

◆ 特约专栏 ◆

基于混合卷积自编码极限学习机的 RGB-D 物体识别

殷云华^{1,2}, 李会方¹

(1. 西北工业大学 电子信息学院, 陕西 西安 710072; 2. 瞬态冲击技术重点实验室, 北京 102202)

摘要: 有效学习丰富的表征信息在 RGB-D 目标识别任务中至关重要, 是实现高泛化性能的关键。针对卷积神经网络训练时间长的问题, 提出了一种混合卷积自编码极限学习机(HCAE-ELM)结构, 包括卷积神经网络(CNN)和自编码极限学习机(AE-ELM), 该结构合并了 CNN 的有效性和 AE-ELM 快速性的优点。它使用卷积层和池化层分别从 RGB 和深度图来有效提取低阶特征, 然后在共享层合并两种模型特征, 输入到自编码极限学习机中以得到高层次的特征, 最终的特征使用极限学习机(ELM)进行分类, 以获得更好的快速泛化能力。文中在标准的 RGB-D 数据集上进行了评估测试, 其实验结果表明, 相比较深度学习和其他的 ELM 方法, 文中的混合卷积自编码极限学习机模型取得了良好的测试准确率, 并且有效地缩减了训练时间。

关键词: 极限学习机; 卷积神经网络; 自编码极限学习机; 物体识别

中图分类号: TP391.4 **文献标志码:** A **DOI:** 10.3788/IRLA201847.0203008

RGB-D object recognition based on hybrid convolutional auto-encoder extreme learning machine

Yin Yunhua^{1,2}, Li Huifang¹

(1. School of Electronics and Information, Northwestern Polytechnical University, Xi'an 710072, China;

2. Science and Technology on Transient Impact Laboratory, Beijing 102202, China)

Abstract: Learning rich representations efficiently plays an important role in RGB-D object recognition task, which is crucial to achieve high generalization performance. For the long training time of convolutional neural networks, a Hybrid Convolutional Auto-Encoder Extreme Learning Machine Structure (HCAE-ELM) was put forward which included Convolutional Neural Network (CNN) and Auto-Encoder Extreme Learning Machine (AE-ELM), which combined the power of CNN and fast training of AE-ELM. It used convolution layers and pooling layers to effectively abstract lower level features from RGB and depth images separately. And then, the shared layer was developed by combining these features from each modality and fed to an AE-ELM for higher level features. The final abstracted features were fed to an ELM classifier, which led to better generalization performance with faster learning speed. The performance of HCAE-ELM was evaluated on RGB-D object dataset. Experimental results show that the

收稿日期: 2017-08-05; 修订日期: 2017-10-03

基金项目: 国家自然科学基金青年科学基金(61402368); 瞬态冲击技术重点实验室基金(61426060103162606007)

作者简介: 殷云华(1979-), 男, 博士生, 主要从事深度学习、物体识别方面的研究。Email: yinyunhua@nuc.edu.cn

导师简介: 李会方(1962-), 男, 教授, 博士生导师, 主要从事图像检测与识别、微电子技术及微电子学、量子信息处理等的研究。

Email: lhuifang@nwpu.edu.cn

proposed method achieves better testing accuracy with significantly shorter training time in comparison with deep learning methods and other ELM methods.

Key words: ELM; CNN; AE-ELM; object recognition

0 引言

物体识别是计算机视觉和人工智能领域中研究的热点,已广泛应用于军事、工业、医学等诸多领域。在实际应用中,物体识别易受光照变化、阴影、视角变化等外界因素的干扰,如何获取图像鲁棒性的特征是物体识别的关键。近年来,三维深度传感器的推出,为物体识别领域的研究提供了新的方向。特别地,以微软的 Kinect^[1]为代表的深度传感设备能快速地获取场景中物体的 RGB 图像和深度(Depth)图像。其中,RGB 图像提供物体表面的颜色和纹理等信息,而深度图像记录的是场景上各点与相机之间的距离信息,能直接反映物体表面的三维特征,是对 RGB 图像信息的有效补充。国内学者利用带深度信息的图像进行了手势识别^[2]、人体识别^[3]、物体识别^[4]等研究,均取得不错效果。基于 RGB-D(RGB 图像+深度图像)图像信息的物体识别已成为当前的研究热点。

1998 年 LeCun 等人构建了深度卷积模型 LeNet,卷积神经网络(CNN)能够从输入图像中自动完成特征提取,成功应用于图像处理的各个任务,因此 CNN 被认为是图像识别领域的主导算法^[5]。在过去的几年里,许多基于 RGB-D 的物体识别深度学习算法被提出。Blum 等人提出了卷积 K 均值描述符^[6],在兴趣点附近自动地学习特征并最终将这些特征进行融合。Socher 等人结合卷积神经网络和递归神经网络提出了 CNN-RNN 深度学习算法^[7],卷积神经网络层学习低维的平移不变性特征并作为 RNN 输入,递归神经网络层学习高维抽象特征。2012 年 Niu XX 等人提出的 CNN-SVM 模型^[8],结合了两种卷积神经网络和支持向量机分类器各自的优点,在图像分类试验中得到了较低的分类错误率。

卷积神经网络的特点主要体现于:(1) 它的神经元间的连接是非全连接的,通过局域感受野来实现,每一网络层的神经元只与上一层的一个小邻域内的

神经单元连接,每个神经元可以提取初级的视觉特征;(2) 同一层中某些神经元之间的权重是共享的。这两个特点使得其对图像平移、缩放、扭曲或其它形变具有高度不变性,同时神经元之间非全连接及权值的共享,减少了权值的数量,降低了网络的复杂度。

目前卷积神经网络在实际应用中取得一定的效果,但仍存在一些共性不足:当特征维数比较小时,难以提取有效的特征;训练过程繁琐困难,需要大量迭代计算,耗费时间。CNN^[9]的全连接相当于是一个一般的单隐层前馈神经网络,其采用的是 BP 算法^[10-11]进行训练,而 BP 算法存在容易陷入局部最小值和泛化性能不好等问题。

极限学习机(Extreme Learning Machine, ELM)是由黄广斌等人建立的一种快速的方法^[12-13],在模型训练的过程中没有调整参数这一步,要做的就是设定网络隐层节点的数量,而且模型还能保证最后得到的输出值是最优的,因此 ELM 不需要消耗过多的训练时间,学习效率高。相比于传统的基于梯度求解的反馈神经网络算法,ELM 具有以下优势:(1) 极限学习机的速度非常快,或者说不需要学习,只需要输出权重;(2) 极限学习机的泛化能力强;(3) 极限学习机一步到位直接构建单隐藏层反馈神经网络,避免了诸如局部最小、合适的学习率、过拟合等难以处理的棘手问题。

因此,为了进一步提高识别精确度和识别效率,文中提出了一种混合卷积自编码极限学习机(HCAE-ELM)学习模型。该模型结构结合了 CNN 的提取特征高效性和 ELM 分类快速性的优点。它使用卷积层和池化层分别从 RGB 和深度图来有效提取低阶特征,然后在共享层合并两种模型特征,输入到自编码极限学习机(AE-ELM)中以得到高层次的特征,最终的特征使用 ELM 进行分类,以获得更好的快速泛化能力。实验结果表明,相比较深度学习和其他的 ELM 方法,笔者的混合卷积自编码极限学习机模型取得了良好的测试准确率,并且大大缩减了训练时间。

1 相关工作

ELM 是一种简单的三层前馈神经网络,输入权值矩阵 W 和偏置向量 b 是随机生成的,只需要计算输出权值矩阵。如图 1 所示,ELM 是由输入层、中间隐层和输出层构成的三层网络结构。

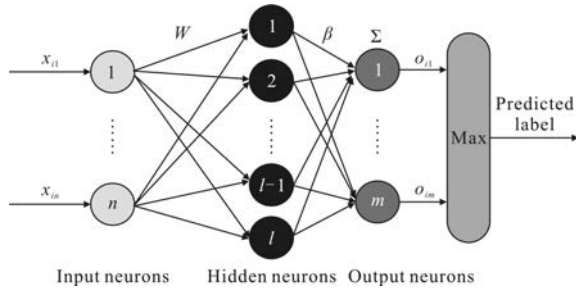


图 1 ELM 分类器结构

Fig.1 ELM classifier architecture

给定 N 个训练样本, $\{X, T\} = \{x_i, t_i\}_{i=1}^N$, 这里 $x_i \in R^n$ 和 $t_i \in R^m$ 分别是第 i 个输入向量和目标向量, 参数 n 和 m 分别输入向量和目标向量的维数。隐藏层节点数为 L , 激活函数为 $g(x)$ 的单隐层前馈神经网络数学模型为:

$$\sum_{i=1}^L \beta_i g(w_i^T \cdot x_j + b_i) = o_j, j=1, \dots, N \quad (1)$$

式中: $o_j \in R^m$ 是第 j 个训练样本的输出向量; $w_i \in R^n$ 是连接第 i 个隐藏节点的输入权重向量; b_i 是第 i 个隐藏节点的偏移量; β_i 是第 i 个隐层到输出层的权值, $g(w_i^T \cdot x_j + b_i)$ 是激活函数。

上式 N 个方程的矩阵形式为:

$$H\beta = T \quad (2)$$

式中: H 为网络的隐含层输出矩阵; β 为输出权重矩阵; T 为期望输出矩阵。

$$H(w_1, \dots, w_L, b_1, \dots, b_L, x_1, \dots, x_L) = \begin{bmatrix} g(w_1 \cdot x_1 + b_1) & \dots & g(w_L \cdot x_1 + b_L) \\ g(w_1 \cdot x_N + b_1) & \dots & g(w_L \cdot x_N + b_L) \end{bmatrix}_{N \times L} \quad (3)$$

$$\beta = \begin{bmatrix} \beta_1^T \\ \vdots \\ \beta_L^T \end{bmatrix}_{L \times m}, T = \begin{bmatrix} t_1^T \\ \vdots \\ t_L^T \end{bmatrix}_{N \times m} \quad (4)$$

在训练极限学习机网络时, 首先随机设置输入权值和偏置, 根据公式(3), 就可以求出确定的隐含层输出矩阵 H , 这样极限学习机的学习训练问题就转化成了一个求解输出权值矩阵的最小二乘范数解的问题, 也就是求解公式(2)的最小二乘范数解 $\hat{\beta}^{[13]}$ 。

$$H\beta = T \Rightarrow \hat{\beta} = H^+ T \quad (5)$$

式中: H^+ 为隐含层输出矩阵 H 的 Moore-Penrose 广义逆矩阵, 简称伪逆。用正交法来计算 H^+ 。

当 $H^T H$ 为非奇异矩阵时, $HH^+ = (H^T H)^{-1} H^T$, 当 HH^T 为非奇异矩阵时, $HH^+ = H^T (H^T H)^{-1}$ 。

为了提高泛化能力和结果的鲁棒性, 加入一个规则项。

$$\beta = \left(\frac{1}{c} + HH^T \right)^{-1} H^T T \quad (L > N) \quad (6)$$

$$\text{或 } \beta = H^T \left(\frac{1}{c} + HH^T \right)^{-1} T \quad (L \leq N) \quad (7)$$

2 算法设计

2.1 模型结构

卷积神经网络的局部感受野和权值共享使之更类似于生物神经网络, 减少了网络模型参数数量, 降低了模型复杂度, 由于图像的局部感受区域允许神经元或处理单元可以访问到最基础的如定向边缘或角点等特征, 所以卷积神经网络对平移缩放和旋转及其他形式的变形具有高度不变性^[14]。尽管 CNN 在提取图像低阶特征方面具有无法比拟的优势, 但由于 CNN 采用梯度下降法来调整各层的权值, 因此训练时需要耗费大量的时间。为了解决这个问题, 文中给出了如图 2 的模型结构, 分成三个独立的部分: (1) CNN 分别从 RGB 和深度图模型进行非监督低层次特征学习; (2) 自编码极限学习机进行高层次特征学习; (3) ELM 进行监督特征分类。

混合卷积自编码极限学习机分别从 RGB 和深度图模型中学习浅层和深层的特征信息。卷积神经网络将图像直接作为网络的输入, 通过卷积和池化操作来完成低层次的特征提取, 自编码极限学习机(AE-ELM)通过前馈神经网络来完成高层次的特征提取, 其隐藏层的输出作为 ELM 分类器的输入, 最后完成特征的标签分类。模型的所有参数均为随机初始化, 不需要进行调整。

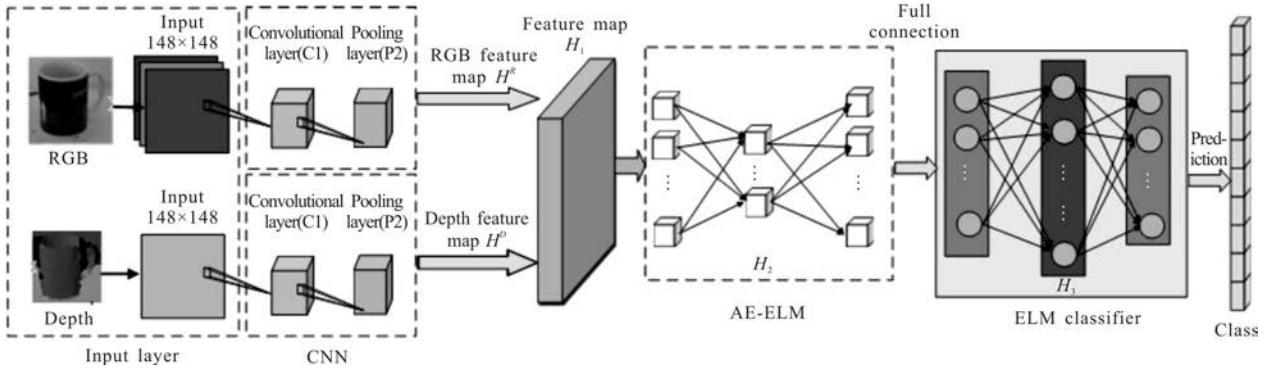


图 2 混合卷积自编码极限学习机模型图

Fig.2 Model diagram of HCAE-ELM

2.2 非监督 CNN 低阶特征提取

2.2.1 随机权值初始化

首先,对输入的 RGB 和深度图的照片进行尺寸处理,统一缩放至 $d \times d = 148 \times 148$ 。

根据连续型概率分布,HCAE-ELM 模型在输入层和卷积层之间随机生成权值,文中选用的是高斯概率函数来生成输入权值的采样分布。

高斯概率分布函数生成卷积层特征图的初始化权值矩阵 $\hat{A}_{init}^R, \hat{A}_{init}^D$, 计算如下:

$$\hat{A}_{init}^R, \hat{A}_{init}^D \in R^{r^2 \times k}$$

$$\hat{a}_k^R, \hat{a}_k^D \in R^r, k=1, 2, 3, \dots, K \quad (8)$$

式中: $r \times r$ 为局部接收域的尺寸, $d \times d$ 为输入图片片的尺寸,得到的特征图的尺寸为 $(d-r+1) \times (d-r+1)$ 。初始化的权值矩阵通过奇异值分解(SVD)方法进行正交化。

2.2.2 卷积

第 k 个特征图的输入权值为 $a_k^R, a_k^D \in R^{r \times r}$, 其正交化后为 $\hat{a}_k^R, \hat{a}_k^D \in R^r$ (列形式)。卷积层从输入的原始图像中通过卷积来提取特征图,在节点 (i, j) 处第 k 个特征图可以计算如下:

$$\begin{cases} C_{i,j,k}^{(R)}(x) = \sum_{m=1}^r \sum_{n=1}^r x_{i+m-1, j+n-1}^R \cdot a_{m,n,k}^R \\ C_{i,j,k}^{(D)}(x) = \sum_{m=1}^r \sum_{n=1}^r x_{i+m-1, j+n-1}^D \cdot a_{m,n,k}^D \end{cases} \quad i, j=1, \dots, (d-r+1) \quad (9)$$

2.2.3 池化

文中采用平方根池化方法,该池化方法的效果已经在一些研究中得到了验证^[15]。实验结果显示平方根池化可以取得比最大池化和平均池化更好的效果。

$C_{i,j,k}^{(R)}, C_{i,j,k}^{(D)}$ 分别表示在节点 (i, j) 处第 k 个 RGB 图和深度图卷积特征, $h_{p,q,k}^{(R)}, h_{p,q,k}^{(D)}$ 分别表示在节点 (p, q) 处第 k 个 RGB 图和深度图的池化特征。

$$\begin{cases} h_{p,q,k}^{(R)} = \sum_{i=p-e}^{p+e} \sum_{j=q-e}^{q+e} C_{i,j,k}^{(R)2} \\ h_{p,q,k}^{(D)} = \sum_{i=p-e}^{p+e} \sum_{j=q-e}^{q+e} C_{i,j,k}^{(D)2} \end{cases} \quad p, q=1, \dots, (d-r+1) \quad (10)$$

如果 (i, j) 超出边界, $C_{i,j,k}^{(R)}, C_{i,j,k}^{(D)} = 0$ 。其中, e 是池化尺寸。

2.3 自编码极限学习机高阶特征学习

H^R 和 H^D 表示单一卷积神经网络层(CNN)分别从 RGB 和深度图中提取低阶特征,通过合并融合得到联合特征,作为自编码极限学习机的输入。其合并过程如下:

$$H_1 = [H^R; H^D]^T \quad (11)$$

接下来,使用自编码极限学习机对合并后的特征进行学习,以得到 RGB-D 图像的高阶特征。

Kasun 等人^[16]基于 ELM 提出自编码极限学习机(AE-ELM),将随机生成的隐含层输入权值和偏置参数正交化能够使得编码器的泛化性能更好。根据前面介绍的极限学习机理论可知 ELM 是一个通用逼近器,所以 AE-ELM 也是一个通用的逼近器。AE-ELM 的网络结构如图 3 所示,在 AE-ELM 中随机生成的隐含层输入权值矩阵和偏置参数向量的的正交化可以将输入数据映射到不同维度的空间,从而实现不同功能的特征表达。

图 3 中 $d > L$, 则 AE-ELM 实现的维度压缩,也就是说将高维度数据映射成低维度特征表达,若是 $d = L$, 表示 AE-ELM 实现的是等维度的特征表达,若

是 $d < L$, 则表示 AE-ELM 实现的是稀疏表达, 即原始数据的高维度特征表达。

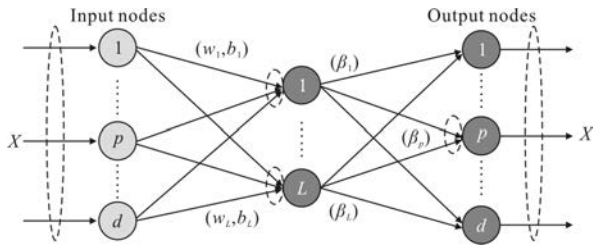


图 3 自编码极限学习机
Fig.3 AE-ELM

AE-ELM 算法通过正交化的隐含层输入权值矩阵和偏置向量将输入数据映射到相同或者不同的数据维度空间, 计算方式如公式(11)所示。

$$H_2 = g(WX + b)$$

$$W^T W = I, b^T b = I \quad (11)$$

式中: $W = [w_1, \dots, w_L]$ 表示正交化的隐含层随机输入权值矩阵; $b = [b_1, \dots, b_L]$ 表示正交化的隐含层偏置向量。

对于稀疏及压缩的特征表达, AE-ELM 的输出权值矩阵片如公式(12)所示。

$$\beta = \left(\frac{I}{C} + H_2^T H_2 \right)^{-1} H_2^T X \quad (12)$$

式中: $H_2 = [h_1, \dots, h_N]$ 表示 AE-ELM 的隐含层输出矩阵; $X = [x_1, \dots, x_M]$ 表示 AE-ELM 的输入和输出矩阵。

对于等维度的特征映射, AE-ELM 的输出权值矩阵 β 通过公式(13)来计算。

$$\beta = H_2^{-1} X$$

$$\beta^T \beta = I \quad (13)$$

通过对公式(12)进行奇异值分解(SVD), 得到公式(14)。

$$H_2 \beta = \sum_{i=1}^N u_i \frac{d_i^2}{d_i^2 + C} u_i^T X \quad (14)$$

式中: u 是 $H_2 H_2^T$ 的特征向量; d 是 H_2 的奇异值。

2.4 监督极限学习机特征分类器

通过无监督的特征学习, 可以得到图像的高阶特征表达。前一个自编码极限学习机的输出层作为后一个极限学习机的输入层, 为了提高泛化性能和结果的鲁棒性, 使用带规则项 C 的极限学习机。

如果训练样本不大, 即远小于隐藏层节点数 ($N < L$), 那么输出的权重矩阵 β 可表示为:

$$\beta = H_3^T \left(\frac{I}{C} + H_3 H_3^T \right)^{-1} T \quad (15)$$

如果训练样本远大于隐藏层节点数 ($N > L$), 那么输出的权重矩阵 β 可表示为:

$$\beta = \left(\frac{I}{C} + H_3 H_3^T \right)^{-1} H_3^T T \quad (16)$$

极限学习机的输出函数可表示为:

$$f(x) = h(x) \beta \quad (17)$$

RGB-D 物体识别属于多分类问题, 给定测试样本输出节点最大值的索引号就是测试集的预测类标签。 $f_j(x)$ 表示输出层第 j 个节点的函数, 即 $f(x) = [f_1(x), \dots, f_m(x)]^T$, 那么预测类标签 x 可表示为:

$$\text{label}(x) = \underset{j \in \{1, \dots, m\}}{\text{argmax}} f_j(x) \quad (18)$$

3 实验验证

RGB-D 物体识别是一个多模态学习问题^[16], 因此混合卷积自编码极限学习机(DCAE-ELM)模型通过合并 RGB 和深度信息, 来达到提高识别精度的目的。

3.1 RGB-D 数据集

为了验证文中所提出混合卷积自编码极限学习机(HCAE-ELM)算法的有效性, 文中实验采用 Lai 等人公布的 RGB-D 数据集^[17]。其中, 该数据库包含 51 个不同类别的家用物品和 300 个这些类别的实体, 每个对象实体均含有从 3 个不同的水平视角高度 (30°, 45° 和 60°) 进行 360° 旋转获取的图像, 使得数据集包含将近 250000 张 RGB-D 图像。对每个实例中的图像以每五帧的间隔进行抽样, 最终得到 9952 张 RGB 图像以及对应的深度图像。图 4 显示了用于实验的 RGB-D 数据集中部分物体对应的 RGB 和深度图像。

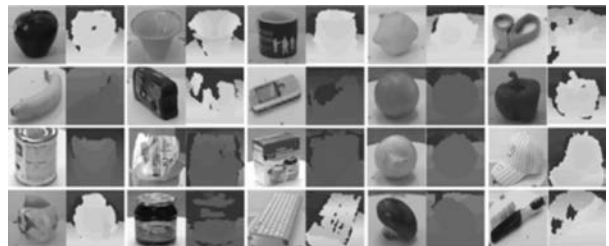


图 4 RGB-D 数据集中部分例子

Fig.4 Some examples from the RGB-D database

物体类别识别是文中的工作重点, 为了和先前

方法的实验结果进行对比, 实验的设置与 Lai K 等人^[17]保持一致以及他们所提供的 10 个随机分割。对于每个分割的测试集, 从每个类中列举一个对象, 形成 51 个测试对象, 每个对象有大约 120 张独立分类的图像, 由此产生 34 000 张图像作为训练模型和 69 000 张图像作为测试模型。在图像被传递给 CNN 前, 调整其大小为 $d=128$ 。CNN 的参数设置为: 滤波器尺寸为 $f=5 \times 5$, 池化尺寸 $e=3$, 规则项 $C=0.01$ 。

设置两个极限学习机的隐藏层节点数 $n=5\ 000$ 。每次试验, 随机地将数据库分成训练样本和测试样本两个部分。从数据库的每一个类别中随机抽取 10 个实例对象作为测试样本, 保留下来的作为训练样本。随机试验 10 次, 取其平均值作为类别识别精度和标准偏差的最终结果。

3.2 结果

在同一数据集上, 文中的混合卷积自编码极限学习机与其他方法进行了比较。之前已有不少学者针对 RGB-D 物体识别提出了相应的算法, 表 1 列出了文中以及先前的一些算法在 RGB-D 数据库上取得的物体识别。比较结果显示了 HCAE-ELM 方法取得了最好的识别效果。对于相似物体的识别, 获取鲁棒性的特征尤为关键, 实验结果显示了合并 RGB 和深度图后提取的特征比单独从 RGB 或深度图提取的特征具有更好的识别效果, 主要原因是不同通道提取的低层次特征不是独立的, 具有互补性。

表 1 不同方法的识别精度

Tab.1 Recognition accuracy for different approaches

| Methods | RGB | Depth | RGB-D |
|-------------------------------|----------|------------|----------|
| Nonlinear SVM ^[17] | 74.5±3.1 | 64.7±2.2 | 83.9±3.5 |
| HKDES ^[18] | 76.1±2.2 | 75.7 ± 2.6 | 84.1±2.2 |
| CNN-RNN ^[7] | 80.8±4.2 | 78.9 ± 3.8 | 86.8±3.3 |
| CNN features ^[19] | 83.1±2.0 | NA | 89.4±1.3 |
| CNN-SPMP-RNN ^[20] | 83.6±2.3 | 85.2±1.2 | 90.7±1.1 |
| MM-ELM-LRF ^[21] | 82.3±3.2 | 82.8±2.1 | 89.3±1.5 |
| HCAE-ELM | 84.3±3.2 | 82.9±2.1 | 90.2±1.5 |

与参考文献[20]中的方法相比, 笔者的方法仍然具有可比性, 虽然其识别精度稍低一些, 但是训练

速度快了近 10 倍, 如表 2 所示。

表 2 训练时间

Tab.2 Training time

| Methods | RGB | Depth | RGB-D/s |
|--------------|----------|----------|---------|
| CNN-SPMP-RNN | NA | NA | 7 100 |
| HCAE-ELM | 571.44 s | 435.36 s | 781 |

如图 5 所示, 与 RGB 或深度图单一模型的网络结构比较, 笔者的混合卷积自编码极限学习机能够有效地学习到更加丰富的特征表达, 因此获得的识别精度也最高。

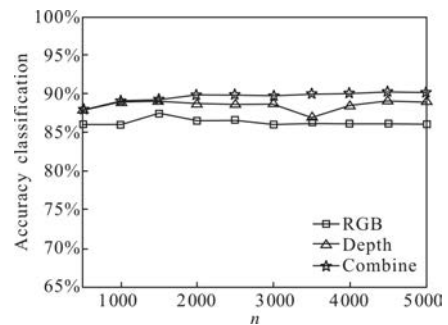


图 5 不同方法下相对不同隐藏层节点数的测试精度

Fig.5 Testing accuracies of different methods vs the number of hidden layer nodes

为了分析参数对算法的影响, 文中做了敏感性分析。混合卷积自编码极限学习机有两个重要参数: 规则项 C 和隐藏层节点数量 n 。因此设置 C 集合 $\{10^{-9}, 10^{-3}, 0, \dots, 10^0\}$, 设置 n 集合为 $\{100, 300, 500, \dots, 2000\}$, 结果如图 6 所示。

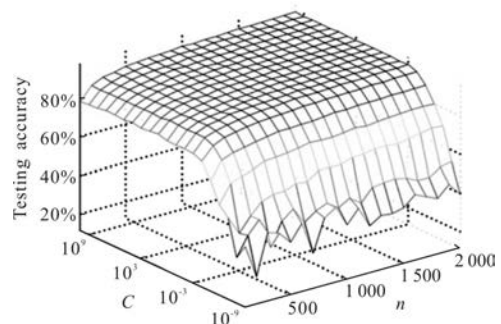


图 6 依据参数 n 和 C 的 HCAE-ELM 的测试精度

Fig.6 Testing accuracy of HCAE-ELM in term of n and C

3.3 错误分析

图 7 为在 RGB-D 数据库上识别结果的混淆矩阵。比较其他方法, 提出的混合卷积自编码极限学习

机有较低的识别错误率。大多数类别的物体都可以被正确分类，意味着提出的方法仅仅使用少量的带标签数据就取得了较高的准确识别率。

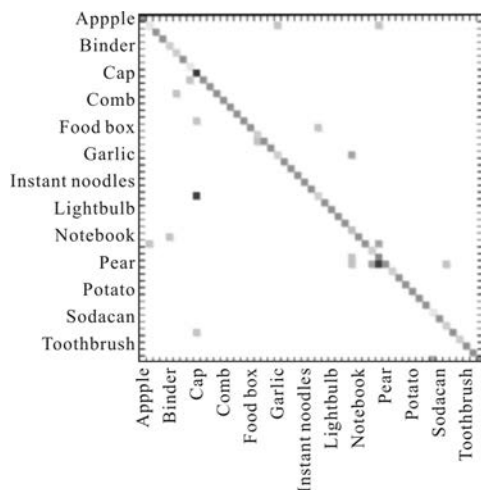


图 7 HCAE-ELM 模型的混淆矩阵

Fig.7 Confusion matrix of the HCAE-ELM model

图 8 显示了四对常常混淆类的例子，洗发水和水杯，大蒜和蘑菇从外观和颜色上都非常相似，苹果和西红柿外形都很近似，白色的帽子和键盘主要是因为纹理和颜色太相似，笔者的方法对这些例子进行识别时有一定的难度。

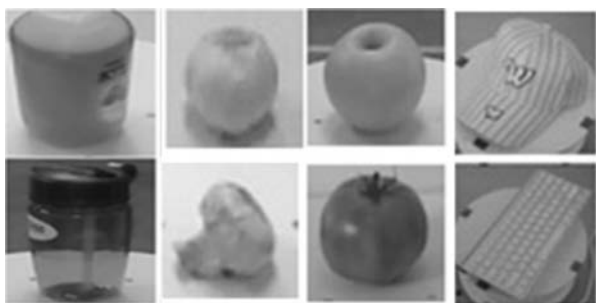


图 8 一些混淆类的例子

Fig.8 Examples of confused classes

4 结 论

为了提高 RGB-D 物体识别的泛化能力和训练速度，提出了一种新型的混合卷积自编码极限学习机，该模型结构合并了 CNN 的提取特征有效性和 ELM 分类快速性的优点。CNN 分别从 RGB 和深度图提取低阶特征，然后进行特征合并，使用自编码极限学习机进行高阶特征学习，最后送入 ELM 进行分类。整个过程的参数均随机选取，不需要精确调节，

大大降低了训练时间，较传统的深度学习，笔者的模型结构具备更高的鲁棒性和更好的识别分类效果。

参考文献：

- [1] Microsoft Kinect[EB/OL]. (2013-03-05). <http://www.kinect.com>.
- [2] Cao Chuqing, Li Ruifeng, Zhao Lijun. Hand posture recognition method based on depth image technology [J]. *Computer Engineering*, 2012, 38(8): 16-21. (in Chinese)
曹维清, 李瑞峰, 赵立军. 基于深度图像技术的手势识别方法[J]. *计算机工程*, 2012, 38(8): 16-21.
- [3] Wang Xin, Wo Bohai, Guan Qiu, et al. Human action recognition based on manifold learning[J]. *Journal of Image and Graphics*, 2014, 19(6): 914-923. (in Chinese)
王鑫, 沃波海, 管秋, 等. 基于流形学习的人体动作识别[J]. *中国图象图形学报*, 2014, 19(6): 914-923.
- [4] Li Changyong, Cao Qixin. Extraction method of shape feature for vegetables based on depth image[J]. *Transactions of the Chinese Society for Agricultural Machinery*, 2012, 43(Z1): 242-245. (in Chinese)
李长勇, 曹其新. 基于深度图像的蔬果形状特征提取[J]. *农业机械学报*, 2012, 43(Z1): 242-245.
- [5] Xu Ke. Study of convolutional neural network applied on image recognition[D]. Hangzhou: Zhejiang University, 2012. (in Chinese)
许可. 卷积神经网络在图像识别的应用研究[D]. 杭州: 浙江大学, 2012.
- [6] Blum M, Springenberg J T, Wulfing J, et al. A learned feature descriptor for object recognition in RGB-D data[C]// IEEE International Conference on Robotics & Automation, 2012, 44(8): 1298-1303.
- [7] Socher R, Huval B, Bath B P, et al. Convolutional- recursive deep learning for 3D object classification [C]// NIPS, 2012: 665-673.
- [8] Niun Xiaoxiao, Suen Ching Y. A novel hybrid CNN-SVM classifier for recognizing handwritten digits [J]. *Pattern Recognition*, 2012, 45(4): 1318-1325.
- [9] Liu Tianhua, Yang Shaoqing, Liu Songtao. Research of sea-aero target detection from photoelectricity image based on cellular neural networks[J]. *Infrared and Laser Engineering*, 2008, 37(S2): 310-313. (in Chinese)
刘天华, 杨绍清, 刘松涛. 基于 CNN 的海空光电目标检测技术研究[J]. *红外与激光工程*, 2008, 37(S2): 310-313.
- [10] Li Junmei, Hu Yihua, Tao Xiaohong. Recognition method

- based on principal component analysis and back-propagation neural network [J]. *Infrared and Laser Engineering*, 2005, 34(6): 719–723. (in Chinese)
- 李军梅, 胡以华, 陶小红. 基于主成分分析与 BP 神经网络的识别方法研究[J]. 红外与激光工程, 2005, 34(6): 719–723.
- [11] Wang Yong, Xu Haisong. Spectral characterization of scanner based on PCA and BP ANN [J]. *Chinese Optics Letters*, 2005, 3(12): 725–728.
- [12] Huang G B, Zhu Q Y, Siew C K. Extreme learning machine: a new learning scheme of feedforward neural networks [C]//IEEE International Joint Conference on Neural Networks, 2004, 2: 985–990.
- [13] Huang G B, Zhu Q Y, Siew C K: Extreme learning machine: theory and applications[J]. *Neurocomputing*, 2006: 70(1–3): 489–501.
- [14] Arel I, Rose D C, Karnowski T P. Deep machine learning—a new frontier in artificial intelligence research [J]. *IEEE Computational Intelligence Magazine*, 2010, 5(4): 13–18.
- [15] Boureau Y L, Ponce J, Lecun Y. A theoretical analysis of feature pooling in visual recognition [C]//International Conference on Machine Learning, 2010, 32 (4): 111–118.
- [16] Kasun L L C, Zhou H, Huang G B, et al. Representational learning with ELMs for big data [C]//Intelligent Systems IEEE, 2013, 28(6): 31–34.
- [17] Lai K, Bo L, Ren X, et al. A large-scale hierarchical multi-view RGB–D object dataset [C]//IEEE International Conference on Robotics and Automation, 2011: 1817–1824.
- [18] Bo L, Lai K, Ren X, et al. Object recognition with hierarchical kernel descriptors [C]//IEEE Int Conf on Computer Vision and Pattern Recognition, 2011: 1729–1736.
- [19] Schwarz M, Schulz H, Behnke S. RGB–D object recognition and pose estimation based on pre-trained convolutional neural network features [C]//IEEE Int Conf on Robotics & Automation, 2015: 1329–1335.
- [20] Cheng Y, Zhao X, Huang K, et al. Semi-supervised learning and feature evaluation for RGB–D object recognition [J]. *Computer Vision & Image Understanding*, 2015, 139 (C): 149–160.
- [21] Li F, Liu H, Xu X, et al. Multi-modal local receptive field extreme learning machine for object recognition [C]//International Joint Conference on Neural Networks, 2016: 1696–1701.