

深度学习技术浅述*

王 敏^{1,2}, 周树道², 杨 忠¹, 刘展华², 任尚书³

(1. 南京金陵科技学院智能科学与控制工程学院, 江苏 南京 211169;

2. 国防科技大学气象海洋学院, 江苏 南京 211101;

3. 解放军95171部队, 广东 广州 510000)

摘 要: 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, 具有强大的学习和高效的特征表达能力, 成为人工智能领域的研究热点, 近年来发展十分迅猛。首先介绍了深度学习的特点及其与浅层学习的区别, 并详细介绍了深度学习的国内外发展现状和典型的深度学习网络模型, 重点介绍了深度学习技术中一些关键环节处理方法。

关键词: 深度学习; 机器学习; 模式识别; 特征提取; 人工神经网络

中图分类号: TP18 文献标志码: A 文章编号: 1003-7241(2019)05-0051-07

Simple Analysis of Deep learning Technology

WANG Min^{1,2}, ZHOU Shu-dao², YANG Zhong¹, LIU Zhan-hua², REN Shang-shu³

(1.School of Intelligent Science and Control Engineering, Jinling Institute of Technology, Nanjing 211169 China;

2.College of Meteorology and Oceanography, National University of Defense Technology, Nanjing 211101 China;

3.Unit 95171 of PLA, Guangzhou 510000 China)

Abstract: Deep learning can form high-level representation of attribute categories or characteristics through combine low-level features, in order to find the distributed characteristic data, it has the strong learning and efficient feature expression ability, and it becomes a research focus in the field of artificial intelligence and develops very rapidly in recent years. The characteristics of deep learning and its difference from shallow learning are introduced at first. The development status and typical deep learning network models of deep learning are introduced in detail, and finally some key links in deep learning technology are introduced.

Key words: deep learning; machine learning; pattern recognition; feature extraction; artificial neural network

1 引言

近十年来,深度学习(Deep Learning)在人工智能领域取得了重要且成功的突破,成为机器学习一个新的和热门的研究方向,具有强大的学习和高效的特征表达能力,在计算机视觉^[1-3]、图像与视频分析^[4-6]、语音识别^[7-10]、多媒体^[11]等诸多领域都取得了巨大成功。

深度学习的“深度”学习能力,是相对人工神经网络^[12](Artificial Neural Network, ANN)、支持向量机^[13](Sup-

port Vector Machine, SVM)、贝叶斯分类等“浅层机器学习”而言的,经深度学习得到的网络模型中,具有更多的非线性操作层数^[14],这种多层网络结构能够更加有效地表达复杂函数,进而自动学习到表征能力更强的特征。因此,将深度学习应用到图像的识别分类中,可以有效地提高识别的精度。本文就目前广泛应用的三种深度神经网络结构,更深入的介绍深度学习技术的基本原理、特点、国内外研究现状及其关键技术等内容。

2 深度学习工作原理

目前的物体自动识别分类技术大多是基于数字图像

*基金项目: 国家自然科学基金(编号 41775165)

收稿日期: 2018-02-09

处理技术的模式识别技术,一般流程如图1所示,其中特征提取和分类器分类是两个独立的操作环节,如图1所示。

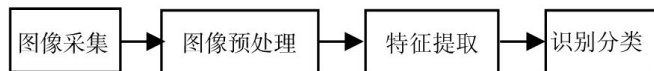


图1 常规模式识别流程

良好的特征能够准确地提取出有利于解决问题的信息,提取的特征的有效性直接决定着识别分类的效果。因此,大多研究者都重点研究如何提取不同种类物体的有效特征,但也仅能识别出有限种物体;另一方面,分类器的选择和设计决定着最终的识别精度。这种常规的将特征提取和识别分类分别作为两个独立研究的操作环节,势必会顾此失彼,导致识别效果较差。

2006年多伦多大学的G.E.Hinton等^[15]首次提出了深度学习的概念,其指出深度学习是通过一定的训练方法训练样本数据得到多个层级网络结构的机器学习过程^[14]。该网络结构中包含众多的神经元,每个神经元都连接着其他的大量神经元,连接权值在学习过程中能够自动调整并决定网络的功能。通过如此学习得到的网络结构符合人工神经网络的特征^[16],因此深度学习网络实际上是一种深层次的人工神经网络。

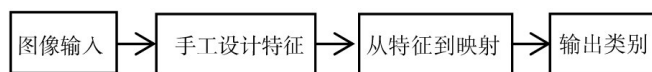
深度学习能够自动学习要建模数据的潜在分布的多层表达,其核心技术是监督和非监督^[17]学习算法,通过这种监督和非监督学习算法自动地学习到不同数据的多种特征,更好地表达数据,如图2所示,图中灰色框表示能从数据中进行学习的组件。相较传统的经典浅层学习而言,深度学习具有以下优势:

(1) 深度学习是从大数据中自动学习更具表征能力的特征,而非手工设计的特征,特别适用于变化多端的自然数据,具有非常优良的鲁棒性和优化能力,这是深度学习与传统模式识别手段的最大不同;

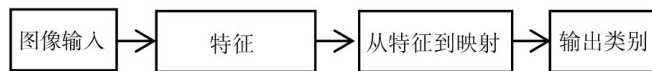
(2) 将特征提取和分类器合并到一个组件中,用样本数据自动去学习表征特征,充分发挥二者联合协作的优势,降低了手工设计特征的繁重工作量。同时利用反向传播算法对网络所有层进行联合优化,使各层之间相互协作达到最优,从而使整个网络的性能得到大幅提升;

(3) 从像素级原始数据到抽象的语义概念逐层提取信息,在提取图像的全局特征和上下文信息方面具有突出的优势;

(4) 非监督学习过程中不用事先知道样本的标签值,无需人工参与也能自动学习和提取到良好的特征。



(a) 经典机器学习



(b) 深度学习

图2 经典机器学习与深度学习的区别

3 国内外研究现状

深度学习是基于人工神经网络发展起来的一项技术,1957年, Frank Rosenblatt 首次提出“感知器”的概念^[18],在当时得到了广泛应用,在许多研究领域都取得了一定的成果。但由于当时计算条件的限制,并没有能够持续很长时间。1986年, DE Rumelhart, GE. Hinton 和 R.J. Williams 共同在 Nature 上发表论文^[19],提出误差后向传播(Error Back Propagation, EBP)算法,该算法解决了“感知器”的“异或”问题,降低了神经网络的计算复杂度,在解决相同问题的前提下,采用误差后向传播算法的人工神经网络的效果要优于简单的“感知器”。误差后向传播算法在各个领域得到广泛应用,并延续至今,重新掀起了神经网络高潮。1989年, Yann Lecun 成功将误差后向传播算法应用到手写邮政编码的识别任务上^[20]。Yann Lecun 在人工神经网络的基础上提出了“卷积神经网络(Convolutional Neural Network, CNN)”,目前卷积神经网络在视频分析和图像识别等领域都取得了很好的成果。

2006年,机器学习泰斗 G.E.Hinton 在 Science 上采用非监督的基于深度信念网络(Deep Belief Networks, DBN)^[21]的逐层训练算法,解决了深度神经网络训练难度大的问题。2012年,深度学习在计算机视觉领域出现了最具影响力的突破, Hinton 团队采用深度学习获得了 ImageNet^[22]图像分类比赛^[23]的冠军。排名第2到第4位的团队采用传统手工设计特征的计算机视觉识别方法,准确率差别都在1%之内。Hinton 研究团队首次参加比赛,且识别精度超出第二名10%以上。此结果在计算机视觉领域内产生了巨大的震动,并因此掀起了深度学习的研究热潮。

人脸识别是计算机视觉领域另一个重要的应用和挑战。Labeled Faces in the Wild(LFW)^[24]是当今最著名的人脸识别测试集,创建于2007年。LFW在互联网上收集了五千多位名人的脸部照片,且具有复杂的光线、表

情、姿态、年龄和遮挡等方面的变化,用于评估人脸识别算法在非可控条件下的性能。LFW的样本测试集包含了6000副人脸图像,随机猜的准确率仅为50%。经典的Eigenface人脸识别算法^[25]用于这个测试集也只有60%的识别率。在其他机器学习算法中,最好的识别率是96.33%^[26]。但采用深度学习可以使人脸识别率达到99.47%^[27]。

深度学习在工业界也产生了巨大的影响。百度和谷歌在Hinton 科研团队赢得ImageNet 竞赛6个月后,均发布了新的基于图像内容的搜索引擎。此类搜索引擎沿用了Hinton提出的深度学习网络模型,并将各自的数据上应用在该网络模型中,大幅提高了图像搜索的准确率。早在2012年百度成立了深度学习研究院,并于2014年5月又在硅谷成立了新的深度学习实验室,斯坦福著名教授吴恩达作为首席科学家。2014年1月,谷歌以4亿美金收购了一家深度学习创业公司DeepMind。Facebook 于2013年12月在纽约成立了新的人工智能实验室,聘请卷积神经网络发明人Yann LeCun担任首席科学家。鉴于深度学习技术给学术界和工业界带来的巨大影响力,2013年被MIT Technology Review 列为世界十大技术突破之首。2015年10月,谷歌旗下DeepMind 公司发明了基于深度学习的人工智能阿尔法围棋(AlphaGo),以5:0的总比分击败了欧洲围棋冠军樊麾;2016年3月,又以4:1的总比分击败了世界围棋冠军、职业九段选手李世石;2017年5月,再以3:0的比分击败了中国围棋第一人柯洁。

百度作为国内最早突破深度学习网络工业化的企业,将深度学习网络模型用于广告CTR 预告、声学建模、语音识别和图像识别等领域,获得了巨大成功,2017年1月百度首席科学家吴恩达带领机器人“小度”在《最强大脑》节目上对人脸识别以及语音识别比赛环节上战胜了人工识别。另外,科大讯飞、阿里巴巴等对深度学习的研究也初见成效。研究所和高校中,2014年,中国科学院自动化研究所模式识别国家重点实验室陈云霁教授等开创性地提出了深度学习处理器架构“寒武纪”。同年,由香港中文大学的汤晓鸥和王晓刚教授等指导的项目团队在ImageNet 竞赛中取得了令人瞩目的成果。

4 典型的深度学习网络类型

(1) 卷积神经网络形式

基于卷积神经网络(CNN)^[28-29]的图像识别分类方法

通过依次交替的卷积层和下采样层模型结构模拟人脑视觉系统对图像进行分类处理。该网络由输入层、隐层和输出层构成。其中隐层是由多个卷积层和子采样层构成的重复结构,如图3所示。

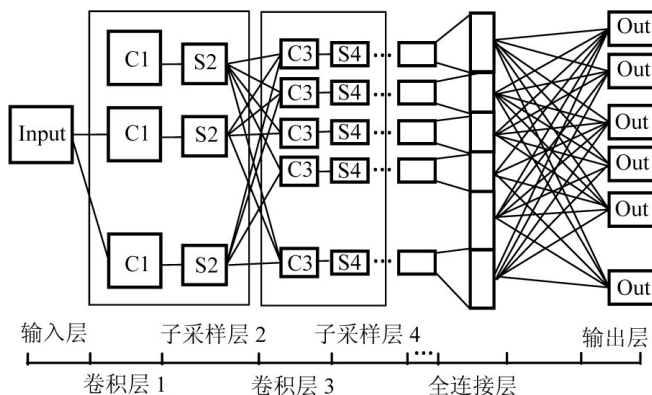


图3 卷积神经网络模型结构

(2) 受限玻尔兹曼机形式

受限制玻尔兹曼机(RBM)^[30-32]由可见层和隐含层组成,是一个基于能量的网络,如图4所示。对于一个RBM模型来说,给定可见层 v ,那么隐含层可以服从任意分布,反之亦然;一般都认为 v 和 h 是二值变量。

常规的基于RBM的深度学习网络都是多层RBM连接形式,通过添加RBM层数的方式不断提高模型的抽象性,从而建立语义的层次结构。

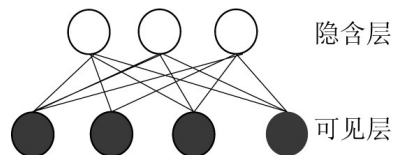


图4 受限玻尔兹曼机模型结构

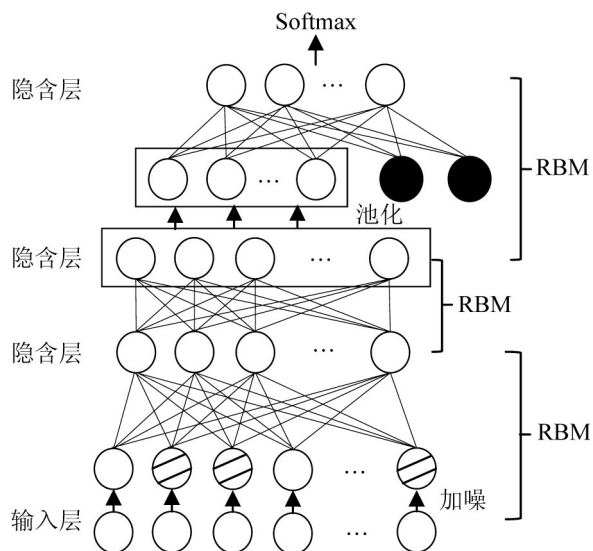


图5 多层受限玻尔兹曼机模型结构

(3) 深度信念网络形式

基于深度信念网络(DBNs)^[15]的图像识别分类方法,其核心是采用分层训练,不仅极大地提高了训练的速度,同时模型对复杂图像语义的表达能力也得到了很大的提升,如图6所示。

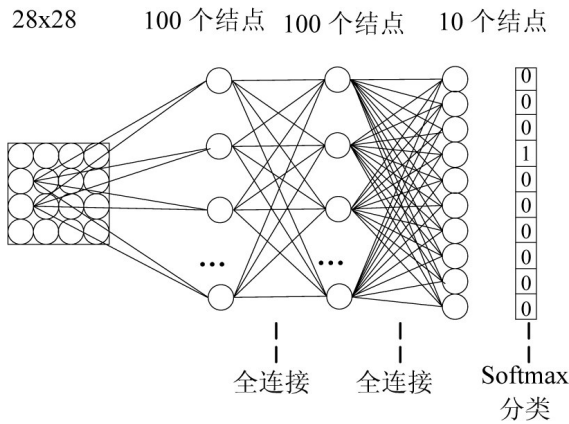


图6 深度信念网络模型结构

5 深度学习网络关键技术

深度学习网络模式识别算法中的关键技术涉及模型结构、网络参数选择、学习算法等方面。其中模型结构包括高维数据降维、内部分类器等,网络参数选择涉及初始化方法、网络层数和激活函数等,而学习算法包括学习率、特征学习方法等。这些关键技术对模式识别深度学习网络的训练时间、收敛速度、识别精度、学习效果有着重要的影响。

5.1 高维数据降维

深度学习网络中的每个样本要充分保留图像信息,维度就会很高,且训练过程中产生的特征图很多,需要进行降维操作,以便提高识别速度。经过池化(Pooling)操作得到的高层特征图,分辨率和维度降低了,还能避免过拟合。典型的池化方法有平均池化和最大池化法,其表达式分别为

$$S_{ij} = \frac{1}{c^2} \left(\sum_{i=1}^c \sum_{j=1}^c F_{ij} \right) + b_2 \quad (1)$$

$$S_{ij} = \max_{i=1, j=1}^c (F_{ij}) + b_2$$

$\max_{i=1, j=1}^c (F_{ij})$ 表示从输入特征F的大小为c的池化域中去除的最大元素。

5.2 内部分类

深度学习系统内部的分类器有多类别的Softmax回归模型、两类别的Logistic回归模型、欧式径向基函数等,

通过全连接方式将网络模型所提取特征映射为最终输出标签。虽然在训练时有效结合了反向传播,方便了模型中参数的更新,但Softmax回归模型的单层全连接方式无法解决非线性分类问题。可以考虑采用支持向量机SVM作为内部分类器,同时考虑训练分类时所用的不同物体的类间代价函数。

一般是以均方误差作为代价函数

$$E_0 = \frac{1}{2} \sum_{k=1}^B (t_k^n - y_k^n)^2 = \frac{1}{2} \|t^n - y^n\|_2^2 \quad (2)$$

式中, t_k^n 为第n个样本中第k个神经元对应的期望值, y_k^n 为第n个样本中第k个神经元实际的输出值。

可以用不同物体的类内距离与类间距离作为不同样本的类别可分性度量,分别表示为

$$E_1 = \frac{1}{2} \|y_c^n - M_c\|_2^2 \quad (3)$$

$$E_2 = \frac{1}{2} \sum_{c'=1}^B \|M_c - M_{c'}\|_2^2 \quad (4)$$

式中, y_c^n 为属于第c类的第n个样本的实际输出值, M_c 和 $M_{c'}$ 分别表示第c类和第c'类样本输出的均值。 E_1 和 E_2 分别为输出特征的类内距离和类间距离。通常,为使网络的输出类别更易可分,在迭代过程中应增大类间距离而减小类内距离。由此,修正的代价函数可表示为

$$E = E_0 + \alpha E_1 - \beta E_2 \quad (5)$$

式中, α 与 β 为权值参数。

5.3 初始化方法

不同的参数初始值使得深度学习网络学习到不同的局部极值,因此,初始值决定着深度学习网络的训练效果以及收敛程度。常规经验是采用较小的随机数初始化权重,如在 $[-4\sqrt{\frac{6}{fan_{in} + fan_{out}}}, 4\sqrt{\frac{6}{fan_{in} + fan_{out}}}]$ 范围内随机生成^[33],其中 fan_{in} 和 fan_{out} 分别为各层的输入和输出特征图个数。实际上,对不同的激活函数应考虑不同的初始化策略,同时应使每一层输出的方差尽量相等。近年来提出的新的初始化方法有:

(1) xavier初始化

xavier初始化^[33]用于tanh等激活函数可以获得比较好的效果。定义参数所在层的输入、输出维度分别为n和m,则初始化参数将均匀分布在 $W \sim U[-\sqrt{\frac{6}{n+m}}, \sqrt{\frac{6}{n+m}}]$ 应用时,可采用下面3种方式:

- ◆ 默认情况,只考虑输入 $W \sim U[-\sqrt{\frac{3}{n}}, \sqrt{\frac{3}{n}}]$

◆ 只考虑输出 $W \sim U[-\sqrt{\frac{3}{m}}, \sqrt{\frac{3}{m}}]$

◆ 取输入和输出层的均值 $W \sim U[-\sqrt{\frac{6}{n+m}}, \sqrt{\frac{6}{n+m}}]$

(2) MSRA 初始化

MSRA 初始化^[34]方法采用预训练的模型去初始化网络的部分层。不同于xavier初始化需要满足激活函数线性条件,MSRA初始化可以应用于ReLU(Rectified Linear Units)和PReLU激活函数。

MSRA初始化的权重分布是一个高斯分布,其中均值为0,均值为 $\frac{2}{n}$,初始化值满足下式

$$W \sim G[0, \sqrt{\frac{2}{n}}]$$

同样有3种方案:

- ◆ 默认情况下,n是输入层的维度;
- ◆ n取输出层的维度;
- ◆ n取输入和输出层的均值。

5.4 网络层数

深度学习网络的网络层数,即网络深度,决定了网络的训练效率和检测结果。层数过多,增加了网络的复杂度,增多了需要训练的参数,延长了训练时间,容易出现过拟合现象;层数过少,提取的特征不能准确地表达所需要的信息。目前,没有固定的网络层数,经典的卷积神经网络一直沿用结构简单的LeNet-5模型,即3层卷积层、2层下采样层、1层全连接层以及1层输出层。但在实际应用中网络层数又不止7层,如2012年AlexNet结构(12层)以及2014年在AlexNet结构上发展的VGG模型已经将网络扩展到16层的深度。实际应用中可以采用经典LeNet-5模型、AlexNet模型以及自行设计新的不同网络

层数的网络结构,然后根据识别率和识别速度来确定最终的网络结构层数。

5.5 激活函数

深度学习网络模型需要对卷积操作得到的特征采取非线性变换进行筛选,用来避免线性模型表达能力不够的问题。传统的深度学习网络采用的非线性变换激活函数有Sigmoid、Tanh和Softsign3种。

◆ sigmoid: $R = \frac{1}{1 + e^{-y}}$

◆ tanh: $R = \frac{e^y - e^{-y}}{e^y + e^{-y}}$

◆ softsign: $R = \frac{y}{1 + |y|}$

但这三种非线性操作都是饱和非线性函数,在传播过程中容易丢弃信息,而修正线性单元ReLU是一种不饱和非线性函数,更容易学习优化,在训练梯度下降时,ReLU的收敛速度比传统的饱和非线性函数更快。

$$R = \max(0, y) \quad (6)$$

5.6 学习率

学习率影响着深度学习网络模型的学习性能,学习率过大,导致网络误差极具增加,权重变得非常大;而学习率过小,虽然可以有效避免不稳定情况的出现,但会导致收敛速度非常慢。为了实现快速且有效的学习收敛过程,可以采用自适应学习率调节方法。目前的自适应学习率调整算法的基本思路主要分为两种:(1)首先初始化学习率初值为一个较大或较小的值,然后在每次迭代过程中按照固定的比率减小或增大学习率,直至训练完成;(2)首先按照经验赋给学习率一个初值,再根据网络误差的变化趋势来调整学习率,误差增大即减小学习率,误差减小则增加学习率。

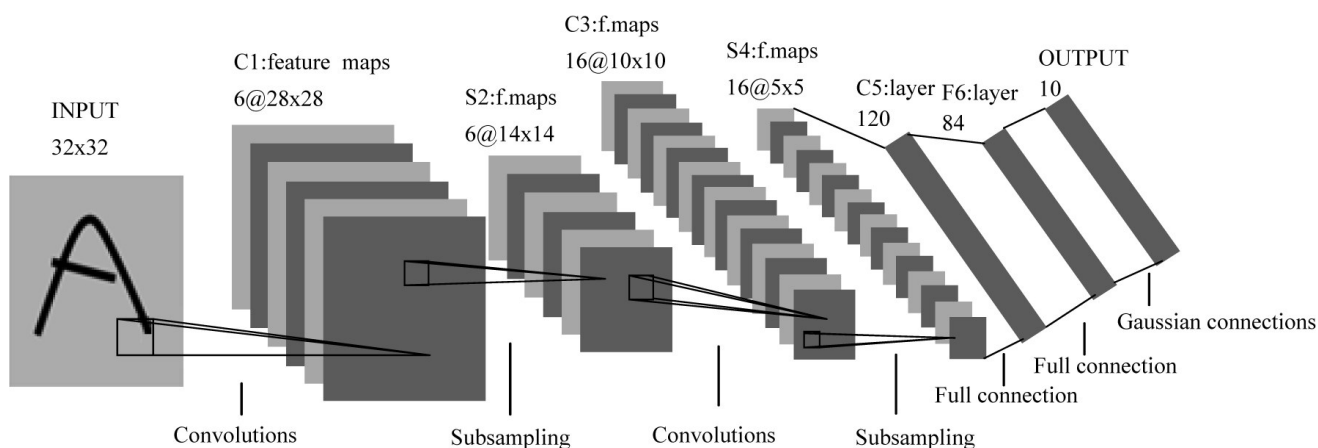


图7 LeNet-5模型结构图

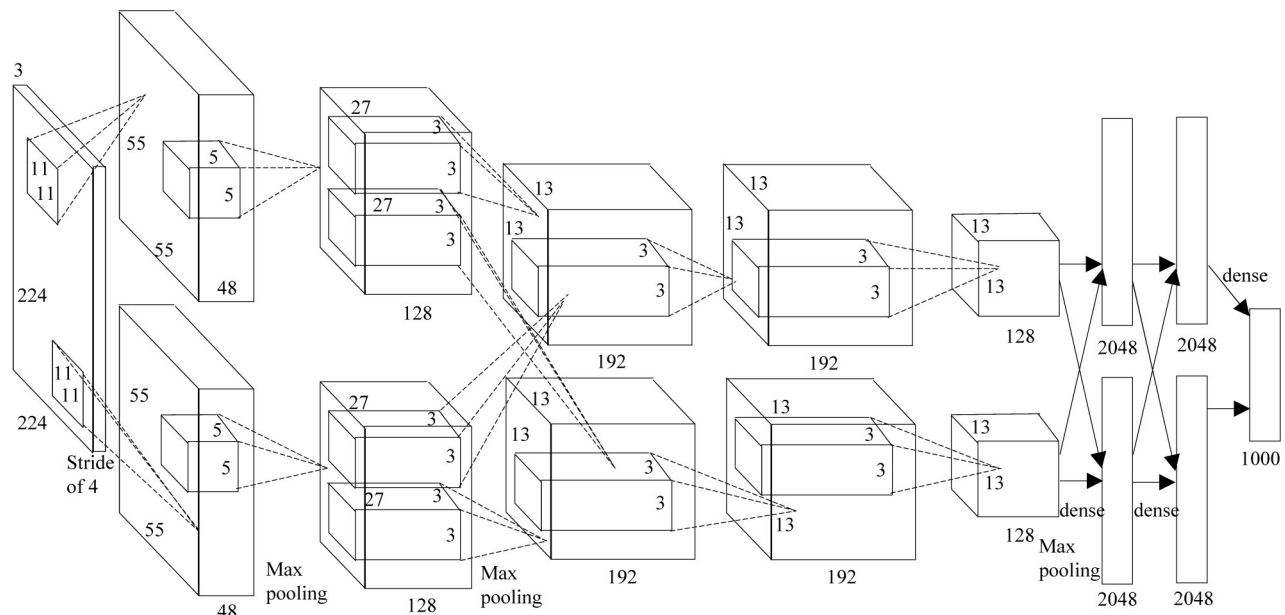


图8 AlexNet模型结构图

5.7 特征学习

正则化可以从众多的特征变量中自动“提取”出更重要的特征变量,以此减小特征变量的数量级,同时还可以防止过拟合、提高网络的稀疏性。正则化方法实际上就是在损失函数上额外添加一个正则化项(regularizer)或惩罚项(penalty term),典型的正则化方法有L1正则化和L2正则化。其中L1正则化是求权值向量中各个元素的绝对值之和,能产生稀疏权值矩阵,适用于特征选择;而L2正则化则是将权值向量中各个元素的平方求和后开平方根,可以防止模型过拟合(overfitting)。

6 结束语

深度学习技术在图像与视频分析、计算机视觉、多媒体、语音识别等诸多领域都发挥着重要作用。本文系统介绍了深度学习的基本工作原理及特点、与浅层机器学习之间的区别、几种深度学习网络模型,包括卷积神经网络、受限波尔兹曼机和深度信念网络三种方法,重点介绍了深度学习技术中的一些关键环节处理方法。

参考文献:

- [1] KRIZHEVSKY A, SUTSKEVER I, HINTON G E. Imagenet classification with deep convolutional neural networks[C]//Advances in Neural Information Processing Systems. Red Hook, NY: Curran Associates, 2012: 1097-1105.
- [2] BENGIO Y, BASTIEN F, BERGERON A, et al. Deep learners benefit more from out-of-distribution examples[C]//Proc of the 14th International Conference on Artificial Intelligence and Statistics. 2011.

ficial Intelligence and Statistics. 2011.

- [3] LEE H, EKANADHAM C, NG A Y. Sparse deep belief net model for visual area V2[C]//Advances in Neural Information Processing Systems. Cambridge: MIT Press, 2008: 873-880.
- [4] RANZATO M, SUSSKIND J, MNH V, et al. On deep generative models with applications to recognition [C]//Proc of IEEE Conference on Computer Vision and Pattern Recognition. 2011: 2857-2864.
- [5] SUSSKIND J, HINTON G E, MEMISEVIC R, et al. Modeling the joint density of two images under a variety of transformations[C]//Proc of IEEE Conference on Computer Vision and Pattern Recognition, 2011: 2793-2800.
- [6] LUO PING, WANG XIAO-GANG, TANG XIAO-OU. Hierarchical face parsing via deep learning[C]//Proc of IEEE Conference on Computer Vision and Pattern Recognition. Washington DC: IEEE Computer Society, 2012: 2480-2487.
- [7] FARABET C, COUPRIE C, NAJMAN L, et al. Learning hierarchical features for scene labeling[J]. Pattern Analysis and Machine Intelligence, IEEE Transactions on, 2013, 35(8): 1915-1929.
- [8] 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.
- [9] ACL 2014 Ninth Workshop on Statistical Machine Translation[DB/OL]. [2014-9-23]. <http://www.statmt.org/wmt14/translation-task.html>.
- [10] KOEHN P, HOANG H, BIRCH A, et al. Moses: open source toolkit for statistical machine translation[C]//Proceedings of the 45th Annual Meeting of the ACL on In-

teractive Poster and Demonstration Sessions. Stroudsburg, PA: Association for Computational Linguistics, 2007: 177–180.

[11] 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): 30–42.

[12] AGATONOVIC K, KUSTRIN S, BERESFORD R. Basic concepts of artificial neural network (ANN) modeling and its application in pharmaceutical research[J]. Journal of pharmaceutical and biomedical analysis, 2000, 22(5): 717–727.

[13] CORTES C AND VAPNIK V. Support-vector networks[J]. Machine Learning, 1995, 20(3): 273–297.

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

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

[16] PSALTIS D, SIDERIS A, YAMAMURA A. A multilayered neural network controller[J]. IEEE Control Systems Magazine, 1988, 8(2): 17–21.

[17] ERHAN D, MANZAGOL P A, BENGIO Y, et al. The difficulty of training deep architectures and the effect of unsupervised pre-training[C]. AISTATS, 2009: 153–160.

[18] ROSENBLATT F. The perceptron: a probabilistic model for information storage and organization in the brain[J]. Psychological Review, 1958, 65(6): 386–408.

[19] RUMELHART D E, HINTON G E, WILLIAMS R J. Learning representations by back-propagating errors[C]. MIT Press, 1986: 533–536.

[20] LECUN Y, BOSE B, DENKER J S, et al. Back-propagation APPLIED to Handwritten Zip Code Recognition [J]. Neural Computation, 1989, 1(4): 541–551.

[21] HINTON G E, SALAKHUTDINOV R R. Reducing the dimensionality of data with neural networks[J]. Science. 2006, 313(5786): 504–507.

[22] DENG J, DONG W, SICKER R, et al. Imagenet: A large-scale hierarchical image database[C]. In IEEE Int'l Conf. Computer Vision and Pattern Recognition, 2009.

[23] KRIZHEVSKY A, SUTSKEVER L, and HINTON G E. Imagenet classification with deep convolutional neural networks[C]. Proc. Neural Information Processing Systems, 2012.

[24] HUANG G B, RAMESH M, BERG T, et al. Labeled faces in the wild: A database for studying face recognition in unconstrained environments[R]. Technical report, University of Massachusetts, Amherst, 2007.

[25] TURK M and PENTLAND A. Eigenfaces for rec-

ognition[J]. Journal of Cognitive Neuroscience, 1991, 3(1): 71–86.

[26] CHEN D, CAO X, WEN F, and SUN J. Blessing of dimensionality: High-dimensional feature and its efficient compression for face verification[C]. Proc. IEEE Int'l Conf. Computer Vision and Pattern Recognition, 2013.

[27] SUN Y, WANG X, and TANG X. Deeply learned face representations are sparse, selective, and robust[J]. Computer Vision and Pattern Recognition, 2015(6): 2892–2900.

[28] LECUN Y, BOSE B, DENKER J S, et al. Handwritten digit recognition with a back-propagation network [C]//Advances in Neural Information Processing Systems. San Francisco, CA: Morgan Kaufmann Publishers, 1990: 396–404.

[29] 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.

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

[31] FREUND Y, HAUSLER D. Unsupervised learning of distributions of binary vectors using two layer networks [C]//Advances in Neural Information Processing Systems. San Francisco, CA: Morgan Kaufmann Publishers, 1994: 912–919.

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

[33] GLOROT X and BENGIO Y. Understanding the difficulty of training deep feedforward neural networks[J]. Journal of Machine Learning Research, 2010(9): 249–256.

[34] KAIMING HE, XIANGYU ZHANG, SHAOQING REN, JIAN SUN. Delving Deep into Rectifiers: Surpassing Human-Level Performance on ImageNet Classification[C]//Proceedings of the IEEE international conference on computer vision, 2015: 1026–1034.

作者简介: 王敏(1983-), 女, 博士, 副教授, 研究方向: 图像处理、模式识别。