文章编号：1000-5013(2016)02-0207-06 **doi**：10.11830/ISSN.1000-5013.2016.02.0207

选择性搜索和多深度学习模型融合的目标跟踪

钟必能[1,2]，潘胜男[1,2]

（1. 华侨大学 计算机科学与技术学院，福建 厦门 361021；
2. 华侨大学 计算机视觉与模式识别重点实验室，福建 厦门 361021）

摘要： 提出一种基于深度学习的多模型（卷积神经网络和卷积深信度网络）融合目标跟踪算法. 该算法在提取候选粒子方面，使用选择性搜索和粒子滤波的方法. CVPR2013 跟踪评价指标（50 个视频序列、30 个跟踪算法）验证了：该算法在跟踪中能有效地缓解目标物体由于遮挡、光照变化和尺度变化等因素造成的跟踪丢失情况的发生.

关键词： 目标跟踪；深度学习；多模型融合；选择性搜索；评价指标
中图分类号： TP 301 **文献标志码**： A

目标跟踪是机器视觉中一个重要的研究分支，然而由于应用场合中的一些不确定因素，要想获得一种稳健、鲁棒又快速的跟踪方法仍具有挑战性. 为了解决这个问题，近年来越来越多的学者采用多层的深度学习模型进行目标的特征提取. Fan 等[1]针对跟踪问题，提出了基于卷积神经网络（CNN）的行人跟踪法. Carneiro 等[2]使用深度神经网络训练目标的表观模型，将该模型用于超声图像中左心房内膜的轮廓跟踪. Wang 等[3]提出了基于降噪自编码器（auto-encoder）的跟踪方法. 虽然深度学习模型具有更强的物体特征表达能力，但是以上提到的跟踪算法都是基于单线索[1-3]的. 单一线索用于跟踪方法对环境变化敏感，鲁棒性不高. 为了提高跟踪算法的性能，本文提出了基于深度学习的多线索（CNN 和 CDBN）目标跟踪算法；在获取候选粒子方面，将选择性搜索（seletive search）方法[4]和粒子滤波[5]相结合用到了跟踪问题中.

1 目标跟踪算法

1.1 算法框架

提出一种基于选择性搜索和多深度模型融合的目标跟踪算法. 文中所用的多线索模型（CNN 模型和 CDBN 模型）的融合办法和具体跟踪算法细节，如图 1 所示.

1.2 目标表观的建模

1.2.1 CNN 建模 卷积神经网络（convolutional neural networks，CNN）是深度学习的一个重要模型. 它是一个多层的神经网络，每层由多个二维平面组成，而每个平面由多个独立神经元组成[6]. 卷积神经网络中的每一个特征提取层（C-层）都紧跟着一个求局部平均与二次提取的下采样层（S-层），这种特有的两次特征提取结构，使网络在识别时对输入样本有较高的畸变容忍能力.

针对输入图片大小设计的 CNN 模型结构，如图 2 所示. 图 2 中：输入图片的大小为 32 px×32 px；Cov$i$ 对应的是第 i 个卷积层；Pool$i$ 对应的是第 i 个下采样层；Kernel_size 是卷积模板的大小；Stride 是每一次卷积滑动的步伐；Relu 是非线性变换函数；Norm 是归一化.

收稿日期： 2015-06-16
通信作者： 钟必能（1981-），男，副教授，博士，主要从事计算机视觉、模式识别、目标跟踪方面的研究. E-mail：bnzhong@hqu.edu.cn.
基金项目： 国家自然科学基金资助项目（61202299）；国家自然科学基金面上资助项目（61572205）；福建省自然科学基金资助项目（2015J01257）；福建省高校杰出青年科研人才培育计划项目（JA13007）

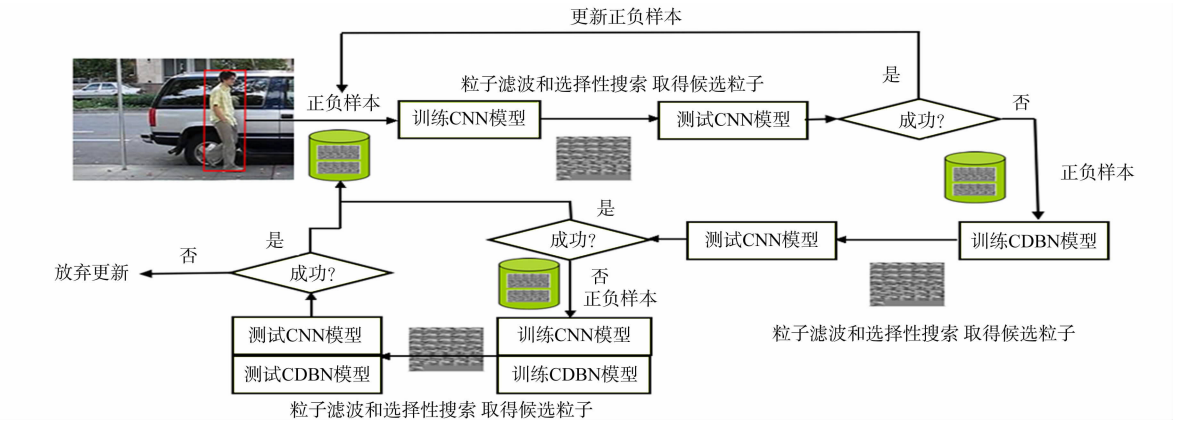


图 1 选择性搜索和多深度模型融合的目标跟踪算法框架

Fig. 1 Framework of the multi-clue fusion target tracking algorithm based on selective search and deep learning

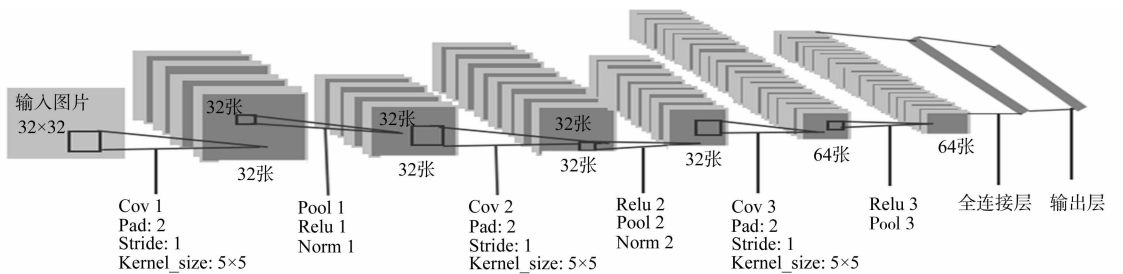


图 2 CNN 模型结构图

Fig. 2 Structure of CNN

1.2.2 CDBN 建模 卷积深信度网络[7-9](convolutional deep belief network,CDBN)是由多层卷积受限波尔兹曼机 CRBM 组成,每一个受限波尔兹曼机 CRBM 的基本机构是由卷积层和采样层构成.

根据输入图片的大小和需要,设计的 CDBN 结构如图 3 所示. 图 3 中:Cov*i* 为卷积层;Kernal_size 为卷积模板的大小;输入图片大小固定为 32 px×32 px.

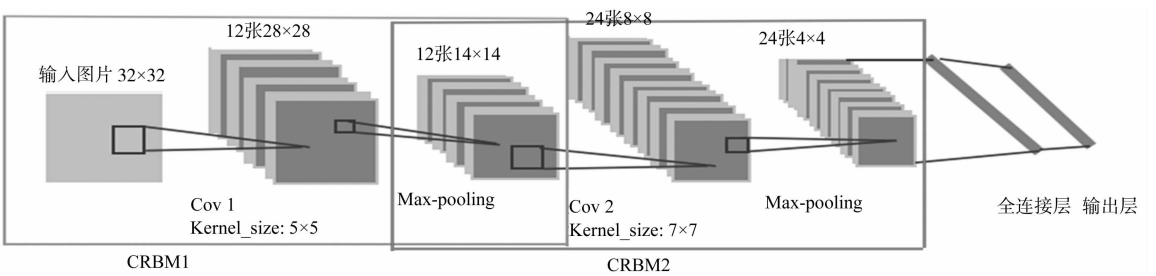


图 3 CDBN 结构图

Fig. 3 Structure of CDBN

1.3 目标运动的搜索

粒子滤波[5]是寻找一组在状态空间中传播的随机样本对概率密度函数进行近似,利用样本均值代替积分运算,进而获得状态最小方差分布的过程. 选择性搜索[4]的前期工作是利用图像分割的方法得到一些原始区域,然后使用一些合并策略将这些区域合并,得到一个层次化的区域结构,而这些结构就包含着可能需要的物体. 选择性搜索意在找出可能的目标位置进行物体的识别和分类. 与传统的单一策略相比,选择性搜索提供了多种策略;与全搜索相比,又大幅度降低了搜索空间.

1.4 多模型离线训练阶段

首先,从 Tiny Images[10]数据集中挑选出 65 类 202 932 张图片,对 CNN 模型进行离线训练;然后,用 cifar-100 数据集对 CDBN 模型进行离线训练. 通过离线训练,就可以得到物体的广义性特征.

1.5 目标跟踪算法细节

选择性搜索和深度学习的目标跟踪算法的详细设计过程(图 1)为

初始化：

1. 离线训练：train_CNN(202 932 张图片)，train_CDBN(60 000 张图片)；
2. 在第一张给出的待跟踪物体的位置处，得出 s^+ 张正例样本和 s^- 张负例样本；
3. train_CNN(正负样本)；

for 1：视频帧的最后一帧

1. 用 selective search 与粒子滤波相结合，在 t 时刻初始化粒子；
2. 测试粒子：conf1＝test_CNN(所有粒子)；
3. 寻找最可信粒子：max1＝max(conf1)；
4. if max1＜Threshold 某个阈值

①启动 CDBN 模型：记作 train_CDBN(正负样本)；

②测试所有粒子：记作 conf2＝test_CNBN(所有粒子)；

③找出最佳位置：记作 max2＝max(conf2)；

If max2＜Threshold 某个阈值

①train_CNN 正负样本)，train_CDBN(正负样本)；

②conf1＝test_CNN(所有粒子)max1＝max(conf1)；

③conf2＝test_CDBN 所有粒子)max2＝max(conf2)；

④max＝max(max1，max2)；

if max＞某个阈值

①得到跟踪目标：X＝max；

②更新正负样本：找出 conf1 与 conf2 排序后的前 10 个较好的图片也作为 s^+，找出 500 张负例 s^-；

else

放弃更新正负样本；

else

①得到跟踪目标：X＝max2；

②更新正负样本：找出 conf1 与 conf2 排序后的前 10 个较好的图片也作为 s^+，找出 500 张负例 s^-；

else

①得到跟踪目标：X＝max1；

②更新正负样本：找出 conf1 中前 10 个较好的图片也作为 s^+，找出 500 张负例 s^-；

end if

5. train_CNN(正负样本)；
6. 进入到下一张图片，以便对这一张新的图片进行寻找到要跟踪的目标；

文中阈值取值为 0.03(对单 CNN 模型进行跟踪实验，通过对实验结果的统计和分布而选取的)；正负样本数取值为 546(正样本 45(在跟踪中得到的正样本)＋1(初始化的样本)，负样本为 500 个).

2　实验部分

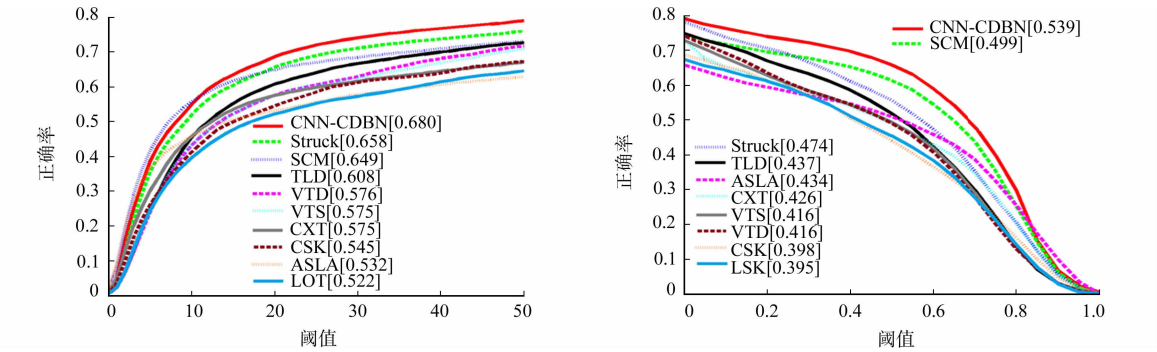
2.1　实验设置

算法采用 Matlab 语言编写，在 Intel(R) Xeon(R) E5620 2.40 GHz 处理机和 12 G 内存的机器上运行. 粒子数设置为 3 000，每一个目标物体的大小设置为 32 px×32 px(用于设定正负样本的大小，通过对视频帧进行一定的缩放和旋转得到)，滑动窗口的大小设置为 45(FIFO 先进先出容器，用于存储正例样本). 在没有 GPU 加速的情况下，对像素大小为 320 px×240 px 的图片，平均处理速度为每秒 5 帧. 1 个视频帧各部分所用时间分别为：selective search 0.051 428 s；particle filtering 0.035 386 s；CNN 0.101 24 s；CDBN 0.094 31 s.

实验对比在CVPR2013跟踪评价指标[11]中进行.在这个评价基准中,有30个不同的跟踪算法.2种测试方法分别为Precision Plot和Success Plot.

2.2 与其他跟踪算法的比较

将文中算法与precision plot和success plot两种评价方法的综合性能进行对比.在50个视频序列和30个跟踪算法中进行综合性能对比实验,结果如图4所示.图4中:横坐标代表不同评价方法对应的阈值;纵坐标代表正确率.在这个对比实验中,挑出了性能较好的排在前10个跟踪算法.由图4可知:文中算法综合性能优于其他跟踪算法.

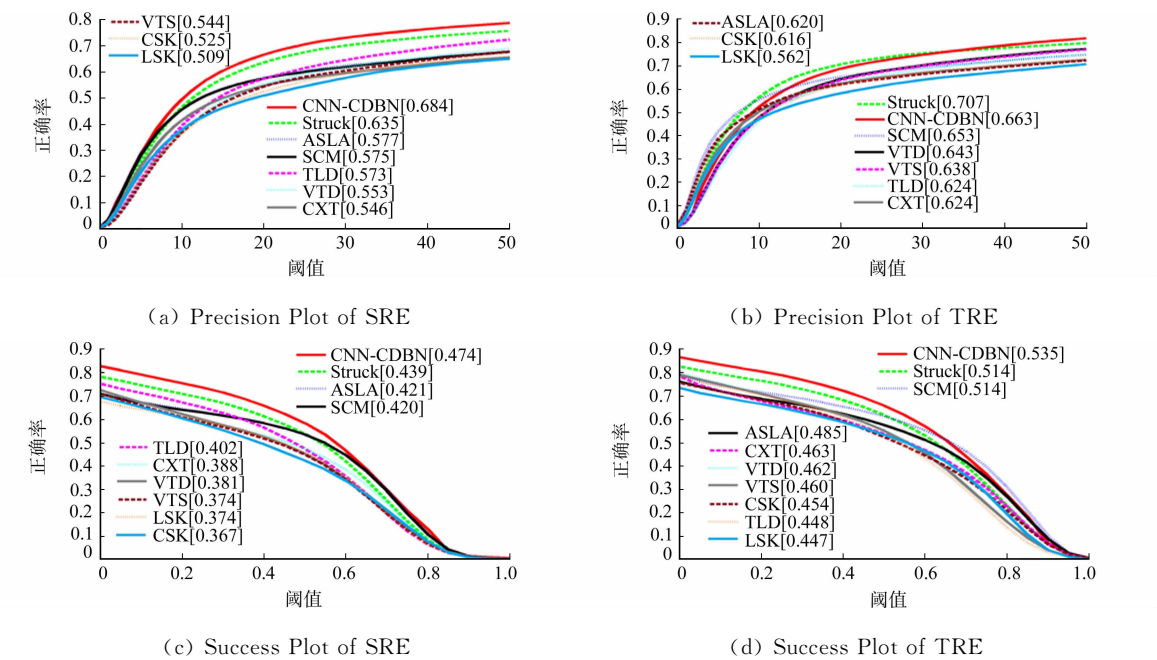


(a) Precision Plot法 (b) Success Plot法

图4 综合性能的对比

Fig. 4 Comparison of comprehensive performance

为了分析初始化对跟踪算法性能的影响,对目标物体初始化进行一定的时间(TRE)和空间(SRE)扰动,具体细节见CVPR2013评价指标[11].在这种情况下,进行对比的实验,如图5所示.由图5可知:基于深度学习的多模型建模方法,能够很好地表达物体的表观特征,适应物体的初始化表观变化.



(a) Precision Plot of SRE (b) Precision Plot of TRE

(c) Success Plot of SRE (d) Success Plot of TRE

图5 初始化对性能的影响

Fig. 5 Effect of initialization on the performance

跟踪的目标物体在不同场景中,其运动属性是不同的.不同的属性也是检验跟踪算法好坏的关键因素.在CVPR2013评价指标中,对11个不同属性进行评价,结果如表1所示.由表1可知:文中的跟踪算法(ours)在物体快速运行、运动路径模糊等场景下,都具有一定的鲁棒性.

从50个数据序列中随机挑选出的6个序列进行显示,结果如图6所示.由于篇幅原因,只显示一部分.每一个数据序列从第一针开始,每间隔40帧选出1张跟踪图片.由图6可知:基于选择性搜索和深度学习的多模型跟踪算法在这些序列中都有着良好表现.

表 1　不同属性性能对比

Tab. 1　Comparing the performance of different attributes

物体属性	第一	第二	第三	第四	第五
快速运动(17)	ours (0.511)	Struck (0.451)	TLD (0.385)	CXT(0.348)	OAB (0.322)
背景复杂 (21)	ASLA (0.410)	ours (0.410)	Struck(0.408)	SCM(0.387)	VTD(0.377)
运动模糊(12)	ours (0.508)	Struck (0.452)	TLD(0.392)	CXT(0.354)	DFT(0.325)
形变(19)	ours (0.430)	Struck (0.398)	ASLA(0.386)	DFT(0.364)	CPF(0.362)
光照变化(25)	ours (0.429)	ASLA(0.405)	Struck(0.396)	SCM(0.389)	VTS(0.378)
平面旋转 (31)	ours (0.429)	CXT(0.410)	Struck(0.410)	ASLA(0.405)	SCM(0.399)
低分辨率 (4)	Struck (0.360)	ours (0.345)	MTT(0.326)	OAB(0.31)	TLD(0.305)
遮挡(29)	Struck (0.405)	ours (0.402)	SCM(0.398)	TLD(0.384)	LSK(0.384)
出平面旋转(39)	ours (0.438)	Struck (0.409)	ASLA(0.404)	SCM(0.396)	VTD(0.392)
离开视线(6)	Struck(0.421)	ours (0.417)	LOT(0.411)	TLD(0.407)	CPF(0.394)
尺度变化(28)	ours (0.468)	ASLA(0.440)	SCM(0.438)	Struck(0.395)	TLD(0.384)

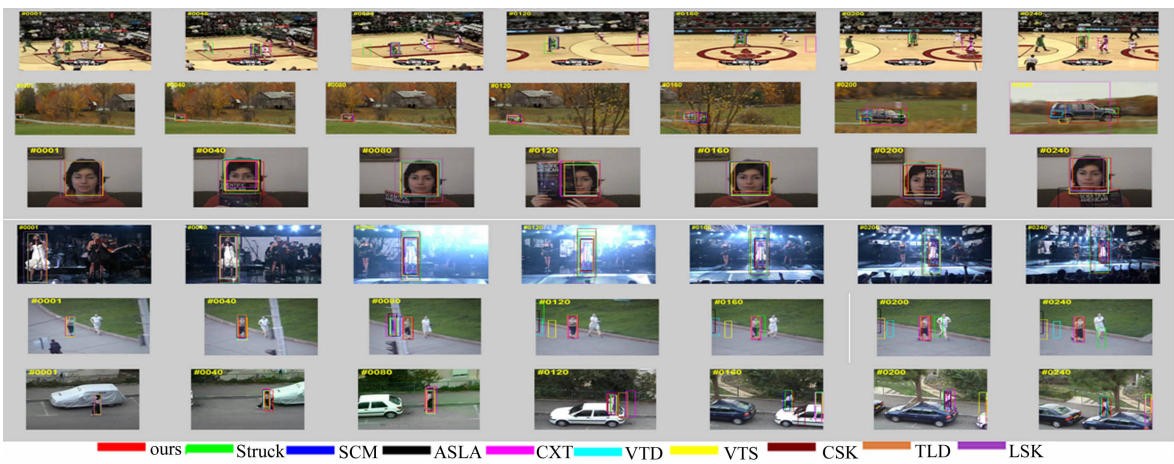
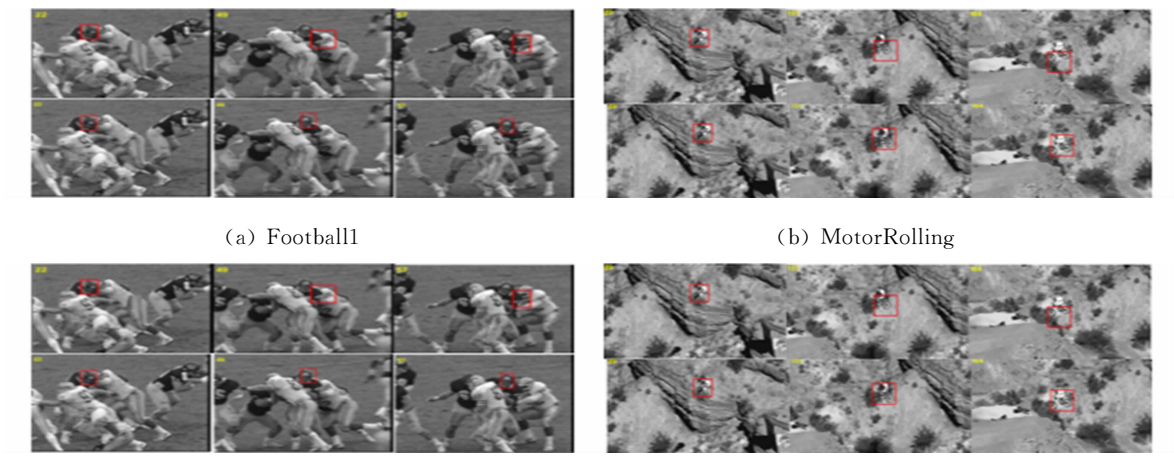


图 6　不同数据集上的跟踪效果对比

Fig. 6　Comparison of the tracking results on different data sets

实验以 CNN 为主模型、CDBN 为辅助模型，当主模型跟踪失败了才会启动辅助模型，这样就可以提高跟踪速度和准确率. 随机挑选 CNN 单模型跟踪丢失的 4 个视频序列(Football1，MotorRolling，Matrix，Soccer)，针对每一个视频序列，随机挑选 3 帧图片进行对比，其单模型与多模型的对比图，如图 7 所示. 图 7 中：上一行是 CNN 单模型跟踪的情况，下一行对应的是 CNN 与 CDBN 模型融合后的跟踪效果图. 由图 7 可知：CNN 与 CDBN 模型的融合挽救了单个模型跟踪丢失的情况的发生.



(a) Football1　　(b) MotorRolling

(c) Matrix　　(d) Soccer

图 7　CNN 单模型和 CNN 与 CDBN 多模型融合跟踪效果对比

Fig. 7　Comparison of the tracking results of CNN single model and multi-clue fusion(CNN and CDBN) model

3 结束语

提出了一种基于深度学习的多模型融合目标跟踪算法,在提出候选粒子方面,采用了性能互补的选择性搜索方法和粒子滤波方法.研究结果表明:基于深度学习的多模型融合方法能够提取表达能力更强的目标物体特征,从而有效地处理跟踪中遮挡、光照变化等问题;同时,采用性能互补的选择性搜索方法和粒子滤波方法,能更准确地在视频序列中搜索到跟踪中的目标粒子,从而减少跟踪漂移问题的发生.在CVPR2013跟踪算法的性能评价指标中验证了文中算法能够取得更好的跟踪性能.

参考文献:

[1] FAN Jialue, XU Wei, WU Ying, et al. Human tracking using convolutional neural networks[J]. IEEE Trans Neural Netw, 2010, 21(10): 1610-1623.

[2] CARNEIRO G, NASCIMENTO J C. Combining multiple dynamic models and deep learning architectures for tracking the left ventricle endocardium in ultrasound data[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2013, 35(11): 1649-1665.

[3] WANG Naiyan, YEUNG D Y. Learning a deep compact image representation for visual tracking[C]// Proceedings of Twenty-Seventh Annual Conference on Neural Information Processing Systems. Nevada: MIT Press, 2013: 5-10.

[4] UIJLINGS J R R, van DE SANDE K E A, GEVERS T, et al. Selective search for object recongnition[J]. International Journal of Computer Vision, 2013, 104(2): 154-171.

[5] CARREIRA J, SMINCHISESCU C. Cpmc: Automatic object segmentation using constrained parametric min-cuts [J]. PAMI, 2012, 34(7): 1312-1328.

[6] KRIZHEVSKY A, SUTSKEVER I, HINTON G E. ImageNet classification with deep convolutional neural networks [C]// Advances in Neural Information Processing Systems. Washington: MIT Press, 2012: 2-8.

[7] LEE H, LARGMAN Y, PHAM P, et al. Unsupervised feature learning for audio classification using convolutional deep belief networks[C]// Advances in Neural Information Processing Systems. New York: MIT Press, 2009: 1-22.

[8] HUANG G B, LEE H, LEARNED-MILLER E. Learning hierarchical representations for face verification with convolutional deep belief[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2012, 8(6): 1836-1844.

[9] LEE H, GROSSE R, RANGANATH R, et al. Unsupervised learning of hierarchical representations with convolutional deep belief networks[J]. Communications of the ACM, 2011, 54(10): 95-103.

[10] TORRALBA A, FERGUS R, FREEMAN W. 80 million tiny images: A large data set for nonparametric object and scene recognition[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2008, 30(11): 1958-1970.

[11] WU Yi, LIM J, YANG M H. Online object tracking: A benchmark[C]// IEEE Conference on Computer Vision and Pattern Recognition. Portland: IEEE Press, 2013: 2-10.

Multi-Clue Fusion Target Tracking Algorithm Based on Selective Search and Deep Learning

ZHONG Bineng[1,2], PAN Shengnan[1,2]

(1. College of Computer Science and Technology, Huaqiao University, Xiamen 361021, China;
2. Computer Vision and Pattern Recognition Laboratory, Huaqiao University, Xiamen 361021, China)

Abstract: A multi-clue tracking algorithm (convolutional neural network and convolutional deep belief network) based on deep learning was proposed. The algorithm used selective search and particle filtering method in extracting candidate particles. CVPR2013 tracking benchmark (50 video sequences, 30 tracking algorithms) verifies: the algorithm can ease the loss of tracking due to the occlusion, the change of illumination and size etc.

Keywords: object tracking; deep learning; multi-clue fusion; selective search; evaluating indicator

(**责任编辑**: 黄晓楠 **英文审校**: 吴逢铁)