计算机系统应用 ISSN 1003-3254, CODEN CSAOBN
Computer Systems & Applications,2018,27(11):78−83 [doi: 10.15888/j.cnki.csa.006607]
©中国科学院软件研究所版权所有.
E-mail: csa@iscas.ac.cn
http://www.c-s-a.org.cn
Tel: +86-10-62661041

基于多任务 CNN 的监控视频中异常行人快速检测①

李俊杰, 刘成林, 朱　明

(中国科学技术大学 信息科学技术学院, 合肥 230027)

摘　要: 在近年来社会公共安全受到广泛关注的情况下, 如何利用监控视频对异常行人进行监督, 预防危险事件的发生成为了一个热门课题. 异常行人是指与普通行人在外观上有明显异常性区别的人, 例如用头盔大面积遮挡面部或低头躲避摄像头, 考虑到异常行人的特征主要集中在头面部, 本文提出一种基于多任务卷积神经网络和单类支持向量机的针对头面部特征的异常行人快速检测方法. 首先进行头面部区域的检测, 然后使用多任务卷积神经网络提取头面部区域的特征, 之后使用单类支持向量机判断是正常行人还是异常行人. 此外, 本文还针对卷积神经网络设计了一种卷积核拆分方法, 加快了特征提取的速度, 最终实验表明, 本文提出的算法能够快速有效的检测出监控视频中的异常行人.

关键词: 监控视频; 异常行人; 多任务卷积神经网络; 卷积核拆分; 单类支持向量机

引用格式: 李俊杰,刘成林,朱明.基于多任务 CNN 的监控视频中异常行人快速检测.计算机系统应用,2018,27(11):78–83. http://www.c-s-a.org.cn/1003-3254/6607.html

Fast Abnormal Pedestrians Detection Based on Multi-Task CNN in Surveillance Video

LI Jun-Jie, LIU Cheng-Lin, ZHU Ming

(School of Information Science and Technology, University of Science and Technology of China, Hefei 230027, China)

Abstract: In case that public safety has already caused extensive social concern in recent years, how to use surveillance video to detect abnormal pedestrians and prevent dangerous events becomes a hot topic. Abnormal pedestrians are those who are distinctly different from ordinary pedestrians in appearance, for example, using helmet to cover the face or ducking from the camera. Considering that the characteristics of abnormal pedestrians are mainly concentrated in head and face, this study proposes a fast detection method for abnormal pedestrians based on multi-task Convolutional Neural Network (CNN) and one-class Support Vector Machine (SVM) for head-facial features. First, we detect head-facial regions in surveillance video, then we use the multi-task CNN to extract features of these regions, and then we use one-class SVM to judge whether it is a normal pedestrian or not. In addition, this study designs a convolution kernel splitting method for CNN to accelerate the feature extraction speed. Finally, the experiment shows that the algorithm proposed in this study can effectively and quickly detect abnormal pedestrians in surveillance video.

Key words: surveillance video; abnormal pedestrians; multi-task CNN (Convolutional Neural Network); convolution kernel splitting method; one-class SVM (Support Vector Machine)

随着社会经济水平和人们安全意识的提高, 视频监控的使用已经变得越来越普及, 如今在商场、校园、街道等公共区域, 我们能够很轻易的发现很多监控探头. 与此同时, 近年来人们对社会公共安全的关注度越来越高, 每当发生公共场所伤害事件都会在社会上引起广泛讨论. 在这种情况下, 以监控视频为载体的异常行人检测技术就成为了一个热门且重要的课题. 异常行人检测技术是指通过图像处理、机器学习等方

① 基金项目: 国家重大科技专项 (2017ZX03001019)
Foundation item: National Science and Technology Major Project of China (2017ZX03001019)
收稿时间: 2018-03-26; 修改时间: 2018-04-24; 采用时间: 2018-04-27; csa 在线出版时间: 2018-10-24

© 中国科学院软件研究所 http://www.c-s-a.org.cn

法在视频中检测是否有异常行人的存在并给出其位置, 其中异常行人是指与普通行人在外观上有明显异常性区别的人, 例如低头躲避摄像头或用帽子口罩等物品大面积遮挡住面部, 如图 1 所示. 该技术能够让管理者更容易的注意到监控视频中的异常行人, 更好的预防和应对各种突发情况, 同时节省了人力成本, 减轻了管理者的工作负担.

图 1 监控场景中的异常行人示例

1 异常行人检测概述

考虑到异常行人的异常特征大都集中在头面部, 本文针对头面部区域进行讨论, 提出了一个异常行人快速检测方法, 首先检测监控视频中的头面部区域, 然后使用多任务卷积神经网络提取头面部区域的特征, 最后用单类支持向量机进行异常行人的判别. 检测系统整体架构如图 2 所示.

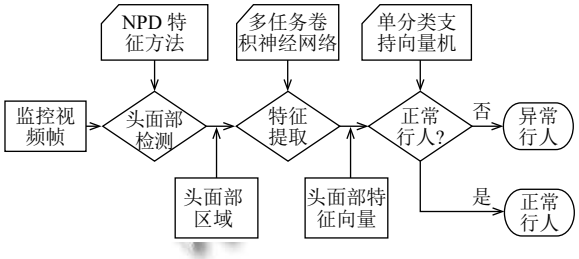

图 2 异常行人检测系统架构

本文提出的异常行人检测方法可分为以下两步: (1) 确定行人头面部区域位置, (2) 判断每个位置对应的行人是否为异常行人.

1.1 头面部区域检测

行人头面部区域检测的方法有两种: (1) 通过人脸检测得到人脸位置, 进一步扩充得到头面部区域, (2) 直接进行头面部区域检测.

人脸检测方法很多, 最经典的是 Viola 和 Jones 提出的基于 Haar-like 特征和级联 AdaBoost 分类器的方法[1], 但在实际场景中由于表情、遮挡、光照等因素的影响, 很多传统检测方法效果并不理想, 近年来针对非约束人脸提出了很多新的方法, 如基于 NPD 特征的方法[2]以及基于深度学习的方法[3,4], 这些方法都取得了很好的效果. 与人脸检测相比, 头面部区域检测的研究较少, 现有的一些方法如基于 FDF 特征[5]以及基于 HOG 特征[6]的方法, 在实际场景中效果也往往并不十分理想.

综合考虑, 本文选择了适用范围广、准确率高且检测速度快的 NPD 特征方法进行头面部区域的检测. 通过 NPD 特征方法获得人脸位置, 进而得到头面部区域位置.

1.2 异常行人判别

异常行人判别的研究主要有三个难点: (1) 异常行人样本非常稀少; (2) 可能出现的异常情况无法穷举; (3) 应用于实际场景中需要保证算法的实时性.

目前对头面部区域异常情况的讨论主要集中在遮挡问题上, 提出了很多面部遮挡检测方法, 例如用肤色比例判断遮挡[7,8], 或通过检测眼睛、嘴巴来间接判断遮挡情况[9–11]. 但是基于肤色比例的方法受光照等环境因素的影响较大, 而对眼睛、嘴巴的检测对图像的分辨率要求又很高, 此外实际场景中的异常情况也无法穷举, 因此这些针对性的方法用于实际场景中的异常行人检测往往无法取得令人满意的效果.

考虑到异常行人样本较少且无法列举出所有可能情况, 传统的分类模型无法应用于该问题, 本文针对性的使用了图像特征与单分类算法相结合的方法进行异常行人的判别.

常见的图像特征, 如 LBP[7], Haar-like[1], HOG[12]都经常用于行人相关的研究, 此外近年来随着深度学习方法的兴起, 用卷积神经网络提取图像特征也成为了一种有效手段[13].

单分类算法是指只用一类样本训练分类器, 进而该分类器能够判断输入是否属于该类. 本文中使用正常行人的头面部样本训练单分类器, 进而实现对异常行人的判别. 常见的单分类算法有单类支持向量机[14]、FAST-MCD 算法[15]和孤立森林算法[16].

2 算法设计与实现

2.1 多任务卷积神经网络

本文设计了一个多任务卷积神经网络用于行人头面部区域的特征提取, 该网络初级模型如图 3 所示.

© 中国科学院软件研究所 http://www.c-s-a.org.cn

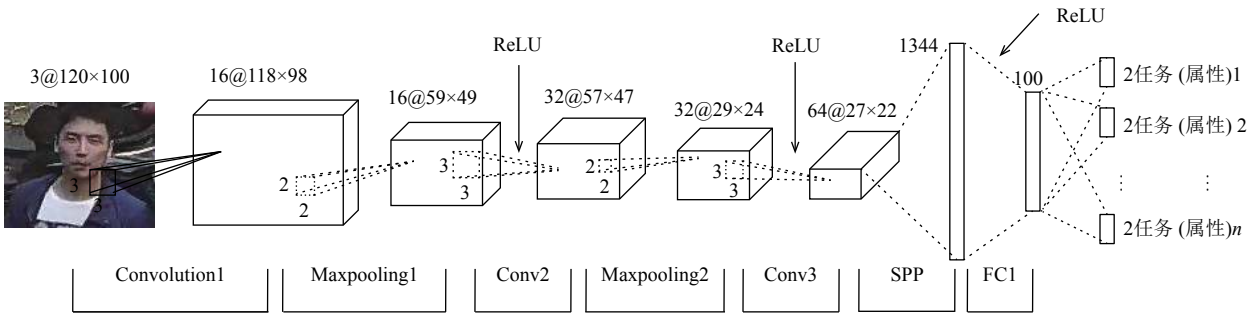

图 3 多任务卷积神经网络初级模型 (输入以 120×100 为例)

该网络包括数据输入部分、三个卷积层模块、一个全连接层和数据输出部分. 其中全连接层即作为输入图像的特征向量表示, 用于后续进一步操作.

上述卷积层模块是由一个卷积层、一个池化层和一个 ReLU 激活函数层组成, 其中前两个使用了最大值池化, 而第三个使用了空间金字塔池化 (SPP)[17], 通过使用空间金字塔池化层, 网络输入不再要求大小相同, 不需对输入图片进行剪切或非等比缩放, 可以尽可能的保留图像特征, 提高网络的特征提取能力.

与普通网络相比, 多任务网络在全连接层之后连接了多个不同的分类任务, 这些分类任务共同使用卷积网络输出的特征向量, 在训练过程中联合更新网络参数. 因此多任务网络提取出的特征向量对所有输出属性都有很好的代表性, 具体到头面部特征提取问题, 将是否戴眼镜、是否戴帽子等多个任务作为网络输出, 可以让该网络提取出的特征向量对整个头面部各部分细节都能进行很好的描述. 此外, 在网络训练过程中, 每一个头面部样本都对应多个标签, 如戴眼镜、不戴帽子等, 这种多标签样本也更适合多任务网络. 因此, 多任务卷积神经网络是最合适有效的头面部特征提取模型.

2.2 卷积核拆分

为了保证算法的实时性, 本文通过拆分卷积核的方法减少卷积网络的参数和计算次数, 加快特征提取速度.

本文参考 GoogLeNet[18]和 MobileNets[19]对卷积核进行了多个维度上的拆分, 示意图如图 4 所示.

图 4 中 M 为输入特征图通道数, N 为输出特征图通道数, 拆分前共有 N 个维度为 $k\times k\times M$ 的卷积核, 将其拆分成两个深度卷积和一个点卷积. 深度卷积将卷积核分别应用到单个输入通道, 点卷积对不同通道进行组合, 显然最终输出维度不会发生改变.

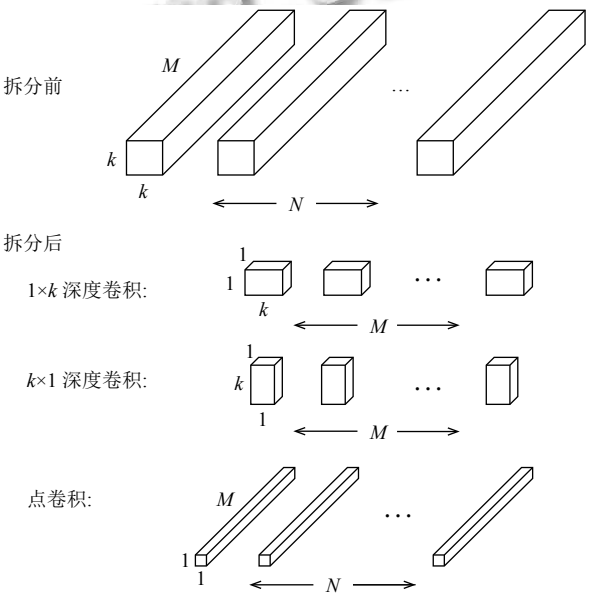

图 4 卷积核拆分

假定特征图尺寸在卷积前后不变, 始终为 $n\times n$, 因为每次卷积操作特征图尺寸减小量为 k–1, 远远小于特征图大小, 且在实际工程中可以通过像素填充方法实现, 故该假定不影响对计算代价的讨论.

卷积核拆分后各层的计算代价依次为 $n\times n\times M\times 1\times k\times 1$、$n\times n\times M\times k\times 1\times 1$ 和 $n\times n\times N\times 1\times 1\times M$, 与拆分前相比得:

$$\frac{n\cdot n\cdot M\cdot k+n\cdot n\cdot M\cdot k+n\cdot n\cdot N\cdot M}{n\cdot n\cdot N\cdot k\cdot k\cdot M}=\frac{2}{kN}+\frac{1}{k^2}$$

相比结果小于 1, 拆分操作成功降低了计算代价. 以初级网络模型中 k 为 3, N 依次为 16、32 和 64 为例, 三个卷积层的计算代价分别降低为拆分前的 15.3%、13.2% 和 12.2%. 最终得到网络模型如图 5 所示.

2.3 训练数据集

该网络的训练过程分为两步:

© 中国科学院软件研究所 http://www.c-s-a.org.cn

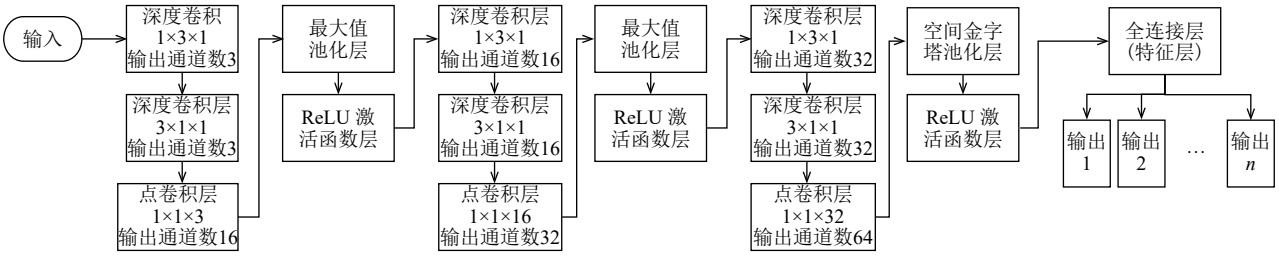

图 5 改进后的多任务卷积神经网络模型

1) 使用公开人脸属性数据集 CelebA[20]进行网络预训练, 选用了其中十二个属性作为多任务网络模型的输出属性, 分别为眼袋、光头、刘海、黑发、金发、眼镜、性别、年龄段、嘴巴张开、胡子、帽子和领带. 部分属性及对应样本如图 6 所示.

2) 在预训练得到的参数基础上, 用实际监控视频中的样本进行微调, 多任务网络的输出部分改为如下四个分类任务: 是否戴眼镜、是否戴帽子、是否露出嘴巴和人脸方位 (正面、侧面和背面), 如图 7 所示.

图 6 CelebA 数据集部分样本示例

图 7 实际监控视频部分样本示例

2.4 单分类算法

单分类算法中应用最广泛的是单类支持向量机. 单类支持向量机是基于支持向量机提出的算法, 与支持向量机相同, 核函数[21]的选择是影响其性能的关键.

目前常用的核函数有高斯核函数、Sigmoid 核函数、多项式核函数、线性核函数和三角核函数[22]等. 本文实验部分将对不同单分类算法以及不同核函数的性能进行详细的比较.

3 实验与分析

本文系统实现部分使用了深度学习框架 Caffe[23] 和基于 Python 的机器学习工具 sklearn[24].

3.1 多任务卷积神经网络的训练与评估

首先, 我们在 CelebA 人脸属性数据集上进行网络预训练, 该数据集包含 202 599 张人脸图片, 并为每张图片提供了多达 40 种属性的标注. 在进行了一定程度的等比例缩放后, 随机选取 2 万张作为测试集, 其他作为训练集. 之后, 我们在实际监控场景的头面部样本集上进行参数微调, 共使用了 1 万张图片, 人工对分类属性进行标注后, 随机选取其中 2 千张作为测试集, 其他作为训练集.

我们分别训练了卷积核拆分前后的网络模型, 计算了所有分类任务的平均准确率和处理 100 张图片所需的平均耗时, 结果如表 1 所示.

表 1 卷积核拆分前后效果对比

	卷积核	平均准确率 (%)	分类耗时 (s/100 张)
预训练 (CelebA)	拆分前	92.05	0.4274
	拆分后	91.64	0.3179
微调 (实际场景样本)	拆分前	94.28	0.4071
	拆分后	94.32	0.2982

实验结果表明, 本文设计的网络结构在多分类问题上有较高的准确率, 提出的卷积核拆分方法也明显的降低了计算代价, 减少了网络分类耗时. 此外我们发现随着进行微调时分类任务数的减少, 耗时也会相应减少, 因此用该网络提取特征时只计算到全连接层会进一步降低耗时.

3.2 图像特征与单分类器的组合

我们在实际监控场景的样本集上评估了不同图像特征与不同单分类器的组合效果. 我们使用了 1 万张正常行人头面部样本进行训练, 各 2 千张正常样本和异常样本进行测试. 准确率对比如表 2 所示.

© 中国科学院软件研究所 http://www.c-s-a.org.cn

表 2 不同组合准确率对比 (%)

		HOG	LBP	Haar	multi-taskCNN
FAST-MCD		76.8	59.2	60.2	55.4
孤立森林		79.9	61.7	79.1	79.2
单类支持向量机	高斯核	75.5	78.3	77.3	92.5
	Sigmoid 核	64.2	53.6	60.4	91.1
	多项式核	54.8	55.3	51.2	90.5
	线性核	64.2	53.5	60.3	91.2
	三角核	90.3	77.5	83.1	91.8

可以看出, 本文提出的多任务卷积神经网络特征提取模型与单类支持向量机的组合准确率明显高于其他方法.

在此基础上, 我们评估了不同核函数的分类速度 (Frames Per Second, FPS) 和召回率, 这两者很大程度上影响着系统的实用性. 其中, 召回率是指单分类器正确判别为异常行人的样本在总异常行人样本中的比例, 在异常检测问题中是一个十分重要的性能指标. 结果如表 3 所示.

表 3 不同核函数相应分类速度和召回率

	高斯核	Sigmoid 核	多项式核	线性核	三角核
FPS(张/s)	4479	4229	5425	6294	4360
召回率 (%)	98.75	99.95	98.35	99.95	94.5

表 3 中, 线性核的分类速度和召回率都明显好于其他核函数, 并且准确率只略低于高斯核和三角核. 因此, 我们选择基于线性核的单类支持向量机作为异常行人检测系统的单分类器.

3.3 异常行人检测系统

如前文所述, 本文设计的异常行人检测系统的实现主要依赖于人脸检测、多任务卷积神经网络特征提取模型和单类支持向量机三个算法.

我们评估了三个典型的人脸检测算法在本系统中的效果. 在分辨率为 960×540 的监控视频中测试, 相应检测速度 (FPS) 和异常行人召回率如表 4 所示.

表 4 异常行人检测系统性能

人脸检测方法	FPS(张/s)	召回率 (%)
Haar-like[1]	36.552	25.36
NPD[2]	15.425	81.64
Cascaded CNN[4]	4.156	91.13

综合考虑检测速度和召回率, 我们选择了基于 NPD 特征的人脸检测算法来实现头面部区域的检测. 最终异常行人检测效果如图 8 所示.

图 8 异常行人检测示例

4 总结与展望

本文设计了一个用于快速提取头面部特征的多任务卷积神经网络, 并结合 NPD 人脸检测算法和基于线性核的单类支持向量机, 实现了监控视频场景中异常行人的快速检测, 取得了令人满意的效果.

下一步的研究工作将致力于设计召回率高且检测速度快的头面部检测算法, 以使得整个异常行人检测系统更加的快速且有效.

参考文献

1 Viola P, Jones M. Rapid object detection using a boosted cascade of simple features. Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. Kauai, HI, USA. 2001.

2 Liao SC, Jain AK, Li SZ. A fast and accurate unconstrained face detector. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2016, 38(2): 211–223. [doi: 10.1109/TPAMI.2015.2448075]

3 Wang K, Dong Y, Bai HL, *et al*. Use fast R-CNN and cascade structure for face detection. Proceedings of 2016 Visual Communications and Image Processing. Chengdu, China. 2016. 1–4.

© 中国科学院软件研究所 http://www.c-s-a.org.cn

4 Li JJ, Karmoshi S, Zhu M. Unconstrained face detection based on cascaded convolutional neural networks in surveillance video. Proceedings of the 2nd International Conference on Image, Vision and Computing. Chengdu, China. 2017. 46–52.
5 Ishii Y, Hongo H, Yamamoto K, *et al*. Face and head detection for a real-time surveillance system. Proceedings of the 17th International Conference on Pattern Recognition. Cambridge, UK. 2004. 298–301.
6 Ding XF, Xu H, Cui P, *et al*. A cascade SVM approach for head-shoulder detection using histograms of oriented gradients. Proceedings of 2009 IEEE International Symposium on Circuits and Systems. Taipei, China. 2009. 1791–1794.
7 Ji PF, Kim Y, Yang Y, *et al*. Face occlusion detection using skin color ratio and LBP features for intelligent video surveillance systems. Proceedings of 2016 Federated Conference on Computer Science and Information Systems. Gdansk, Poland. 2016. 253–259.
8 Zhang XH, Zhou L, Zhang T, *et al*. A novel efficient method for abnormal face detection in ATM. Proceedings of 2014 International Conference on Audio, Language and Image Processing. Shanghai, China. 2014. 695–700.
9 张伟峰, 朱明. 基于巡逻小车的人脸遮挡异常事件实时检测. 计算机系统应用, 2017, 26(12): 175–180.
10 Zhang YL, Lu Y, Wu HT, *et al*. Face occlusion detection using cascaded convolutional neural network. Proceedings of the 11th Chinese Conference on Biometric Recognition. Chengdu, China. 2016. 720–727.
11 Xia YZ, Zhang BL, Coenen F. Face occlusion detection based on multi-task convolution neural network. Proceedings of the 12th International Conference on Fuzzy Systems and Knowledge Discovery. Zhangjiajie, China. 2015. 375–379.
12 Dalal N, Triggs B. Histograms of oriented gradients for human detection. Proceedings of 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. San Diego, CA, USA. 2005. 886–893.
13 Razavian AS, Azizpour H, Sullivan J, *et al*. CNN features off-the-shelf: An astounding baseline for recognition. Proceedings of 2014 IEEE Conference on Computer Vision and Pattern Recognition Workshops. Columbus, OH, USA. 2014. 512–519.
14 Schölkopf B, Platt JC, Shawe-Taylor J, *et al*. Estimating the support of a high-dimensional distribution. Neural Computation, 2001, 13(7): 1443–1471. [doi: 10.1162/089976601750264965]
15 Rousseeuw PJ, Van Driessen K. A fast algorithm for the minimum covariance determinant estimator. Technometrics, 1999, 41(3): 212–223. [doi: 10.1080/00401706.1999.10485670]
16 Liu FT, Ting KM, Zhou ZH. Isolation-based anomaly detection. ACM Transactions on Knowledge Discovery from Data, 2012, 6(1): 1–39. [doi: 10.1145/2133360.2133363]
17 He KM, Zhang XY, Ren SQ, *et al*. Spatial pyramid pooling in deep convolutional networks for visual recognition. Proceedings of the 13th European Conference on Computer Vision– ECCV 2014. Zurich, Switzerland, 2014: 346–361.
18 Szegedy C, Vanhoucke V, Ioffe S, *et al*. Rethinking the inception architecture for computer vision. Proceedings of 2016 IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas, NV, USA. 2016. 2818–2826.
19 Howard AG, Zhu ML, Chen B, *et al*. Mobilenets: Efficient convolutional neural networks for mobile vision applications. arXiv preprint arXiv:1704.04861, 2017.
20 Liu ZW, Luo P, Wang XG, *et al*. Deep learning face attributes in the wild. Proceedings of 2015 IEEE International Conference on Computer Vision. Santiago, Chile. 2015. 3730–3738.
21 汪廷华, 陈峻婷. 核函数的选择研究综述. 计算机工程与设计, 2012, 33(3): 1181–1186. [doi: 10.3969/j.issn.1000-7024.2012.03.068]
22 Utkin LV, Chekh AI. A new robust model of one-class classification by interval-valued training data using the triangular kernel. Neural Networks, 2015, 69: 99–110. [doi: 10.1016/j.neunet.2015.05.004]
23 Jia YQ, Shelhamer E, Donahue J, *et al*. Caffe: Convolutional architecture for fast feature embedding. Proceedings of the 22nd ACM International Conference on Multimedia. Orlando, FL, USA. 2014. 675–678.
24 Pedregosa F, Varoquaux G, Gramfort A, *et al*. Scikit-learn: Machine learning in Python. Journal of Machine Learning Research, 2011, 12(10): 2825–2830.

© 中国科学院软件研究所 http://www.c-s-a.org.cn