

## 基于深度学习的医学图像识别研究进展

刘 飞 张俊然\* 杨 豪

(四川大学电气信息学院, 成都 610065)

**摘 要:** 近年来, 随着医学影像技术的快速发展, 医学图像分析步入大数据时代, 如何从海量的医学图像数据中挖掘出有用信息, 对医学图像识别带来巨大的挑战。深度学习是机器学习的一个新领域, 传统的机器学习方法不能有效地挖掘到医学图像中蕴含的丰富信息, 而深度学习通过模拟人脑建立分层模型, 具有强大的自动特征提取、复杂模型构建以及高效的特征表达能力, 更重要的是深度学习能从像素级的原始数据中逐级提取从底层到高层的特征, 这为解决医学图像识别面临的新问题提供了新思路。首先阐述深度学习方法, 列举深度学习方法的三种常见的实现模型, 并介绍深度学习的训练过程; 随后总结了深度学习方法在疾病检测与分类和病变识别两方面的应用情况, 以及深度学习应用在医学图像识别中的两个共性问题; 最后对深度学习在医学图像识别中存在的问题进行分析及展望。

**关键词:** 医学图像识别; 机器学习; 深度学习

中图分类号: R318 文献标志码: A 文章编号: 0258-8021(2018) 01-0086-09

## Research Progress of Medical Image Recognition Based on Deep Learning

Liu Fei Zhang Junran\* Yang Hao

(School of Electrical Engineering and Information, Sichuan University, Chengdu 610065, China)

**Abstract:** In recent years, with the rapid development of medical imaging technology, medical image analysis has entered the era of big data. How to extract useful information from a large number of medical image data has become one great challenge to medical image recognition. Deep learning is a new field of machine learning, conventional machine learning method can't effectively extract enough information contained in the medical image, while the deep learning has the power of establishing a hierarchical model, powerful automatic feature extraction, complex model building and efficient feature expression through the simulation of the human brain. More importantly, deep learning method can extract the features from the bottom to the top level from the original data of the pixel level, which provides a new way to solve the new problems faced by medical image recognition. Based on a large number of domestic and foreign literatures, this paper elaborated the three methods of depth learning, enumerated three common implementation models of deep learning methods, and introduced the training process of depth learning. We summarized the application of deep learning in two aspects of disease detection and classification and lesions recognition, and summarized the two common problems in the application of deep learning in medical image recognition. The analysis and prospects of deep learning in medical image recognition problems were proposed and discussed as well.

**Key words:** medical image recognition; machine learning; deep learning

doi: 10.3969/j.issn.0258-8021.2018.01.012

收稿日期: 2016-12-02, 录用日期: 2017-10-14

基金项目: 广西高校重点实验室科学基金(GXSCHP201411); 四川省科技计划资助项目(2015HH0036); 成都市科技惠民资助项目(2015-HM01-00561-SF)

\* 通信作者( Corresponding author) E-mail: zhangjunran@126.com

## 引言

医学图像识别是综合医学影像、数学建模、计算机技术等学科的交叉领域,在医学图像大数据时代,海量而复杂的图像数据带来两个方面的新问题:一方面要处理的医学图像数据维数更高,要求有更强学习适应能力的模型;另外一方面医学图像大数据更加分散破碎,数据结构更加复杂,常常需要整合不同的信息<sup>[1]</sup>。传统的数据分析方法常常不能满足人们的要求,因此在医学大数据时代,如何从海量医学图像数据中挖掘出有用信息,已成为学术界和工业界的研究热点。

深度学习<sup>[2]</sup>是机器学习<sup>[3]</sup>中的新领域,旨在通过模拟人脑自动地学习数据各个层次的抽象特征,从而更好地反映数据的本质特征<sup>[4]</sup>。自2006年,Hinton<sup>[2]</sup>提出一种基于概率图模型的多层受限玻尔兹曼机(restricted Boltzmann machine,RBM)后,深度学习已成为图像处理和计算机视觉领域的主导工具。近年来,深度学习在图像识别<sup>[5]</sup>、语音识别<sup>[6-7]</sup>、自然语言处理<sup>[8-9]</sup>、计算机视觉<sup>[10-11]</sup>等领域取得了巨大的成功,并引发了更多领域利用深度学习进行数据挖掘和分析的热潮,在医学领域也引起相应专家及公司的重视,如研发出AlphaGo的Google子公司Deep Mind宣布DeepMind Health健康<sup>[12]</sup>;IBM公司提出Watson for Oncology<sup>[13]</sup>,通过学习大量数据分析病人的肿瘤影像信息,从而为医生制定可靠的医疗方案提供帮助。

首先介绍深度学习网络的分类方式、常见的3种深度学习模型及深度学习的训练过程;然后介绍深度学习在疾病检测与分类和病变识别中的应用;最后分析深度学习方法应用在医学图像识别中面临的问题并对未来进行展望。

## 1 深度学习方法

深度学习从人工神经网络发展而来,20世纪80年代用于人工神经网络的BP算法的提出,开启了基于统计学习的机器学习热潮,但在随后的训练过程发现,BP算法存在收敛速度慢、易陷入局部最小等缺陷<sup>[14]</sup>。20世纪90年代,各种浅层机器学习模型如boosting<sup>[15]</sup>、SVM<sup>[16-17]</sup>等被提出,这些模型在理论上和应用中都获得了巨大的成功,但由于训练方法需要很多经验和技巧,使得浅层的机器学习长时间处于平静期。直到2006年深度信任网络的提出<sup>[4]</sup>,开启了对深度学习研究的新篇章,其中2012

年Hinton采用CNN模型赢得ImageNet图像分类的冠军,准确率比第二名提高10%以上<sup>[18]</sup>,在计算机视觉领域取得了突破性的进展。此后随着用于序列数据建模的递归神经网络(recurrent neural networks,RNN)<sup>[19-20]</sup>、自然语言处理和音素识别的深度条件随机场(deep-structured conditional random fields,DCRF)<sup>[21-22]</sup>、图像处理的深层残差网络(deep residual networks,DRN)<sup>[23-24]</sup>等模型的出现和各种深度学习算法的提出及GPU计算能力的提升,使得深度学习在图像识别、语音识别等众多领域取得了巨大的成功,2016年Goole Alpha与李世石的人机围棋大赛使得深度学习技术广为人知。

深度学习的目的在于通过构建多层隐层的机器学习模型和海量训练数据来学习更有用的特征,从而提高分类或预测的准确性<sup>[25]</sup>。与传统的浅层学习相比,深度学习具有如下特点:一是特征学习,深度学习方法能够根据不同的应用自动从海量数据中学习所需的高级特征表示,更能表达数据的内在信息。二是深层结构,深度学习模型结构深,通常拥有5层甚至更多层的隐层节点,包含更多的非线性变换,使得拟合复杂模型的能力大大增强。三是无监督学习,输入的数据只有本身数据信息,没有标签信息,深度学习未标记数据的模式;通过数据内在的一些特征和联系将数据自动分类。通过在训练过程中加入无监督学习作为预训练,使得深度学习模型相比人工神经网络具有更好的分类能力。

### 1.1 深度学习网络的分类

深度学习是通过多层非线性信息处理方法来构建深层网络<sup>[26]</sup>,根据结构的不同分为如下3类<sup>[27]</sup>:生成式深层网络、有监督(判别式)深层网络、混合深层网络。

1) 生成式深层网络:通过学习观测数据高阶相关性,或观测数据和关联类别之间的统计特征分布来实现模式分类的一类深层结构<sup>[28]</sup>,用于在没有目标类标签信息的情况下捕捉观测到的或可见数据的高阶相关性。常见的生成式模型是深层玻尔兹曼机(deep Boltzmann machine,DBM)<sup>[29]</sup>、和积网络(sum-product network,SPN)<sup>[30]</sup>、递归神经网络(recurrent neural networks,RNN)<sup>[19-20]</sup>等。

2) 有监督(判别式)深层网络:描述在可见数据条件下的类别后验分布<sup>[31]</sup>,目标类别标签以直接或者间接的形式给出,因此,有监督深度网络也称为判别式深度网络。深度堆叠网络(deep stacking

network ,DSN)<sup>[32]</sup>、深度结构条件随机场 ( deep-structured conditional random fields , DCRF)<sup>[33-34]</sup> 是典型的有监督学习深度网络。

3) 混合深层网络: 将生成式深层结构与判别式深层结构相结合的一类深层结构。通常情况下, 数据被用于作为预训练网络的权重, 以加快监督阶段学习过程, 无监督深度网络的结果作为重要辅助<sup>[35]</sup>。预训练深度神经网络 ( pre-trained deep neural networks , PDNN)<sup>[36]</sup> 是一种混合深度网络。

## 1.2 深度学习模型

深度学习方法的模型种类较多, 其中比较常用的模型是深度自编码网络 ( deep auto-encoder network , DAN)、深度信念网络 ( deep belief network , DBN)、卷积神经网络 ( convolution neural network , CNN)。近年来还出现了许多新的深度模型, 如循环神经网络模型 ( recurrent neural nets , RNN)<sup>[37]</sup>、张量堆叠网络模型 ( tensor deep stacking network , TDSN)<sup>[38]</sup>。本节介绍常见的 3 种深度学习模型及其构建方法。

### 1.2.1 深度自动编码器 DAE

Bengio 等通过改进原型自动编码器结构 ( 如图 1 所示), 产生了深度自编码器 ( deep auto-encoder , DAE)<sup>[39]</sup>。深度自编码器的基本元件是 AE, AE 包含一个输入层、一个隐层、一个输出层, AE 主要用于学习压缩的或过完备的特征表示, 当自编码器包含多个隐层时就形成了 DAE。DAE 是一类经过无监督逐层贪心预训练和系统性参数优化的多层非线性网络, 从无标签数据中提取高维输入数据的分层特征, 并得到原始数据的分布式特征表示的深度神经网络结构<sup>[40]</sup>。

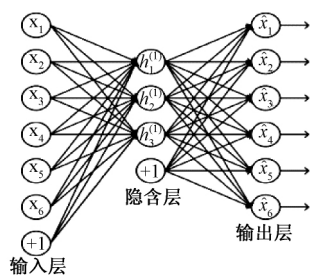


图 1 原型自动编码器<sup>[41]</sup>

Fig. 1 Prototype Automatic Encoder<sup>[41]</sup>

DAE 的构建主要有两步: 第一步, 改进原型自动编码器, 通过增加隐含层和神经元的数量、调整隐含层的节点分布、改变权值的分享方式等途径构建 DAE 的基本结构; 第二步, 根据不同的任务选取

合适的代价函数及优化策略、隐含层品质因数、系统性参数优化的性能指数等确定 DAE 的训练方案<sup>[41]</sup>。编码器在输入数据中加入含有一定统计特性的噪声, 构成基于统计理论的 DAE, 在原型自动编码器的代价函数表达式中加入解析性收缩惩罚因子, 就构成了基于鲁棒理论的 DAE。

### 1.2.2 深度信念网络 DBN

2006 年, Hinton 提出 DBN<sup>[4]</sup>, 开启了机器学习第二次浪潮——深度学习。将多个限制玻尔兹曼机 ( restricted Boltzmann machine , RBM) 堆叠得到深度玻尔兹曼机 ( deep Boltzmann machine , DBM), 如果靠近数据层的部分层之间的连接为有向连接, 即为 DBN<sup>[42]</sup>, 如图 2 所示。DBN 的基本单元是 RBM, 单个 RBM 包含一个由随机的隐单元构成的隐层 ( 一般是伯努利分布) 和一个由随机的可见单元构成的可见层 ( 一般是伯努利分布或高斯分布): 其中隐层和可见层之间是双向连接, 隐单元两两之间、可见单元两两之间无连接; RBM 采用对比梯度算法 ( contrastive divergence , CD) 对无标签样本进行训练, 属于无监督学习算法; DBN 通过组合许多 RBM, 把上一层 RBM 的特征激励作为下一层的训练数据, 从而可以对隐层的数据特征进行高效的学习<sup>[35]</sup>。

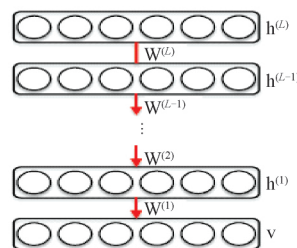


图 2 深度信念网络<sup>[43]</sup>

Fig. 2 Deep belief network<sup>[43]</sup>

DBN 的构建步骤如下:

步骤 1, 首先充分训练第一个 RBM, 将训练得到的权重和偏移量固定, 并将隐层作为第二个 RBM 的输入向量。

步骤 2, 采用同样的方法训练第二个 RBM, 并将第二个 RBM 堆叠在第一个 RBM 的上方。

步骤 3, 重复前面两个步骤得到多个 RBM, 最后将多个 RBM 按顺序堆叠在一起就得到 DBN。其中堆叠的过程是: 训练一个高斯-伯努利 RBM 或者伯努利-伯努利 RBM 后, 将隐层的激活概率作为下一层伯努利-伯努利 RBM 的可见层输入数据, 依次类推, 这种堆叠过程达到最大似然学习效果, 且是无监督学习过程, 因此不需要标签信息。

### 1.2.3 卷积神经网络 CNN

CNN 是受生物学上的感受野机制而提出的,经过不断的改进最终发展成一个特别适合图像处理的深度学习模型,同时 CNN 是一种特殊的神经网络模型:一方面,CNN 除了全连接层与输出层之外的神经元之间采用部分连接,而传统的神经网络都是采用全连接的方式,这就使得传统的神经网络训练耗时且难以训练;另一方面,CNN 在同一层的神经元之间共享权值,通过权值的共享既减少了权值的数量又降低了网络模型的复杂度。目前在国际标准的 ImageNet 数据集上,许多成功的模型都是基于 CNN,如大规模图像识别的深度学习网络 GoogLeNet<sup>[44]</sup>和 Adam<sup>[45]</sup>以及 LeNet-5<sup>[46]</sup>等。

CNN 的基本单元是 stage,其结构如图 3 所示,每个 stage 模块都是由卷积层(convolution layer)和池化层(pooling player)组成<sup>[47]</sup>,卷积层用于增强原始信号、提高信噪比,且通过权值的共享减少了模型的训练参数和计算的复杂度;池化层通过减少卷积层之间的连接,进一步减少训练的数据量,同时对卷积层的输出进行降采样,达到减少下一层的数据的效果<sup>[48]</sup>。通过将多个 stage 堆叠在一起,并在模型的末端加入全连接层和分类器就构成 CNN。

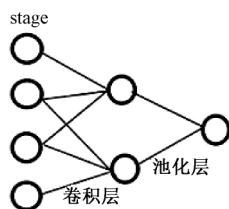


图 3 CNN 的基本结构单元 stage<sup>[47]</sup>

Fig.3 Stage is the basic unit of the convolutional neural network<sup>[47]</sup>

### 1.3 深度学习训练过程

深度学习从人工神经网络发展而来,训练方法继承了人工神经网络的反向传播方法和梯度下降方法,反向传播算法<sup>[49]</sup>(back propagation,BP)是从大量样本数据中学习统计规律,从而对测试样本做出判别。相比人工提取特征,反向传播算法消除了手工设计的影响,具有很大的优越性,但采用 BP 算法训练深层结构的网络时存在以下两个问题:一是 BP 算法主要解决复杂的非线性问题,网络的权值沿着局部方向逐渐调整,使得权值收敛到局部极小点,从而导致整个网络训练失败;二是训练速度慢,而且在训练深层次结构的网络时效果不明显。

批量梯度下降法、随机梯度下降法是对 BP 算

法进行改进的优化方法。批量梯度下降法<sup>[50]</sup>是最原始的梯度下降法,通过最小化所有训练样本的损失函数使得最终求解的是全局最优解,它的优点是得到一个全局最优解,而且易于实现并行,但批量梯度下降每次学习都要使用整个训练集,可能导致非凸函数收敛于局部极值点,同时训练过程会随着样本数量的加大导致训练时间过长。随机梯度下降法<sup>[51]</sup>是通过对所有样本进行随机选择最小化,每条样本的损失函数来求解最优解,在更新模型参数时只选择一个样本。其优点是训练速度快,但随机梯度下降法需要人为调整很多超参数,如学习速率、收敛准则、层数以及每层的单元个数等,这些超参数若选择不当可能导致每次更新不会按照全局的方向进行。

在深度学习模型中涉及多个非线性处理单元层,优化目标为非凸函数,当使用批量梯度下降法、随机梯度下降法来训练深层网络时,会出现训练时间过长、梯度不稳定、目标函数常常陷入局部最优等问题,同时随着网络层数的增加,局部最优的情况越来越严重<sup>[26]</sup>。为了克服这一问题,Hinton 提出一种贪婪逐层预训练方法<sup>[4]</sup>:首先逐层构建单层神经元,并每次训练一个单层网络,然后在所有层都训练完成后采用 wake-sleep 算法<sup>[52]</sup>进行调优,该方法通过在非监督数据上建立多层神经网络的方法有效克服了训练过程中梯度下降法的局部最小值和梯度不稳定的缺点。

Hinton 提出的贪婪逐层预训练方法在训练过程中加入无监督学习作为预训练,这是目前深度学习模型的训练过程与人工神经网络训练的主要区别。总结起来,深度学习训练过程大致可以分为以下两步:

1) 自下而上的非监督学习:采用无标签数据从底层开始逐层向上分层训练各层参数,具体来说就是先采用无标签数据训练第一层,训练时先学习第一层的参数,然后将第一层的输出作为第二层的输入,依次类推,直至训练到最顶层,由此得到各层的参数,由于模型容量限制以及稀疏性约束,使得得到的模型能够学习到数据本身的结构,从而更具表示能力的特征,这个过程可以看作是一个无监督训练过程,是与传统神经网络区别最大的部分;

2) 自上而下的监督学习:在第一步学习各层参数的基础上,在网络的最顶层添加一个分类器,通过带标签的数据训练,使误差自上向下传输,从而对网络进行调优,然后再利用第一步中得到的各层参数进一步微调整个多层模型的参数,这一过程可

以看成是一个有监督训练的过程。

## 2 深度学习在医学图像识别中的应用

深度学习广泛应用于医学图像识别中,通过在给定的数据集上训练模型来完成新数据上的特定任务,而在传统的医学图像识别方法是:基于多特征融合方法、基于奇异值分解和小波变换方法,对于特征的提取效率低且挖掘到的信息有限,识别效果不理想。相比传统的医学图像识别方法,深度学习能够挖掘到医学图像中潜在的非线性关系,特征提取效率更高。近年来,已有不少的研究人员将深度学习应用在医学图像识别中,这些工作为进一步的临床应用研究提供了重要的依据。疾病检测与分类是针对一批样本人群进行的,以确定某个样本是否患病或者其患病程度如何;而病变识别一般是针对某个样本自身医学图像中某个病变部位和其他部分的识别。目前深度学习方法在医学图像领域的上述两方面中被广泛应用,具体研究成果见表1,同时深度学习在图像配准、分割等图像预处理过程中也得到了广泛应用,由于篇幅有限对此部分不做综述。因此本节主要从疾病检测与分类和病变识别两个方面来介绍深度学习在医学图像识别中的研究进展。

表1 近年来深度学习在疾病检测与分类和病变识别中的应用

Tab.1 In recent years, the application of deep learning in disease detection and classification and lesion recognition

类别	时间	问题	使用模型
疾病检测与分类	2016	AD/正常人 <sup>[53]</sup>	CNN
	2015	AD/MCI 分类 <sup>[54]</sup>	RBM + SVM
	2016	恶性肿瘤检测 <sup>[55]</sup>	CNN
	2014	间质性肺疾病分类 <sup>[65]</sup>	CNN
	2016	肺部结节分类 <sup>[64]</sup>	CNN
	2016	大肠腺癌检测、分类 <sup>[56]</sup>	SC-CNN
	2016	脑微出血检测 <sup>[57]</sup>	3D CNN
	2016	腹部淋巴结检测 <sup>[58]</sup>	CNN
	2015	硬化转移、淋巴结等检测 <sup>[61]</sup>	CNN
	2013	低鳞状上皮内病变 <sup>[67]</sup>	DBN
病变识别	2015	核性白内障的病变 <sup>[68]</sup>	CNN + SVM
	2013	Huntington 舞蹈病 <sup>[70]</sup>	DBN
	2014	多并发硬化症病变 <sup>[71]</sup>	DBN
	2016	乳腺病变 <sup>[73]</sup>	CNN
	2013	肿瘤细胞识别 <sup>[74]</sup>	CNN + SVM
	2016	糖尿病视网膜病变 <sup>[75]</sup>	CNN
	2014	淋巴结识别 <sup>[76]</sup>	CNN
	2015	食道癌 <sup>[77]</sup>	3S-CNN

### 2.1 疾病检测与分类

医学图像中包含着大量的反映人体健康水平的信息,目前这部分数据主要依靠人工进行分析,易受主观因素的干扰且效率不高,容易造成数据资源的浪费。深度学习通过多层非线性变化,从海量数据中自动提取抽象特征,既消除了主观因素的影响又能提取到更加高级的抽象特征。

深度学习在阿尔茨海默病(alzheimer disease, AD)和轻度认知障碍(mild cognitive impairment, MCI)中有大量的研究。Sarraf 等使用 CNN 分类患有 AD 病的大脑和正常大脑,该模型对患有 AD 病的大脑和正常大脑的分类准确率达 96.85%<sup>[53]</sup>,该模型还能够扩展到更加复杂的分类任务;Li 等将 RBM 作为基本单元构建深度学习模型,并用于从 MRI 和 PET 扫描图像中分类 AD/MCI 患者,该模型分类准确率平均提高 5.9%<sup>[54]</sup>。

在实际的应用中基于深度学习的计算机辅助诊断将医学图像中与疾病诊断相关的特征提取出来,结合临床知识在很大程度上减少医生的工作量,得到十分精确的诊断或分类结果。Enlitic 公司开发出基于 CNN 的恶性肿瘤检测系统<sup>[55]</sup>,对放射师检查过的大量医学图像数据进行学习,自动总结出代表恶性肿瘤形状的“特征”,从而识别图像中是否存在恶性肿瘤,该系统识别肝癌的精度是放射师检查精度的 5 倍。Sirinukunwattana 使用空间约束卷积神经网络(spatially constrained convolutional neural network, SC-CNN)来检测和分类大肠腺癌细胞<sup>[56]</sup>,在分类问题上使用邻近集成预测(neighboring ensemble predictor, NEP)方法,该方法相对基于经典特征分类的方法有更好的分类效果;Dou 使用 3D CNN 从 MR 图像中自动检测脑微出血(cerebral microbleeds, CMBs)<sup>[57]</sup>,该方法从 MRI 图像中提取更具代表性的高级特征,相对手工提取特征和 2D CNN 提取特征,3D CNN 检测精度高达 93.16%。

相对于数字图像和灰度图像来说,医学图像不易获得且数据量少,这是深度学习应用在医学图像领域的一个共性问题。数据的缺少易造成过拟合问题,进而导致检测和分类结果不理想,当前有不少研究者在这方面做了很多方面的探索:如 Shin 将深度学习应用于 CT 图像中腹部淋巴结(三维图像)的检测和间质性肺病的分类(二维图像)<sup>[58]</sup>,他通过 transfer learning 方法<sup>[59]</sup>增加数据量,这项研究表明,transfer learning 能减少因数据的缺乏而带来的影响,有助于提高分类的准确率;Roth 针对 CNN 训

练过程中数据不足,通过 data augmentation 方法<sup>[60]</sup>扩充训练样本和测试样本的数据量,该模型对硬化转移检测的准确率提高了 13%,淋巴结检测的准确率提高了 27%,结肠息肉检测的准确率提高了 17%<sup>[61]</sup>;数据的缺少不仅容易出现过拟合问题,而且容易导致模型在测试数据上的泛化能力难以得到保证。Srivastava 等提出 dropout 技术<sup>[62]</sup>,通过在训练过程中随机剔除神经元来避免出现拟合问题从而提高模型的泛化能力;Wan 在 dropout 思想的基础上提出 dropconnect 方法<sup>[63]</sup>。Setio 使用包含多个 2D 卷积神经网络的多视图卷积网络(multi-view convolutional networks, MVCN)检测肺结节疾病<sup>[64]</sup>,采用 data augmentation 和 dropout 方法避免出现拟合问题,准确率高达 90.1%。Li 等使用 CNN 模型来分类间质性肺病<sup>[65]</sup>,该模型采用 dropout 方法和单卷积层结构来避免出现拟合问题。

## 2.2 病变识别

病变识别是深度学习在医学图像中的重要应用之一,传统的病变识别如小波变换方法等对病变识别的准确率不高,将深度学习应用于病变识别具有独特的优点:深度学习模型能够更快地处理数据,通过深度学习模型预测异常病变可以降低病变的机率,同时可提高医生诊断的准确率和效率。

相比普通的图像识别,医学图像识别问题更加复杂,对某些复杂的医学图像识别问题,可以通过构造更加深层、更加复杂的深度学习模型来解决。Chakdar 使用 DBN 进行基于子宫抹片识别低级别鳞状上皮内的病变(low grade squamous intraepithelial lesion, LGSIL),该方法将 DBN 提取出来的特征和原始特征共同作用于 SVM 模型使得分类准确率达到 100%<sup>[66]</sup>;Kondo 利用主成分回归分析算法结合深层 GMDH 型神经网络来识别医学图像中的左右肾区的病变<sup>[67]</sup>,该模型通过自动适应网络中的结构参数来提高识别的准确率;Gao 使用 CNN 和 SVM 构建一个自动学习特征的系统,并用于识别图像中核性白内障的病变,该模型的准确率提高了 5.6%<sup>[68]</sup>;Yan 等设计了一个多阶段深度学习框架并用于身体部位病变识别,在训练阶段,通过 CNN 来提取最具差异性的特征和从训练切片中提取局部信息;在增强阶段,经过预训练的 CNN 进一步增强图像的局部信息,相对基于全局图像上下文的方法,局部方法鲁棒性更好,准确率高达 92.23%<sup>[69]</sup>。

目前也有不少研究者根据某些病理特征来进行病变识别。Plis<sup>[70]</sup>根据头颅 CT 或 MRI 图像中的

尾状核萎缩程度与疾病的严重程度有关,将 DBN 应用于大脑结构和功能磁共振成像来识别 Huntington 病变,实验结果表明,深度学习方法能够学习重要的生理特征从而提高 Huntington 病变的识别;Brosch 使用三个 DBN 对大脑形态变化建立模型,该模型能够自动捕捉到脑形态和脑白质的病变情况,从而对脑白质进行病变识别<sup>[71]</sup>;Xu 使用堆栈稀疏自动编码器(stacked sparse autoencoder, SAE)来识别乳腺癌组织病理学图像上的细胞核,通过去噪自动编码器(denoising autoencoder, DA)来提高噪声的鲁棒性,准确率高达 88.84%<sup>[72]</sup>。

此外,有研究者通过借助医学图像中的识别对象物的技术来进行病变识别,如 Kooi 将 CNN 用于乳腺恶性病变识别,该方法在低灵敏度下识别结果比传统的计算机辅助方法更优,在高灵敏度下准确率更高<sup>[73]</sup>;Cruz-Roa 将自编码神经网络用于识别图像中的肿瘤细胞,该模型增加两个有助于区分癌组织和正常组织的可判断层,相比传统方法,该方法的准确率提高了 7%<sup>[74]</sup>。

深度学习应用在医学图像领域中训练时间普遍过长,对硬件要求高,模型可移植性差,这是深度学习应用在医学图像领域的另一个共性问题。有研究者通过改进基本的深度学习模型,可以减少训练时间,van 针对训练过程中时间过长提出一种改进的 CNN,用于检测彩色眼底图像出血病变,在每个训练过程中从训练数据里随机选择样本进行训练,迭代次数从 170 次减少到 60 次,从而大大减少训练时间<sup>[75]</sup>。

## 3 结论

机器学习方法广泛地应用于医学图像识别中,通过在给定的数据集上的训练模型来完成新数据上的特定任务。然而,一方面传统的机器学习算法常常需要利用先验知识从原始数据中人工提取特征来训练模型,此方法难以提取到复杂特征,而且由于特征选取难度大,可能出现过拟合问题,模型的泛化能力难以得到保证;另一方面,随着医学图像产出量的增大,传统方法难以适应大规模的数据集,模型可移植能力差。深度学习作为机器学习的新兴领域,在图像处理和计算机视觉方面的成功为医学图像的识别提供了新的思路。尽管当前深度学习在医学图像识别中已取得一定经验性的研究成果,但就总体而言,深度学习在医学图像识别中的应用还处于起步阶段,未来还有许多的问题需要



深入研究:

1) 在计算机视觉领域通过利用大量训练数据取得了突破性研究进展,如2015年何凯明设计一个具有152层的ResNet模型将错误率刷新到3.6%,该模型在ImageNet2012分类数据集上训练了128万张图像<sup>[18-19]</sup>。但在医学图像中由于数据的采集和疾病罕见等原因使得大规模医学图像数据的获取异常困难,可以借助图像处理中的迁移学习(transfer learning)和微调(fine tuning)来有效解决这方面的问题,但最好的解决方法还是应该建立更多公共可用的医学图像数据集,通过在公共数据集上提取更为抽象的特征,从而实现在医学图像识别上取得突破性的研究进展。<sup>[80-81]</sup>

2) 深度学习本质上是模拟人脑进行自动学习,从这一角度来看,深度学习是一种数据驱动型模型,但当前深度学习对无监督数据的学习能力严重不足,目前无监督学习算法主要面临两个方面的困难:一方面是高维数据通常具有数据维度高、数据量大等特性,在高维空间中进行相似度量会遇到低维空间中不曾遇到的问题,而相似性度量是无监督学习方法的一个重要指标;另一方面是数据噪声和不完全数据会影响分析过程,使得通过无监督学习方法所发现的模式准确性差。在未来可以致力于探索新的无监督学习算法,诸如为了提高可操作性,无监督学习算法应该具有交互能力,可交互的无监督学习算法的研究是一个重要的方向;为了更好地研究疾病,基于医学图像独特的复杂性、丰富性、重要性,需要针对这方面的无监督学习方法进行深入研究,如针对癌细胞、大脑疾病的无监督学习。

另外,不同医院采集数据的设备之间的异同使得采集到的图像质量有所差异,采集到的不同图像会影响特征的提取并对最终的结果起着决定性作用,为此开发出一种新颖的算法,有效克服不同设备获取的图像差异带来的影响,这也是未来的一个重要的研究方向。

## 参考文献

- [1] May M. Life science technologies: Big biological impacts from big data[J]. Science, 2014, 344(6189): 1298-1300.
- [2] Hinton GE, Ruslan R. Salakhutdinov. Reducing the dimensionality of data with neural networks[J]. Science, 2006, 313(5786): 504-507.
- [3] 周志华. 机器学习[M]. 北京: 清华大学出版社, 2016: 1-48.
- [4] Hinton GE, Osindero S, Teh YW. A fast learning algorithm for

deep belief nets[J]. Neural Computation, 2006, 18(7): 1527-1554.

- [5] Bengio Y, Lamblin P, Dan P, et al. Greedy layer-wise training of deep networks[C] // International Conference on Neural Information Processing Systems. Kitakyushu: Computer Science, 2007: 153-160.
- [6] Suk HI, Lee SW, Shen D, et al. Latent feature representation with stacked auto-encoder for AD/MCI diagnosis[J]. Brain Structure and Function, 2015, 220(2): 841-859.
- [7] Hinton G, Deng L, Yu D, et al. Deep neural networks for acoustic modeling in speech recognition: The shared views of four research groups[J]. IEEE Signal Processing Magazine, 2012, 29(6): 82-97.
- [8] Wei Xu, Rudnicky A. Language modeling for dialog system[C] // International Conference on Spoken Language Processing. Beijing: DBLP, 2000: 118-121.
- [9] Mikolov T, Deoras A, Povey D, et al. Strategies for training large scale neural network language models[C] // Automatic Speech Recognition and Understanding. Providence: IEEE, 2012: 196-201.
- [10] Hinton GE. Modeling pixel means and covariances using factorized third-order Boltzmann machines[C] // 2010 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). San Francisco: IEEE, 2010: 2551-2558.
- [11] Szegedy C, Vanhoucke V, Ioffe S, et al. Rethinking the inception architecture for computer vision[C] // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas: IEEE, 2016: 2818-2826.
- [12] Kivinen J, Williams C, Heess N. Visual boundary prediction: A deep neural prediction network and quality dissection[J]. Journal of Machine Learning Research: Workshop and Conference Proceedings, 2014, 33: 512-521.
- [13] Kris MG, Gucalp A, Epstein AS, et al. Assessing the performance of Watson for oncology, a decision support system, using actual contemporary clinical cases[J]. 2015 33(15): 8023-8023.
- [14] Haykin S, 著, 申富饶, 徐烨, 郑俊, 等译. 神经网络与机器学习[M]. 第三版. 北京: 机械工业出版社, 2011: 1-25.
- [15] Mason L, Baxter J, Bartlett P, et al. Boosting algorithms as gradient descent[C] // International Conference on Neural Information Processing Systems. Cambridge: MIT Press, 2000: 512-518.
- [16] Suykens, Johan AK, and Joos Vandewalle. Least squares support vector machine classifiers[J]. Neural Processing Letters, 1999, 9(3): 293-300.
- [17] Huang, Fu Jie, and Yann LeCun. Large-scale Learning with SVM and Convolutional for Generic Object Categorization[C] // IEEE Computer Society Conference on Computer Vision and Pattern Recognition. New York: IEEE, 2006: 284-291.
- [18] Krizhevsky A, Sutskever I, Hinton GE. ImageNet classification with deep convolutional neural networks[C] // International Conference on Neural Information Processing Systems. Lake Tahoe: Curran Associates Inc, 2012: 1097-1105.
- [19] Chen J, Deng L. A primal-dual method for training recurrent

- neural networks constrained by the echo-state property [J]. Proc Int Conf Learning Representations, 2013, 2013(420): 629201-629201.
- [20] Graves A, Jaitly N. Towards end-to-end speech recognition with recurrent neural networks [C] // Proceedings of the 31st International Conference on Machine Learning. Beijing: ICML, 2014: 1764-1772.
- [21] Yu D, Wang S, Karam Z, et al. Language recognition using deep-structured conditional random fields [C] // IEEE International Conference on Acoustics Speech and Signal Processing. Dallas: IEEE, 2010: 5030-5033.
- [22] Seltzer Michael L, Droppo J. Multi-task learning in deep neural networks for improved phoneme recognition [C] // 2013 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). Vancouver: IEEE, 2013: 6965-6969.
- [23] Zhang K, Zuo W, Chen Y, et al. Beyond a Gaussian Denoiser: Residual Learning of Deep CNN for Image Denoising [J]. IEEE Transactions on Image Processing, 2017, 26(7): 3142-3154.
- [24] Kim JH, Lee SW, Kwak D, et al. Multimodal residual learning for visual QA [C] // Advances in Neural Information Processing Systems. Barcelona: MIT Press, 2016: 361-373.
- [25] Wang Xiao Gang. Deep learning in image recognition [J]. Communications of the CCF, 2015, 11(8): 15-23.
- [26] 邓力, 俞栋. 深度学习: 方法及应用 [M]. 北京: 机械工业出版社, 2016: 3-4.
- [27] Lecun Y, Bengio Y, Hinton G. Deep learning [J]. Nature, 2015, 521(7553): 436-444.
- [28] 孙志远, 鲁成祥, 史忠植, 等. 深度学习研究与进展 [J]. 计算机科学, 2016, 43(2): 1-8.
- [29] Goodfellow I, Mirza M, Courville A, et al. Multi-prediction deep Boltzmann machines [C] // Advances in Neural Information Processing Systems. Lake Tahoe: MIT Press, 2013: 548-556.
- [30] Gens R, Domingos P. Discriminative learning of sum-product networks [C] // Advances in Neural Information Processing Systems. Lake Tahoe: MIT Press, 2012: 3239-3247.
- [31] Deng L, Li X. Machine learning paradigms for speech recognition: An overview [J]. IEEE Transactions on Audio, Speech, and Language Processing, 2013, 21(5): 1060-1089.
- [32] Vinyals O, Jia Y, Deng L, et al. Learning with recursive perceptual representations [C] // Advances in Neural Information Processing Systems. Lake Tahoe: MIT Press, 2012: 2825-2833.
- [33] Graves A, Jaitly N, Mohamed A. Hybrid speech recognition with deep bidirectional LSTM [C] // Automatic Speech Recognition and Understanding. Olomouc: IEEE, 2013: 273-278.
- [34] Yu D, Deng L. Deep-Structured Hidden Conditional Random Fields for Phonetic Recognition [C] // Conference of the International Speech Communication Association. Makuhari: BBLP, 2010: 2986-2989.
- [35] Deng L, Yu D. Deep learning: methods and applications [J]. Foundations & Trends in Signal Processing, 2013, 7(3): 197-387.
- [36] Dahl GE, Yu D, Deng L, et al. Context-dependent pre-trained deep neural networks for large-vocabulary speech recognition [J]. IEEE Transactions on Audio, Speech, and Language Processing, 2012, 20(1): 30-42.
- [37] Pascanu R, Mikolov T, Bengio Y. On the difficulty of training Recurrent Neural Networks [J]. Computer Science, 2012, 52(3): III-1310.
- [38] Imseng D, Motlicek P, Garner PN, et al. Impact of deep MLP architecture on different acoustic modeling techniques for under-resourced speech recognition [C] // Automatic Speech Recognition and Understanding. Olomouc: IEEE, 2013: 332-337.
- [39] Bengio Y. Learning deep architectures for AI [J]. Foundations and Trends in Machine Learning, 2009, 2(1): 1-127.
- [40] Bengio Y, Delalleau O. On the expressive power of deep architectures [M] // Algorithmic Learning Theory. Berlin: Springer, 2011: 18-36.
- [41] 曲建岭, 杜辰飞, 邱亚洲, 等. 深度自动编码器的研究与展望 [J]. 计算机与现代化, 2014, 8(228): 128-134.
- [42] Salakhutdinov R, Hinton GE. Deep Boltzmann Machines [J]. Journal of Machine Learning Research, 2009, 5(2): 1967-2006.
- [43] 山世光, 阚美娜, 刘昕, 等. 深度学习: 多层神经网络的复兴与变革 [J]. 科技导报, 2016, 34(14): 60-70.
- [44] Szegedy C, Liu Wei, Jia Y, et al. Going deeper with convolutions [C] // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Boston: IEEE, 2015: 1-9.
- [45] Chilimbi T, Suzue Y, Apacible J, et al. Project Adam: building an efficient and scalable deep learning training system [C] // Usenix Conference on Operating Systems Design and Implementation. Berkeley: USENIX Association, 2014: 571-582.
- [46] Vincent P, Larochelle H, Bengio Y, et al. Extracting and composing robust features with denoising autoencoders [C] // Proceedings of the 25th international conference on Machine learning. Helsinki: ICML, 2008: 1096-1103.
- [47] 李渊, 骆志刚, 管乃洋, 等. 生物医学数据分析中的深度学习应用 [J]. 生物化学与生物物理进展, 2016, 43(5): 472-483.
- [48] Dean J, Corrado GS, Monga R, et al. Large scale distributed deep networks [C] // International Conference on Neural Information Processing Systems. New York: Curran Associates Inc, 2012: 1223-1231.
- [49] McClelland J. Learning internal representations by error propagation [J]. Readings in Cognitive Science, 1988, 1(2): 399-421.
- [50] Burges C, Shaked T, Renshaw E, et al. Learning to rank using gradient descent [C] // Proceedings of the 22nd International Conference on Machine Learning. Bonn: ICML, 2005: 89-96.
- [51] Johnson R, Zhang T. Accelerating stochastic gradient descent using predictive variance reduction [C] // International Conference on Neural Information Processing Systems. New York: Curran Associates Inc, 2013: 315-323.
- [52] Hinton GE, Dayan P, Frey BJ, et al. The "wake-sleep" algorithm for unsupervised neural networks [J]. Science, 1995, 268(5214): 1158-1169.
- [53] Sarraf S, Tofighi G. Classification of alzheimer's disease using



- fmri data and deep learning convolutional neural networks [J]. IEEE Transactions on Medical Imaging , 2016 , 29 ( 3 ) : 1026-1031
- [54] Li F , Tran L , Thung KH , et al. A robust deep model for improved classification of AD/MCI patients [J]. IEEE Journal of Biomedical & Health Informatics , 2015 , 19 ( 5 ) : 1610-1616.
- [55] Summers RM. Progress in fully automated abdominal CT interpretation [J]. American Journal of Roentgenology , 2016 , 207 ( 1 ) : 67-79.
- [56] Sirinukunwattana K , Raza SEA , Tsang YW , et al. Locality sensitive deep learning for detection and classification of nuclei in routine colon cancer histology images [J]. IEEE Transactions on Medical Imaging , 2016 , 35 ( 5 ) : 1196-1206.
- [57] Dou Q , Chen H , Yu L , et al. Automatic Detection of Cerebral Microbleeds From MR Images via 3D Convolutional Neural Networks [J]. IEEE Transactions on Medical Imaging , 2016 , 35 ( 5 ) : 1182-1195.
- [58] Shin HC , Roth HR , Gao M , et al. Deep convolutional neural networks for computer - aided detection: CNN architectures , dataset characteristics and transfer learning [ J ]. IEEE Transactions on Medical Imaging , 2016 , 35 ( 5 ) : 1285-1298.
- [59] Dai W , Yang Q , Xue GR , et al. Boosting for transfer learning [ C ] // Proceedings of the 24th International Conference on Machine Learning. Corvallis: ICML , 2007: 193-200.
- [60] Charalambous CC , Bharath AA. A data augmentation methodology for training machine/deep learning gait recognition algorithms [J]. IEEE Transactions on Medical Imaging , 2016 , 35 ( 10 ) : 1016-1027.
- [61] Roth H , Lu L , Liu J , et al. Improving computer - aided detection using convolutional neural networks and random view aggregation [J]. IEEE Transactions on Medical Imaging , 2015 , 35 ( 5 ) : 1170-1181.
- [62] Srivastava N. Improving Neural Networks with Dropout [ D ]. Toronto: University of Toronto , 2013.
- [63] Wan L , Zeiler M , Zhang S , et al. Regularization of neural networks using dropconnect [ C ] // Proceedings of the 30th International Conference on Machine Learning. Atlanta: ICML , 2013: 1058-1066.
- [64] Setio A A , Ciompi F , Litjens G , et al. Pulmonary nodule detection in CT images: false positive reduction using multi-view convolutional networks [ J ]. IEEE Transactions on Medical Imaging , 2016 , 35 ( 5 ) : 1160-1169.
- [65] Li Q , Cai W , Wang X , et al. Medical image classification with convolutional neural network [ C ] // International Conference on Control Automation Robotics & Vision. Marina Bay: IEEE , 2014: 844-848.
- [66] Chakdar K , Potetz B. Deep learning for the semiautomated analysis of pap smears [J]. Medical Applications of Artificial Intelligence , 2014 , 18 ( 1 ) : 193-213.
- [67] Kondo T , Takao S , Ueno J. The 3 - dimensional medical image recognition of right and left kidneys by deep GMDH-type neural network [ C ] // Intelligent Informatics and Biomedical Sciences ( ICHBMS ). Rhodes: IEEE , 2015: 313-320.
- [68] Gao X , Lin S , Wong TY. Automatic feature learning to grade nuclear cataracts based on deep learning [J]. IEEE Transactions on Biomedical Engineering , 2015 , 62 ( 11 ) : 2693-2701.
- [69] Yan Z , Zhan Y , Peng Z , et al. Multi - instance deep learning: discover discriminative local anatomies for bodypart recognition [J]. IEEE Transactions on Medical Imaging , 2016 , 35 ( 5 ) : 1332-1343.
- [70] Plis SM , Hjelm DR , Salakhutdinov R , et al. Deep learning for neuroimaging: a validation study [J]. Frontiers in Neuroscience , 2013 , 8 ( 8 ) : 229-240.
- [71] Brosch T , Yoo Y , Li DKB , et al. Modeling the variability in brain morphology and lesion distribution in multiple sclerosis by deep learning [ M ] // Medical Image Computing and Computer-Assisted Intervention-MICCAI 2014. Beilin: Springer International Publishing , 2014: 462-469.
- [72] Xu J , Xiang L , Liu Q , et al. Stacked sparse autoencoder (SSAE) for nuclei detection on breast cancer histopathology images [J]. IEEE Transactions on Medical Imaging , 2016 , 35 ( 1 ) : 119-130.
- [73] Kooi T , Litjens G , Van GB , et al. Large scale deep learning for computer aided detection of mammographic lesions. [J]. Medical Image Analysis , 2017 , 35 ( 24 ) : 303-312.
- [74] Cruz-Roa , Angel Alfonso , Ovalle , et al. A deep learning architecture for image representation , visual interpretability and automated basal-cell carcinoma cancer detection [ C ] // International Conference on Medical Image Computing and Computer-Assisted Intervention. Nagoya: Springer-Verlag , 2013: 403-410.
- [75] Grinsven MJJPV , Ginneken BV , Hoyng CB , et al. Fast convolutional neural network training using selective data sampling: Application to hemorrhage detection in color fundus images [J]. IEEE Transactions on Medical Imaging , 2016 , 35 ( 5 ) : 1273-1284.
- [76] Roth HR , Lu L , Seff A , et al. A new 2.5D representation for lymph node detection using random sets of deep convolutional neural network observations [J]. Medical Image Computing and Computer - Assisted Intervention , 2014 , 17 ( 1 ) : 520-527.
- [77] Ypsilantis PP , Siddique M , Sohn HM , et al. Predicting response to neoadjuvant chemotherapy with PET imaging using convolutional neural networks [J]. PLoS ONE , 2015 , 10 ( 9 ) : 1-18.
- [78] He K , Zhang X , Ren S , et al. Deep residual learning for image recognition [ C ] // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas: IEEE , 2015: 770-778.
- [79] Russakovsky O , Deng J , Su H , et al. Image net large scale visual recognition challenge [ J ]. International Journal of Computer Vision , 2015 , 115 ( 3 ) : 211-252.
- [80] Huynh BQ , Li H , Giger ML. Digital mammographic tumor classification using transfer learning from deep convolutional neural networks [J]. Journal of Medical Imaging , 2016 , 3 ( 3 ) : 034501.
- [81] Tajbakhsh N , Shin JY , Gurudu SR , et al. Convolutional neural networks for medical image analysis: Full training or fine tuning? [J]. IEEE Transactions on Medical Imaging , 2016 , 35 ( 5 ) : 1299-1312.