基于优化卷积深度信念网络的智能手机身份认证方法

张义超[1]，孙子文[1,2*]

[1]江南大学物联网工程学院，江苏 无锡 214122；

[2]物联网技术应用教育部工程研究中心，江苏 无锡 214122

摘要 针对智能手机面临的信息安全问题，研究了一种优化卷积深度信念网络的智能手机身份认证方法。先对采集的原始数据进行预处理，再引入稀疏自编码器进行预训练，预训练后的权重作为卷积深度信念网络模型的卷积核，选用逐层贪婪算法用于模型的正式训练；训练后，经均方根连接层对提取的特征进行整合，并利用监督学习算法调节均方根连接层与输出层之间的权重；最后，由 Softmax 分类器输出分类结果。该方法可直接处理高维手势数据，建立手势模型进行特征提取。仿真结果表明，与隐马尔科夫算法、深度信念网络算法相比，该方法可显著提高身份认证的准确率。

关键词 图像处理；稀疏自编码器；卷积深度信念网络；均方根连接层；Softmax 分类器

中图分类号 TP391 **文献标志码** A **doi**：10.3788/LOP57.081009

Identity Authentication for Smart Phones Based on an Optimized Convolutional Deep Belief Network

Zhang Yichao[1], Sun Ziwen[1,2*]

[1] *School of Internet of Things, Jiangnan University, Wuxi, Jiangsu* 214122, *China*;

[2] *Engineering Research Center of Internet of Things Technology Applications of Ministry of Education, Wuxi, Jiangsu* 214122, *China*

Abstract In this study, we propose an intelligent identity authentication method for an optimized convolutional deep belief network to address the information security problem faced by smart phones. First, the collected raw data is preprocessed and then input into the sparse autoencoder for pretraining. The pretrained weight is used as the convolution kernel of convolutional deep belief networks, and the layer-by-layer greedy algorithm is adopted to formally train the model. Subsequent to the training, the extracted features are integrated with the root mean square layer, and the weight between the root mean square layer and the output layer is adjusted using the supervised learning algorithm. Finally, the classification results are output through the Softmax classifier. The proposed method can directly process high-dimensional gesture data and establish a gesture model for feature extraction. Simulation results show that compared with the hidden Markov algorithm and the deep belief network algorithm, the proposed method can significantly improve the accuracy of identity authentication.

Key words image processing; sparse autoencoder; convolutional deep belief network; root mean square layer; Softmax classifier

OCIS codes 100.4996; 120.1880; 150.1135; 200.4260

1 引 言

随着网络技术和电子商务的快速发展，智能手机和平板计算机等移动终端设备的应用越来越广泛。日常生活中，智能手机不仅能储存文件和图片等私密信息，还新增网络购物和手机转账等功能，如

收稿日期：2019-07-23；**修回日期**：2019-08-24；**录用日期**：2019-09-11

基金项目：国家自然科学基金(61373126)、中央高校基本科研业务费专项资金(JUSRP51510)、江苏省自然科学基金(BK20131107)

* **E-mail**：sunziwen@jiangnan.edu.cn

果被非法使用者入侵，可能会导致个人信息泄露和通讯录资料公布，甚至财产丢失等后果[1]。因此，为了防止未经授权的用户访问私人手机，使用手机前进行用户的身份认证显得愈发重要。

对于智能手机的安全防护，国内外学者大多集中在利用人体的生理和行为特征对智能手机建立身份认证系统[2]。Ganesh 等[3]采用基于安全手势的智能手势认证方案，结合手指压力和手机在解锁时的倾斜度构成双重认证系统，使用模糊推理系统(FIS)进行特征训练及分类，实现用户的认证过程，获得了较高的正确肯定率。刘颖[4]利用一组图像预处理算法实现指纹检测，由局部图像中的分叉数目及人工神经网络(ANN)进行分类。宋成等[5]通过智能手机内置的触摸屏传感器获取用户的手势特征并建立身份认证模板，选用动态时间规整(DTW)算法对采集的手势特征数据模板进行相似度匹配，最后由分类器投票原则匹配用户身份。Miao 等[6]引入深度信念网络(DBN)解决基于惯性传感器的手势识别问题，根据实验性能调整 DBN 的最优体系结构和超参数，在短时间内获得较高的识别精度，并在北京航空航天大学(BUAA)移动手势数据库中获得满意结果。

上述身份认证方法均在一定程度上保护用户隐私，但仍存在一些不可避免的问题。传统个人身份认证识别码(PIN)识别的优缺点自相矛盾，一方面简单密码易被木马或其他恶意程序窃取，另一方面复杂密码很难记忆且操作繁琐；指纹识别准确率高且识别速度快，但需要额外或高要求的硬件，对指纹的清晰度、手指的湿度和清洁度等均有不同程度要求；基于 DBN 的手势识别对硬件要求低，且正确率较高，可用于中长期预测，但无法直接处理高维输入问题，模型的泛化能力较低[7]。

针对智能手机的信息安全问题，本文研究一种基于优化卷积深度信念网络(CDBN)[8]的智能手机手势身份认证方法。先采集基于人体行为难以模仿和复制的动态手势数据，并进行平滑去噪及归一化的预处理；再构建基于手势数据的 CDBN 模型并进行训练；训练结束后经输出认证模型的均方根(RMS)连接层[9]对提取的手势特征进行整合，选用监督学习算法调节 RMS 连接层与输出层之间的权重，最后经 Softmax 分类器判决测试手势数据的合法身份。

2 手势身份认证理论与方法

在 DBN 固有结构的基础上加入卷积神经网络(CNN)[10]，构成 CDBN 结构，图 1 为手势身份认证的总体框架。在训练阶段，先对移动终端采集的原始数据进行预处理，再传入 CDBN 模型进行训练(步骤①)，其中模型的卷积核为稀疏自编码器(SAE)[11]预训练后固定的权重，训练结束后经 RMS 连接层对提取特征进行整合(步骤②)，并选用监督学习算法调整 RMS 连接层与输出层之间的权重(步骤③)，固定手势身份认证的模型参数；在测试阶段，预处理的测试数据传入已固定的 CDBN 模型计算输出(步骤④)，并由 RMS 连接层整合(步骤⑤)，最后经 Softmax 分类器检测用户为真实或假冒。本认证方法的性能评估可通过身份认证的准确率(ACC)、错误拒绝率(FRR)、错误接受率(FAR)实现[12]。

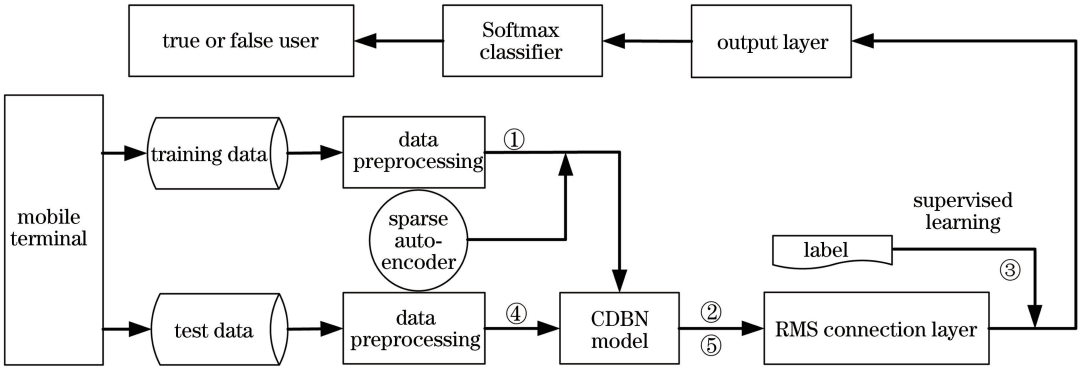

图 1 手势身份认证的总体框架

Fig. 1 Overall framework of gesture identity authentication

2.1 数据预处理

原始数据采集阶段，移动终端的触摸屏传感器用于收集用户的触摸手势，原始手势数据序列为

$$\bar{\boldsymbol{A}}^k=\{(\bar{x}_1^k,\bar{y}_1^k,\bar{z}_1^k),\cdots,(\bar{x}_l^k,\bar{y}_l^k,\bar{z}_l^k),\cdots,(\bar{x}_{n_k}^k,\bar{y}_{n_k}^k,\bar{z}_{n_k}^k)\},\tag{1}$$

式中：$\bar{x}_l^k$、$\bar{y}_l^k$、$\bar{z}_l^k$ 分别为第 k 次运动手势第 l 个数据

点 x 轴坐标值、y 轴坐标值以及指尖压力值 z；n_k 为第 k 次手势样本点的总点数，满足 $1\leqslant l\leqslant n_k$（$k$，$l$，$n_k\in\mathbf{Z}$）。

采集数据过程中，由于手势动作前后一致性，不可避免地存在用户手指不同幅度抖动等客观因素的干扰，使得采集的部分样本存在较大误差。为了去除原始数据中无关因素的干扰，保证数据质量和检测的准确性，需对原始数据进行预处理。

采集的原始手势数据经过滤处理，删除短距离滑动和往复滑动两种异常操作类型数据[13]，对保留的手势数据进行平滑去噪、手势序列平移及归一化的预处理。

2.1.1 平滑去噪处理

为了去除干扰噪声的影响，选用滑动平均滤波器对数据序列进行平滑去噪处理[14]，以第 k 次采样手势第 l 个数据点经平滑去噪处理后的 x 轴坐标值为例，去噪表达式为

$$\tilde{x}_l^k=\sum_{m=0}^{D-1}\bar{x}_{l-m}^k/D,\tag{2}$$

式中：$\bar{x}_{l-m}^k$ 为第 k 次运动手势第 $l-m$ 个数据点的 x 轴坐标值；D 为取平均值点的个数。同理可得平滑去噪处理后的 y 轴坐标值 $\tilde{y}_l^k$ 和指尖压力值 $\tilde{z}_l^k$。

2.1.2 手势序列平移及归一化处理

滑动手势的过程中，手势书写位置和滑动轨迹大小均具有不确定性，可选用离差标准化对手势进行归一化处理消除其干扰[15]。以第 k 次手势第 l 个数据点为例，采集数据的压力值 $z\in[0,1]$，故只需对 x、y 轴坐标值进行归一化处理，归一化后的 x 轴坐标值 x_l^k 为

$$x_l^k=\frac{\hat{x}_l^k-\min\limits_{1\leqslant l\leqslant n_k}(\hat{x}^k)}{\max\limits_{1\leqslant l\leqslant n_k}(\hat{x}^k)-\min\limits_{1\leqslant l\leqslant n_k}(\hat{x}^k)},\tag{3}$$

式中：$\hat{x}_l^k=\tilde{x}_l^k-\sum\limits_{l=1}^{n_k}\tilde{x}_l^k/n_k$，$\hat{x}_l^k$ 为第 k 次手势第 l 个采样点坐标平移后 x 轴的坐标值。同理可得归一化后的 y 轴坐标值 y_l^k。

每次滑动手势轨迹采集的数据点数不一定相同，预处理后，取所有手势中采集数据点数的平均值 n，即

$$n=\mathrm{mean}(n_1,n_2,\cdots,n_k)\tag{4}$$

用于手势特征提取，少于平均值点数时进行补 0 处理。式中：n_1，n_2，…，n_k 分别为第 1 次，2 次，…，k 次手势的总数据点数。

第 k 次原始手势数据序列经预处理后可表示为

$$\boldsymbol{A}^k=\{(x_1^k,y_1^k,z_1^k),\cdots,(x_l^k,y_l^k,z_l^k),\cdots,(x_n^k,y_n^k,z_n^k)\}。\tag{5}$$

手势认证模型分为 CDBN 模型和输出认证模型两部分，其中 CDBN 模型用来提取手势特征，输出认证模型用来认证分类。

2.2 CDBN 模型

CDBN 模型的主体是 CDBN 结构，其中 CDBN 结构由多层卷积受限玻尔兹曼机（CRBM）堆叠组成，一般的 CDBN 结构和带有池化层的 CDBN 结构如图 2 所示。

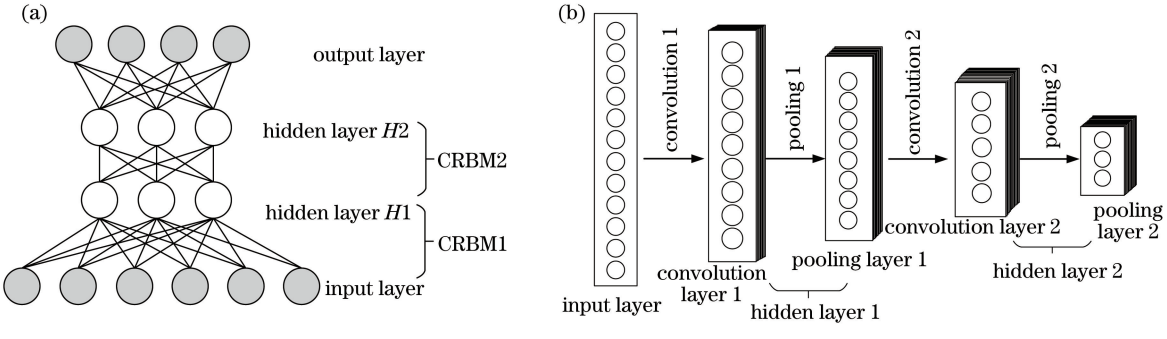

图 2 不同类型的 CDBN 模型结构。(a)一般结构；(b)带有池化层

Fig. 2 Different types of CDBN models structure. (a) General structure; (b) with pooling layer

2.2.1 CRBM 结构

1）由受限玻尔兹曼机（RBM）到 CRBM

RBM[16]是一种特殊类型的马尔可夫随机场，由随机可见层和随机隐藏层组成，结构如图 3(a)所示。CRBM 是在 RBM 结构的基础上新增池化层（最大概率层），结构如图 3(b)所示。

CRBM 由输入层 V 和隐藏层 H 组成，隐藏层 H 又进一步分为卷积层 C 和池化层 P。输入层的输入为数据预处理后的 $N_v\times N_v$ 手势序列，隐藏层有 M 组，均为 $N_H\times N_H$ 的高斯变量，共有 N_H^2M 个隐藏层神经元，每组隐藏层与大小为 $N_w\times N_w$ 的卷积核相连接（其中 $N_H=N_v-N_w+1$）。池化

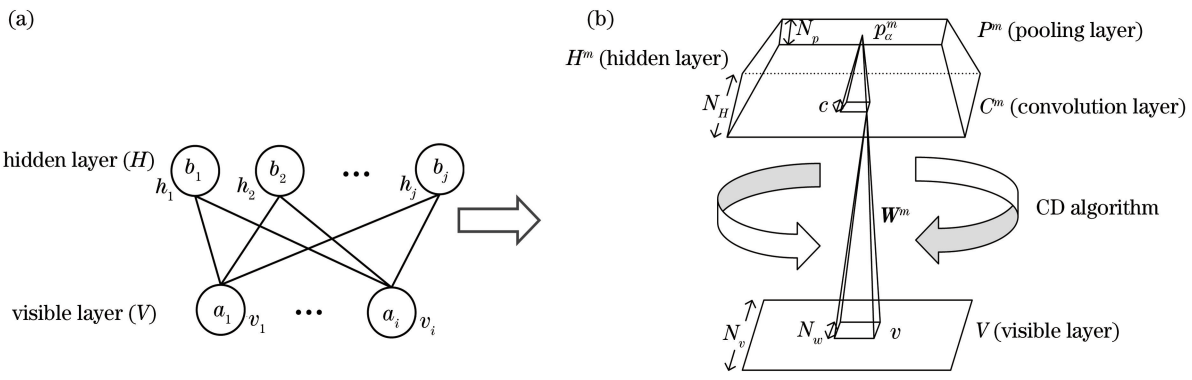

图 3 不同模型结构。(a) RBM 结构;(b) CRBM 结构

Fig. 3 Different model structures. (a) RBM structure; (b) CRBM structure

层每组有 $N_p \times N_p$ 个二值矩阵,对于每个 $m \in \{1,2,\cdots,M\}$,池化层 P^m 是通过每个 $c \times c$ 维的卷积层 C^m 降维获得,每个卷积层 C^m 被分成大小为 $c \times c$ 的块,每个块 α 恰好与池化层的一个高斯单元 p_α^m 连接(其中 $N_p = N_H / c$)。CRBM 模型的特点为局部感受野和权值共享[17],即隐藏层与可见层之间是局部连接,每个组内所有隐藏层神经元的卷积核权值共享。此外,每组隐藏层的偏置量为 $\boldsymbol{b}_m$,且所有可见单元共享偏置量为 $\boldsymbol{a}$,可见层与池化层的第 m 组卷积核为 $\boldsymbol{W}^m$。

2) 池化层

由于隐藏层的输出数据量较大,且存在冗余信息,直接学习参数使得网络复杂度高,增大计算量及内存消耗,引入池化层后,可将隐藏层输出的数据降维、去除冗余信息,而且身份认证的准确率基本不会受到影响。常用的池化降维方法有最大值池化和均值池化等[18]。

最大值池化的原理是在卷积层 C^m 中依步长取每个 $c \times c$ 块的最大值。假设 $h_1, h_2, \cdots, h_{16}$ 数值依次增大,选用 2×2 的池化块进行池化降采样,步长为 2,其前向池化和反向重构运算过程如图 4 所示。

前向传播时对每个不重叠的 2×2 区域进行降采样,选出每个区域中的最大值作为输出,并记录最大值在每个小区域中的位置。池化层进行反向重构时,将最大值及对应位置向前一层传递,其他位置均置零,其中 4×4 矩阵中非零的位置即为前向运算所得每个小区域最大值的位置。其中 h_j 为隐藏层神经元的输出值。

均值池化与最大值池化方法相似,不同点为均值池化时卷积层 C^m 依次取每个 $c \times c$ 区域块的均值,且反向重构时 2×2 区域内四个元素均为该区域前向池化时的均值。

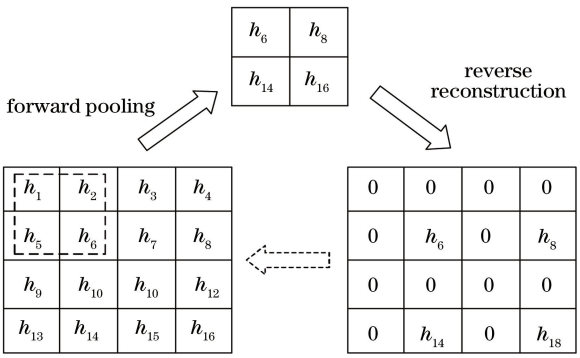

图 4 最大值池化运算

Fig. 4 Maximum pooling operation

2.2.2 CDBN 的训练

对于高维手势数据的输入,选用多个卷积核进行特征提取,使得隐藏层节点数大于输入层节点数,但并不是每一组隐藏层的每个隐藏节点都反映了输入层的特征。引入 SAE,对隐藏层节点加入稀疏性限制,经 SAE 训练后的权重作为 CDBN 的卷积核进行训练,训练后明显提高身份认证模型的泛化能力。

1) 引入 SAE

对于一般的自编码器神经网络,假设有 N 个训练样本,用 v 表示样本输入,$\tilde{v}$ 表示输出层期望的输出,可表示为 $\{(v^{(1)},\tilde{v}^{(1)}),\cdots,(v^{(d)},\tilde{v}^{(d)}),\cdots,(v^{(N)},\tilde{v}^{(N)})\}$。

对于某个样本,其损失函数[10]为

$$J[\boldsymbol{W},\boldsymbol{b};v^{(d)},\tilde{v}^{(d)}] = \| \tilde{v}^{(d)} - h_{\boldsymbol{W},\boldsymbol{b}}(v)^{(d)} \|^2/2, \tag{6}$$

式中:$h_{\boldsymbol{W},\boldsymbol{b}}(v)^{(d)}$ 为第 d 个样本的输出层实际输出值。

对于 N 个训练样本,其损失函数可表示为

$$J(\boldsymbol{W};\boldsymbol{b})=\sum_{d=1}^{N}\{J[\boldsymbol{W},\boldsymbol{b};v^{(d)},\tilde{v}^{(d)}]\}/N+\lambda\sum_{g=1}^{n_g-1}\sum_{i=1}^{s_g}\sum_{j=1}^{s_{g+1}}[\boldsymbol{W}_{ji}^{(g)}]^2/2=\sum_{d=1}^{N}[\|\tilde{v}^{(d)}-h_{\boldsymbol{W},\boldsymbol{b}}(v)^{(d)}\|^2/2]/N+\lambda\sum_{g=1}^{n_g-1}\sum_{i=1}^{s_g}\sum_{j=1}^{s_{g+1}}[\boldsymbol{W}_{ji}^{(g)}]^2/2,\tag{7}$$

式中：n_g 为输出层；s_g 为第 g 层隐藏层神经元的个数；$\boldsymbol{W}_{ji}^{(g)}$ 为第 g 层与第 $g+1$ 层的连接权重；λ 为权重衰减系数，用来控制式中前后两项在整个优化目标中所占的比重。第一项表示所有样本误差的均值；第二项表示归一化项（又称权重衰减项），目的是为了降低连接层权重的更新速度，防止过拟合。

将稀疏性限制加入自编码神经网络中，当给定输入 v 时，假定 $h_j^{(g)}$ 表示 SAE 第 g 层第 j 个隐藏层神经元的响应，则可定义隐藏层神经元 j 在 N 个训练样本集合上的平均活跃程度，即

$$\hat{\rho}_j=\sum_{d=1}^{N}[h_j^{(g)}v^{(d)}]/N。\tag{8}$$

令 ρ 表示稀疏性参数，一般取值很小（比如 0.02）。引入 K_{L} 离散度来衡量某个隐藏层节点的平均活跃程度与设定的稀疏度 ρ 之间的相似性，即

$$K_{\mathrm{L}}(\rho\|\hat{\rho}_j)=\rho\log\frac{\rho}{\hat{\rho}_j}+(1-\rho)\log\left(\frac{1-\rho}{1-\hat{\rho}_j}\right)。\tag{9}$$

K_{L} 离散度相当于额外的惩罚项，K_{L} 离散度值越大代表 ρ 和 $\hat{\rho}_j$ 之间相差越大，K_{L} 离散度等于 0 代表两者完全相等，即 $\rho=\hat{\rho}_j$。将这个稀疏限制惩罚因子加入总的损失函数中，得到

$$J_{\mathrm{sparse}}(\boldsymbol{W},\boldsymbol{b})=J(\boldsymbol{W};\boldsymbol{b})+\beta\sum_{j=1}^{s_g}\left[\rho\log\frac{\rho}{\hat{\rho}_j}+(1-\rho)\log\left(\frac{1-\rho}{1-\hat{\rho}_j}\right)\right],\tag{10}$$

式中：β 为稀疏惩罚项的权重参数。

选用梯度下降法学习损失函数，不断更新模型的权重和偏置，使得 $J_{\mathrm{sparse}}(\boldsymbol{W},\boldsymbol{b})$ 达到最小值。利用无监督的稀疏自编码器对预处理后的手势数据进行预训练，训练结束后，将权值作为 CDBN 的卷积核进行正式训练，可提高 CDBN 模型的泛化能力。

2）CRBM 的对比散度（CD）算法

假定输入是由 0、1 组成的二值矩阵，为了更好地反映系统稳定性，对于一组给定的状态 (v,h)，定义能量函数[19]为

$$E(v,h)=-\sum_{m=1}^{M}\sum_{i,j=1}^{N_H}\sum_{r,s=1}^{N_w}h_{i,j}^{m}\boldsymbol{W}_{r,s}^{m}v_{i+r-1,j+s-1}-\sum_{m=1}^{M}\boldsymbol{b}_m\sum_{i,j=1}^{N_H}h_{i,j}^{m}-\boldsymbol{a}\sum_{i,j=1}^{N_v}v_{i,j},\tag{11}$$

式中：$h_{i,j}^{m}$ 为隐藏层 H 中第 m 个特征图节点 (i,j) 的值；$\boldsymbol{W}_{r,s}^{m}$ 为与输入层相连的第 m 个卷积核节点 (r,s) 的值；$v_{i+r-1,j+s-1}$ 为输入层 V 中节点 $(i+r-1,j+s-1)$ 的值。将(11)式简化为

$$E(v,h)=-\sum_{m=1}^{M}h_{i,j}^{m}\cdot(\tilde{\boldsymbol{W}}^m*v)-\sum_{m=1}^{M}\boldsymbol{b}_m\sum_{i,j=1}^{N_H}h_{i,j}^{m}-\boldsymbol{a}\sum_{i,j=1}^{N_v}v_{i,j},\tag{12}$$

式中：$*$ 为卷积运算；$\cdot$ 为矩阵的乘法；$\tilde{\boldsymbol{W}}^m$ 为对矩阵 $\boldsymbol{W}^m$ 逆时针旋转 180°。

令模型的联合概率分布为 $P(v,h)=\exp[-E(v,h)]/Z$，其中 $Z=\sum_{v,h}\exp[-E(v,h)]$，由联合概率分布可得条件概率分布为[19]

$$P(h_{i,j}^{m}=1\,|\,v)=\sigma[(\tilde{\boldsymbol{W}}^m*v)_{i,j}+\boldsymbol{b}_m],\tag{13}$$

$$P(v_{i,j}=1\,|\,h)=\sigma\left[\left(\sum_m\boldsymbol{W}^m*h^m\right)_{i,j}+\boldsymbol{a}\right],\tag{14}$$

式中：$\sigma(u)=1/[1+\exp(-u)]$ 为激活函数，$u\in(-\infty,+\infty)$，$\sigma(u)\in(0,1)$。由条件概率分布计算 CRBM 训练后神经元的激活概率。

经稀疏自编码器处理后，权值作为 CDBN 的卷积核，进行逐层贪婪，每层 CRBM 选用对比散度（CD）算法[16]进行训练。为了尽量减少状态转移次数，缩短训练时间，将数据分批进行训练，每一批数据训练完成后改变一次权重和偏置量，详细的方法及步骤如下。

1）模型参数初始化。

选用 M 组卷积核，令模型参数 $\beta_1=(Q_1,a,b_1)$，其中 Q_1 为第一层 CRBM 的卷积核，卷积核为稀疏自编码器训练后的权重，设可见层和隐藏层的起始偏置量均为 0。

2）单步吉布斯采样。

经预处理的原始数据作为输入，确定模型的卷积核和学习率等参数，采用单步吉布斯采样进行特征提取，提取过程如图 5 所示。

每一张特征图相当于一个卷积核，代表输入手势数据序列不同位置的一个局部特征，即每一张特征图代表一种特征。因此，不同特征图的隐藏层单元分别表示不同位置的可见层手势数据序列的不同特征。

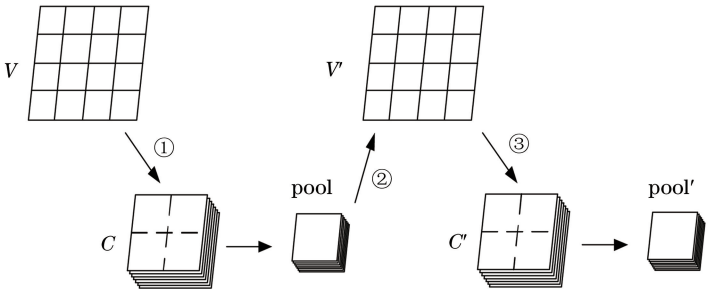

图 5 单步吉布斯采样

Fig. 5 Single step Gibbs sampling

①前向计算卷积层和池化层的输出。

设 $\boldsymbol{v}_{\text{init}}$ 表示初始可见层 V 的输入，$\boldsymbol{c}_{\text{init\_input}}$ 表示前向传输时卷积层 C 的输出，即

$$\boldsymbol{c}^{m}_{\text{init\_input}}=(\boldsymbol{v}_{\text{init}}*\boldsymbol{W}^{m})+\boldsymbol{b}_{m}。\tag{15}$$

选用 $c\times c$ 维矩阵对卷积层的输出进行最大值或均值池化处理，池化方法见 3.1.2 节。池化后的矩阵为

$$\boldsymbol{B}^{m}_{\text{init}}=\text{pool}(\boldsymbol{c}^{m}_{\text{init\_input}}),\tag{16}$$

式中：pool 为对卷积层的输出进行池化处理；$\boldsymbol{B}^{m}_{\text{init}}$ 为第 m 组卷积核池化层的初始输出。将池化前的输入与池化后的初始输出进行最大概率计算，得到池化层的最终输出，即

$$\boldsymbol{B}^{m}_{\text{init\_out}}=\exp(\boldsymbol{c}^{m}_{\text{init\_input}})/(1+\boldsymbol{B}^{m}_{\text{init}}),\tag{17}$$

式中：$\boldsymbol{B}^{m}_{\text{init\_out}}$ 为第 m 组卷积核池化层的最终输出。

②重构可见层的输入。

由激活函数 $\sigma(u)=1/[1+\exp(-u)]$ 计算隐藏层神经元的激活状态，被激活的概率为

$$h^{m}_{\sigma}=\sigma(\boldsymbol{B}^{m}_{\text{init\_out}})=1/[1+\exp(\boldsymbol{B}^{m}_{\text{init\_out}})],\tag{18}$$

式中：h^{m}_{σ} 为第 m 组卷积核隐藏层神经元被激活的概率。由隐藏层神经元被激活的概率重构可见层的输入为

$$\boldsymbol{v}_{\text{re\_input}}=(h^{m}_{\sigma}\otimes\tilde{\boldsymbol{W}}^{m})+\boldsymbol{a},\tag{19}$$

式中：⊗为反卷积运算。

③重构卷积层和池化层的输出。

由重构的可见层输入继续重构卷积层和池化层的输出，利用(15)式计算重构后卷积层的输出为 $\boldsymbol{c}^{m}_{\text{re\_input}}$，经池化和最大概率计算后，池化层的最终输出为 $\boldsymbol{B}^{m}_{\text{re\_out}}$。

④更新模型参数。

重构可见层输入 V 和隐藏层输出 H 后，模型的参数 β_1 可根据

$$\Delta\boldsymbol{W}^{m}=\delta(\boldsymbol{v}_{\text{init}}*\boldsymbol{B}^{m}_{\text{init\_out}}-\boldsymbol{v}_{\text{re\_input}}*\boldsymbol{B}^{m}_{\text{re\_out}}),\tag{20}$$

$$\Delta\boldsymbol{a}=\delta(\boldsymbol{v}_{\text{init}}-\boldsymbol{v}_{\text{re\_input}}),\tag{21}$$

$$\Delta\boldsymbol{b}^{m}=\delta(\boldsymbol{B}^{m}_{\text{init\_out}}-\boldsymbol{B}^{m}_{\text{re\_out}}),\tag{22}$$

进行更新。式中：δ 为学习率；$\Delta\boldsymbol{W}^{m}$、$\Delta\boldsymbol{a}$ 和 $\Delta\boldsymbol{b}^{m}$ 分别为第 m 个卷积核值、可见层向量偏置量和第 m 个卷积核对应偏置量的更新值。

3) 结束 CRBM 训练，输出底层 CRBM 模型参数，即 $\beta_1=(Q_1,a,b_1)$。

逐层贪婪学习算法利用 CD 算法先训练底层 CRBM，固定模型参数，再将底层 CRBM 训练后池化层的输出作为较高层 CRBM 训练的可见层的输入，继续执行 CD 算法，且较高层的训练不会影响底层已固定的模型参数，直至所有 CRBM 训练结束。

模型参数初始化后，通过循环迭代步骤 2)进行参数更新，设最大迭代更新次数为 T。当迭代更新次数小于 T 时，将更新后模型参数作为初始值，重复步骤①～④；如果等于 T，则执行步骤 3)。

2.3 输出认证模型

CDBN 训练结束后，将每层 CRBM 池化层的输出连接 RMS 连接层，训练阶段经监督学习算法调整 RMS 连接层与分类器层之间的权重，验证和测试阶段由固定的 CDBN 参数计算输出后经 RMS 连接层特征整合，再经固定的认证权重到达分类器层，最后由 Softmax 分类器输出分类结果，输出认证模型结构如图 6 所示。

2.3.1 RMS 连接层

手势数据传入可见层后，经 CRBM 的卷积层和池化层进行特征提取，不同的卷积核提取到不同的局部特征，之后需将提取的特征进行重组整合才能进行认证测试，且所选整合方法不同实验效果不同。传统的全连接(FC)层[9]将多次卷积和池化后高度抽象化的特征直接进行整合，再利用监督学习算法调节全连接层与输出层之间的权重，经分类器得到真实、假冒用户的概率，但直接将抽象后的特征连接至全连接层会造成参数太多，往往存在冗余特征，且耗费过多功率。将提取的特征经 RMS 连接层处理，选用全局平均池化(GAP)方法进行特征提取，直

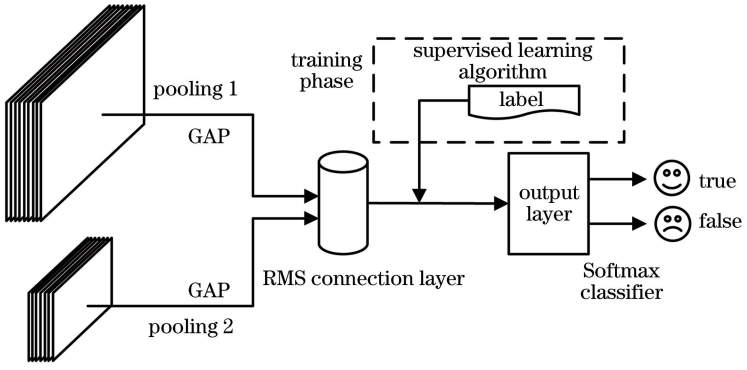

图 6 输出认证模型的结构

Fig. 6 Structure diagram of the output authentication model

接实现降维，更重要的是极大地减少网络参数（全连接层参数是 CRBN 中占比最大的参数）。

图 6 是将两层 CRBM 池化层的输出进行特征整合，对池化层输出的每一张特征图进行 GAP 处理，即对每张特征图的大量零散高斯特征值取 RMS，其运算公式为

$$f(m)=\sqrt{\frac{(u_1^m)^2+(u_2^m)^2+\cdots+(u_q^m)^2}{q}},\quad(23)$$

式中：$f(m)$为第 m 张特征图的 GAP 值；u^m 为第 m 张特征图的高斯特征值。

经 GAP 处理后，原有的高维特征矩阵整合为一维高斯数值，大大减少网络参数，缩短训练时间。将 RMS 连接层整合后的手势特征传入输出层。

2.3.2 监督学习算法

经 RMS 连接层整合特征后，仍无法判断训练的模型参数符合哪种用户的特征，故引入反向传播（BP）算法[20]，利用用户的标签信息进行监督学习，调节 RMS 连接层与输出层之间的权重，以确定模型用户类别。

BP 算法的原理是利用输出层的误差估计该输出层的直接前导层的误差，再用该误差继续估计更前层的误差，一层一层反传下去，即可获得每一层的误差估计值，通过不断更新模型的参数值减小实际输出值与期望输出值的误差。

损失函数表示实际输出与期望输出的相差程度，衡量模型预测的好坏，实验选用均方根误差损失函数，其表达式为

$$C=\sum_{d=1}^{N}\sum_{f=1}^{2}\|\mathrm{ex}_f^d-\mathrm{ao}_f^d\|^2/2N,\quad(24)$$

式中：ex_f^d 为第 d 组数据的第 f 个输出节点期望输出；ao_f^d 为第 d 组数据的第 f 个输出节点实际输出。

当损失函数值不满足最小值条件时，执行 BP 算法，选用梯度下降法求解误差 C 的最小值，并更新模型之间的连接权重。当达到最小值条件后，固定 RMS 连接层与输出层的连接权重，即可进行认证分类。

2.3.3 Softmax 分类器

经监督学习算法固定 RMS 连接层与输出层之间的权重后，即可进行身份认证验证与测试。添加 Softmax 分类器，并选用 ACC、FRR 和 FAR 评估算法的准确度。其中，ACC 表示用户身份认证的准确率，FRR 表示真实用户被拒绝的概率，FAR 表示假冒用户被接受的概率，FRR、FAR 值越小，算法模型的识别精度越高。

Softmax 分类器的原理是将输出层的输出值转换为预测用户的概率值，再由概率值判断用户分类。由图 6 可知，输出层有两个神经元，经 Softmax 分类器转换为两类用户的概率值，转换公式为

$$P(z_f\mid V)=\frac{\exp(\mathrm{ao}_f)}{\exp(\mathrm{ao}_1)+\exp(\mathrm{ao}_2)},\quad(25)$$

式中：ao_f 为第 f 个神经元的输出值；$\exp(\mathrm{ao}_f)$表示对输出值取以 e 为底的指数；$P(z_f|V)$为第 f 类用户的概率值。

由输出概率值的大小，采用分类器投票原则即可判断样本用户为真实或假冒。

3 仿真及结果分析

3.1 数据集

选用 HUAWEI P10 手机作为移动终端设备采集原始数据，以 MATLAB R2016b、Microsoft Visual Studio2015 为仿真平台，选用 ACC、FRR 和 FAR 判断用户身份认证的准确率。

由 7 名不同用户采集 15 天所得数据建立实验数据库，每人每天采集 100 组，共 10500 组数据。

将样本分为训练集、测试集和验证集，三者所占比例为 8∶1∶1。其中，训练集用来训练身份认证模型，模拟拟合的数据样本；验证集用于调整模型的超参数；测试集用于认证模型准确率以及检测模型的泛化能力。普通参数为各连接层的权重及偏置量；超参数包括网络深度（CRBM 的层数）、池化方法的选择、稀疏性指数、迭代次数及连接层的选择等。

在智能手机上采集原始触摸手势数据序列后，经应用程序接口（API）函数读取，进行数据过滤、平滑去噪及归一化预处理。令用户 1 为真实用户，其余用户为假冒用户。用户 1 原始手势轨迹与预处理后的手势轨迹如图 7 和图 8 所示。

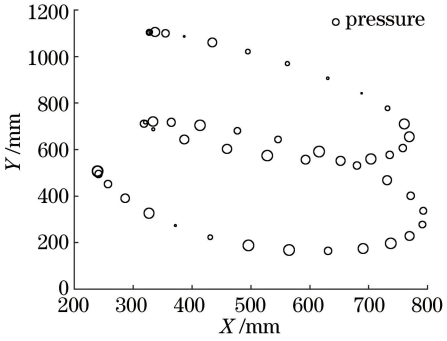

图 7 用户 1 原始运动轨迹

Fig. 7 User 1 original motion trajectory

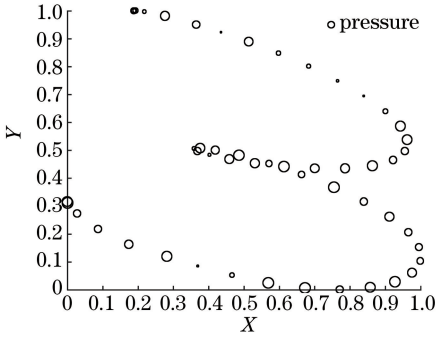

图 8 用户 1 预处理后的运行轨迹

Fig. 8 User 1 pre-processed trajectory

图 7 表示原始运动轨迹，图 8 表示经过滤、平滑去噪、归一化处理后的运动轨迹。图中“°”表示压力值，压力值越大，图中“°”的半径越大。

3.2 仿真过程

原始数据预处理后，将 8400 组训练数据的手势特征整理为 40×40×8400 的矩阵，传入稀疏自编码器进行预训练，当输出层的输出接近输入层的输入时，模型的权重为 7×7，选取 9 组权重值作为 CRBM 的卷积核。改变网络深度、池化方法、稀疏性指数、迭代次数及连接层方法等超参数，得到不同身份认证的准确率。

以选用网络深度为 2、最大值池化、稀疏性指数为 0.02、迭代次数为 10^{ϵ}（表示每层 CRBM 的迭代次数都是 10）及 RMS 连接层为例，训练及验证过程如下。

1）初始化模型的基本参数和超参数，分批传入训练数据中，每批 100 组，每一批数据执行完一次 CD 算法后，权重和偏置量更新一次。经稀疏自编码器预训练后的权重作为第一层 CRBM 的卷积核，令池化层区域块大小为 2×2，步长为 1，进行 CD 算法训练，当迭代次数为 10 时，结束 CD 算法，卷积层的输出为 34×34×9×8400，经池化降维后变为 17×17×9×8400。同理经稀疏自编码器预训练后，较高层 CRBM 的卷积核为 5×5，选取 16 组，将第一层 CRBM 的输出作为第二层 CRBM 的输入继续训练，训练结束后池化层的输出为 6×6×16×8400。

2）CDBN 训练结束后，将每层 CRBM 池化层的输出传入 RMS 连接层，进行 GAP 处理，输出矩阵为 25×8400。令真实用户的标签值为 1，假冒用户的标签值为 0，构建标签矩阵并执行反向传播算法，调整 RMS 连接层与输出层之间的连接权重，当损失函数取得最小值时，固定模型参数，结束身份认证训练。

3）训练结束后，预处理（与训练数据同一种预处理方法）后的验证数据传入 CDBN 模型，经固定模型参数输出 ACC、FRR 和 FAR 的值。结果表明，模型的泛化能力较好，可用于身份认证测试。

3.3 仿真结果及分析

将建立的模型进行横向和纵向对比分析，确定最优模型参数。

3.3.1 横向对比实验

预处理后假冒用户还原后的手势轨迹如图 9 所示，每一行表示还原一组用户的 10 次触摸手势，还原后各个用户的手势特征较明显。

以网络深度为 2、最大值池化、稀疏性指数为 0.02、迭代次数为 10^{ϵ} 及 RMS 连接层为标准，每组仿真仅改变一个超参数，其结果如表 1～5 所示。

1）网络深度

不同网络深度的仿真结果如表 1 所示，Time 表示模型的训练时间。当网络深度为 2 时，ACC 值最大，同时其对应的 FAR 和 FRR 值最小，表明并不是网络深度越深，模型越复杂，身份认证的效果越好，网络深度过高可能会造成过拟合现象。

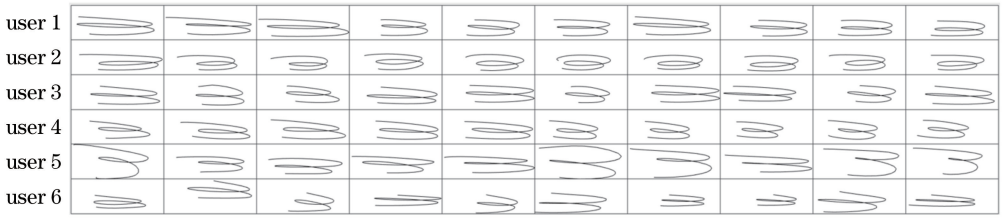

图 9 假冒用户的手势还原

Fig. 9 Fake user′s gesture recovery diagram

表 1 不同网络深度的仿真结果

Table 1 Simulation results of different network depths

Depth	ACC /%	FAR /%	FRR /%	Time /s
1	93.524	6.22	8.00	60.90
2	97.667	2.22	3.33	130.56
3	92.333	7.11	10.67	251.12

2）稀疏性指数

加入稀疏性限制，用少量神经元提取手势模型更本质的特征，同时降低网络复杂度。不同稀疏性指数的仿真结果如表 2 所示。当稀疏性指数过低时，激活神经元较少，不足以提取全部手势特征，导致模型欠拟合；当稀疏性指数过高时，训练数据过拟合，泛化能力较低，测试数据准确率下降。当稀疏性指数为 0.020时，ACC 值最大，同时 FRR 值最小。

表 2 不同稀疏性指数的仿真结果

Table 2 Simulation results of different sparsity indices

Index	ACC /%	FAR /%	FRR /%	Time /s
0.010	96.333	2.80	8.00	109.36
0.015	97.000	2.00	8.00	115.98
0.020	97.667	2.22	3.33	130.56
0.025	95.000	0	30.00	134.02
0.200	83.333	0	100.00	139.88

3）池化方法

不同池化方法的仿真结果如表 3 所示，其中 None、Mean、Max 分别对应选择无池化、平均值池化、最大值池化方法。手势特征经卷积层输出后，未经池化层反而增加模型整体的训练时间，且卷积后的输出用于模型测试时包含干扰数据，导致模型泛化能力较低。最大值池化和均值池化的 ACC 值、训练时间均接近，但均值池化需在池化块求和后取平均值，与最大值池化直接取池化块最大值相比时间复杂度更高，故最大值池化效果更好。

表 3 不同池化方法的仿真结果

Table 3 Simulation results of different pooling methods

Method	ACC /%	FAR /%	FRR /%	Time /s
None	97.333	2.00	6.00	400.31
Mean	97.652	2.33	3.35	130.77
Max	97.667	2.22	3.33	130.56

4）迭代次数

不同迭代次数的仿真结果如表 4 所示。当模型训练达到一定次数后，模型已达到饱和，如果继续训练，会产生过拟合，导致身份认证准确率下降。

表 4 不同迭代次数的仿真结果

Table 4 Simulation results of different iteration times

Epoch	ACC /%	FAR /%	FRR /%	Time /s
$5^{\in}$	96.000	0.80	20.00	70.33
$10^{\in}$	97.667	2.22	3.33	130.56
$25^{\in}$	96.000	2.80	10.00	375.61
$50^{\in}$	97.000	2.00	8.00	643.80

5）连接层的选择

不同连接层的仿真结果如表 5 所示。将各层池化后的抽象特征连接传统全连接层时，参数多且存在冗余特征，而连接 RMS 连接层时，参数少、运算快的同时，准确率高。

表 5 不同连接层的仿真结果

Table 5 Simulation results of different connection layers

Layer	ACC /%	FAR /%	FRR /%	Time /s
Fully	97.333	2.67	2.67	367.44
RMS	97.667	2.22	3.33	130.56

综上所述，同时考虑认证的准确率及训练所需时间，选用网络深度为 2、最大值池化、稀疏性指数为 0.02、迭代次数为 $10^{\in}$ 及 RMS 连接层的超参数时，系统的认证性能最好。

3.3.2 纵向对比实验

将 CDBN 算法与其他相近算法进行仿真比较。采集的数据经预处理后进行特征提取，选取手势轨迹 X 轴坐标、Y 轴坐标、滑动速度、加速度、点曲率及手势轨迹的一阶导数作为手势特征，分别传入 BP 算法、隐马尔可夫算法（HMM）和 DBN 算法，训练后建立手势身份认证模型。与优化 CDBN 算法相比，其性能差异如表 6 所示。与 HMM 算法、BP 算法、DBN 算法相比，提出的优化 CDBN 算法进行身份认证时，其准确率分别提高 5.007 个百分点、4.417个百分点、1.037 个百分点，虽然所用训练时间

有所提高，但幅度较小。

表 6 CDBN 算法与 BP、HMM、DBN 算法的性能比较

Table 6 Performance comparison among CDBN algorithm, BP, HMM, and DBN algorithms

Method	ACC /%	FAR /%	FRR /%	Time /s
BP	92.660	4.53	5.01	56.84
HMM	93.250	3.44	4.46	85.69
DBN	96.630	2.37	3.79	119.31
CDBN	97.667	2.22	3.33	130.56

BP 算法基于随机初始点的局部梯度信息，模型的训练时间较短，但准确率较低，常陷入局部最优点；HMM 算法相比于 BP 算法准确率提高，且训练时间增幅较小，但 HMM 只依赖于每一个状态和其对应的观察对象，相对于较长手势轨迹来说，范围较窄，使得身份认证准确率有限；DBN 算法的准确率较高，训练时间可观，但 DBN 模型不能直接处理高维输入数据，输入手势特征需降维处理后才能传入 DBN 模型，且 DBN 的初始权重由随机初始化得到，具有不确定性，不能反映数据的特征结构，导致预训练效果较差。在 DBN 的基础上引入 CNN 构成 CDBN 结构后，可直接选择高维特征量作为可见层的输入，SAE 训练后的权重用来初始化 CDBN 的卷积核，使网络各层权重具有较强的鲁棒性。由表 6 可知，模型的准确率提高 1 个百分点的同时，训练时间增幅较小。

综合横向和纵向对比结果可知，构建的身份认证模型可有效降低合法用户的 FAR 及非法用户的 FRR，提高身份认证的准确率。

4 结 论

研究了一种基于优化卷积深度信任网络模型的移动终端身份认证的新方法，在 DBN 结构的基础上，加入 CNN 和 SAE 部分，由模型提取手势特征，通过监督学习算法与无监督学习算法的结合，构造更加完备的身份认证网络模型。该方法与同类型的其他方法相比，一是可以解决多维输入问题，直接将采集的原始多维数据直接传入网络模型；二是选用稀疏自编码器初始化模型的卷积核，提高网络各层结构的鲁棒性，且正确率明显提高，但该模型仍有提升空间。仿真中固定卷积核个数，故无法判断选取的卷积核对准确率的影响，另外，所提方法只检测真实和假冒两种用户，更细致的判断规则为下一步继续研究的内容。

参 考 文 献

[1] Wang D, Cheng H B, He D B, et al. On the challenges in designing identity-based privacy-preserving authentication schemes for mobile devices [J]. IEEE Systems Journal, 2018, 12(1): 916-925.

[2] Meng Y X, Wong D S, Schlegel R, et al. Touch gestures based biometric authentication scheme for touchscreen mobile phones [M] // Kutyłowski M, Yung M. Information security and cryptology. Lecture notes in computer science. Berlin, Heidelberg: Springer, 2013, 7763: 331-350.

[3] Ganesh S M, Vijayakumar P, Deborah L J. A secure gesture based authentication scheme to unlock the smartphones [C] // 2017 Second International Conference on Recent Trends and Challenges in Computational Models (ICRTCCM), February 3-4, 2017, Tindivanam, Tamilnadu, India. New York: IEEE, 2017: 153-158.

[4] Liu Y. Study on the cost risk evaluation model of power transmission project based on bill-of-quantity model [J]. Automation & Instrumentation, 2018 (10): 23-26.
刘颖. 指纹识别算法的研究与改进[J]. 自动化与仪器仪表, 2018(10): 23-26.

[5] Song C, Zhang Y D, Wang L, et al. Trajectory privacy protection scheme based on DTW exchange query[J]. Journal of Beijing University of Posts and Telecommunications, 2018, 41(6): 97-102.
宋成, 张亚东, 王磊, 等. 基于 DTW 交换查询的轨迹隐私保护方案[J]. 北京邮电大学学报, 2018, 41(6): 97-102.

[6] Miao Y Q, Wang L N, Xie C Y, et al. Gesture recognition based on deep belief networks[M] // Zhou J, Wang Y H, Sun Z N, et al. Gesture recognition based on deep belief networks. Lecture notes in computer science. Cham: Springer, 2017, 10568: 439-446.

[7] Ma Y X, Hao Y X, Chen M, et al. Audio-visual emotion fusion (AVEF): a deep efficient weighted approach[J]. Information Fusion, 2019, 46: 184-192.

[8] Lee H, Grosse R, Ranganath R, et al. Convolutional deep belief networks for scalable unsupervised learning of hierarchical representations [C] // Proceedings of the 26th Annual International Conference on Machine Learning-ICML'09, June 14-18, 2009, Montreal, Quebec, Canada. New York: ACM, 2009: 609-616.

[9] Bell S, Zitnick C L, Bala K, et al. Inside-outside net: detecting objects in context with skip pooling and recurrent neural networks [C] // 2016 IEEE

Conference on Computer Vision and Pattern Recognition (CVPR), June 27-30, 2016, Las Vegas, NV, USA. New York: IEEE, 2016: 2874-2883.

[10] Tan G H, Hou J, Han Y P, et al. Low-parameter real-time image segmentation algorithm based on convolutional neural network [J]. Laser & Optoelectronics Progress, 2019, 56(9): 091003.
谭光鸿, 侯进, 韩雁鹏, 等. 基于卷积神经网络的低参数量实时图像分割算法[J]. 激光与光电子学进展, 2019, 56(9): 091003.

[11] Zhao R, Yan R Q, Chen Z H, et al. Deep learning and its applications to machine health monitoring[J]. Mechanical Systems and Signal Processing, 2019, 115: 213-237.

[12] Sun Z W, Zhang Y C. Deep belief network model for mobile terminal identity authentication[J]. Netinfo Security, 2019(3): 34-42.
孙子文, 张义超. 移动终端身份认证的深度信念网络模型[J]. 信息网络安全, 2019(3): 34-42.

[13] Rzecki K, Pławiak P, Niedzwiecki M, et al. Person recognition based on touch screen gestures using computational intelligence methods[J]. Information Sciences, 2017, 415/416: 70-84.

[14] Wang Y W, Lei H N, Bu M, et al. Distribution characteristics and identification of several typical blood cells under optical phase models[J]. Chinese Journal of Lasers, 2009, 36(10): 2629-2635.
王亚伟, 雷海娜, 卜敏, 等. 几种典型血细胞的光学相位模型及其分布特征与识别方法[J]. 中国激光, 2009, 36(10): 2629-2635.

[15] Khaw M W, Glimcher P W, Louie K. Normalized value coding explains dynamic adaptation in the human valuation process[J]. Proceedings of the National Academy of Sciences, 2017, 114(48): 12696-12701.

[16] Lee H, Pham P, Largman Y, et al. Unsupervised feature learning for audio classification using convolutional deep belief networks[C]// Advances in Neural Information Processing Systems, December 7-10, 2009, Vancouver, British Columbia, Canada. Canada: NIPS, 2009: 1096-1104.

[17] Norouzi M. Convolutional restricted Boltzmann machines for feature learning[D]. Vancouver: Simon Fraser University, 2009.

[18] He Z B. Singer identification using convolutional deep belief networks [D]. Guangzhou: South China University of Technology, 2015.
何灼彬. 基于卷积深度置信网络的歌手识别[D]. 广州: 华南理工大学, 2015.

[19] Norouzi M, Ranjbar M, Mori G. Stacks of convolutional restricted Boltzmann machines for shift-invariant feature learning[C]// 2009 IEEE Conference on Computer Vision and Pattern Recognition, June 20-25, 2009, Miami, FL, USA. New York: IEEE, 2009: 2735-2742.

[20] Gao H, Xue L Y. Back propagation neural network based on improved genetic algorithm fitting LED spectral model [J]. Laser & Optoelectronics Progress, 2017, 54(7): 072302.
高航, 薛凌云. 基于改进遗传算法的反向传播神经网络拟合 LED 光谱模型[J]. 激光与光电子学进展, 2017, 54(7): 072302.