图 2, 主要完成线性减少特征图 I 的数据量. 处理流

程是将原始特征图 I 进行线性缩减尺度, 将得到的结果与原始特征进行比较 (方法是引入逆双边权重), 判断出特征丢失程度. 输入特征图 I 经过激励函数得到的输出 O 特征公式 (2):

$$D_{\alpha,\lambda}(O)[p]=\frac{1}{\sum_{q'\in\Omega_p}\omega_{\alpha,\lambda}[p,q']}\sum_{q\in\Omega_p}\omega_{\alpha,\lambda}[p,q]I[q] \quad (2)$$

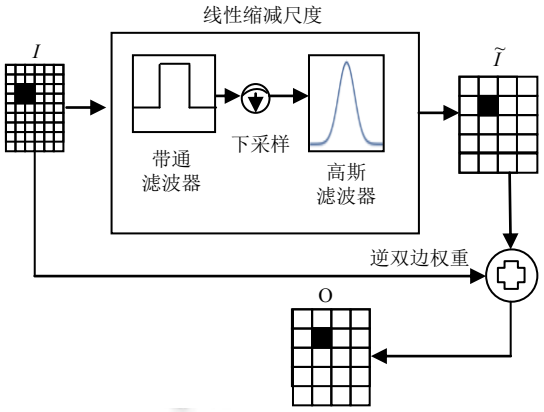

图 2 DPP 池化层逻辑结构示意图

式 (2) 计算的是输入相邻点$I[q]_{q\in\Omega_p}$的空间加权平均值, 作为池化的输出结果. 其中 $I[q]$为输入 DPP 池化层的图片特征图, $O[p]$为输出池化层的图片特征图, α, λ 为神经网络回报参数, 是根据不同数据集训练得到的, 该逆双边权重公式 (为了解决下采样之后特征损失) 如式 (3):

$$\omega_{\alpha,\lambda}[p,q]=\alpha+\rho_\lambda(I[q]-\tilde{I}[p]) \quad (3)$$

在网络反馈学习中, 通过优化$\log(\alpha)$和$\log(\lambda)$确保参数非负, 对于 α 参数是为了确保输入的特征不被网络训练完全清除, 保存细节特征, 并最后作用于输出结果. λ 为调节奖励函数形状的参数. 对于$I[q]>\tilde{I}[p]$时采用非对称的$\rho_{\text{Asym}}(x)=\left(\sqrt{(\max(0,x)^2+\varepsilon^2}\right)^\lambda$作为奖励函数. 反之采用对称的$\rho_{\text{Sym}}(x)=\left(\sqrt{x^2+\varepsilon^2}\right)^\lambda$作为网络的奖励函数 ($\varepsilon$是修正因数, 减少 x 的浮动带来的影响, 使函数图像从 0 开始).

2 DPP-SiamFC 网络架构

2.1 网络架构

本文为了实现视频中目标相似性干扰、旋转、快速移动、遮挡和变形等问题处理能力. 对 SiamFC 网络进行改进, 改进之后的网络结构如图 3, 融合网络 (Concatenation) 是 3 条分支的加权平均值, 再通过深度卷积层对特征进行融合.

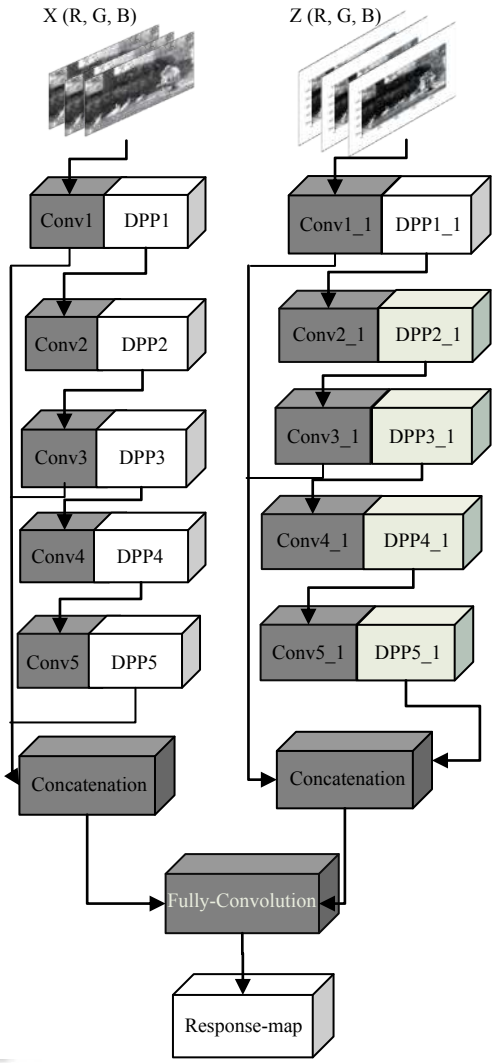

图 3 DPP-SiamFC 网络架构

Conv 和 Conv_1 是对称的卷积层, 它们卷积核大小, 通道数和步长并不相同, 相同的是两个卷积层使用的卷积核的个数. 这使得输出特征图的个数一致. DPP 池化层的结构如图 2 所示, 目的是更好的保留目标细节特征.

Fully-Convolution 是将两个分支的结果进行卷积处理, 生产下一帧的目标位置, 从而得到最终的特征输出.

2.2 DPP-SiamFC 网络各层参数

DPP-SiamFC 网络各层参数并不相同, 其中 DPP 层提供Ω_P, $\tilde{\Omega}_P$, ε三个参数. 其中Ω_P通常取 3×3 相同如$\tilde{\Omega}_P$, ε=0.1, 则网络的各层参数如表 1 所示.

表 1 DPP-SiamFC 网络各层参数

卷积层	宽高	通道	个数	池化层	Ω_P	$\tilde{\Omega}_P$	ε
原帧	640×480	3	/	DPP1	3×3	3×3	0.1
Conv1	3×123	123	96	DPP2	3×3	3×3	0.1
Conv2	3×57	57	256	DPP3	3×3	3×3	0.1
Conv3	3×53	53	384	DPP4	3×3	3×3	0.1
Conv4	3×51	51	384	DPP5	3×3	3×3	0.1
Conv5	3×49	49	32	DPP1_1	3×3	3×3	0.1
Conv1_1	1×59	59	96	DPP2_1	3×3	3×3	0.1
Conv2_1	1×25	25	256	DPP3_1	3×3	3×3	0.1
Conv3_1	1×21	21	384	DPP4_1	3×3	3×3	0.1
Conv4_1	3×51	51	384	DPP5_1	3×3	3×3	0.1
Conv5_1	1×17	17	32				

3 实验分析

我们将 DPP-SiamFC 网络于 ILSVRC2015 数据集上进行训练, 实现对每个分类特征的离线训练. 在线追踪于 VOT2017 追踪数据集, 观察在各个分类追踪的效果[19–21].

3.1 运动目标追踪

实验展示 DPP-SiamFC 在 VOT2017 各个分类效果, 尤其在含有复杂背景, 有众多干扰物、遮挡、快速移动、和目标变形的数据集.

3.2 DPP-SiamFC 与经典网络实验效果对比

图 4 展示了 DPP-SiamFC 在有很多干扰物且存在部分遮挡条件下追踪单个目标物的效果, 整个视频的标定区域和预测区域重叠面积比平均约为 79.1%, 高于 80% 预测精度的视频帧约占总数的 83%.

图 5 是 DPP-SiamFC 在目标快速移动任务中效果. 该数据集是摩托车比赛, 途中有树木的遮挡.

(a) 第一帧 (b) 第二帧 (c) 第三帧 (d) 第四帧

图 4 groundtruth(蓝色)、DPP-SiamFC(红色)、KCF(相关滤波算法黄色) 和 SiamFC(绿色) 在相似物干扰数据集的效果

(a) 起始帧 (b) 通过障碍物之前 (c) 通过障碍物之后

图 5 groundtruth(蓝色)、DPP-SiamFC(红色)、KCF(相关滤波算法黄色) 和 SiamFC(绿色) 在快速移动 (包含部分遮挡数据集) 的效果

图 6 是目标形变, 和背景复杂的夜间街道数据集中 DPP-SiamFC 追踪效果. 追踪效果较为良好, 能实现对目标持续追踪的目的.

如图 7 所示 SiamFC 很难追踪快速上升并旋转的特技摩托. 而 DPP-SiamFC 能很好的将目标捕捉, 达到旋转物体追踪的效果.

通过图 8 中 SiamFC、DPP-SiamFC 和 KCF 算法预测区域和 groundtruth 标定的重叠面积比 (IOU) 在 60 个追踪数据集上的平均精确度 (例如: 图 8 中 KCF′表示 KCF 算法在 60 个数据集上的精度平均值) 78%, 87%, 70%(如表 2) 可以看出, 改进之后的 DPP-SiamFC 神经网络在大多数数据集上效果优于 SiamFC

和 KCF 网络, 本文在 SiamFC 网络中引入 DPP 池化层和残差网络能很好保留数据集上的细节特征, 提升在追踪任务中的准确度, 但在综合的任务数据集中稳定性还需提高.

(a) 体操运动第一帧 (b) 体操运动第二帧 (c) 体操运动第三帧 (d) 体操运动第四帧

(e) 夜间车辆第一帧 (f) 夜间车辆第二帧 (g) 夜间车辆第三帧

图 6 groundtruth(蓝色)、DPP-SiamFC(红色)、KCF(相关滤波算法黄色) 和 SiamFC(绿色) 在目标变形和背景复杂条件下的追踪图像

(a) 飞车表演第一帧 (b) 飞车表演第二帧 (c) 飞车表演第三帧

图 7 groundtruth(蓝色)、DPP-SiamFC(红色)、KCF(相关滤波算法黄色) 和 SiamFC(绿色) 在摩托车特技比赛中的对比

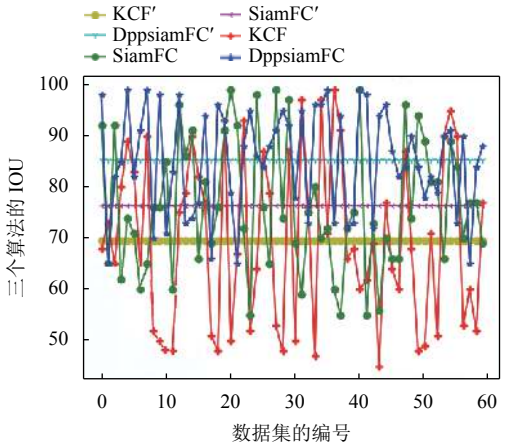

图 8 SiamFC、DPP-SiamFC 和 KCF 的 IOU 比较

4 结论与展望

实验结果证明, 通过在 SiamFC 孪生网络上引入 DPP 池化层和残差网络, 有利于网络细节特征的保留, 在 VOT2017 追踪数据集中 DPP-SiamFC 有更高精确度, 同时在背景复杂、物体变形、快速移动、遮挡等数据集中目标追踪有一定改善. 但是在多重任务追踪集的效果还有待提高. 今后我们的工作将致力于网络与数据集之间的对抗性研究.

表 2 SiamFC、DPP-SiamFC 和 KCF 精度比较 (单位: %)

算法	Siamfc	DPP-SiamFC	KCF
60 个数据集平均精度	78	87	70

参考文献

1 Hua Y, Alahari K, Schmid C. Online object tracking with proposal selection. Proceedings of 2015 IEEE International Conference on Computer Vision. Santiago, Chile. 2015. 3092–3100.

2 Zhu G, Porikli F, Li HD. Beyond local search: Tracking objects everywhere with instance-specific proposals. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas, NV, USA. 2016. 943–951.

3 Yang TY, Chan AB. Recurrent filter learning for visual tracking. Proceedings of 2017 IEEE International Conference on Computer Vision Workshops. Venice, Italy. 2017. 2010–2019.
4 Kosiorek AR, Bewley A, Posner I. Hierarchical attentive recurrent tracking. arXiv preprint arXiv: 1706.09262, 2017.
5 He AF, Luo C, Tian XM, *et al*. A twofold Siamese network for real-time object tracking. Proceedings of 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City, UT, USA. 2018. 4834–4843.
6 Hao ZX, Li Y, You SD, *et al*. Detail preserving depth estimation from a single image using attention guided networks. Proceedings of 2018 International Conference on 3D Vision. Verona, Italy. 2018. 304–313.
7 Zhu G, Porikli F, Li HD. Tracking randomly moving objects on edge box proposals. arXiv: 1507.08085, 2015.
8 Kristan M, Matas J, Leonardis A, *et al*. The visual object tracking vot2015 challenge results. Proceedings of the IEEE International Conference on Computer Vision Workshops. Santiago, Chile. 2015. 564–586.
9 Weber N, Waechter M, Amend SC, *et al*. Rapid, detail-preserving image downscaling. ACM Transactions on Graphics, 2016, 35(6): 205.
10 Choi J, Chang HJ, Jeong J, *et al*. Visual tracking using attention-modulated disintegration and integration. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas, NV, USA. 2016. 4321–4330.
11 Bertinetto L, Valmadre J, Henriques JF, *et al*. Fully-convolutional Siamese networks for object tracking. Proceedings of European Conference on Computer Vision. Amsterdam, The Netherlands. 2016. 850–865.
12 Wang MM, Liu Y, Huang ZY. Large margin object tracking with circulant feature maps. Proceedings of 2017 IEEE Conference on Computer Vision and Pattern Recognition. Honolulu, HI, USA. 2017. 4800–4808.
13 Huang C, Lucey S, Ramanan D. Learning policies for adaptive tracking with deep feature cascades. Proceedings of 2017 IEEE International Conference on Computer Vision. Venice, Italy. 2017. 105–114.
14 Hu J, Shen L, Sun G. Squeeze-and-excitation networks. Proceedings of 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City, UT, USA. 2018. 7132–7141.
15 Wang LJ, Ouyang WL, Wang XG, *et al*. Visual tracking with fully convolutional networks. Proceedings of 2015 IEEE International Conference on Computer Vision. Santiago, Chile. 2015. 3119–3127.
16 Wu Y, Lim J, Yang MH. Object tracking benchmark. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2015, 37(9): 1834–1848. [doi: 10.1109/TPAMI.2014.2388226]
17 Tao R, Gavves E, Smeulders AWM. Siamese instance search for tracking. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas, NV, USA. 2016. 1420–1429.
18 Cen MB, Jung C. Fully convolutional Siamese fusion networks for object tracking. Proceedings of 2018 25th IEEE International Conference on Image Processing. Athens, Greece. 2018. 3718–3722.
19 Liang PP, Blasch E, Ling HB. Encoding color information for visual tracking: Algorithms and benchmark. IEEE Transactions on Image Processing, 2015, 24(12): 5630–5644. [doi: 10.1109/TIP.2015.2482905]
20 Danelljan M, Häger G, Khan FS, *et al*. Convolutional features for correlation filter based visual tracking. Proceedings of the IEEE International Conference on Computer Vision Workshops. Santiago, Chile. 2015. 621–629.
21 Abadi M, Agarwal A, Barham P, *et al*. Tensorflow: Large-scale machine learning on heterogeneous distributed systems. arXiv preprint arXiv: 1603.04467, 2016.