第41卷第1期

2015年1月

北京工业大学学报

JOURNAL OF BEIJING UNIVERSITY OF TECHNOLOGY

Vol.41 No. 1  
Jan. 2015

深度学习研究综述

尹宝才，王文通，王立春

(北京工业大学城市交通学院多媒体与智能软件技术北京市重点实验室，北京100124)

摘要：鉴于深度学习在学术界和工业界的重要性，依据数据流向对目前有代表性的深度学习算法进行归纳和总  
结，综述了不同类型深度网络的结构及特点.首先介绍了深度学习的概念;然后根据深度学习算法的结构特征，概  
述了前馈深度网络、反馈深度网络和双向深度网络3类主流深度学习算法的网络结构和训练方法;最后介绍了深  
度学习算法在不同数据处理中的最新应用现状及其发展趋势.可以看到:深度学习在不同应用领域都取得了明显  
的优势，但仍存在需要进一步探索的问题，如无标记数据的特征学习、网络模型规模与训练速度精度之间的权衡、  
与其他方法的融合等.

关键词：深度学习；深度神经网络；卷积神经网络；反卷积网络；深度玻尔兹曼机

中图分类号：TP391.41 文献标志码：A 文章编号： 0254 -0037(2015)01 -0048 -12

doi: 10.11936/bjutxb2014100026

Review of Deep Learning

YIN Bao-cai，WANG Wen-tong，WANG Li-chun  
(Beijing Key Laboratory of Multimedia and Intelligent Software Technology，College of Metropolitan Transportation，  
Beijing University of Technology，Beijing 100124，China)

Abstract: Considering deep learning’s importance in academic research and industry application，this  
paper reviews methods and applications of deep learning. First， the concept of deep learning is  
introduced，and the main stream deep learning algorithms are classified into three classes : feed-forward  
deep networks，feed-back deep networks and bi-directional deep networks according to the architectural  
characteristics. Second，network architectures and training methods of the three types of deep networks  
are reviewed. Finally， state-of-the-art applications of mainstream deep learning algorithms is illustrated  
and trends of deep learning is concluded. Although deep learning algorithms outperform traditional  
methods in many fields， there are still many issues， such as feature learning on unlabeled data; the  
balance among network scale，training speed and accuracy; and model fusion.

Key words: deep learning ； deep neural networks; convolutional neural network ； deconvolutional  
network ； deep Boltzmann machines

**1深度学习**

深度学习是机器学习领域一个新的研究方向，  
近年来在语音识别、计算机视觉等多类应用中取得

突破性的进展[1-20].其动机在于建立模型模拟人类  
大脑的神经连接结构，在处理图像、声音和文本这些  
信号时，通过多个变换阶段分层对数据特征进行描  
述[21-22]，进而给出数据的解释.以图像数据为例，灵

收稿日期： 2014-09-05

基金项目：国家自然科学基金资助项目（61390512)

作者简介：尹宝才（1963—），男，教授，主要从事数字多媒体技术、多功能感知技术、虚拟现实与图形学方面的研究，E-mail:

ybc@ bjut. edu. cn

第1期

尹宝才，等：深度学习研究综述

49

长类的视觉系统中对这类信号的处理依次为:首先  
检测边缘、初始形状，然后再逐步形成更复杂的视觉  
形状[22]，同样地，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，给出数据的  
分层特征表示.

深度学习之所以被称为“深度”，是相对支撑向  
量机（support vector machine, SVM )、提升方法  
(boosting)、最大熵方法等“浅层学习”方法而言的,  
深度学习所学得的模型中,非线性操作的层级数[21]更多.浅层学习依靠人工经验抽取样本特征，网络  
模型学习后获得的是没有层次结构的单层特  
征[23—25];而深度学习通过对原始信号进行逐层特征  
变换，将样本在原空间的特征表示变换到新的特征  
空间，自动地学习得到层次化的特征表示,从而更有  
利于分类或特征的可视化[26].深度学习理论的另  
外一个理论动机是:如果一个函数可用k层结构以  
简洁的形式表达,那么用k -1层的结构表达则可能  
需要指数级数量的参数（相对于输人信号），且泛化  
能力不足[21，27].

深度学习的概念最早由多伦多大学的G. E.  
Hinton等[26]于2006年提出，指基于样本数据通过  
一定的训练方法得到包含多个层级的深度网络结构  
的机器学习过程[21].传统的神经网络随机初始化  
网络中的权值,导致网络很容易收敛到局部最小值,  
为解决这一问题,Hinton提出使用无监督预训练方  
法优化网络权值的初值,再进行权值微调的方法，拉  
开了深度学习的序幕.

深度学习所得到的深度网络结构包含大量的单  
一元素（神经元），每个神经元与大量其他神经元相  
连接,神经元间的连接强度（权值）在学习过程中修  
改并决定网络的功能.通过深度学习得到的深度网  
络结构符合神经网络的特征[28],因此深度网络就是  
深层次的神经网络，即深度神经网络（deep neural  
networks, DNN).

深度神经网络是由多个单层非线性网络叠加而  
成的[21，29],常见的单层网络按照编码解码情况分为  
3类:只包含编码器部分、只包含解码器部分、既有  
编码器部分也有解码器部分.编码器提供从输人到  
隐含特征空间的自底向上的映射,解码器以重建结  
果尽可能接近原始输人为目标将隐含特征映射到输  
人空间[30].深度神经网络分为以下3类（如图1所  
示) .

1. 前馈深度网络（feed-forward deep networks,  
   FFDN),由多个编码器层叠加而成,如多层感知机

(multi-layer perceptrons, MLP)[31-32]、卷积神经网络  
(convolutional neural networks, CNN)[33-34]等.

1. 反馈深度网络（feed-back deep networks,  
   FBDN),由多个解码器层叠加而成,如反卷积网络  
   (deconvolutional networks, DN)[30]、层次稀疏编码网  
   络（hierarchical sparse coding, HSC)[35]等.
2. 双向深度网络（bi-directional deep networks,  
   BDDN)，通过叠加多个编码器层和解码器层构成  
   (每层可能是单独的编码过程或解码过程,也可能  
   既包含编码过程也包含解码过程），如深度玻尔兹  
   曼机（deep Boltzmann machines, DBM )[36-37]、深度信  
   念网络（deep belief networks , DBN)[26]、找式自编码  
   器(stacked auto-encoders , SAE)[38]等.

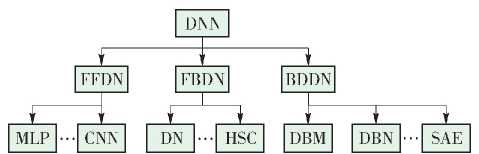


图1深度神经网络分类结构  
Fig. 1 Classification of deep neural networks

2前馈深度网络

前馈神经网络是最初的人工神经网络模型之  
一.在这种网络中，信息只沿一个方向流动,从输人  
单元通过一个或多个隐层到达输出单元，在网络中  
没有封闭环路.典型的前馈神经网络有多层感知  
机[29-30]和卷积神经网络[32-33]等.

F. Rosenblatt[39]提出的感知机是最简单的单层  
前向人工神经网络,但随后M. Minsky等[40]证明单  
层感知机无法解决线性不可分问题（如异或操作），  
这一结论将人工神经网络研究领域引人到一个低潮  
期，直到研究人员认识到多层感知机可解决线性不  
可分问题[31-32],以及反向传播算法与神经网络结合  
的研究[41—^]使得神经网络的研究重新开始成为热  
点.但是由于传统的反向传播算法[41—«]具有收敛速  
度慢、需要大量带标签的训练数据、容易陷人局部最  
优等缺点，多层感知机的效果并不是十分理想.

1984年日本学者K. Fukushima等基于感受野  
概念[45]提出的神经认知机可看作卷积神经网络的  
一种特例[45] , Y. Lecun等[33-34]提出的卷积神经网  
络是神经认知机的推广形式.卷积神经网络是由多  
个单层卷积神经网络组成的可训练的多层网络结  
构.每个单层卷积神经网络包括卷积、非线性变换

50

北京工业大学学报

2015年

和下采样3个阶段[46]，其中下采样阶段不是每层都  
必需的.每层的输人和输出为一组向量构成的特征  
图（feature map)(第一层的原始输人信号可以看作  
一个具有高稀疏度的高维特征图）.例如,输人部分  
是一张彩色图像，每个特征图对应的则是一个包含  
输人图像彩色通道的二维数组（对于音频输人，特

征图对应的是一维向量;对于视频或立体影像,对应  
的是三维数组）；对应的输出部分，每个特征图对应  
的是表示从输人图片所有位置上提取的特定特征.  
2.1单层卷积神经网络

卷积、非线性变换和下采样3个阶段构成的单  
层卷积神经网络如图2所示.

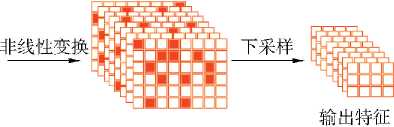


图2单层卷积神经网络的3个阶段

Fig. 2 Three phases of a single layer convolutional neural network

卷积阶段,通过提取信号的不同特征实现输人  
信号进行特定模式的观测.其观测模式也称为卷积  
核,其定义源于由D. H. Hubel等[44]基于对猫视觉  
皮层细胞研究提出的局部感受野概念.每个卷积核  
检测输人特征图上所有位置上的特定特征，实现同  
一个输人特征图上的权值共享[34].为了提取输人  
特征图上不同的特征，使用不同的卷积核进行卷积  
操作.

卷积阶段的输人是由~个伊％大小的二维  
特征图构成的三维数组.每个特征图记为&该阶  
段的输出y也是个三维数组，由叫个m2 X m3大小  
的特征图构成.在卷积阶段,连接输人特征图^和  
输出特征图y.的权值记为％.,即可训练的卷积核  
(局部感受野[44,46])，卷积核的大小为& X 输出  
特征图为

y. = b +移wy.\*① ⑴

式中：\*为二维离散卷积运算符;b.是可训练的偏置  
参数.

非线性阶段,对卷积阶段得到的特征按照一定  
的原则进行筛选，筛选原则通常采用非线性变换的  
方式，以避免线性模型表达能力不够的问题.

非线性阶段将卷积阶段提取的特征作为输人,  
进行非线性映射R = h(y).传统卷积神经网络中非  
线性操作采用sigmoid、tanh或softsign等饱和非线  
性（saturating nonlinearities)函数[47],近几年的卷积  
神经网络中多采用不饱和非线性（non-saturating  
nonlinearity)函数 ReLU( rectified linear units)[1,48-50].  
在训练梯度下降时，ReLU比传统的饱和非线性函

数有更快的收敛速度，因此在训练整个网络时，训练  
速度也比传统的方法快很多[1]. 4种非线性操作函  
数的公式为

sigmoid：

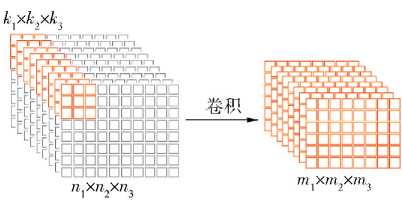
|  |  |  |  |
| --- | --- | --- | --- |
|  | R- | 1 | (2) |
|  | =1 +e-y |
| tanh: | R = | ey -e-y | (3) |
|  | "ey +e-y |
| softsign: | R | y | (4) |
|  | '1 + 1 y 1 |
| ReLU： | R= | max(0 ,y) | (5) |

其函数形态如图3所示.

下采样阶段,对每个特征图进行独立操作，通常  
采用平均池化（average pooling)或者最大池化（max  
pooling)的操作.平均池化依据定义的邻域窗口计  
算特定范围内像素的均值PA,邻域窗口平移步长大  
于1(小于等于池化窗口的大小）；最大池化则将均  
值Pa替换为最值Pm输出到下个阶段.池化操作后,  
输出特征图的分辨率降低,但能较好地保持高分辨  
率特征图描述的特征.一些卷积神经网络完全去掉  
下采样阶段，通过在卷积阶段设置卷积核窗口滑动  
步长大于1达到降低分辨率的目的[33,51].

2.2卷积神经网络

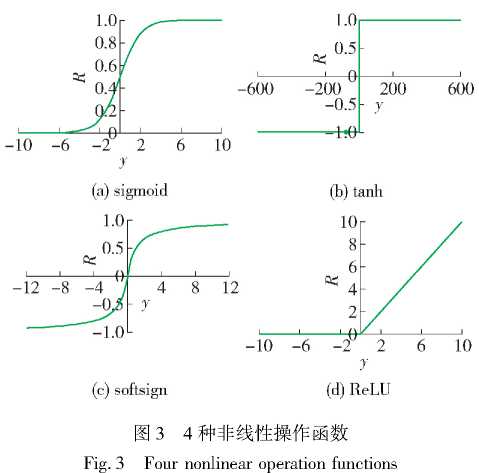
如图4所示,将单层的卷积神经网络进行多次  
堆叠，前一层的输出作为后一层的输人，便构成卷积  
神经网络.其中每2个节点间的连线,代表输人节



第1期

尹宝才，等：深度学习研究综述

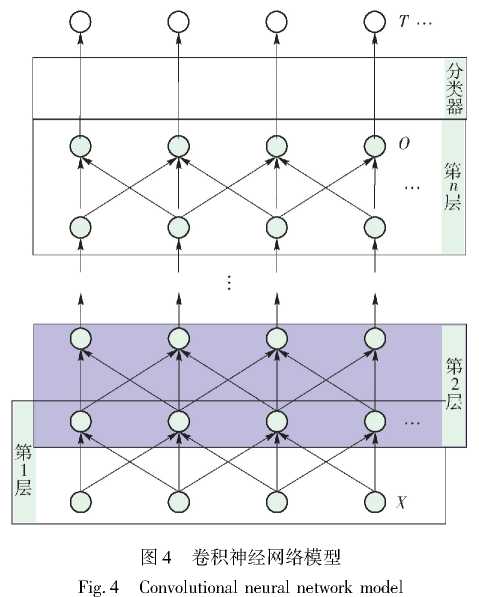
51



点经过卷积、非线性变换、下采样3个阶段变为输出  
节点，一般最后一层的输出特征图后接一个全连接  
层和分类器.为了减少数据的过拟合,最近的一些  
卷积神经网络，在全连接层引人“ Dropout，41,52]或  
“DropConnect冶[53]的方法，即在训练过程中以一•定  
概率P将隐含层节点的输出值(对于“DropConnect”  
为输人权值）清0,而用反向传播算法更新权值时,  
不再更新与该节点相连的权值.但是这2种方法都  
会降低训练速度[1,48,53].

在训练卷积神经网络时,最常用的方法是采用  
反向传播法则[4243,54]以及有监督的训练方式,算法  
流程如图5所示.网络中信号是前向传播的，即从  
输人特征向输出特征的方向传播，第1层的输人X,  
经过多个卷积神经网络层,变成最后一层输出的特

征图O.将输出特征图O与期望的标签T进行比  
较，生成误差项E.通过遍历网络的反向路径，将误



差逐层传递到每个节点，根据权值更新公式（式  
(6))，更新相应的卷积核权值％.在训练过程中，  
网络中权值的初值通常随机初始化（也可通过无监  
督的方式进行预训练[55]),网络误差随迭代次数的  
增加而减少,并且这一过程收敛于一个稳定的权值  
集合，额外的训练次数呈现出较小的影响.

对于卷积网络的任意一层L,其第i个输人特征  
X,和第j个输出特征Y.之间的权值w.的更新公  
式[43]为

驻*w*.二琢啄.*X*,(6)

当L层是卷积网络的最后一层时，如图6(a)所  
示，啄为

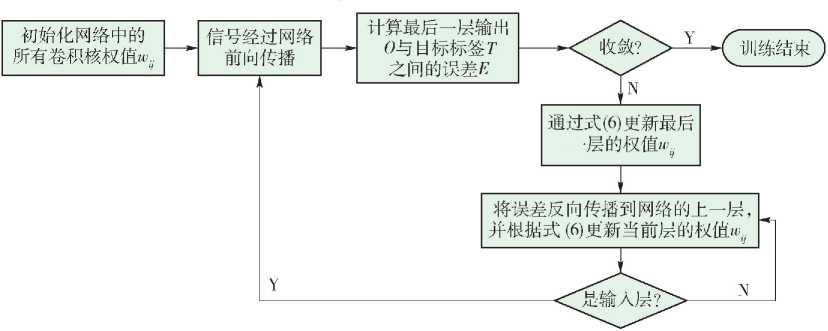


图5卷积神经网络训练过程  
Fig. 5 Training convolutional neural network

52

北京工业大学学报

2015年

啄.=(T,. - Yj) h'L (Xi) (7)

式中：T为第j个预期标签;h'(x)为非线性映射函数  
的导数;j = 1,2,…,Nl.

式(6)中，当L层不是最后一层时,如图6(b)  
所示,L + 1层是其下一层，则啄为

Nl+1

啄=hL(Xi) X 啄 mWm(8)

式中：Nl + i为第L + 1层输出特征的数目；m = 1,  
2，…,Nl + i ;w.m为L层的第j个输出（作为L + 1层的  
第j个输人）与L + 1层第m个输出之间的权值.

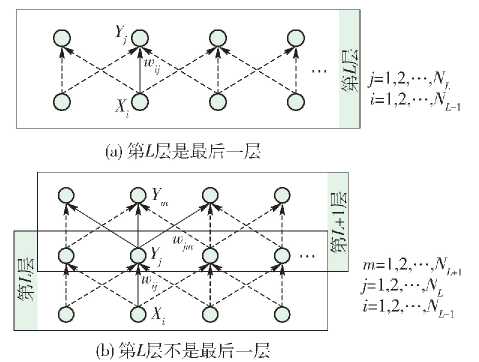


图6卷积神经网络第L层权值w.更新  
(实线为与计算相关的连接关系）  
Fig. 6 Update w-^, weight of layer L in CNN

2.3卷积神经网络的特点

卷积神经网络的特点在于,采用原始信号（一  
般为图像）直接作为网络的输人,避免了传统识别  
算法中复杂的特征提取和图像重建过程;局部感受  
野方法获取的观测特征与平移、缩放和旋转无关.  
卷积阶段利用权值共享结构减少了权值的数量进而  
降低了网络模型的复杂度,这一点在输人特征图是  
高分辨率图像时表现得更为明显.同时,下采样阶  
段利用图像局部相关性的原理对特征图进行子抽  
样,在保留有用结构信息的同时有效地减少数据处  
理量.

3反馈深度网络

与前馈网络不同，反馈网络并不是对输人信号  
进行编码，而是通过解反卷积[30]或学习数据集的  
基[35,56]，对输人信号进行反解.前馈网络是对输人  
信号进行编码的过程，而反馈网络则是对输人信号  
解码的过程.

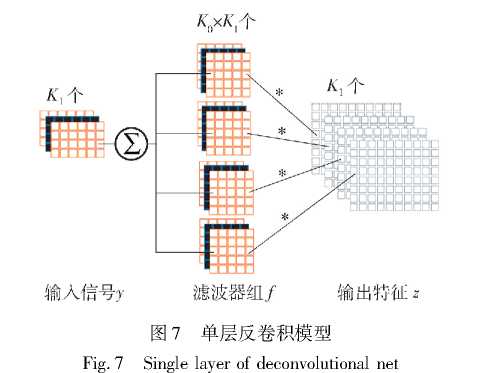
典型的反馈深度网络有反卷积网络[30]、层次稀

疏编码网络[35]等.

以反卷积网络为例，M. D. Zeiler等[30]提出的  
反卷积网络模型和Y. LeCun等[33鄄34]提出的卷积神  
经网络思想类似,但在实际的结构构件和实现方法  
上有所不同.卷积神经网络是一种自底向上的方  
法,该方法的每层输人信号经过卷积、非线性变换和  
下采样3个阶段处理,进而得到多层信息.相比之  
下,反卷积网络模型的每层信息是自顶向下的，组合  
通过滤波器组学习得到的卷积特征来重构输人信  
号.层次稀疏编码网络和反卷积网络非常相似，只  
是在反卷积网络中对图像的分解采用矩阵卷积的形  
式,而在稀疏编码中采用矩阵乘积的方式[35].

3.1单层反卷积网络

反卷积网络是通过先验学习，对信号进行稀疏  
分解和重构的正则化方法.图7所示是一个单层反  
卷积网络模型，输人信号y由心个特征通道yi,  
y2,…,Tk。组成，其中任意一个通道yc可看作&个隐  
层特征图Zk与滤波器组fk,c (个数为K。伊Ki)的  
卷积.



Xzk\*/k,c=yc(9)

由于式(9)是一个欠定（未知数的个数多于方  
程个数）的函数，为了求得其唯一解，需要引人一个  
关于特征图A的正则项，且该正则项使得特征图&  
趋于稀疏.于是代价函数为

姿K2

Ci(y)=姿移

2 c = 1

Ki

移 Zk \*fk,c -Tc

k = 1

移 IZk IP

k=1

(10)

式中:第1项为输人图像与重建结果的误差;第2项  
为特征图的稀疏程度，为p范数,一般取p = 1;姿为  
平衡重建误差和特征图稀疏度的权重系数.

第1期

尹宝才，等：深度学习研究综述

53

3.2反卷积网络

通过将3. 1节所述单层反卷积网络进行多层叠  
加，可得到反卷积网络，如图8所示.多层模型中，  
在学习滤波器组的同时进行特征图的推导，第L层  
的特征图和滤波器是由第L -1层的特征图通过反  
卷积计算分解获得.

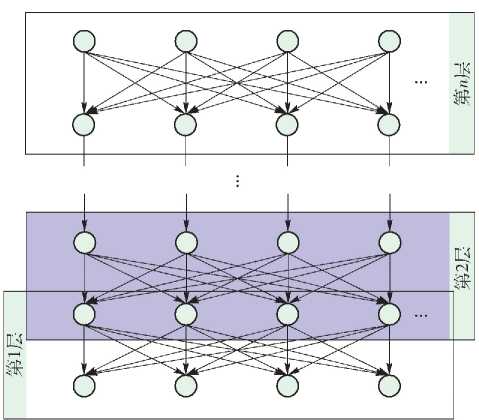


图8反卷积网络模型

Fig. 8 Decovolutional network model

反卷积网络训练时,使用一组不同的信号y =

! y1 ,y2，…,y11，求解 argmin^C/y)，利用式（11)，进  
行滤波器组/和特征图z的迭代交替优化[3°].训练  
从第1层开始，采用贪心算法,逐层向上进行优化,  
各层间的优化是独立的.

在反卷积网络中，单层网络的代价函数（为当  
前层所有输人信号的代价函数之和）为

Ci(y)=务移移移 gk，c(zi，. \*/k，c)-zC，；-1 +

移移 Izk’ （11)

式中第1项为前一层与当前层重建目标的误差.其  
中:zk，i是当前层的特征图;/k，c是当前层的滤波器组;  
z：，;-,是前一层的特征图;gk，c表示同一层中输人特征  
图与输出特征图之间的连通情况,是一个固定的二  
值矩阵.通常假定第1层是全连接的,后边的层为  
稀疏连接.第2项为特征图的稀疏程度;姿是平衡  
重建误差和特征图稀疏度的权重系数.

3.3反卷积网络的特点

反卷积网络的特点在于,通过求解最优化输人  
信号分解问题计算特征，而不是利用编码器进行近  
似,这样能使隐层的特征更加精准,更有利于信号的

分类或重建.

4双向深度网络

双向网络由多个编码器层和解码器层叠加形  
成,每层可能是单独的编码过程或解码过程,也可能  
同时包含编码过程和解码过程.双向网络的结构结  
合了编码器和解码器2类单层网络结构，双向网络  
的学习则结合了前馈网络和反馈网络的训练方法,  
通常包括单层网络的预训练和逐层反向迭代误差2  
个部分，单层网络的预训练多采用贪心算法:每层使  
用输人信号11与权值w计算生成信号/i + 1传递到下  
一层，信号/i + 1再与相同的权值w计算生成重构信  
号1L映射回输人层,通过不断缩小与1L间的误  
差，训练每层网络；网络结构中各层网络结构都经过  
预训练之后,再通过反向迭代误差对整个网络结构  
进行权值微调.其中单层网络的预训练是对输人信  
号编码和解码的重建过程,这与反馈网络训练方法  
类似;而基于反向迭代误差的权值微调与前馈网络  
训练方法类似.

典型的双向深度网络有深度玻尔兹曼机[36鄄37]、  
深度信念网络[26]、栈式自编码器[38]等.

以深度玻尔兹曼机为例，深度玻尔兹曼机由R.  
Salakhutdinov等[36]提出，它由多层受限玻尔兹曼机  
(restricted Boltzmann machine， RBM ) [57—59]叠加

构成.

4.1受限玻尔兹曼机

玻尔兹曼机（Boltzmann machine，BM )是一•种  
随机的递归神经网络，由G. E. Hinton等[6°-62]提  
出，是能通过学习数据固有内在表示、解决复杂学习  
问题的最早的人工神经网络之一.玻尔兹曼机由二  
值神经元构成，每个神经元只取0或1两种状态，状  
态1代表该神经元处于激活状态，0表示该神经元  
处于抑制状态.然而，即使使用模拟退火算法，这个  
网络的学习过程也十分慢.

Hinton等提出的受限玻尔兹曼机[57鄄59]去掉了玻  
尔兹曼机同层之间的连接,从而大大提高了学习效  
率.受限玻尔兹曼机分为可见层v以及隐层h，可见  
层和隐层的节点通过权值w相连接，2层节点之间  
是全连接，同层节点间互不相连，如图9所示.

受限玻尔兹曼机一种典型的训练方法如图10  
所示,首先随机初始化可见层,然后在可见层与隐层  
之间交替进行吉布斯采样:用条件分布概率P(h I v)  
计算隐层;再根据隐层节点，同样用条件分布概率  
P( v I h)来计算可见层;重复这一采样过程直到可见

54

北京工业大学学报

2015年

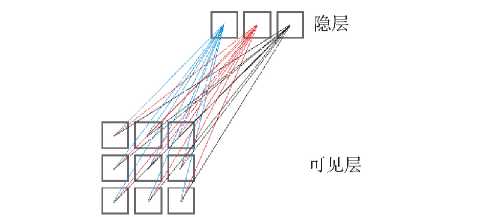
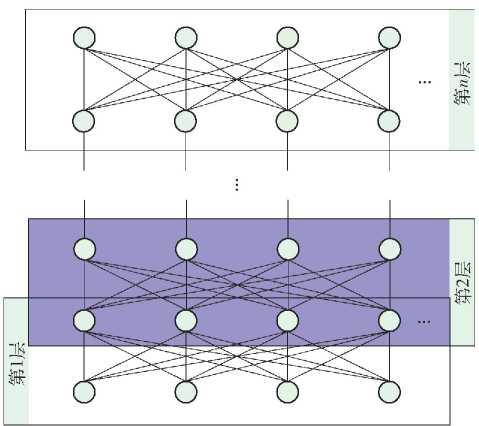


图9受限玻尔兹曼机(单层深度玻尔兹曼机）

Fig. 9 Restricted Boltzmann machine ( single layer  
of deep Boltzmann machines)

层和隐层达到平稳分布.而Hinton提出了一种快速  
算法，称作对比离差（contrastive diverg ence, CD)学  
习算法[37,59,63].这种算法使用训练数据初始化可见



层，只需迭代k次上述采样过程（即每次迭代包括从  
可见层更新隐层，以及从隐层更新可见层），就可获  
得对模型的估计(通常k = 1).

*PW P(h^)*

*p(hk)*

隐层 Coo) (oo) Coo) (oo)

*US …X、*

*(QOO)*

可见层(550 *(ooQ (pm*

*P^I^)*

图10受限玻尔兹曼机的训练过程  
Fig. 10 Training procedure of restricted Boltzmann machine

4.2深度玻尔兹曼机

将多个受限玻尔兹曼机堆叠，前一层的输出作  
为后一层的输人,便构成了深度玻尔兹曼机,如图

11所示.网络中所有节点间的连线都是双向的.

图 11 深度玻尔兹曼机

Fig. 11 Deep Boltzmann machines

深度玻尔兹曼机训练分为2个阶段:预训练阶  
段和微调阶段，如图12所示.

在预训练阶段，采用无监督的逐层贪心训练方  
法来训练网络每层的参数，即先训练网络的第1个  
隐含层，然后接着训练第2,3，…个隐含层,最后用  
这些训练好的网络参数值作为整体网络参数的初始  
值.预训练之后，将训练好的每层受限玻尔兹曼机  
叠加形成深度玻尔兹曼机,利用有监督的学习对网  
络进行训练（一般采用反向传播算法）.

由于深度玻尔兹曼机随机初始化权值以及微调  
阶段采用有监督的学习方法,这些都容易使网络陷

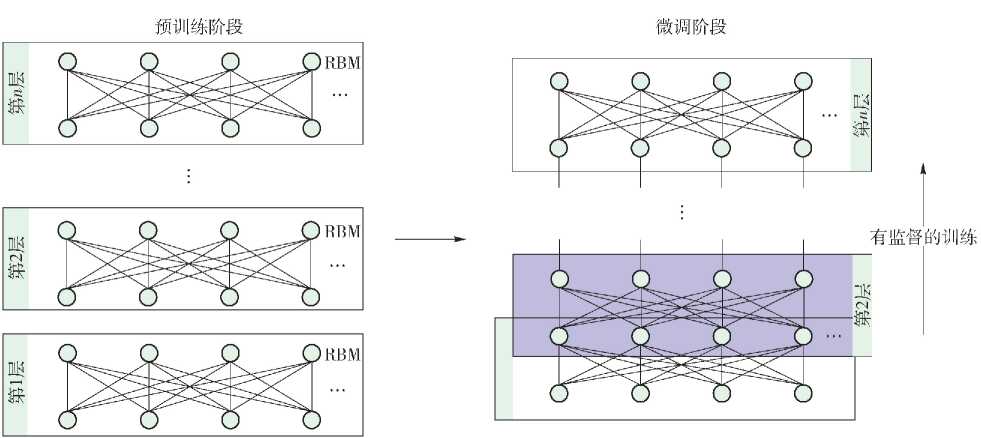


图12深度玻尔兹曼机逐层贪心训练方法

Fig. 12 Greedy layer-wise pre-training of DBM

第1期

尹宝才，等：深度学习研究综述

55

于避免陷人局部最小值问题[64].

5深度学习应用

深度学习目前在很多领域都优于过去的方法,  
下面根据所处理数据类型的不同，对深度学习的应  
用进行介绍.

**5.1深度学习在语音识别、合成及机器翻译中的应用**

微软研究人员使用深度信念网络对数以千计的  
senones ( 一种比音素小很多的建模单兀）直接建模,  
提出了第1个成功应用于大词汇量语音识别系统的  
上下文相关的深层神经网络-隐马尔可夫混合模型  
(CD-DNN-HMM)[2],比之前最领先的基于常规CD-  
GMM-HMM的大词汇量语音识别系统相对误差率减  
少16%以上.

随后又在含有300 h语音训练数据的  
Switchboard标准数据集上对CD-DNN-HMM模型进  
行评测[65].基准测试字词错误率为18. 5%，与之前  
最领先的常规系统相比，相对错误率减少了 33%.

H. Zen等[3]提出一种基于多层感知机的语音  
合成模型.该模型先将输人文本转换为一个输人特  
征序列，输人特征序列的每帧分别经过多层感知机  
映射到各自的输出特征,然后采用文献[66]中的算  
法，生成语音参数,最后经过声纹合成生成语音.训  
练数据包含由一名女性专业演讲者以美国英语录制  
的3. 3万段语音素材，其合成结果的主观评价和客  
观评价均优于基于HMM方法的模型.

K. Cho等[67]提出一种基于循环神经网络  
(recurrent neural network, RNN)的向量化定长表示  
模型（RNNenc模型），应用于机器翻译.该模型包  
含2个RNN, —个RNN用于将一组源语言符号序  
列编码为一组固定长度的向量，另一个RNN将该向  
量解码为一组目标语言的符号序列.

在该模型的基础上，D. Bahdanau等[4]克服了  
文献[67]中固定长度的缺点（固定长度是其效果提  
升的瓶颈），提出了 RNNsearch的模型.该模型在翻  
译每个单词时,根据该单词在源文本中最相关信息  
的位置以及已翻译出的其他单词，预测对应于该单  
词的目标单词.该模型包含一个双向RNN作为编  
码器，以及一个用于单词翻译的解码器.在进行目  
标单词位置预测时,使用一个多层感知机模型进行  
位置对齐.采用BLEU评价指标,RNNsearch模型在  
ACL2014机器翻译研讨会（ACL WMT 2014)提供的  
英/法双语并行语料库[68]上的翻译结果评分均高于  
RNNenc模型的评分，略低于传统的基于短语的翻

译系统Moses[69](本身包含具有4. 18亿个单词的多  
语言语料库）.另外,在剔除包含未知词汇语句的测  
试预料库上，RNNsearch的评分甚至超过了 Moses.  
5.2深度学习在图像分类及识别中的应用  
5. 2.1深度学习在大规模图像数据集中的应用

A. Krizhevsky等[1]首次将卷积神经网络应用  
于ImageNet大规模视觉识别挑战赛（ImageNet large  
scale visual recognition challenge , ILSVRC)[70]中，所  
训练的深度卷积神经网络[1]在ILSVRC— 2012挑战  
赛中，取得了图像分类和目标定位任务的第一.其  
中，图像分类任务中，前5选项错误率为15.3% ,远  
低于第2名的26. 2%的错误率;在目标定位任务  
中，前5选项错误率34%，也远低于第2名的50%.

在 ILSVRC—2013比赛中，M. D. Zeiler 等[5]采  
用卷积神经网络的方法,对文献[1 ]的方法进行了  
改进,并在每个卷积层上附加一个反卷积层用于中  
间层特征的可视化[5,30]，取得了图像分类任务的第  
一名.其前5选项错误率为11.7%,如果采用  
ILSVRC—2011数据进行预训练,错误率则降低到  
11.2%.在目标定位任务中，P. Sermanet等[6]采用  
卷积神经网络结合多尺度滑动窗口的方法,可同时  
进行图像分类、定位和检测，是比赛中唯一一个同时  
参加所有任务的队伍.多目标检测任务中，获胜队  
伍的方法在特征提取阶段没有使用深度学习模型，  
只在分类时采用卷积网络分类器进行重打分[7].

在ILSVRC—2014比赛中，几乎所有的参赛队  
伍都采用了卷积神经网络及其变形方法[7].其中  
GoogLeNet小组采用卷积神经网络结合Hebbian理  
论提出的多尺度的模型，以6. 7%的分类错误,取得  
图形分类“指定数据”组的第一名；CASIAWS小组采  
用弱监督定位和卷积神经网络结合的方法,取得图形  
分类“额外数据”组的第一名，其分类错误率为11%.

在目标定位任务中，VGG小组在深度学习框架  
Caffe的基础上，采用3个结构不同的卷积神经网络  
进行平均评估，以26%的定位错误率取得“指定数  
据”组的第一名；Adobe组选用额外的2 000类  
ImageNet数据训练分类器,采用卷积神经网络架构  
进行分类和定位，以30%的错误率，取得了“额外数  
据”组的第一名.

在多目标检测任务中，NUS小组采用改进的卷

积神经网络网中网（network in network,

NIN)[8]与多种其他方法融合的模型，以37%的平均  
准确率（mean average precision, mAP)取得“提供数  
据”组的第一名；GoogLeNet以44%的平均准确率取

56

北京工业大学学报

2015年

得“额外数据”组的第一名.

从深度学习首次应用于ILSVRC挑战赛并取得  
突出的成绩，到2014年挑战赛中几乎所有参赛队伍  
都采用深度学习方法，并将分类识错率降低到  
6.7%，可看出深度学习方法相比于传统的手工提取  
特征的方法在图像识别领域具有巨大优势.

5.2.2深度学习在人脸识别中的应用

基于卷积神经网络的学习方法,香港中文大学  
的 DeepID 项目[9]以及 Facebook 的 DeepFace 项  
目[10]在户外人脸识别（labeled faces in the wild,  
LFW)数据库[71]上的人脸识别正确率分别达  
97郾45%>和97郾35% ,只比人类识别97郾5%>[72]的正  
确率略低一点点.DeepID项目采用4层卷积神经网  
络（不含输人层和输出层）结构,DeepFace采用5层  
卷积神经网络（不含输人层和输出层,其中后3层  
没有采用权值共享以获得不同的局部统计特征）  
结构.

之后,采用基于卷积神经网络的学习方法,香港  
中文大学的DeepID2项目[11]将识别率提高到了  
99, 15%，超过目前所有领先的深度学习[9—10]和非深  
度学习算法[73]在LFW数据库上的识别率以及人类  
在该数据库的识别率[72]. DeepID2项目采用和  
DeepID项目类似的深度结构,包含4个卷积层，其  
中第3层采用2伊2邻域的局部权值共享，第4层没  
有采用权值共享,且输出层与第3、4层都全连接.  
5.3深度学习在视频分类及行为识别中的应用

A. Karpathy等[12]基于卷积神经网络提供了一  
种应用于大规模视频分类上的经验评估模型，将  
Sports-1M数据集[12]的100万段YouTube视频数据  
分为487类.该模型使用4种时空信息融合方法用  
于卷积神经网络的训练，融合方法包括单帧（single  
frame)、不相邻两帧（late fusion)、相邻多帧（early  
fusion)以及多阶段相邻多帧（slow fusion);此外提出  
了一种多分辨率的网络结构，大大提升了神经网络  
应用于大规模数据时的训练速度.该模型在Sports -  
1M上的分类准确率达63. 9% ,相比于基于人工特  
征的方法(55. 3%)，有很大提升.此外，该模型表现  
出较好的泛化能力，单独使用slow fusion融合方法  
所得模型在UCF-101动作识别数据集[74]上的识别  
率为65. 4%，而该数据集的基准识别率为43. 9%.

S. Ji等[13]提出一个三维卷积神经网络模型用  
于行为识别.该模型通过在空间和时序上运用三维  
卷积提取特征,从而获得多个相邻帧间的运动信息.  
该模型基于输人帧生成多个特征图通道,将所有通

道的信息结合获得最后的特征表示.该三维卷积神  
经网络模型在TRECVID数据上优于其他方法，表明  
该方法对于真实环境数据有较好的效果;该模型在  
KTH数据上的表现,逊于其他方法,原因是为了简  
化计算而缩小了输人数据的分辨率.

M. Baccouche等[14]提出一•种时序的深度学习  
模型，可在没有任何先验知识的前提下，学习分类人  
体行为.模型的第一步，是将卷积神经网络拓展到  
三维，自动学习时空特征.接下来使用RNN方法训  
练分类每个序列.该模型在KTH上的测试结果优  
于其他已知深度模型，KTH1和KTH2上的精度分  
别为94. 39% 和92. 17%.

事实上，深度学习的应用远不止这些,但是本文  
只是分别从数据的维度上（音频文本,一维；图像,  
二维;视频,三维）对深度学习的典型应用进行详细  
介绍，目的在于突出深度学习带来的优越性能以及  
其对不同数据的应用能力.其他应用还包括图像超  
分辨率重建[15—16]、纹理识别[17]、行人检测[18]、场景  
标记[19]、门牌识别[20]等.

**6深度学习的问题及趋势**

深度学习算法在计算机视觉（图像识别、视频  
识别等）和语音识别中的应用，尤其是大规模数据  
集下的应用取得突破性的进展,但仍有以下问题值  
得进一步研究：

1. 无标记数据的特征学习.目前,标记数据的  
   特征学习仍然占据主导地位[1，7],而真实世界存在  
   着海量的无标记数据，将这些无标记数据逐一添加  
   人工标签，显然是不现实的.所以，随着数据集和存  
   储技术的发展，必将越来越重视对无标记数据的特  
   征学习，以及将无标记数据进行自动添加标签技术  
   的研究.
2. 模型规模与训练速度、训练精度之间的权  
   衡.一般地,相同数据集下,模型规模越大,训练精  
   度越高，训练速度会越慢.例如一些模型方法采用  
   ReLU非线性变换、GPU运算，在保证精度的前提  
   下,往往需要训练5~7 d[1,4].虽然离线训练并不影  
   响训练之后模型的应用,但是对于模型优化,诸如模  
   型规模调整、超参数设置、训练时调试等问题，训练  
   时间会严重影响其效率.故而,如何在保证一定的  
   训练精度的前提下,提高训练速度，依然是深度学习  
   方向研究的课题之一.
3. 与其他方法的融合.从上述应用实例中可  
   发现,单一的深度学习方法，往往并不能带来最好的

第1期

尹宝才，等：深度学习研究综述

57

效果，通常融合其他方法或多种方法进行平均打分,  
会带来更高的精确率.因此,深度学习方法与其他  
方法的融合，具有一定的研究意义.

**参考文献：**

1. KRIZHEVSKY A, SUTSKEVER I, HINTON G E.

Imagenet classification with deep convolutional neural  
networks [C]椅 Advances in Neural Information Processing  
Systems. Red Hook, NY: Curran Associates, 2012 : 1097-  
1105.

1. DAHL G E, YU D, DENG L, et al. Context-dependent  
   pre-trained deep neural networks for large-vocabulary  
   speech recognition [ J]. Audio, Speech, and Language  
   Processing, IEEE Transactions on, 2012, 20(1) : 3042.
2. ZEN H, SENIORA, SCHUSTER M. Statistical parametric  
   speech synthesis using deep neural networks [ C]椅  
   Acoustics, Speech and Signal Processing (ICASSP), 2013  
   IEEE International Conference on. Piscataway, NJ: IEEE,  
   2013: 7962-7966.
3. BAHDANAU D, CHO K, BENGIO Y. Neural machine  
   translation by jointly learning to align and translate [J].

CoRR, 2014: abs/1409. 0473.

1. ZEILER M D, FERGUS R. Visualizing and understanding  
   convolutional neural networks [ J]. CoRR, 2013: abs/  
   1311. 2901.
2. SERMANET P, EIGEN D, ZHANG X, et al. Overfeat:  
   integrated recognition, localization and detection using  
   convolutional networks [ J]. CoRR, 2013: abs/1312.  
   6229.
3. RUSSAKOVSKY O, DENG J, SU H, et al. ImageNet  
   large scale visual recognition challenge [J]. CoRR, 2014:  
   abs/1409. 0575.
4. LIN M, CHEN Q, YAN S. Network in network [J].  
   CoRR, 2013: abs/1312. 4400.
5. SUN Y, WANG X, TANG X. Deep learning face  
   representation from predicting 10, 000 classes [ C]椅  
   Proceedings of the IEEE Conference on Computer Vision  
   and Pattern Recognition. Piscataway, NJ: IEEE, 2014:  
   1891-1898.
6. TAIGMAN Y, YANG M, RANZATO M A, et al.

Deepface: closing the gap to human-level performance in  
face verification [ C ]椅 Proceedings of the IEEE  
Conference on Computer Vision and Pattern Recognition.  
Piscataway, NJ: IEEE, 2014: 1701-1708.

1. SUN Y, WANG X, TANG X. Deep learning face  
   representation by joint identification-verification [ J ] .

CoRR, 2014: abs/1406. 4773.

1. KARPATHY A, TODERICIG, SHETTYS, et al. Large-

scale video classification with convolutional neural  
networks [ C]椅 IEEE Conference on Computer Vision and

Pattern Recognition ( CVPR). Piscataway, NJ: IEEE,  
2014: 1725-1732.

1. JI S, XU W, YANG M, et al. 3D convolutional neural  
   networks for human action recognition [ J ] . Pattern  
   Analysis and Machine Intelligence, IEEE Transactions  
   on, 2013, 35(1): 221-231.
2. BACCOUCHE M, MAMALET F, WOLF C, et al.

Sequential deep learning for human action recognition[ C]椅  
Human Behavior Understanding. Berlin: Springer, 2011:  
29-39.

1. DONG C, LOY C C, HE K, et al. Learning a deep  
   convolutional network for image super-resolution [C]椅  
   Computer Vision-ECCV 2014. Cham: Springer  
   International Publishing, 2014: 184-199.
2. CUI Z, CHANG H, SHAN S, et al. Deep network  
   cascade for image super-resolution [ C ]椅 Computer  
   Vision-ECCV 2014. Cham: Springer International  
   Publishing, 2014: 49-64.
3. BADRI H, YAHIA H, DAOUDI K. Fast and accurate  
   texture recognition with multilayer convolution and  
   multifractal analysis[C]椅 Computer Vision-ECCV 2014.  
   Cham: Springer International Publishing, 2014: 505-  
   519.
4. ZENG X, OUYANGW, WANG M, et al. Deep learning  
   of scene-specific classifier for pedestrian detection[ C]椅  
   Computer Vision-ECCV 2014. Cham: Springer  
   International Publishing, 2014: 472-487.
5. FARABETC, COUPRIEC, NAJMAN L, et al. Learning

hierarchical features for scene labeling [ J ] . Pattern  
Analysis and Machine Intelligence, IEEE Transactions  
on, 2013, 35(8) : 1915-1929.

1. GOODFELLOW I J, BULATOV Y, IBARZ J, et al.

Multi-digit number recognition from street view imagery  
using deep convolutional neural networks [ J]. CoRR,  
2013: abs/1312. 6082.

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

[22] SERRE T, KREIMAN G, KOUH M, et al. A

quantitative theory of immediate visual recognition [ J] .  
Progress in Brain Research, 2007, 165: 33-56.

[ 23 ] LOWE D G. Distinctive image features from scale-  
invariant keypoints [J]. International Journal of Computer

Vision, 2004, 60(2) : 91-110.

1. DALAL N, TRIGGS B. Histograms of oriented gradients  
   for human detection [ C]椅 Computer Vision and Pattern  
   Recognition, 2005. IEEE Computer Society Conference  
   on. Piscataway, NJ: IEEE, 2005: 886-893.
2. OJALA T, PIETIKAINEN M, MAENPAA T.

Multiresolution gray-scale and rotation invariant texture  
classification with local binary patterns [ J ] . Pattern

58

北京工业大学学报

2015年

Analysis and Machine Intelligence, IEEE Transactions  
on, 2002, 24(7) : 971-987.

1. HINTON G E, OSINDERO S, TEH YW. Afast learning  
   algorithm for deep belief nets [J]. Neural Computation,

2006, 18(7) : 1527-1554.

1. HASTAD J, GOLDMANN M. On the power of small-  
   depth threshold circuits [J]. Computational Complexity,

1991, 1(2) : 113-129.

1. PSALTIS D, SIDERIS A, YAMAMURA A. A  
   multilayered neural network controller [J]. IEEE Control  
   Systems Magazine, 1988, 8(2) : 17-21.
2. BENGIO Y, LAMBLIN P, POPOVICI D, et al. Greedy  
   layer-wise training of deep networks [C]椅 Advances in  
   Neural Information Processing Systems. Cambridge, MA:  
   MIT Press, 2007: 153-160.
3. ZEILER M D, KRISHNAN D, TAYLOR G W, et al.  
   Deconvolutional networks [ C]椅 Computer Vision and  
   Pattern Recognition ( CVPR), 2010 IEEE Conference  
   on. Piscataway, NJ: IEEE, 2010: 2528-2535.
4. HORNIK K, STINCHCOMBE M, WHITE H. Multilayer  
   feedforward networks are universal approximators [ J ] .  
   Neural Networks, 1989, 2(5) : 359-366.
5. GARDNER M W, DORLING S R. Artificial neural

networks ( the multilayer perceptron )——a review of  
applications in the atmospheric sciences [J]. Atmospheric  
Environment, 1998, 32(14/15): 2627-2636.

1. LeCun Y, BOSER B, DENKER J S, et al. Handwritten  
   digit recognition with a back-propagation network [C]椅  
   Advances in Neural Information Processing Systems. San  
   Francisco, CA: Morgan Kaufmann Publishers, 1990:  
   396404.
2. 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.
3. YU K, LIN Y, LAFFERTY J. Learning image  
   representations from the pixel level via hierarchical sparse  
   coding [C]椅 Computer Vision and Pattern Recognition  
   (CVPR), 2011 IEEE Conference on. Piscataway, NJ:  
   IEEE, 2011: 1713-1720.
4. SALAKHUTDINOV R, HINTON G E. Deep Boltzmann  
   machines [ C ]椅 JMLR Workshop and Conference  
   Proceedings Volume 5: AISTATS 2009. Brookline, MA:  
   Microtome Publishing, 2009 : 448 -455.
5. 刘建伟，刘媛，罗雄麟.玻尔兹曼机研究进展[J].计  
   算机研究与发展，2014, 51(1): 1-16.

LIU Jian-wei, LIU Yuan, LUO Xiong-lin. Research and  
development on Boltzmann machine [ J ] . Journal of  
Computer Research and Development, 2014, 51(1): 1-  
16. ( in Chinese)

1. VINCENT P, LAROCHELLE H, BENGIO Y, et al.

Extracting and composing robust features with denoising

autoencoders [ C]椅 Proceedings of the 25 th international  
conference on Machine learning. New York, NY: ACM,  
2008: 1096-1103.

[ 39] ROSENBLATT F. The perceptron: a probabilistic model  
for information storage and organization in the brain [J].  
Psychological Review, 1958, 65(6) : 386.

1. MINSKY M, PAPERT S. Perceptrons [M]. Cambridge,  
   MA: MIT Press, 1969: 105-110.
2. RUMELHART D E, HINTON G E, WILLIAMS R J.

Learning representations by back-propagating errors [ J] .

Nature, 1986, 323: 533-536.

1. LeCun Y. Une procedure d’apprentissage pour reseau a  
   seuil assymetrique [J]. Proceedings of Cognitiva, 1985,  
   85: 599-604.
2. HINTON G E. How neural networks learn from  
   experience [J]. Scientific American, 1992, 267 (3 ):  
   145-151.
3. HUBEL D H, WIESEL T N. Receptive fields, binocular  
   interaction and functional architecture in the cat’s visual  
   cortex [J]. The Journal of Physiology, 1962 , 160 (1):  
   106.
4. FUKUSHIMA K, MIYAKE S. Neocognitron: a new  
   algorithm for pattern recognition tolerant of deformations  
   and shifts in position[J]. Pattern Recognition, 1982, 15  
   (6) : 455469.
5. LeCun Y, KAVUKCUOGLU K, FARABET C.  
   Convolutional networks and applications in vision [C]椅  
   Circuits and Systems ( ISCAS ) , Proceedings of 2010  
   IEEE International Symposium on. Piscataway, NJ:  
   IEEE, 2010: 253-256.
6. GLOROT X, BENGIO Y. Understanding the difficulty of  
   training deep feedforward neural networks [ C ]椅  
   International Conference on Artificial Intelligence and  
   Statistics. Brookline, MA: 2010: 249-256.
7. DAHL G E, SAINATH T N, HINTON G E. Improving

deep neural networks for LVCSR using rectified linear  
units and dropout [C]椅 Acoustics, Speech and Signal  
Processing ( ICASSP ) , 2013 IEEE International

Conference on. Piscataway, NJ: IEEE, 2013: 8609-  
8613.

1. NAIR V, HINTON G E. Rectified linear units improve  
   restricted Boltzmann machines [C]椅 Proceedings of the  
   27th International Conference on Machine Learning  
   (ICML-10). Madison, WI: Omnipress, 2010: 807-814.
2. GLOROT X, BORDES A, BENGIO Y. Deep sparse  
   rectifier networks[ C]椅JMLR Workshop and Conference  
   Proceedings Volume 15: AISTATS 2011. Brookline,  
   MA: Microtome Publishing, 2011: 315-323.
3. SIMARD P Y, STEINKRAUS D, PLATT J C. Best

practices for convolutional neural networks applied to  
visual document analysis [C]椅 Document Analysis and

第1期

尹宝才，等：深度学习研究综述

59

Recognition， 2003. Proceedings Seventh International  
Conference on. Washington, DC: IEEE Computer  
Society, 2003,2: 958-963.

1. HINTON G E，SRIVASTAVA N，KRIZHEVSKY A，et

al. Improving neural networks by preventing co-adaptation  
of feature detectors[J]. CoRR，2012: abs/1207. 0580.

1. WAN L，ZEILER M，ZHANGS，et al. Regularization of  
   neural networks using dropconnect [ C ]椅 Proceedings of  
   the 30th International Conference on Machine Learning.  
   Brookline， MA: Microtome Publishing，2013，28(3):  
   1058-1066.

[ 54 ] BOUVRIE J. Notes on convolutional neural networks  
[ R ] . Massachusetts: Center for Biological and  
Computational Learning，2006: 3844

1. JARRETT K，KAVUKCUOGLU K，RANZATO M，et al.

What is the best multi-stage architecture for object  
recognition? [ C ]椅 Computer Vision， 2009 IEEE 12th  
International Conference on. Piscataway， NJ: IEEE，  
2009: 2146-2153.

1. OLSHAUSEN B A，FIELD D J. Sparse coding with an  
   overcomplete basis set: a strategy employed by V1 ? [ J].  
   Vision Research，1997, 37(23) : 3311-3325.

[ 57 ] SMOLENSKY P. Information processing in dynamical  
systems: foundations of harmony theory [ M ]椅 Rumelhart  
D E， McClelland J L. Parallel Distributed Processing，  
Cambridge，MA: MIT Press，1986: 194-281.

[58] FREUND Y，HAUSSLER D. Unsupervised learning of  
distributions of binary vectors using two layer networks  
[C ]椅 Advances in Neural Information Processing  
Systems. San Francisco，CA: Morgan Kaufmann

Publishers，1994: 912-919.

[ 59 ] HINTON G E. Training products of experts by minimizing  
contrastive divergence [ J ]. Neural Computation， 2002，  
14(8) : 1771-1800.

1. HINTON G E，SEJNOWSKI T J. Optimal perceptual  
   inference [ C ]椅 Proceedings of the IEEE Conference on  
   Computer Vision and Pattern Recognition. Piscataway，  
   NJ: IEEE，1983: 448453.
2. HINTON G E，SEJNOWSKI T J. Analysing cooperative  
   computation [ C ]椅 Proceedings of the Fifth Annual  
   Conference of the Cognitive Science Society. Rochester，  
   NY: Lawrence Erlbaum Associates， 1983.
3. HINTON G E，SEJNOWSKI T J，ACKLEY D H.

Boltzmann machines: constraint satisfaction networks that  
learn [ M ] . Pennsylvania: Department of Computer  
Science， 1984.

[ 63 ] HINTON G. A practical guide to training restricted  
Boltzmann machines [ R ] . Toronto : University of  
Toronto， 2010.

1. ERHAN D，BENGIO Y，COURVILLE A，et al. Why

does unsupervised pre-training help deep learning? [ J].  
The Journal of Machine Learning Research， 2010，11:  
625-660.

1. SEIDE F，LI G，YU D. Conversational speech

transcription using context-dependent deep neural  
networks [ C ]椅 International Speech Communication  
Association. Annual Conference. 12th 2011.

(Interspeech 2011 ). Red Hook， NY: Curran

Associates， 2011: 437-440.

1. TOKUDA K，YOSHIMURA T，MASUKO T，et al.

Speech parameter generation algorithms for HMM-based  
speech synthesis [ C ]椅 Acoustics， Speech， and Signal  
Processing， 2000. Proceedings 2000 IEEE International  
Conference on. Piscataway， NJ: IEEE， 2000: 1315-  
1318.

1. CHO K，van MERRIENBOER B，GULCEHRE C，et al.  
   Learning phrase representations using RNN encoder-  
   decoder for statistical machine translation [ J ]. CoRR，  
   2014: abs/1406. 1078.
2. ACL 2014 Ninth Workshop on Statistical Machine

Translation [ DB/OL ]. [ 2014-9-23 ]. http:椅 www.

statmt. org / wmt14 / translation-task. html.

1. KOEHN P，HOANG H，BIRCH A，et al. Moses: open

source toolkit for statistical machine translation [ C ]椅  
Proceedings of the 45 th Annual Meeting of the ACL on  
Interactive Poster and Demonstration Sessions.

Stroudsburg， PA: Association for Computational

Linguistics， 2007: 177-180.

1. DENG J，DONG W，SOCHER R，et al. Imagenet: a  
   large-scale hierarchical image database [ C ]椅 Computer  
   Vision and Pattern Recognition，2009. IEEE Conference  
   on. Piscataway，NJ: IEEE，2009 : 248-255.
2. HUANG G B，MATTAR M，BERG T，et al. Labeled  
   faces in the wild: a database forstudying face recognition  
   in unconstrained environments [ C]椅 Workshop on Faces  
   in ‘ Real-Life ’ Images: Detection， Alignment， and  
   Recognition. Marseille: Erik Learned-Miller and Andras  
   Ferencz and Frederic Jurie，2008.
3. KUMAR N，BERG A C，BELHUMEUR P N，et al.  
   Attribute and simile classifiers for face verification [ C]椅  
   Computer Vision， 2009 IEEE 12th International  
   Conference on. Piscataway，NJ: IEEE，2009: 365-372.

[ 73 ] LU C，TANG X. Surpassing human-level face verification  
performance on LFW with GaussianFace [ J ]. CoRR，  
2014: abs/1404. 3840.

[74] SOOMRO K，ZAMIR A R，SHAH M. Ucf101: a dataset  
of 101 human actions classes from videos in the wild[ J].  
CoRR，2012: abs/1212. 0402.

(责任编辑吕小红）