深度学习研究综述

孙志军[1]，薛 磊[1,2]，许阳明[1,2]，王 正[1,2]
（1. 电子工程学院，合肥 230037；2. 安徽省电子制约技术重点实验室，合肥 230037）

摘 要：深度学习是一类新兴的多层神经网络学习算法，因其缓解了传统训练算法的局部最小性，引起机器学习领域的广泛关注。首先论述了深度学习兴起渊源，分析了算法的优越性，并介绍了主流学习算法及应用现状，最后总结了当前存在的问题及发展方向。

关键词：深度学习；分布式表示；深信度网络；卷积神经网络；深凸网络
中图分类号：TP181 **文献标志码**：A **文章编号**：1001-3695(2012)08-2806-05
doi:10.3969/j.issn.1001-3695.2012.08.002

Overview of deep learning

SUN Zhi-jun[1], XUE Lei[1,2], XU Yang-ming[1,2], WANG Zheng[1,2]
(1. *Electronic Engineering Institute*, *Hefei* 230037, *China*; 2. *Key Laboratory of Electronic Restriction*, *Hefei* 230037, *China*)

Abstract: Deep learning is a new way of training multi-layer neural network. The optimization difficulty associated with the deep models can be alleviated, it has induced great attention of machine learning community. Firstly, this paper discussed the origin of deep learning, then analyzed virtue brought by deep learning. It introduced the main stream deep learning algorithm and their applications. Finally it concluded the problem remaining and development orientation.

Key words: deep learning; distributed representation; deep belief network; convolutional neural network; deep convex network

0 引言

深度学习的概念源于人工神经网络的研究，含多隐层的多层感知器(MLP)就是一种深度学习结构。深度学习通过组合低层特征形成更加抽象的高层表示(属性类别或特征)，以发现数据的分布式特征表示[1]。BP 算法作为传统训练多层网络的典型算法，实际上对于仅含几层网络，该训练方法就已很不理想[2]。深度结构(涉及多个非线性处理单元层)非凸目标代价函数中普遍存在的局部最小是训练困难的主要来源。

Hinton 等人[3~5]基于深信度网(DBN)提出非监督贪心逐层训练算法，为解决深层结构相关的优化难题带来希望，随后提出多层自动编码器深层结构。此外 Lecun 等人提出的卷积神经网络(CNNs)是第一个真正多层结构学习算法[6]，它利用空间相对关系减少参数数目以提高 BP 训练性能。此外深度学习还出现许多变形结构如去噪自动编码器[7]、DCN[8]、sum-product[9]等。

当前多数分类、回归等学习方法为浅层结构算法，其局限性在于有限样本和计算单元情况下对复杂函数的表示能力有限，针对复杂分类问题其泛化能力受到一定制约[2]。深度学习可通过学习一种深层非线性网络结构，实现复杂函数逼近，表征输入数据分布式表示，并展现了强大的从少数样本集中学习数据集本质特征的能力[1,10]。本文意在向读者介绍这一刚刚兴起的深度学习新技术。

1 深度学习神经学启示及理论依据

1.1 深度学习神经学启示

尽管人类每时每刻都要面临着大量的感知数据，却总能以一种灵巧方式获取值得注意的重要信息。模仿人脑那样高效准确地表示信息一直是人工智能研究领域的核心挑战。神经科学研究人员利用解剖学知识发现哺乳类动物大脑表示信息的方式：通过感官信号从视网膜传递到前额大脑皮质再到运动神经的时间，推断出大脑皮质并未直接地对数据进行特征提取处理，而是使接收到的刺激信号通过一个复杂的层状网络模型，进而获取观测数据展现的规则[11~13]。也就是说，人脑并不是直接根据外部世界在视网膜上投影，而是根据经聚集和分解过程处理后的信息来识别物体。因此视皮层的功能是对感知信号进行特征提取和计算，而不仅仅是简单地重现视网膜的图像[14]。人类感知系统这种明确的层次结构极大地降低了视觉系统处理的数据量，并保留了物体有用的结构信息。对于要提取具有潜在复杂结构规则的自然图像、视频、语音和音乐等结构丰富数据，深度学习能够获取其本质特征。

受大脑结构分层次启发，神经网络研究人员一直致力于多层神经网络的研究。BP 算法是经典的梯度下降并采用随机选定初始值的多层网络训练算法，但因输入与输出间非线性映射使网络误差函数或能量函数空间是一个含多个极小点的非线性空间，搜索方向仅是使网络误差或能量减小的方向，因而经

收稿日期：2012-03-09；**修回日期**：2012-04-10

作者简介：孙志军(1985-)，男，吉林磐石人，博士研究生，主要研究方向为机器学习、模式识别(robotman@126.com)；薛磊(1963-)，男，安徽霍丘人，教授，博导，主要研究方向为通信系统、通信信号处理；许阳明(1964-)，男，安徽舒城人，副教授，主要研究方向为无线通信、通信信号处理；王正(1973-)，男，福建莆田人，讲师，博士研究生，主要研究方向为数据融合、智能信号处理.

常收敛到局部最小,并随网络层数增加情况更加严重。理论和实验表明 BP 算法不适于训练具有多隐层单元的深度结构[15]。此原因在一定程度上阻碍了深度学习的发展,并将大多数机器学习和信号处理研究从神经网络转移到相对较容易训练的浅层学习结构。

传统机器学习和信号处理技术探索仅含单层非线性变换的浅层学习结构。浅层模型的一个共性是仅含单个将原始输入信号转换到特定问题空间特征的简单结构。典型的浅层学习结构包括传统隐马尔可夫模型(HMM)、条件随机场(CRFs)、最大熵模型(MaxEnt)、支持向量机(SVM)、核回归及仅含单隐层的多层感知器(MLP)等。

1.2　浅层结构函数表示能力的局限性

早期浅层结构局限性结论是关于利用逻辑门电路实现函数奇偶性问题。利用一个深度为 $O(\log d)$ 的网络用 $O(d)$ 个计算节点去计算一个 d 比特和的奇偶性,而对于两层网络则需要指数倍数目的计算单元。随后又有学者指出可以利用深度为 K 的多项式级的逻辑门电路实现的函数,对于 $K-1$ 层电路需要指数倍的计算节点。文献[10]指出深度学习结构可以很简洁地表示复杂函数,否则一个不合适的结构模型将需要数目非常大的计算单元。这里简洁包含三方面内容:a)需要的数据量特别是带类标记的样本;b)需要的计算单元的数目;c)需要的人为先验知识。例如多项式 $\prod_{i=1}^{n}\sum_{j=1}^{m}a_{ij}x_j$ 可以高效地(相对于需训练的计算单元数目)利用 $O(mn)$ 运算量表示成和积(sum-product)结构,如果表示成积和结构,将需要 $O(n^m)$ 计算量。此外文献[16]指出存在一大类函数不能用浅层电路表示。这些数学结果指出了浅层学习网络的局限性,激发了利用深度网络对复杂函数建模的动机。

1.3　局部表示、分布式表示和稀疏表示

最近许多研究者已经研究了分布式表示的一个变体,它介于纯粹局部表示和稠密分布式表示之间——稀疏表示。它的思想是尽量要求所获取表示中只有少数维是有效的,使绝大多数维设为 0 或接近于 0 的无效维。目的是尽量找出信号的主要驱动源。

基于模板匹配的模型可认为含两层计算单元,第一层构建对输入数据进行匹配的多个模板,每一匹配单元可输出一个匹配度;第二层采用特定机制融合第一层的输出匹配度。典型基于局部匹配的例子是核方法。

$$f(x)=b+\sum_i \alpha_i K(x,x_i) \tag{1}$$

这里 b 和 α_i 形成第二计算层。核函数 $K(x,x_i)$ 将输入 x 匹配到训练样本 x_i,并在全局范围求和。式(1)的结果可作为分类器的区分类标签,或者回归预测器的预测值。利有局部核函数的核方法能获取泛化性能,因其利用光滑性的先验知识,即目标函数可利用光滑函数逼近。在监督学习中,由训练样本 (x_i,y_i) 组建预测器,当输入 x 与 x_i 靠近时,输出接近 y_i。通常这是合理假设,但文献[10]中指出当目标函数非常复杂时,这样的模型泛化能力很差。其原因是利用局部估计学习算法表示函数时,一个局部估计子将输入空间进行切分,并需要不同自由度参数来描述目标函数在每一区域的形状。当函数较为复杂时,需要利用参数进行描述的区域数目也是巨大的。固定核函数的这种局限性已引起基于先验知识设计核函数的研究,而如果缺乏足够的先验知识是否可通过学习获取一个核函数?该问题同样引起大量研究。Lanckriet 等人[17]提出利用半正定规划技术学习数据的核矩阵,然后利用该核矩阵获取较好的泛化性能。然而当学习到的核函数相互关联时,能否获取更加简洁的表示?深度学习即基于这种思想并通过多次网络学习输入样本的分布式表示,被认为是较有前景的方法。

分布式表示[18]是在机器学习和神经网络研究中可以处理维数灾难和局部泛化限制的一个古老的思想。如图 1 所示,分布式表示由一系列有可能是统计独立的显著特征组成,与局部泛化的方法对比,基于分布式表示的可区分模式的数目与分布式表示的维数(学习到的特征)是指数倍关系的。参数数目上的减少对统计机器学习是非常有意义的,因为不仅可以降低运算量,同时仅需相对较少的样本即可避免过拟合现象的发生。而聚类算法和最近邻算法等局部表示算法将输入空间切分如图 1 左侧所示,不同局部之间是互斥的,不能形成简洁的分布式表示。ICA 、PCA 和 RBM 等算法用较少的特征将输入空间切分如图 1 右侧所示,并构建分布式表示,参数数目和需要的样本数要比子区域的数目少得多,这也是为什么会对未观测数据泛化的原因。PCA 和 ICA 可以获取输入的主要分量信息,但对于输出信号数目小于输入信号数目时,不能很好地解决欠定问题。文献[19]中提出了利用自联想神经网络来提取数据的非线性主分量的方法,该学习方法的目的是通过事物的部分信息或者带噪声的信息来还原事物的本来信息。自联想神经网络的隐层节点数目少于输入节点数目时,可认为在自联想过程中,这些隐层能够保留数据集中的主要信息。多层神经网络和 Boltzmann 机已被用于学习分布式表征。文献[20]已证明利用 DBN 学习特征空间对高斯过程回归的性能进行提高。深度学习算法可以看成核机器学习中一个优越的特征表示方法。文献[2]指出单个决策树的泛化性能随目标函数变量增加而降低。多个树的集成(森林)比单个树更加强大,也是因为增加了一个第三层,并潜在地形成分布式表示,可表达与子树数目指数倍个的分布。

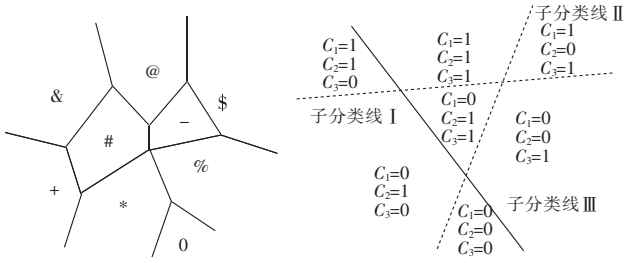

图 1　数据样本的局部式表示(左)和分布式表示(右)

1.4　深度学习成功的关键

深度学习具有多层非线性映射的深层结构,可以完成复杂的函数逼近是深度学习优势之一;此外深度学习理论上可获取分布式表示,即可通过逐层学习算法获取输入数据的主要驱动变量。该优势是通过深度学习的非监督预训练算法完成,通过生成性训练可避免因网络函数表达能力过强而出现过拟合情况。但由于单层有限的计算能力,通过多层映射单元可提取出主要的结构信息。文献[15]深入分析并通过实验验证了贪婪层次非监督深度学习方法的优势所在。

2　典型的深度学习结构

深度学习涉及相当广泛的机器学习技术和结构,根据这些结构和技术应用的方式,可以将其分成如下三类:

a)生成性深度结构。该结构描述数据的高阶相关特性,

或观测数据和相应类别的联合概率分布。

b)区分性深度结构。目的是提供对模式分类的区分性能力,通常描述数据的后验分布。

c)混合型结构。它的目标是区分性的,但通常利用了生成型结构的输出会更易优化。

2.1 生成型深度结构

文献[3]首次提出的 DBN 是目前研究和应用都比较广泛的深度学习结构。与传统区分型神经网络不同,可获取观测数据和标签的联合概率分布,这方便了先验概率和后验概率的估计,而区分型模型仅能对后验概率进行估计。DBN 解决传统 BP 算法训练多层神经网络的难题:a)需要大量含标签训练样本集;b)较慢的收敛速度;c)因不合适的参数选择陷入局部最优。

DBN 由一系列受限波尔兹曼机(RBM)单元组成。RBM 是一种典型神经网络,如图 2 所示。该网络可视层和隐层单元彼此互连(层内无连接),隐单元可获取输入可视单元的高阶相关性。相比传统 sigmoid 信度网络,RBM 权值的学习相对容易。为了获取生成性权值,预训练采用无监督贪心逐层方式来实现。在训练过程中,首先将可视向量值映射给隐单元;然后可视单元由隐层单元重建;这些新可视单元再次映射给隐单元,这样就获取了新的隐单元。反复执行这种步骤叫做吉布斯采样。

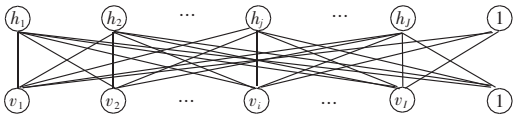

图 2 RBM 模块

RBM 在给定模型参数条件下的联合分布为

$$p(v,h;\theta)=\exp(-E(v,h;\theta))/Z \tag{2}$$

其中:$Z=\sum_v\sum_h\exp(-E(v,h;\theta))$是归一化因子或剖分函数。模型赋予可视向量 v 的边际概率为

$$p(v;\theta)=\sum_h\exp(-E(v,h;\theta))/Z \tag{3}$$

对伯努利(可视)—伯努利(隐藏)RBM 能量函数定义为

$$E(v,h;\theta)=-\sum_{i=1}^{I}\sum_{j=1}^{J}w_{ij}v_ih_j-\sum_{i=1}^{I}b_iv_i-\sum_{j=1}^{J}a_jh_j$$

其中:w_{ij}为可视单元和隐单元连接权值;b_j 和 a_j 是偏置量;I 和 J 是可视单元和隐单元的数目。条件概率如式(4)计算:

$$\begin{aligned}p(h_j=1\mid v;\theta)&=\delta(\sum_{i=1}^{I}w_{ij}v_i+a_j)\\p(v_i=1\mid h;\theta)&=\delta(\sum_{i=1}^{I}w_{ij}h_j+b_i)\end{aligned} \tag{4}$$

这里 $\delta(x)=1/(1+\exp(x))$。相似地,对于高斯(可视)—伯努利(隐)RBM 能量函数为

$$E(v,h;\theta)=-\sum_{i=1}^{I}\sum_{j=1}^{J}w_{ij}v_ih_j+\frac{1}{2}\sum_{i=1}^{I}(v_i-b_j)^2-\sum_{j=1}^{J}a_jh_j \tag{5}$$

对应的条件概率变成:

$$\begin{aligned}p(h_j=1\mid v;\theta)&=\delta(\sum_{i=1}^{I}w_{ij}v_i+a_j)\\p(v_i=1\mid h;\theta)&=N(\sum_{i=1}^{I}w_{ij}h_j+b_i,1)\end{aligned} \tag{6}$$

其中:v_i 是满足均值为$\sum_{j=1}^{J}w_{ij}h_j+b_i$、方差为 1 的高斯分布的实数值。高斯—伯努利 RBMs 可将实值随机变量转换到二进制随机变量,然后再进一步利用伯努利—伯努利 RBMs 处理。利用对数似然概率 $\log(p(v;\theta)$梯度可推导出 RBM 的权值更新准则:

$$\Delta w_{ij}=E_{\text{data}}(v_ih_j)-E_{\text{model}}(v_ih_j) \tag{7}$$

其中:$E_{\text{data}}(v_ih_j)$是在观测数据训练集中的期望;$E_{\text{model}}(v_ih_j)$是模型中定义的期望。精心训练 RBM 对成功应用深度学习是一个关键。文献[21]提供了对 RBM 实际训练的指导。

通过自底向上组合多个 RBM 可以构建一个 DBN,如图 3 所示。应用高斯—伯努利 RBM 或伯努利—伯努利 RBM,可用隐单元的输出作为训练上层伯努利—伯努利 RBM 的输入,第二层伯努利和伯努利的输出作为第三层的输入等。这个逐层高效的学习策略理论证明可参见文献[3],它指出上述逐层学习程序提高了训练数据基于混合模型的似然概率的变化下界。

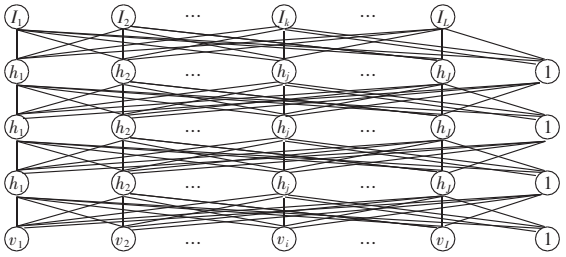

图 3 DBN 模型

2.2 区分性深度结构

卷积神经网络(CNNs)是第一个真正成功训练多层网络结构的学习算法,与 DBNs 不同,它属于区分性训练算法。受视觉系统结构的启示,当具有相同参数的神经元应用于前一层的不同位置时,一种变换不变性特征就可获取了。后来 LeCun 等人沿着这种思路,利用 BP 算法设计并训练了 CNNs。CNNs 作为深度学习框架是基于最小化预处理数据要求而产生的。受早期的时间延迟神经网络影响,CNNs 靠共享时域权值降低复杂度。CNNs 是利用空间关系减少参数数目以提高一般前向 BP 训练的一种拓扑结构,并在多个实验中获取了较好性能[6,22]。在 CNNs 中被称做局部感受区域的图像的一小部分作为分层结构的最底层输入。信息通过不同的网络层次进行传递,因此在每一层能够获取对平移、缩放和旋转不变的观测数据的显著特征。

文献[6,22]描述了 CNNs 在 MNIST 数据库中的手写体识别应用情况。如图 4 所示,本质上,输入图形与一系列已训练的滤波器系数进行卷积操作;后经加性偏置和压缩、特征归一化等,最初阶段伴随进一步降维的下采样(C_x)提供对空域变化的鲁棒性;下采样特征映射经加权后的可调偏置,最终利用激活函数进行传递。组合多个上述映射层(图 5)可获取层间关系和空域信息,这样 CNNs 适于图像处理和理解。国内学者夏丁胤[23]将这种网络应用于网络图像标注中。最近 CNNs 已应用于包括人脸检测、文件分析和语音检测等不同机器学习的问题中。

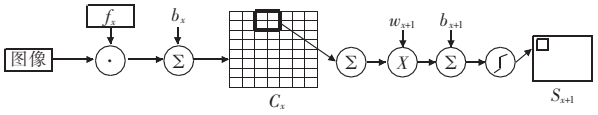

图 4 CNN 中卷积和采样过程

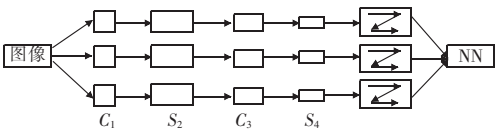

图 5 卷积神经网络的原理

文献[8]近期提出一新的深度学习算法。DCN 如图 6 所

示,每层子模块是含单隐层和两个可训练的加权层神经网络。DCN 是由一系列分层子模块串联组成。模块第一个线性输入层对应输入特征维数,隐层是一系列非线性参数可调单元,第二线性输出包含线性输出单元及原始输入数据,最顶模块的输出代表分类目标单元。例如,如果 DCN 设定用于实现数字识别,输出可表示成 1 ~ 10 的 0-1 编码。如用于语音识别,输入对应语音波形采样或波形提取特征;如功率谱或倒谱系数,输出单元代表不同音素。

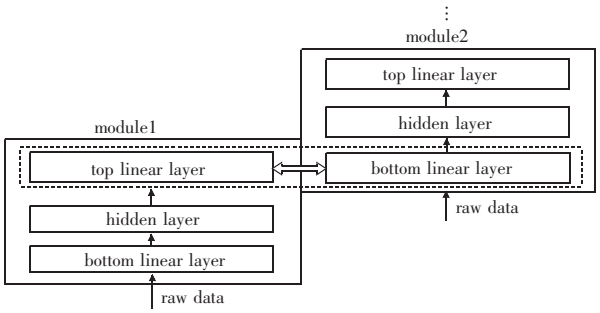

图 6　DCN 的结构

2.3　混合型结构

混合型结构的学习过程包含两个部分,即生成性部分和区分性部分。现有典型的生成性单元通常最终用于区分性任务,生成性模型应用于分类任务时,预训练可结合其他典型区分性学习算法对所有权值进行优化。这个区分性寻优过程通常是附加一个顶层变量来表示训练集提供的期望输出或标签。BP 算法可用于优化 DBN 权值,它的初始权值通过在 RBM 和 DBN 预训练中得到而非随机产生,这样的网络通常会比仅通过 BP 算法单独训练的网络性能优越。可以认为 BP 对 DBNs 训练仅完成局部参数空间搜索,与前馈型神经网络相比加速了训练和收敛时间。

最近,基于 DBNs 的研究包括应用层叠自动编码器取代传统 DBNs 中的 RBMs。该方法可采用和 DBNs 相同的训练准则,不同的是自动编码器利用区分性模型。去噪自动编码器在训练中引入随机变化过程可以产生与传统的 DBNs 相比拟的泛化性能;对单个去噪自动编码器的训练与 RBMs 生成性模型一致。

3　深度学习应用现状

深度学习在信号处理中的应用对象不仅包含语音、图像和视频,同样也包含文本、语言和传递人类可获知的语义信息。传统的 MLP 已经在语音识别领域应用多年,在单独使用的情况下它们的性能远低于利用 GMM-HMM 的系统。最近,凭借具有很强区分性能力的 DBNs 和序列建模能力的 HMMs,深度学习技术成功应用于语音、大词汇量连续语音识别(LVCSR)[24]任务。文献[25]利用五层 DBN 来替换 GMM-HMM 中的高斯混合模型,并利用单音素状态作为建模单元进行语音识别。文献[26]中,Nair 等人提出在顶层利用三阶波尔兹曼机的改进型 DBN,并将该 DBN 应用于三维物体识别任务 NORB 数据库,给出了接近于历史最好识别误差结果,特别地,它指出 DBN 实质上优于 SVMs 等浅层模型。文献[27]提出了 tRBM,并利用自动编码器对舌轮廓进行实时提取。与一般训练不同的是,它首先利用样本数据和人工提取的轮廓数据同时作为训练样本输入,经正常的自动编码器输出;训练完毕后,利用提出的 tRBM 对顶层进行改进,以使仅有感知图像作为输入对舌轮廓进行预测。此外深度学习在语言文件处理的研究日益受到普遍关注。利用神经网络对语言建模已有很长的历史,在语音识别、机器翻译、文本信息检索和自然语言处理方面具有重要应用。最近,深层网络已经开始吸引语言处理和信息检索方面的研究人员的注意。文献[28]利用基于 DBN 的多任务学习技术来解决机器字译问题,这可以推广到更困难的机器翻译问题。利用 DBN 和深度自动编码器对文件检索可以显示基于单词特征,与广泛应用的语义分析相比具有明显优势,可令文献检索更容易,这一思想已被初步扩展到音频文件检索和语音识别类问题[29]。

4　结束语

深度学习已成功应用于多种模式分类问题。这一领域虽处于发展初期,但它的发展无疑会对机器学习和人工智能系统产生影响。同时它仍存在某些不适合处理的特定任务,譬如语言辨识,生成性预训练提取的特征仅能描述潜在的语音变化,不会包含足够的不同语言间的区分性信息;虹膜识别等每类样本仅含单个样本的模式分类问题也是不能很好完成的任务。

深度学习目前仍有大量工作需要研究。模型方面是否有其他更为有效且有理论依据的深度模型学习算法,探索新的特征提取模型是值得深入研究的内容。此外有效的可并行训练算法也是值得研究的一个方向。当前基于最小批处理的随机梯度优化算法很难在多计算机中进行并行训练。通常办法是利用图形处理单元加速学习过程,然而单个机器 GPU 对大规模数据识别或相似任务数据集并不适用。在深度学习应用拓展方面,如何充分合理地利用深度学习在增强传统学习算法的性能仍是目前各领域的研究重点。

参考文献:

[1] BENGIO Y, DELALLEAU O. On the expressive power of deep architectures[C]//Proc of the 14th International Conference on Discovery Science. Berlin: Springer-Verlag, 2011: 18-36.

[2] BENGIO Y. Learning deep architectures for AI[J]. Foundations and Trends in Machine Learning, 2009, 2(1): 1-127.

[3] HINTON G, OSINDERO S, TEH Y. A fast learning algorithm for deep belief nets[J]. Neural Computation, 2006, 18(7): 1527-1554.

[4] BENGIO Y, LAMBLIN P, POPOVICI D, *et al*. Greedy layer-wise training of deep networks[C]//Proc of the 12th Annual Conference on Neural Information Processing System. 2006: 153-160.

[5] LECUN Y, BOTTOU L, BENGIO Y, *et al*. Gradient-based learning applied to document recognition[J]. Proceedings of the IEEE, 1998, 86(11): 2278-2324.

[6] VINCENT P, LAROCHELLE H, BENGIO Y, *et al*. Extracting and composing robust features with denoising autoencoders[C]//Proc of the 25th International Conference on Machine Learning. New York: ACM Press, 2008: 1096-1103.

[7] VINCENT P, LAROCHELLE H, LAJOIE I, *et al*. Stacked denoising autoencoders: learning useful representations in a deep network with a local denoising criterion[J]. Journal of Machine Learning Research, 2010, 11(12): 3371-3408.

[8] YU Dong, DENG Li. Deep convex net: a scalable architecture for speech pattern classification[C]//Proc of the 12th Annual Confe-

rence of International Speech Comunication Association. 2011:2285-2288.

[9] POON H, DOMINGOS P. Sum-product networks: a new deep architecture[C]//Proc of IEEE International Conference on Computer Vision. 2011:689-690.

[10] BENGIO Y, LECUN Y. Scaling learning algorithms towards AI[M]// BOTTOU L, CHAPELLE O, DeCOSTE D, *et al*. Large-Scale Kernel Machines. Cambridge: MIT Press, 2007:321-358.

[11] LEE T S, MUMFORD D. Hierarchical Bayesian inference in the visual cortex[J]. Optical Society of America, 2003, 20(7):1434-1448.

[12] SERRE T, WOLF L, BILESCHI S, *et al*. Robust object recognition with cortex-like mechanisms[J]. IEEE Trans on Pattern Analysis and Machine Intelligence, 2007, 29(3):411-426.

[13] LEE T S, MUMFORD D, ROMERO R, *et al*. The role of the primary visual cortex in higher level vision[J]. Vision Research, 1998, 38(15):2429-2454.

[14] ROSSI A F, DESIMONE R, UNGERLEIDER L G. Contextual modulation in primary visual cortex of macaques[J]. Journal of Neuroscience, 2001, 21(5):1689-1709.

[15] ERHAN D, BENGIO Y, COUVILLE A, *et al*. Why does unsupervised pre-training help deep learning[J]. Journal of Machine Learning Research, 2010, 11(3):625-660.

[16] BRAVERMAN M. Poly-logarithmic independence fools bounded-depth boolean circuits[J]. Communications of the ACM, 2011, 54(4):108-115.

[17] LANCKRIET G R G, CRITIANINI N, BARTLETT P, *et al*. Learning the kernel matrix with semidefinite programming[J]. Journal of Machine Learning Research, 2004, 5(1):27-72.

[18] HINTON G E. Learning distributed representations of concepts[C]// Proc of the 8th Annual Conference of the Cognitive Science Society. 1986:1-12.

[19] KRAMER M. Nonlinear principal component analysis using autoassociative neural networks[J]. AIChE Journal, 1991, 37(2):233-243.

[20] SALAKHUTDINOV R. Learning deep generative models[D]. Toronto: Graduate Department of Computer Science, University of Toronto, 2009.

[21] HINTON G. A practical guide to training restricted boltzmann machines[D]. Toronto: University of Toronto, 2010:1-20.

[22] HUANG Fu-jie, LECUN Y. Large-scale learning with SVM and convolutional for generic object categorization[C]//Proc of IEEE Computer Society Conference on Computer Vision and Pattern Recognition. Washington DC: IEEE Computer Society, 2006:284-291.

[23] 夏丁胤. 互联网图像高效标注和解译的关键技术研究[D]. 杭州: 浙江大学, 2010.

[24] DAHL G E, YU Dong, DENG Li, *et al*. Large vocabulary continuous speech recognition with context-dependent DBN-HMMS[C]//Proc of IEEE International Conference on Acoustics, Speech and Signal Processing. 2011:4688-4691.

[25] MOHAMED A, SAINATH T N, DAHL G E, *et al*. Deep belief networks using discriminative features for phone recognition[C]//Proc of IEEE International Conference on Acoustics, Speech, and Signal Processing. 2011:5060-5063.

[26] NAIR V, HINTON G E. 3D object recognition with deep belief nets [C]//Advances in Neural Information Processing Systems. 2009: 1339-1347.

[27] FASEL I, BERRY J. Deep belief networks for real-time extraction of tongue contours from ultrasound during speech[C]//Proc of the 20th International Conference on Pattern Recognition. Stroudsburg, PA: Association for Computational Linguistics, 2010:1493-1496.

[28] DESELAERS T, HASAN S, BENDER O, *et al*. A deep learning approach to machine transliteration[C]//Proc of the 4th Workshop on Statistical Machine Translation. 2009:233-241.

[29] DENG Li, SELTZER M L, YU Dong, *et al*. Binary coding of speech spectrograms using a deep auto-encoder[C]//Proc of the 11th Annual Conference of International Speech Communication Association. 2010: 1692-1695.

(上接第 2805 页)

[40] TRIGGS B, McLAUCHLAN P F, HARTLEY R I, *et al*. Bundle adjustment: a modern synthesis[C]//Proc of International Workshop on Vision Algorithms: Theory and Practice. London: Springer-Verlag, 2000: 298-372.

[41] MOURAGNON E, LHUILLIER M, DHOME M, *et al*. Real time localization and 3D reconstruction[C]//Proc of IEEE Conference of Computer Vision and Pattern Recognition. Washington DC: IEEE Computer Society, 2006:363-370.

[42] NARODITSKY O, ZHOU X S, GALLIER J, *et al*. Structure from motion with directional correspondence for visual odometry, MS-CIS-11-15[R]. Pennsylvania: GRASP Laboratory, 2010.

[43] ALISMAIL H, BROWNING B, DIAS M B. Evaluating pose estimation methods for stereo visual odometry on robots[C]//Proc of the 11th International Conference on Intelligent Autonomous Systems. 2010.

[44] TORR P H S, ZISSERMAN A. MLESAC: a new robust estimator with application to estimating image geometry[J]. Computer Vision and Image Understanding, 2000, 78(1):138-156.

[45] KITT B, GEIGER A, LATEGAHN H. Visual odometry based on stereo image sequences with RANSAC-based outlier rejection scheme[C]// Proc of IEEE Intelligent Vehicles Symposium. 2010:486-492.

[46] TICK D, SHEN Jing-lin, GANS N. Fusion of discrete and continuous epipolar geometry for visual odometry and localization[C]//Proc of IEEE International Workshop on Robotic and Sensors Environments. 2010:1-6.

[47] Van HAMME D, VEELAERT P, PHILIPS W. Robust visual odometry using uncertainty models[C]//Proc of the 13th International Conference on Advanced Concepts for Intelligent Vision Systems. Berlin: Springer-Verlag, 2011:1-12.

[48] CALONDER M, LEPETIT V, STRECHA C, *et al*. BRIEF: binary robust independent elementary features[C]//Proc of European Conference on Computer Vision. Berlin: Springer-Verlag, 2010:778-792.

[49] LEUTENEGGER S, CHLI M, SIEGWART R. BRISK: binary robust invariant scalable keypoints[C]//Proc of International Conference on Computer Vision. Berlin: Springer-Verlag, 2011:2548-2555.

[50] RUBLEE E, RABAUD V, KONOLIGE K, *et al*. ORB: an efficient alternative to SIFT or SURF[C]//Proc of International Conference on Computer Vision. Berlin: Springer-Verlag, 2011:2564-2571.

[51] GUIZILINI V, RAMOS F. Visual odometry learning for unmanned aerial vehicles[C]//Proc of IEEE International Conference on Robotics and Automation. 2011:6213-6220.

[52] MIKOLAJCZYK K, TUYTELAARS T, SCHMID C, *et al*. A comparison of affine region detectors[J]. International Journal of Computer Vision, 2005, 65(1-2):43-72.