

# 深度学习最新研究进展综述

王蕾<sup>1</sup>, 张宝昌<sup>2</sup>

(1. 北京航空航天大学自动化科学与电气工程学院, 北京 100191;  
2. 意大利理工学院模式分析与计算机视觉实验室, 热那亚 16163)

**摘要:** 深度学习作为机器学习领域的新兴技术, 给人工智能及相关领域带来了生机与活力。首先, 对深度学习的重要性、概念及特点进行了详细阐释, 说明深度学习的价值及意义。然后对深度学习目前较成熟的 2 个典型模型——卷积神经网络 (convolutional neural networks, CNNs) 和自动编码器进行详细综述, 并对其最新研究进展应用进行了概括。之后对深度学习中比较有潜力及实际意义的 2 个模型进行了介绍: 多层核函数机 (multilayer kernel machines, MKMs) 及深度时空推理网模型 (deep spatio-temporal inference network, DeSTIN), 为深度学习模型的发展方向注入新鲜力量。最后指出深度学习目前存在的缺点, 并对未来发展方向进行阐述。  
**关键词:** 信息处理技术; 深度学习; 综述; 神经网络; 网络结构; 模型比较

中图分类号: TP181 文献标识码: A 文章编号: 1674-2850(2015)06-0510-08

## Review on deep learning

WANG Lei<sup>1</sup>, ZHANG Baochang<sup>2</sup>

(1. School of Automation Science and Electrical Engineering, Beihang University, Beijing 100191, China;  
2. Pattern Analysis and Computer Vision (PAVIS), Italian Institute of Technology, Genoa 16163, Italy)

**Abstract:** Deep learning, which is an emerging technology of machine learning field, has brought vitality and vigor to artificial intelligence and related fields. Firstly, this paper gives the detailed elaboration of deep learning about importance, concept and characters. Then we illustrate the value and significance of deep learning. Secondly, it describes two kinds of typical deep learning models which are relatively mature at present in detail: convolutional neural networks (CNNs) and sparse auto-encoder network. We also summarize the latest application of them. Thirdly, we introduce two deep learning models which relatively have the potential and practical significance: multilayer kernel machines (MKMs) and deep spatio-temporal inference network (DeSTIN), which inject fresh energy to deep learning model development. At last, it concludes the existing shortcomings and summarizes the development trend of deep learning model.

**Key words:** information processing technology; deep learning; review; neural network; network structure; model comparison

## 0 引言

在统计机器学习领域, 值得关注的问题是如何对输入样本进行特征空间的选择。例如对行人检测问题, 需要寻找表现人体不同特点的特征向量。一般来说, 当输入空间中的原始数据不能被直接分开时, 则将其映射到一个线性可分的间接特征空间。而此间接空间通常可由 3 种方式获得: 定义核函数映射到高维线性可分空间, 如支持向量机 (support vector machine, SVM)、手工编码或自动学习。前 2 种方式

作者简介: 王蕾 (1989—), 女, 硕士研究生, 主要研究方向: 行人检测与图像处理

通信联系人: 张宝昌, 副教授, 主要研究方向: 模式识别, 机器学习. E-mail: bczhang@139.com

对专业知识要求很高,且耗费大量的计算资源,不适合高维输入空间。而第3种方式利用带多层非线性处理能力的深度学习结构进行自动学习,经实际验证被普遍认为具有重要意义与价值。深度学习结构相对于浅层学习结构[如SVM、人工神经网络(artificial neural networks, ANN)],能够用更少的参数逼近高度非线性函数。例如,文献[1]证明一个 $n$ 位奇偶校验函数能够被 $O(\lg n)$ 个隐层以及 $O(n)$ 个神经元的前馈神经网络编码,而如果采用单隐层的前馈神经网络,则需要指数数量级的相同神经元来完成该任务。

深度学习算法打破了传统神经网络对层数的限制,可根据设计者需要选择网络层数。它的训练方法与传统的神经网络相比有很大区别,传统神经网络随机设定参数初始值,采用BP算法利用梯度下降算法训练网络,直至收敛。但深度结构训练很困难,传统对浅层有效的方法对于深度结构并无太大作用,随机初始化权值极易使目标函数 $f(w,b)$ 收敛到局部极小值,且由于层数较多,残差向前传播会丢失严重,导致梯度扩散,因此深度学习过程中采用贪婪无监督逐层训练方法[2-3]。即在一个深度学习设计中,每层被分开对待并以一种贪婪方式进行训练,当前一层训练完后,新的一层将前一层的输出作为输入并编码以用于训练;最后每层参数训练完后,在整个网络中利用有监督学习进行参数微调[4]。

## 1 CNNs

CNNs是一种有监督深度的模型架构,尤其适合二维数据结构。目前研究与应用都较广泛,在行人检测[5]、人脸识别[6]、信号处理[7]等领域均有新的成果与进展。它是带有卷积结构的深度神经网络,也是首个真正意义上成功训练多层网络的识别算法。CNNs与传统ANN算法的主要区别在于权值共享以及非全连接[8]。权值共享能够避免算法过拟合,通过拓扑结构建立层与层间非全连接空间关系来降低训练参数的数目,同时也是CNNs的基本思想。CNNs的实质是学习多个能够提取输入数据特征的滤波器,通过这些滤波器与输入数据进行逐层卷积及池化,逐级提取隐藏在数据中拓扑结构特征[9]。随网络结构层层深入,提取的特征也逐渐变得抽象,最终获得输入数据的平移、旋转及缩放不变性的特征表示。较传统神经网络来说,CNNs将特征提取与分类过程同时进行,避免了两者在算法匹配上的难点。

CNNs主要由卷积层( $C_i$ )与下采样层( $S_j$ )交替重复出现构建网络结构,卷积层用来提取输入神经元数据的局部特征,下采样层用来对其上一层提取的数据进行缩放映射以减少训练数据量,也使提取的特征具有缩放不变性。一般来说,可以选择不同尺度的卷积核来提取多尺度特征,使提取的特征具有旋转、平移不变性[10]。用于图像识别的二维CNNs基本框架如图1所示,其中卷积层采用2种卷积核。

输入图像与可学习的核进行卷积,卷积后的数据经过激活函数得到一个特征图,卷积层的计算公式如下所示:

$$M_k^l = S \left( \sum_{n \in I_k} \omega_{nk} * M_n^{l-1} + b_n^l \right), \quad (1)$$

其中, $M_k^l$ 为第 $l$ 层的第 $k$ 个特征图; $I_k$ 为第 $k$ 个特征图的所有输入特征图集合;

$\omega_{nk}$ 为对应卷积核的可学习参数; $b_n^l$ 为第 $l$ 层的第 $n$ 个输入图像所对应的加权偏置; $S(\cdot)$ 为激活函数,如Sigmoid函数。由式(1)可以看出,卷积层的特征图可以由多个输入图组合获得,但对于同一幅输入图其卷积核参数是一致的,这也是权值共享的意义所在。卷积核的初始值

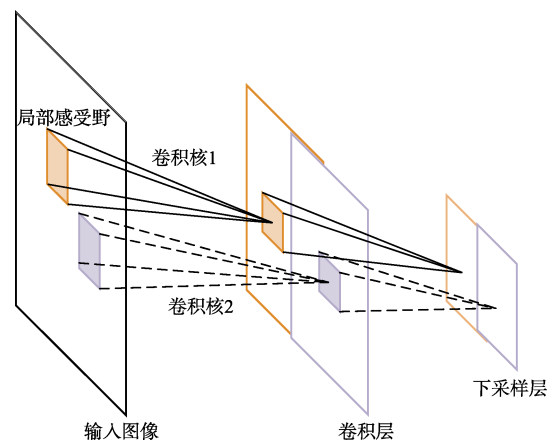


图1 CNNs结构示意图

Fig. 1 Structure of CNNs

并非随机设置，而是通过训练或者按照一定标准预先给定，如仿照生物视觉特征用 Gabor 滤波器进行预处理<sup>[11]</sup>。

下采样层通过降低网络空间分辨率来增强缩放不变性，下采样层的计算公式如下所示：

$$M_k^l = S \left( \beta \sum_{n \in I_k} M_n^{l-1} + b_n^l \right), \quad (2)$$

其中， $\beta$  为可训练的标量参数，其值随下采样方法变化而不同，例如采用 Mean-Pooling， $\beta = \frac{1}{m}$ ， $m$  表示对  $m \times m$  像素块进行下采样（常用大小为  $2 \times 2$ ），如此输出图像每个维度均缩小为原图的  $m$  倍，每个输出图均有一个加权偏置  $b_n^l$ ，然后将输出结果输入到一个非线性函数（如 Sigmoid 函数）。

CNNs 的输出层一般采用线性全连接，目前最常用的就是 Softmax 分类方法。CNNs 的参数训练过程与传统的人工神经网络类似，采用反向传播算法，包括前向传播与反向传播 2 个重要阶段。目标函数如下所述：

$$L = \|Y^i - T^i\|^2, \quad (3)$$

其中， $Y^i$  为网络输出值； $T^i$  为目标输出值；目标函数  $L$  为两者方差。参数训练目标为最小化式 (3)，训练过程一般采用 SGD 算法，但是文献[12]中提到的优化算法 LBFGS 在 CNNs 参数优化中较 SGD 算法效果有明显提高。CNNs 实际应用中会遇到诸多问题，如网络权值的预学习问题，收敛条件以及非全连接规则等，这些均需要实际应用中进一步解决与优化。

CNNs 在许多模式识别领域已取得较好性能，文献[11]中，将 CNNs 应用到行人检测中，并提出一种联合深度神经网络模型（unified deep net, UDN）。模型结构如图 2 所示，输入层有 3 个通道，均为对 YUV 空间进行相关变换得到，实验结果表明在此实验平台前提下，此输入方式较灰色像素输入方式正确率提高 8%。第一层卷积采用 64 个不同卷积核，初始化采用 Gabor 滤波器，第二层卷积采用不同尺度的卷积核，提取人体的不同部位的具体特征，训练过程作者采用联合训练方法。最终实验结果在 Caltech 及 ETH 数据集上错失率较传统的人体检测 HOG-SVM 算法均有明显下降，在 Caltech 库上较目前最好的算法错失率降低 9%。

文献[13]中对 CNNs 所处理数据的空间维度进行了延伸，采用三维 CNNs 模型，如图 3 所示，对人体运动行为进行识别。三维卷积将连续的帧首先组成立方体，然后在立方体内运用卷积核。进而通过三维特征提取器来提取视频中不同帧在空间和时间维度上的特征信息。这也是深度 CNNs 一个新的发展空

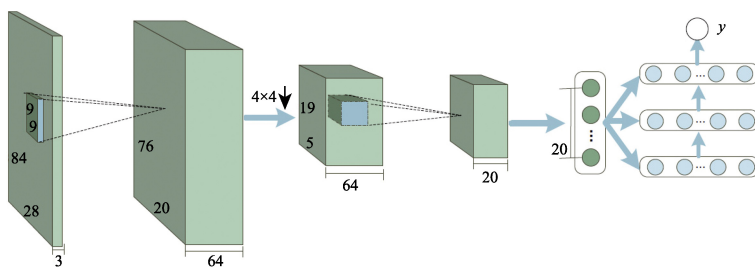


图 2 UDN 行人检测模型

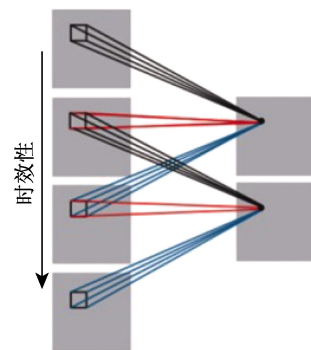


图 3 三维卷积模型

Fig. 2 UDN pedestrian detection model

间，随着研究的深入，相信这个领域会给社会带来很大利益。

## 2 稀疏自动编码器模型

与 CNNs 不同，深度自动编码器<sup>[14]</sup>是一种无监督的神经网络学习架构。此类架构的基本结构单元为自动编码器，它通过对输入特征  $\mathbf{X} = \{x_1 x_2 \cdots x_{n-1} x_n\}$  按照一定规则及训练算法进行编码，将其原始特征利用低维向量  $\mathbf{X}^* = \{x_1^* x_2^* \cdots x_{m-1}^* x_m^*\}$  重新表示，其中  $m < n$ 。自动编码器通过构建类似传统神经网络的层次结构，并假设输出  $\mathbf{Y} = \{y_1 y_2 \cdots y_{n-1} y_n\}$  与输入  $\mathbf{X}$  相等，反复训练调整参数得到网络参数值。

上述自编码器若仅要求  $\mathbf{X} \approx \mathbf{Y}$ ，且对隐藏神经元进行稀疏约束，从而使大部分节点值为 0 或接近 0 的无效值，便得到稀疏自动编码算法。一般情况下，隐含层的神经元数应少于输入  $\mathbf{X}$  的个数，因为此时才能保证这个网络结构的价值。正如主成分分析（principal component analysis, PCA）算法，通过降低空间维数去除冗余，利用更少的特征来尽可能完整的描述数据信息。实际应用中将学习得到的多种隐层特征（隐层数通常多个）与原始特征共同使用，可以明显提高算法的识别精度。

稀疏自动编码算法的基本结构如图 4 所示，输入特征  $\mathbf{X}$  经过编码器得到隐藏层编码特征  $\mathbf{X}^*$ ，再通过解码器获得输入  $\mathbf{X}$  的重构特征  $\mathbf{Y}$ ，其中编码过程参数为  $\omega_e$ ，解码过程参数为  $\omega_d$ ，通过反复调整参数使重构误差最小即可，即式（4）成立即可：

$$\min_{\text{err}} \|\mathbf{Y} - \mathbf{X}\|. \tag{4}$$

此过程为无监督学习，将第一层编码器训练输出的数据输入到下一层，重复上述参数训练过程，最终所有编码器的参数均被初始化。将最后一层编码器输出结果输入到一个分类器（如 SVM，逻辑回归）中，最后进行有标签学习。整个稀疏自动编码器的目标函数如下所示：

$$L_{\text{sparse}}(\omega, b) = L(\omega, b) + \beta \sum_{i=1}^{S_2} KL(\rho | \bar{\rho}_i), \tag{5}$$

其中， $L(\omega, b)$  为传统神经网络的目标函数； $\sum_{i=1}^{S_2} KL(\rho | \bar{\rho}_i)$  为惩罚因子； $KL(\rho | \bar{\rho}_i)$  为 KL 距离； $\rho$  为稀疏性参数； $\bar{\rho}_i$  为隐藏单元  $i$  的平均激活值，通常值接近于 0。此学习过程也类似于反向传播算法，利用梯度下降算法去微调整个网络参数即可。

自动编码器参数训练方法有很多，几乎可以采用任何连续化训练方法来训练参数<sup>[15]</sup>。但由于其模型结构不偏向生成型，无法通过联合概率等定量形式确定模型合理性。稀疏性约束在深度学习算法优化中的地位越来越重要，主要与深度学习特点有关。大量的训练参数使训练过程复杂，且训练输出的维数远比输入的维数高，会产生许多冗余数据信息。加入稀疏性限制，会使学习到的特征更加有价值，同时这

Fig. 3 Three-dimensional convolution model

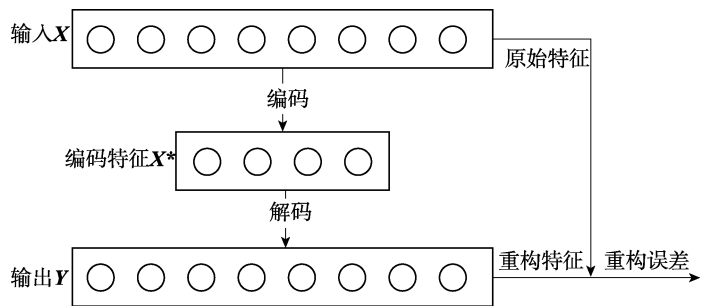


图 4 稀疏编码算法基本单元结构示意图

Fig. 4 Basic unit structure of sparse coding algorithm



也符合人脑神经元响应稀疏性特点<sup>[16]</sup>。

### 3 其他深度模型

#### 3.1 MKMs

受 SVM 算法中核函数的启发,在深度模型结构中加入核函数,构建一种基于核函数的深度学习模型。文献<sup>[17]</sup>提出一种 MKMs 深度模型,如同深度信念网络 (deep belief network, DBNs),反复迭代核 PCA 来逼近高阶非线性函数,每一层核 PCA 的输出作为下一层核 PCA 的输入。作者模拟大型神经网络计算方法创建核函数族,并将其应用在训练多层深度学习模型中。 $L$  层 MKMs 深度模型的训练过程如下:

第 1 步 去除输入特征中无信息含量的特征。

第 2 步 重复  $L$  次:

- a 计算有非线性核产生特征的主成分;
- b 去除无信息含量的主成分特征。

第 3 步 采用 Mahalanobis 距离进行最近邻分类。

在参数训练阶段,采用核主成分分析法 (kernel principal component analysis, KPCA) 进行逐层贪婪无监督学习,并提取第  $k$  层数据特征中的前  $n_k$  主成分,此时第  $k+1$  层便获得第  $k$  层的低维空间特征。为进一步降低每层特征的维数,采用有监督的训练机制进行二次筛选:首先,根据离散化特征点边缘直方图,估计它与类标签之间的互信息,将  $n_k$  主成分进行排序;其次,对于不同的  $k$  和  $w$  采用 KNN 聚类方法,每次选取排序最靠前的  $w$  验证集上的特征并计算其错误率,最终选择错误率最低的  $w$  个特征。

该模型由于特征选取阶段无法并行计算,导致交叉验证阶段需耗费大量时间。据此,文献<sup>[18]</sup>中提出了一种改进方法,通过在隐藏层采用有监督的核偏最小二乘法 (kernel partial least squares, KPLS) 来优化此问题。

#### 3.2 DeSTIN

目前较成熟的深度学习模型大多建立在空间层次结构上,很少对时效性 (temporal) 有所体现。相关研究表明,人类大脑的运行模式是将感受到的模式与记忆存储的模式进行匹配<sup>[19]</sup>,并对下一时刻的模式进行预测,反复进行上述步骤,这个过程包含了时空信息。因此在深度结构中将时效性考虑在内,会更接近人脑的工作模式。DeSTIN<sup>[20]</sup>便是基于这种理念被提出的。

DeSTIN 是一种基于贝叶斯推理理论、动态进行模式分类的深度学习架构,它是一种区分性的层次网络结构。在该深度模型中,数据间的时空相关性通过无监督方式来学习。网络的每一层的每个节点结构一致,且包含多个聚类中心,通过聚类和动态建模来模拟输入。每个节点通过贝叶斯信念推理输出该节点信念值  $b$ ,根据信念值  $b$  提取整个 DeSTIN 网络的模式特征,最后一层网络输出特征可以输入分类器如 SVM 中进行模式分类。

DeSTIN 模型的每一个节点都用来学习一个模式时序,底层节点通过对输入数据的时间与空间特征进行提取,改变其信念值  $b$ ,输入到下一层。由于每一个节点结构相同,训练时可采样并行计算,节约运算资源。该模型最重要的步骤就是信念值更新算法,其公式如下所示:

$$b^1(s^1) = \frac{P(o|s^1) \sum_{s^1 \in S} P(s^1|s,a)b(s)}{\sum_{s^2 \in S} P(o|s^2) \sum_{s^2 \in S} P(s^2|s,a)b(s)}, \quad (6)$$

其中,  $o$  为当前聚类中心, 信念值为  $b$ , 其状态为  $s$ , 更高层节点为  $a$ , 更新后的信念和状态为  $b^1(s^1)$ ,  $P(o|s^1)$  为聚类中心的条件概率值, 该值从聚类过程中获得, 而  $\sum_{s^1 \in S} P(s^1 | s, a) b(s)$  需从实验过程中参数调整获得。

由式(6)可以看出, 信念值更新算法同时考虑了数据的时间与空间特征。目前将时效性考虑在内的深度学习架构虽然不是很成熟, 但也逐渐应用在不同领域, 也是深度学习模型未来发展的一个新方向。

### 3.3 深度学习模型比较

不同深度模型的网络结构模式、参数训练方式等都存在明显不同。根据深度学习模型结构及所采用的技术方式不尽相同, 可将其分为区分型深度结构、生成型深度结构及混合型结构3种<sup>[21]</sup>。区分型结构是指通过描述数据的后验概率分布  $P(v_j|D)$  度量分类的区分型能力; 生成型结构描述观测数据与对应类别的联合概率分布或数据的高阶相关性。根据 CNNs 及自动编码算法的原理及技术理解, 可知前者属于区分型结构, 而后者属于生成型结构。深度学习训练过程有些需要无监督预训练, 而有些模型无监督预训练对模型性能提高无关紧要。

将文中提到的几种深度学习模型结构进行简单比较, 以便更好了解模型特征, 如表1所示。

表1 深度学习模型对比  
Tab. 1 Comparison of deep learning models

算法	结构类型	预训练阶段是否采用无监督	特点描述
CNNs	区分型	否	非全连接, 权值共享, 输入直接为二维图像, 无需处理
稀疏自动编码	区分型	是	最小化重构误差实现特征的压缩表示
DBNs	生成型	是	由若干限制玻尔兹曼机 (restricted Boltzmann machine, RBM) 堆叠而成
MKMs	区分型	是	以 KPCA 方法逐层构建深度学习网络
DeSTIN	区分型	否	基于贝叶斯推理, 无监督的动态模式多层次结构

## 4 深度学习模型发展方向

深度学习概念从提出至今, 深度网络模型及其变体模型不断被提出, 应用领域涉及甚广。结合目前所存在的深度学习模型以及诸多实验测试结果, 对深度学习模型的发展方向作如下阐述。

### 4.1 网络规模

传统 ANN 由于可训练层数有限, 导致提取特征的深度受限制, 进而影响最终识别效果。神经网络是模仿人类大脑神经网络进行工作的, 然而人类大脑是一个庞大而复杂的结构, 不是简单的神经网络就能达到目的。从理论上讲, 可能确实需要堆积大量神经元结构。深度学习正是通过训练一种多层深度非线性结构, 对样本数据的隐含特征进行提取。2012年, 由谷歌 XLab 团队投资的 Google Brain 项目<sup>[22]</sup>曾建立包含 10 亿个神经元的庞大深度学习网络, 推动谷歌地图、语音识别等产品的发展。真正想模拟人类大脑, 实现人工智能, 深度学习网络结构可能仍然需要不断扩大。

### 4.2 混合模型方向

一般来说, 机器学习领域算法最终很可能趋向模型混合化, 结合不同基本模型的优点产生性能更加优良的学习算法。据此, 深度学习模型很有可能从单一结构朝着混合结构发展。目前已出现许多基于 CNNs, RBM, Auto-encoder 3 种基本结构的混合算法模型, 例如卷积 DBNs<sup>[23]</sup>, 由于深度信念网络最初并没有将二维结构信息考虑在内, 将其与卷积概念相结合可以很有效解决此问题。文献<sup>[24]</sup>中作者将

Auto-encoder 结构应用到卷积网络中得到卷积自动编码器, 最后输入到 K-means 分类器中, 可对无标签数据进行分类。又有研究用自动编码器代替 DBNs 中的基本结构单元 RBM。也可以考虑将其他因素加入深度学习模型中, 例如时效性、核函数、基因调控等<sup>[21]</sup>。

### 4.3 无监督学习方向

在机器学习领域, 大多数算法是有监督学习的, 而实际生活中可用的数据可能无标签或者根本无法定义标签。可以说提出有效的严格意义上的无监督学习算法很有必要。所谓无监督学习即通过给定无标签的数据训练一个可以描述给定数据位置概率的函数  $f$ , 典型的算法有密度估计, 聚类分析等, 这些算法适应范围有限。当然, 在深度模型提出以后, 无监督学习似乎又引起研究者的重视, 尽管深度结构模型诸如 Autoencoder, DBNs, RBM 等均为无监督学习算法, 但实际应用中仍需要通过有标签数据进行微调。深度学习模型训练的成功很大一部分原因是依赖有标签的数据, 诸如 Facebook, 百度等大数据库均拥有海量数据库的存在。2014年7月, 百度公司首席科学家 ANDREW<sup>[25]</sup>在中国科学院自动化研究所关于深度学习的报告中说到, 在未来两年内, 深度学习仍然以有标签学习为主导方向, 但之后3至5年内, 将逐渐致力于解决无标签数据, 最终出现有标签与无标签数据共同发展的局面。再者, 生物认知世界的过程大部分也为无标签学习。因此, 深度学习算法很有可能朝着无标签数据发展。

## 5 结论

深度结构的优势在于多个非线性处理层的堆叠, 这种结构有利于学习高阶非线性函数的多个紧凑型参数。深度学习训练采用基于贪婪的逐层无监督预训练, 允许深度网络定位在一个通过有监督调整能够避免局部极值的参数空间。深度学习方法能够达到非常好的精度, 尤其在拥有大量数据集时, 甚至只有少量带标签数据。当然, 深度学习仍存在诸多理论与实际上的问题, 它们正在被一批充满活力并成长中的深度学习研究者们探索着, 并且给予了神经网络与深度学习新的诠释, 相信在未来不久, 机器学习与人工智能领域会有翻天覆地的变化。

### [参考文献] (References)

- [1] BENGIO Y, LECUN Y. Scaling learning algorithms towards AI[M]. Cambridge: MIT Press, 2007.
- [2] BENGIO Y, LAMBLIN P, POPOVICI D, et al. Greedy layer-wise training of deep networks[C]//SCHÖLKOPF B, PLATT J C, HOFFMAN T. Advances in Neural Information Processing Systems 19, Proceedings of the Twentieth Annual Conference on Neural Information Processing Systems. Vancouver: MIT Press, 2007: 153-160.
- [3] HINTON G E, OSINDERO S, TEH Y W. A fast learning algorithm for deep belief nets[J]. Neural Computation, 2006, 18(7): 1527-1554.
- [4] 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.
- [5] OUYANG W L, WANG X G. Joint deep learning for pedestrian detection[C]//ICCV. New York: IEEE, 2013: 2056-2063.
- [6] CHEN Y N, HAN C C, WANG C T, et al. The application of a convolution neural network on face and license plate detection[C]//The 18th International Conference on Pattern Recognition. Hongkong: IEEE, 2006, 4: 552-555.
- [7] REHN M, SOMMER F T. A network that uses few active neurones to code visual input predicts the diverse shapes of cortical receptive fields[J]. Journal of Computational Neuroscience, 2007, 22(2): 135-146.
- [8] HUBEL D H, WIESEK T N. Receptive fields, binocular interaction and functional architecture in the cat's visual cortex[J]. J. Physiol., 1962, 160(12): 106-154.

- [9] KAVUKCUOGLU K, SERMANET P, BOUREAU Y, et al. Learning convolutional feature hierarchies for visual recognition[C]// Proceedings of Advances in Neural Information Processing Systems, 2010. New York: Curran Associates, Inc., 2010: 1090-1098.
- [10] JAKE B. Notes on convolutional neural networks[OL]. Cogprints, 2006[2011-3-11]. <http://cogprints.org/5869/>.
- [11] KWOLEK B. Face detection using convolutional neural networks and Gabor filters[J]. Lecture Notes in Computer Science. 2005, 3696: 551-556.
- [12] LE Q V, NGIAM J Q, COATES A, et al. On optimization methods for deep learning[C]// Proceedings of 28th International Conference on Machine Learning. New York: ACM Press, 2011: 209-216.
- [13] JI S W, XU W, YANG M, et al. 3D convolutional neural networks for human action recognition[J]. IEEE Transactions on Pattern Analysis & Machine Intelligence, 2010, 35(1): 221-231.
- [14] BENGIO Y. Learning deep architectures for AI[J]. Foundations and Trends in Machine Learning, 2009, 2(1): 1-127.
- [15] 秦胜君, 卢志平. 稀疏自动编码器在文本分类中的应用研究[J]. 科学技术与工程, 2013, 13 (31): 9422-9426.
- QIN S J, LU Z P. Research of text categorization based on sparse autoencoder algorithm[J]. Science Technology and Engineering, 2013, 13(31): 9422-9426. (in Chinese)
- [16] WAIBEL A, HANAZAWA T, HINTON G, et al. Phoneme recognition using time-delay neural networks[J]. IEEE Transactions on Pattern Analysis & Machine Intelligence, 1989, 37(3): 328-339.
- [17] CHO Y, SAUL L K. Kernel methods for deep learning[C]// BENGIO Y, SCHUURMANS D, LAFFERTY J D. In Advances in Neural Information Processing Systems 22. Cambridge: MIT Press, 2009: 342-350.
- [18] YGER F, BERAR M, GASSO G. A supervised strategy for deep kernel machine[C]// 2011 Proceedings of European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. ESANN, 2011: 501-506.
- [19] FELLEMAN D J, van ESSEN D C. Distributed hierarchical processing in the primate cerebral cortex[J]. Cerebral Cortex, 1991, 1(1): 1-47.
- [20] KARNOWSKI T P, AREL I, ROSE D. Deep spatiotemporal feature learning with application to image classification[C]// The 9th International Conference on Machine Learning and Applications (ICMLA'10). New York: IEEE, 2010: 883-888.
- [21] 孙志军, 薛磊, 许阳明, 等. 深度学习研究综述[J]. 计算机应用研究, 2012, 29 (8): 2806-2810.
- SUN Z J, XUE L, XU Y M, et al. Overview of deep learning[J]. Computer application research, 2012, 29(8): 2806-2810. (in Chinese).
- [22] MARKOFF J. How many computers to identify a cat?[N]. The New York Times, 2012-6-26.
- [23] HUANG G B, LEE H, LEARNED-MILLER E. Learning hierarchical representations for face verification with convolutional deep belief networks[C]. CVPR. 2012: 2518-2525.
- [24] 段宝彬, 韩立新. 改进的深度卷积网络及在碎纸片拼接中的应用[J]. 计算机工程与应用, 2014, 50 (9): 176-270.
- DUAN B B, HAN L X. Improved convolutional neural networks and its application in stitching of scrapped paper[J]. Computer Engineering and Applications, 2014, 50(9): 176-270. (in Chinese).
- [25] ANDREW N. Deep learning: overview and trends[R]. Beijing: Automatization Institute, 2014.