

# 融合时序相关性的课堂异常行为识别<sup>①</sup>

王明芬, 卢宇

(福建师范大学协和学院 信息技术系, 福州 350117)

通讯作者: 王明芬, E-mail: 605342295@qq.com



**摘要:** 针对人体行为最重要的 motion 特征, 提出了基于时间上下文的二级递推异常行为识别方法. 不同于传统深度学习训练方法, 本文方法不是直接从图像数据中学习特征, 而是把提取的形状信息 HOG 特征作为训练输入. 首先提取基于 HOG 算法的图像形状特征, 采用提取到的特征训练 DBN 网络. 其次利用已经训练好的 DBN 网络和 Softmax 分类器识别出人体粗目标区, 然后根据粗目标区域的时序上下文信息, 计算质心加速度. 最后判断加速度的阈值, 识别出异常行为的精目标区. 本文将粗细目标结合的二级递推方法应用到课堂行为识别中, 通过实验结果表明, 该方法在运动模糊和目标密集遮挡的场景下都能较好地识别出课堂行为, 识别率相比其他方法有较大提升. 课堂异常行为数据分析, 可在课堂动态管理和学习效果评估等方面发挥辅助作用.

**关键词:** DBN 网络; HOG 特征; 时序相关性; 二级递推; 异常行为

引用格式: 王明芬, 卢宇. 融合时序相关性的课堂异常行为识别. 计算机系统应用, 2020, 29(3): 173-179. <http://www.c-s-a.org.cn/1003-3254/7305.html>

## Classroom Abnormal Behavior Recognition Based on Sequential Correlation

WANG Ming-Fen, LU Yu

(Concord University College Fujian Normal University, Fuzhou 350117, China)

**Abstract:** Aiming at the most important motion characteristics of human behavior, a second-level recursive anomaly behavior recognition method based on time context is proposed. Different from traditional deep learning training methods, this method does not directly learn features from image data, but extracts them. The shape information HOG feature is used as the training input. Firstly, the image shape feature based on the HOG algorithm is extracted, and the extracted feature is used to train the DBN network. Secondly, the trained DBN network and the Softmax classifier are used to identify the human body coarse target region, and then according to the coarse The time-series context information of the target area, calculate the centroid acceleration. Finally, the threshold of the acceleration is judged, and the precise target area of the abnormal behavior is identified. This paper applies the two-level recursive method combining the weight and the target to the classroom behavior recognition, and the experimental results show that the The method can better recognize the classroom behavior in the scenes of motion blur and target dense occlusion, and the recognition rate is greatly improved compared with other methods. Classroom abnormal behavior data analysis can play a supporting role in classroom dynamic management and learning effect evaluation.

**Key words:** DBN network; HOG feature; time series correlation; two-level recursion; abnormal behavior

① 基金项目: 福建省重点项目 (2017H0011); 福建省教育厅项目 (JT180823); 福建省省级教改项目 (FBJG20180130)

Foundation item: Major Program of Fujian Province (2017H0011); Program of Education Bureau, Fujian Province (JT180823); Fujian Provincial Education Reform Project (FBJG20180130)

收稿时间: 2019-08-04; 修改时间: 2019-09-02; 采用时间: 2019-09-10; csa 在线出版时间: 2020-02-28

随着计算机视觉技术的发展以及各类视频摄像终端的普及,研究表明人工对海量视频信息的处理具有容易疲劳和自动忽略的缺点.人工智能领域的一个重要研究热点是目标行为识别,要求系统智能检测和识别出感兴趣的目标,减轻人工的工作量<sup>[1]</sup>.近年来在智能监控报警安防等方面,深度学习已有诸多成功的应用.深度置信神经网络模型(DBN)是一个概率生成模型,由Hinton等提出<sup>[2]</sup>,该模型建立一个观察数据和标签之间的联合分布,通过无监督预训练和有监督调优训练达到理想的网络模型.传统的深度学习不对特征进行提取方法的设计,直接利用图像信息进行训练得到目标表示法<sup>[3]</sup>.但是图像具有丰富的特征可以用来描述关键信息,这些特征的训练在系统识别中会发挥重要的作用,因为特征的好坏直接影响到最终的识别效果<sup>[4]</sup>.在视频中时序的相关性是个不可忽视的信息特征,利用时间上下文信息能为系统识别带来增益<sup>[5]</sup>.因此本文提出了将提取的HOG特征作为输入,通过深度置信网络训练得到更高层的抽象特征,利用训练好的DBN网络识别人体区域,最后利用区域的质心的时序相关性特征判断课堂异常行为.通过实验数据验证了课堂行为识别算法的有效性,实验结果表明在训练数据比较少的情况下,也能获得较好的识别效果.

## 1 特征描述

梯度特征可以很好地描述局部目标的形状边缘,梯度方向直方图被用来描述HOG特征,能够有效地对形状特征检测,主要用于解决人体目标检测<sup>[6]</sup>.

### 1.1 HOG特征算法

HOG采用了统计的方式进行提取.首先将图像颜色空间归一化,然后计算梯度,接着将图像分成小的Cell,然为每个Cell中各像素点的梯度方向直方图,最后把每个Block(扫描窗口)的特征进行联合以形成最终的特征<sup>[6]</sup>.具体计算流程图如图1所示.

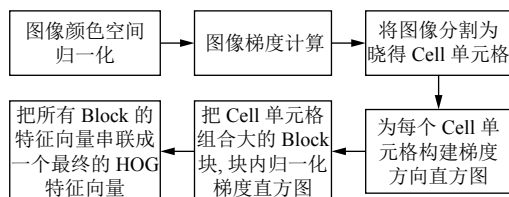


图1 HOG特征提取流程图

HOG先计算各个单元灰度直方图,然后进行归一化处理,降低对光照和阴影的敏感性<sup>[7]</sup>.因此其在人体检测方面有着有独特较多优点,适用于做图像及视频中的人体检测特征.

### 1.2 特征提取

HOG特征最小单位是Cell,计算块区域Block和检测窗口的计算步长就是一个Cell的宽度,因此先把整个图像分割为多个的Cell单元格<sup>[8]</sup>,按特征算法结果共有128个单元格.

实验中我们把梯度图通过分解提取变为机器容易理解的特征向量.将Cell的梯度方向360度分成9个方向块得到特征,每个块包含4个Cell,一个检测窗口特征向量是36.一个64×128大小的图像计算后,它的特征数为36×7×15=3780个.可视化的HOG特征提取显示如图2所示.



图2 可视化特征提取

### 1.3 几何特征

上文的HOG特征是基于形状边缘梯度的特征,在此基础上别出来的目标很容易用几何特征来进一步识别行为动作.本文选取质心的变化加速度来判断课堂的异常行为.

目标区域在坐标系轴上进行投影,接着进行区域扫描那么目标区域就可以用 $P1$ 和 $P2$ 描述,记作 $R[P1, P2]$ ,用外接矩形框表示目标区域如图3所示.

图3中, $P1$ 坐标为 $(x_{\min}, y_{\min})$ , $P2$ 坐标为 $(x_{\max}, y_{\max})$ ,该目标区域记作 $M$ ,则其质心可以表示成:

$$x_{\text{cor}} = \frac{\sum_{(x,y) \in M} xH(x,y)}{\sum_{(x,y) \in M} H(x,y)} \quad (1)$$

$$y_{\text{cor}} = \frac{\sum_{(x,y) \in M} yH(x,y)}{\sum_{(x,y) \in M} H(x,y)} \quad (2)$$

式中, $H(x,y)$ 表示人的目标区域在 $(x,y)$ 位置的像素点灰度值信息.

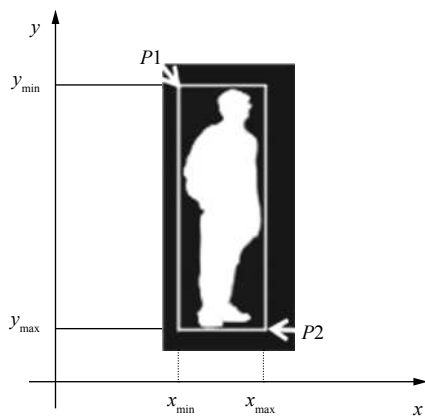


图3 目标矩形区域

### 1.4 时序相关性特征

在视频中, 时序特征能够很好地表示目标的运动趋势. 本文从视频序列的几何特征中计算目标质心的位移和时间, 然后计算前后  $K$  帧间隔的质心加速度, 把质心加速度变化设置为时序相关性特征.

假设第  $N$  帧的质心是  $(x_n, y_n)$ , 第  $M$  帧的质心是  $(x_m, y_m)$ , 计算出质心在  $x, y$  两个方向上的位移  $S1$ :

$$S1 = \sqrt{(x_m - x_n)^2 + (y_m - y_n)^2} \quad (3)$$

第  $N$  帧与第  $M$  帧的时间差是  $t$ , 就可以得到质心的位移速度  $V_1 = S1/t$ .

同理可以得到第  $M$  帧与第  $L$  帧之间的质心速度  $V_2 = S2/t$ . 即可以求出质心加速度:

$$a = \frac{v_2 - v_1}{t} \quad (4)$$

当目标的质心加速度突然加快, 说明目标在短时间内位置发生了变化. 当这个加速度超过设置的阈值,  $z$  则判定为课堂异常行为, 触发警报信息.

## 2 深度置信网络

深度置信网络是由多个限制玻尔兹曼机堆叠以及一个 BP 层组合而成的深度置信网络. 在深度置信网络中, 每个隐含层接收来自低层的神经元的输入, 通过层与层之间非线性关系, 将底层特征组合成高层的信息表示, 并建立观测数据的分布式特征. 它贪婪的前向学习<sup>[9]</sup>, 通过逐层学习可以逐步收敛. 并结合梯度下降<sup>[10]</sup>的反向微调机制, 可以得到更高的收敛精度, 从而达到最佳的模型训练. 根据学习到的网络结构, 系统将输入的样本数据映射到输出特征, 然后采用 Softmax 分类

器识别.

### 2.1 限制玻尔兹曼机

对每层玻尔兹曼机 (RBM) 进行训练是一个深层置信网络的开始. 训练 RBM 的过程简单来说就是寻找可视层节点和隐藏层节点之间连接的最优权值参数. RBM 由一层可视层  $v$  和一层隐藏层  $h$  组成. 该网络的可视层  $v$  和隐藏层  $h$  神经元彼此双向互联, 但同一层内神经元无连接. RBM 中神经元有两种状态, “激活”和“未激活”, 一般用二进制的 1 和 0 表示<sup>[11]</sup>. 每一层可用一个向量表示, 向量的维数由每层神经元的个数决定, 每一个神经元代表数据向量的一维, 具体结构图如图 4 所示.

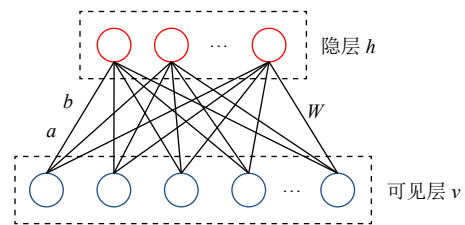


图4 RBM 结构图

RBM 是的可视层神经元向量  $v$  和隐藏层神经元向量  $h$  联合配置的函数为:

$$E(v, h|\theta) = - \sum_{i=1}^n a_i v_i - \sum_{j=1}^m b_j h_j - \sum_{i=1}^n \sum_{j=1}^m w_{ij} v_i h_j \quad (5)$$

式中,  $\theta = (w_{ij}, a_i, b_j)$  为 RBM 的参数,  $a_i$  为可视层单元的偏置值,  $b_j$  为隐含层单元的偏置值,  $w_{ij}$  为可视层与隐含层之间的连接权重,  $n$  和  $m$  分别为可视层与隐含层的神经元数目. 由能量函数可以得到可视层与隐含层的联合概率分布为:

$$p = (v, h|\theta) = \frac{\exp[-E(v, h|\theta)]}{Z(\theta)} \quad (6)$$

式中,  $Z(\theta) = \sum_v \sum_h \exp[-E(v, h|\theta)]$  是归一化因子. 可视层与隐含层节点之间无连接让每个节点之间具有独立性. 当可视层节点状态知道时, 激活隐含层节点的概率为:

$$P(h_j = 1|v) = \frac{1}{1 + \exp(-\sum_i w_{ij} v_i - a_j)} \quad (7)$$

在给定隐含层  $h$  的前提下, 求得其可视层第  $i$  个节点激活概率为:

$$P(v_i|h) = \frac{1}{1 + \exp(-\sum_j W_{ij}h_j - b_i)} \quad (8)$$

RBM 采用极大对数似然函数迭代方法训练出可以拟合目标的参数  $\theta$ . 然后以经典的对比散度算法更新权值, 可以得到优的参数  $\theta$ . 逐层训练完成的 RBM 可以从高维数据中提取出更有表征意义的特征<sup>[12]</sup>.

### 2.2 训练 DBN 网络

自底向上的逐层学习, 通过底层特征训练得到高层特征是 DBN 模型的训练学习方式<sup>[13]</sup>, 在顶层设计一个 BP 网络, 根据识别的误差进行网络参数的反向微调.

首先是预训练, 用贪婪学习算法训练波尔兹曼机, 一次一个直到所有的波尔兹曼机都被训练完成为止. 低层的隐含层的输出将作为高层可视层的输入, 经过不断调整网络权值, 网络状态达到和谐. 经过训练之后得到 DBN 的初始参数  $\theta = (w_{ij}, a_i, b_j)$ . 通过 BP 网络梯度下降算法实现反向微调, 将误差自顶向下地反向传播到每一层, 通过梯度下降算法对整个网络进行微调, 整个网络的参数达到理想. 如图 5 所示.

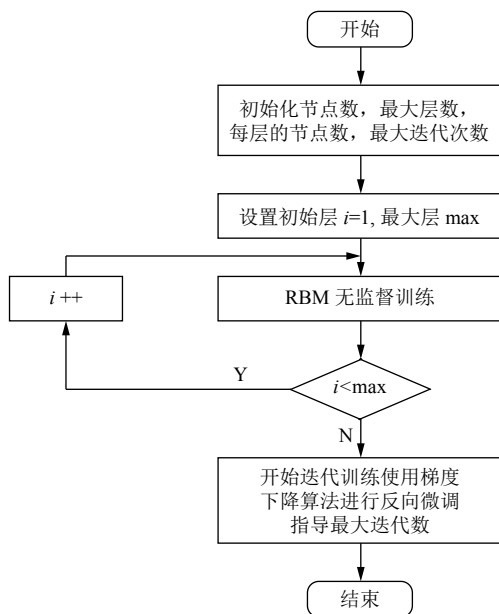


图 5 DBN 训练流程

从一层神经网络开始训练一个网络的方法是可行的, 且可以节约网络资源避免过度计算. 在第一个隐藏层和标签输出层之间插入第二个隐藏层, 然后对整个网络通过 BP 网络反向调整网络的权值. 以此类推, 一层层地设计网络的层数, 这种判别式预训练在能够取

得很好的效果.

## 3 系统设置

对于深度学习的网络模型, 训练迭代次数、网络隐含层的层数是重要参数. 在用 DBN 网络模型对训练样本进行训练时, 采用 BP 算法将训练所得结果与结果标签数据进行误差分析, 根据相关差异进行反向微调, 实现对网络结构中各层间权值的更新, 逐步达到提升网络模型识别精确度的目的<sup>[14]</sup>. 在时间上下文信息中我们需要计算质心的加速度, 因此合适的帧间隔也是一个重要的参数.

### 3.1 系统参数配置

结合实际实验采用包含 1-3 层 RBM 的深度置信网络结构模型. 设置预训练的学习率 0.01, 设置 BP 神经网络的学习率为 0.01, 迭代次数设置为 2000. 采用批训练的方式初始化节点数, 批训练样本数设置为 200. 通过实验分析的方式把网络中 RBM 的层数确定下来, 文中设置 DBN 模型中 RBM 层数为 2.

视频播放的帧是 25 f/m, 我们通过实验对比选择  $K$  帧间隔, 取  $K=5$  为实验参数, 即帧间隔为 1/5 s, 每秒计算 5 次质心的位移速度, 4 次质心加速度. 当质心加速度特征  $a > 4 \text{ m/s}^2$  时, 认为是课堂异常行为.

### 3.2 系统算法流程

从提取的形状特征中提取更为抽象的高层特征作为 DBN 网络的输入, 能更好地让 DBN 网络理解图像特征的分布, 提高 DBN 的表征能力. 本文先采用基于 HOG 算法的图像形状特征提取, 采用提取到的特征训练 DBN 网络. 其次利用二级递推算法, 首先识别出人体目标, 其次利用视频的时序相关性运动特征计算人体区域前后帧的质心速度, 判断课堂行为算法流程如图 6 所示.

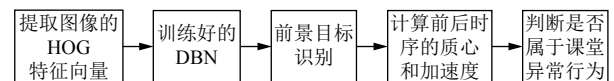


图 6 系统框架流程

## 4 实验结果与分析

硬件实验环境为 CPU 型号 Intel i9 9900X, 内存 32 GB, 显存 11 GB, 集成显卡 GTX 1080 Ti 的工作站, 软件环境为 Ubuntu 14.04 操作系统, Python 3.7+OpenCV 集成系统.

#### 4.1 课堂训练样本库

本系统训练的样本为课堂采集的小样本库, 命名为 Classroom 数据集, 类别是 book, chair person, table

等 4 类. 系统主要分析课堂目标中的时序相关性特征, 因此本文只计算 person 类的质心加速度.

Classroom 数据集训练库的部分图片如图 7 所示.

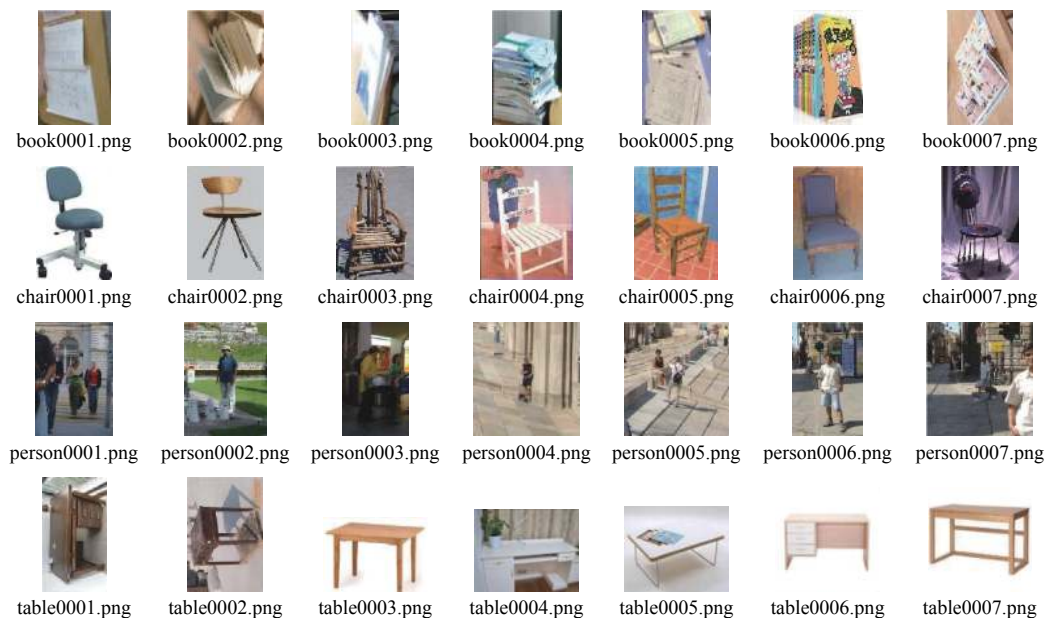


图 7 课堂训练样本库

#### 4.2 课堂检测结果

通过 Classroom 小样本数据集训练的深度置信网络模型, 通过 Softmax 分类器识别效果如图 8, 图 9 所示, 可以识别出 person, table, book, chair 等 4 类目标. 从实验可以看出在目标模糊和目标密集的复杂场景下, 目标也可以被有效地识别出. 这对今后的由于摄像头晃动造成的运动模糊和运动遮挡有很好的应用参考.

文的分析可知, 目标的质心加速度是一个重要的时序相关性特征. 因此计算的帧间隔是个关键参数, 间隔太小则增加系统的计算量, 太大则容易产生漏检. 根据实验调试, 系统设置  $N=5$  的帧间隔, 检测阈值设置为  $4 \text{ m/s}^2$ , 当加速度  $a$  超出设置的阈值时则认为是异常行为. 系统中采用粗框对异常行为目标进行预警.

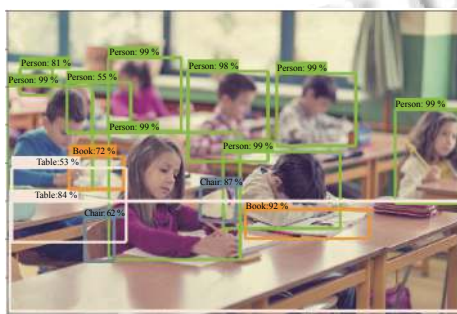


图 8 模糊场景识别效果

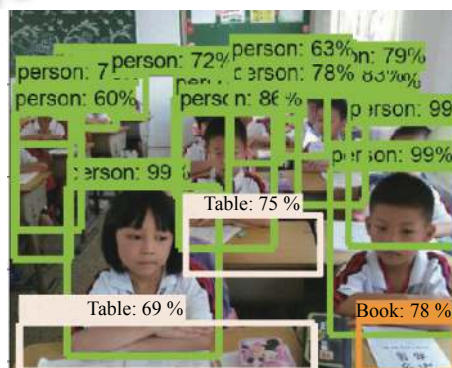


图 9 密集场景识别效果

本系统关心的是课堂 person 类的行为, 因此根据识别结果选择人体目标, 其他类的目标在视频中不再标出. 在正常情况下, 人体目标都是细框图显示. 由前

测试学校提供的课堂视频, 当学生课堂出现了睡觉、趴在桌子上等负面异常行为时, 质心加速度超出了阈值, 系统认定为异常行为目标如图 10, 图 11 所示.

应用网络上的视频测试系统,结果如图12所示.显然突然起立,目标质心加速度也会超过阈值.这时系统也将其标识为异常行为,这个属于正面异常行为.

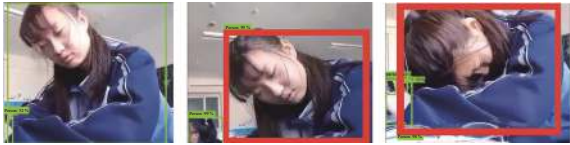


图10 单目标课堂异常行为识别

异常行为有正面异常行为和负面异常行为,但是两者并不是绝对对立的.如在智能监考系统中,起立代表负面异常,但在教学课堂中则代表正面异常.

### 4.3 系统算法分析

不同RBM网络层数的DBN模型率如图13所示.采用的数据样本集为Classroom数据集,KTH人体行为数据库,INRIA Person库,其中第一个数据库为实际采集的小样本数据库,后两者为测试公开库.由于实验所用的样本数较小,涉及到的类别也不多,因此DBN型所需的RBM层数和隐藏层节点不需要设置太多.本文将隐含层RBM的隐藏层节点数量设置为30个.通过实验测试RBM层数与识别率的关系如图13所示.实验结果可知当设置2层RBM时目标识别率均较高,当层数再增加时,网络的识别率反而有所下降.在RBM层数为2时学生课堂的人体目标识别率为98%,符合我们系统的指标要求.

在INRIA Person数据集上进行测试,得到的检



图12 网络课堂视频测试结果

加速度是一个很好的物理特征,计算加速度的时间间隔是一个重要的参数.帧间隔太大,无法检测出理想的目标,间隔太小影响系统的实时性.合理的帧间隔不仅可以检测出速度变化的快慢,而且可以有效的降低系统的运算开销.因此本文采用基于HOG特征输入的2层RBM结构的DBN模型,顶层采用Softmax分

测率如表1所示.可以看出,和未进行特征提取训练的DBN相比,加入HOG特征提取的DBN在准确度上有较好的提升<sup>[15]</sup>,因为HOG可以增强目标的局部特征.同时在输出层设置Softmax分类器,在目标类别不是特别多的情况下,可明显提升目标的识别率.



图11 多目标课堂异常行为识别

类器识别出目标.在视频序列中采用帧间隔为5的参数计算时序相关性特征,最后标识出课堂异常行为目标.

## 5 结论与展望

针对人体行为最重要的motion特征,提出了基于时序相关性的二级递推异常行为识别方法.不仅能解

决传统 DBN 不能处理视频序列的问题,而且可以充分利用视频中目标前后帧提供的质心加速度信息识别出异常目标,提高了系统的识别准确率.实验结果表明本文设计的方案在运动模糊和目标遮挡等复杂场景下都可以识别出目标,这对今后的实际应用中由于摄像头晃动造成的运动模糊和运动遮挡有很好的应用参考.系统后续可以展开联动模块的设计,把课堂行为中异常数据传输到云端进行分析,可在评估学习效果、课堂动态趋势等方面发挥作用.

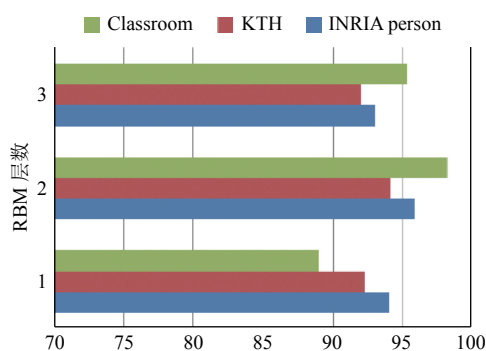


图 13 RBM 层数与识别率

表 1 不同模型识别率

模型结构	识别率 (%)
DBN <sup>[16]</sup>	93.40
HOG+DBN	95.80
HOG+DBN+Softmax	98.20

### 参考文献

- Saligrama V, Chen Z. Video anomaly detection based on local statistical aggregates. Proceedings of 2012 IEEE Conference on Computer Vision and Pattern Recognition. Providence, RI, USA. 2012. 2112–2119.
- Hinton GE, Salakhutdinov RR. Reducing the dimensionality of data with neural networks. Science, 2006, 313(5786): 504–507. [doi: 10.1126/science.1127647]
- 朱煜, 赵江坤, 王逸宁, 等. 基于深度学习的人体行为识别算法综述. 自动化学报, 2016, 42(6): 848–857.
- Hu YJ, Ling ZH. DBN-based spectral feature representation for statistical parametric speech synthesis. IEEE Signal Processing Letters, 2016, 23(3): 321–325.
- Hrasko R, Pacheco AGC, Krohling RA. Time series prediction using restricted Boltzmann machines and backpropagation. Procedia Computer Science, 2015, 55: 990–999. [doi: 10.1016/j.procs.2015.07.104]
- Yadav RP, Senthilarasu V, Kutty K, et al. Implementation of robust HOG-SVM based pedestrian classification. International Journal of Computer Applications, 2015, 114(19): 10–16. [doi: 10.5120/20084-2026]
- Cao H, Yamaguchi K, Naito T, et al. Pedestrian recognition using second-order HOG feature. Proceedings of the 9th Asian conference on Computer Vision. Xi'an, China. 2009. 628–634.
- 卢伟明. 基于单目视觉的道路检测与跟踪算法研究[硕士学位论文]. 长沙: 湖南大学, 2018.
- Bengio Y, Lamblin P, Popovici D, et al. Greedy layer-wise training of deep networks. Proceedings of the 19th International Conference on Neural Information Processing Systems. Canada. 2006. 153–160.
- Ruder S. An overview of gradient descent optimization algorithms. arXiv: 1609.04747, 2016.
- 刘方圆, 王水花, 张煜东. 深度置信网络模型及应用研究综述. 计算机工程与应用, 2018, 54(1): 11–18, 47. [doi: 10.3778/j.issn.1002-8331.1711-0028]
- 刘凯, 张立民, 孙永威. 基于遗传算法的 RBM 优化设计. 微电子学与计算机, 2015, 32(6): 96–100.
- 王琳琳, 刘敬浩, 付晓梅. 融合局部特征与深度置信网络的人脸表情识别. 激光与光电子学进展, 2018, 55(1): 011002.
- 曾志, 吴财贵, 唐权华, 等. 基于多特征融合和深度学习的商品图像分类. 计算机工程与设计, 2017, 38(11): 3093–3098.
- 刘斌, 赵兴, 胡春海, 等. 面向颜色深度图像手脸近距遮挡的手势识别. 激光与光电子学进展, 2016, 53(6): 061001.
- 刘德雨. 基于深度学习的行人检测技术研究[硕士学位论文]. 长春: 长春工业大学, 2018.