

基于增强 AlexNet 的深度学习阿尔茨海默病的早期诊断

吕鸿蒙^{1,2} 赵 地¹ 迟学斌¹

(中国科学院计算机网络信息中心 北京 100190)¹ (中国科学院大学 北京 100049)²

摘 要 在中国,越来越多的老人正在忍受着阿尔茨海默病(Alzheimer's Disease,AD)的痛苦。阿尔茨海默病俗称老年痴呆症,临床上表现为失忆、丧失语言能力等。目前,中国的阿尔茨海默病患者人数已居世界第一。因此,早期诊断阿尔茨海默病变得十分急迫。研究表明,轻度认知障碍(Mild Cognitive Impairment,MCI)转化为阿尔茨海默病的概率很高,它是介于阿尔茨海默病和正常(Healthy Control,HC)之间的一种状态。随着大数据时代的来临,机器学习方法在疾病诊断方面受到热捧。所以,研究提出使用深度学习方法实现对阿尔茨海默病、轻度认知障碍和健康人群的诊断。数据库来自网络公开数据库 ADNI。原始的核磁共振图像(Magnetic Resonance Imaging,MRI)的预处理得到首都医科大学附属北京天坛医院的指导。使用卷积神经网络对降维后的实验数据进行训练。因为目前的网络模型不是针对医学图像的,所以实验的重点在于改进现有网络模型,使之达到良好的诊断效果。改进的网络模型是在图像分类方面十分出色的 AlexNet 网络模型。实验根据阿尔茨海默病的特点提出改进原始模型的 4 种算法,采用并行方式计算,使用曙光 W780-G20 服务器,利用 8 块 NVIDIA Tesla K80 进行 GPU 计算,获得 4 个分类器:AD vs. HC,AD vs. MCI,MCI vs. HC 和 AD vs. MCI vs. HC。数据集中图像总数量超过 7 万张,耗时不超过 30 分钟。最终,通过绘制 ROC 曲线,计算敏感度、特异度、精确度,对测试结果进行评估,得到了较好的测试结果。

关键词 阿尔茨海默病,轻度认知障碍,深度学习,卷积神经网络,增强的 AlexNet 网络模型,脑图像,核磁共振图像
中图法分类号 TP399 文献标识码 A

Deep Learning for Early Diagnosis of Alzheimer's Disease Based on Intensive AlexNet

LV Hong-meng^{1,2} ZHAO Di¹ CHI Xue-bin¹

(Computer Network Information Center, Chinese Academy of Sciences, Beijing 100190, China)¹

(University of Chinese Academy of Sciences, Beijing 100049, China)²

Abstract More and more people suffer from Alzheimer's disease(AD) in China. AD is characterized by loss of memory and language ability, associated with aging. Currently, the number of Chinese patients has ranked first in the world. So, early diagnosis of AD is particularly urgent. Studies have shown that mild cognitive impairment (MCI) has a high probability converted to AD. MCI may be a transition between healthy control (HC) and AD. With the advent of the era of big data, the machine learning algorithm is more and more popular in the diagnosis of disease. The method of deep learning helps us classify AD, MCI and HC. The data set of magnetic resonance imaging (MRI) is from Alzheimer disease neuroimaging initiative (ADNI) as the data set. The pre-treatment of the raw brain MRI is directed by Beijing Tiantan Hospital affiliated to Capital Medical University. Images after dimensionality reduction are learned by deep convolutional neural network (CNN) automatically. The current architecture of network is not for medical images. So, experiments focus on improving existing networks, so as to achieve good diagnostic results. AlexNet is an excellent architecture for images classification which the experiments choose to improve. In this paper, we proposed 4 algorithms to improve the original model according to the characteristic of AD. Data ran in parallel with 8 GPUs of NVIDIA Tesla K80 of W780-G20 by Sugon. Then, we obtained 4 classifiers, AD vs. HC, AD vs. MCI, MCI vs. HC and AD vs. MCI vs. HC. Models were trained no more than 30 minutes with more than 70000 images. Finally, algorithms were evaluated by drawing ROC curve and computing sensitivity and specificity, and the better results were showed.

Keywords Alzheimer's disease, Mild cognitive impairment, Deep learning, Convolutional neural network, Intensive Alex Net, Brain imaging, MRI

本文受国家高技术研究发展计划(863 计划)(2014AA01A302),国家自然科学基金重点项目(91530324),北京市自然科学基金重点项目(4161004),中国科学院计算机网络信息中心主任基金(CNIC_ZR_201502)资助。

吕鸿蒙(1992-),女,硕士生,主要研究领域为脑科学、高性能计算,E-mail:lvhm@sccas.cn;赵 地(1978-),男,博士,副教授,CCF 会员,主要研究领域为脑科学、大数据分析,E-mail:zhaodi@cnic.cn(通信作者);迟学斌(1963-),男,博士,教授,主要研究领域为高性能计算,E-mail:chi@sccas.cn。

1 引言

阿尔茨海默病(Alzheimer's Disease, AD)是典型的老年退行性疾病,俗称老年痴呆症,临床上表现为失忆,丧失行动、语言能力等。国际老年痴呆协会颁布的“2015 全球阿尔茨海默病报告”指出,2050 年,全球 AD 的患病人数将从现在的 4 千 7 百万增加到 1 亿 3 千 2 百万。报告表明了疾病发展速率的迅速及可能造成的严重后果,足见此病对人类的伤害。仅 2015 年,AD 的新增病例就达 1 千万左右^[1]。而且因为 AD 的患病原因不详,一般发现时已到晚期,即使治疗也不会有太大的效果。所以,对 AD 进行早期诊断是抑制病情快速发展甚至避免患病的较好方式。

轻度认知障碍(Mild Cognitive Impairment, MCI)可能是介于 AD 和健康的中间状态。研究表明,从 MCI 向 AD 的年转化率为 10%~15%^[2]。罹患 MCI 的患者比之前没有得过 MCI 的人更可能发展成 AD^[3]。所以,将更多的注意力放在 MCI 和 AD 的差异研究上对阿尔茨海默病的早期诊断有很大的帮助。因为 MCI 的症状比 AD 要轻很多,病人的日常生活能力并不会受到大的影响,只是记忆上出现衰退,所以在病情初期常常不易被发现,这也是延误病情甚至发展成 AD 的重要原因。

核磁共振成像(Magnetic Resonance Imaging, MRI)作为一种当下很流行的神经成像方法,已广泛用于 AD 的诊断^[4]。目前,机器学习方法在分析 MRI 图像时有很好的效果。最近的研究显示,使用机器学习算法预测 AD 比有经验的临床医生还要准确^[5]。机器学习,顾名思义,就是让机器模拟或实现人类的行为。机器学习方法,例如支持向量机(Support Vector Machines, SVM)^[6]、随机森林(Random Forest, RF)^[7]、boosting 算法^[8]等,对 MRI 分析起到了很重要的作用。此外,在以往的研究中,进行机器学习时需要从 MRI 中手工提取 AD 特征,如:海马、颞叶、杏仁核等兴趣区域(Regions Of Interest, ROI),这种方法固然有效,但也存在很大的局限性。首先,它可能出现一些人为的错误。而且因为 AD 的标志物不明,所以在划分兴趣区的时候很有可能出现疏漏,这将极大影响早期诊断的结果。其次,对疾病做诊断需要大量的实验数据进行训练,以保证预测的正确性。这项工作将产生极大的时间成本和人力成本。如果机器可以自动识别兴趣区,并把人类从这项重复性很强的工作中解脱出来,不仅可以保证诊断的正确性,还可以提升工作效率。

实验使用深度学习(Deep Learning, DL)方法,用网络模型训练大量的核磁共振图像,找出 AD 的特征,进而进行预测。深度学习是机器学习的重要组成部分,它通过模拟人脑来分析、解释数据,其重点在于建立与人脑相似的神经网络。人脑的复杂难以想象,所以建立相匹配的网络模型并非易事。通过大量的研究,实验将选用 AlexNet 模型作为 AD 和 MCI 的基础训练模型。AlexNet 是 2012 年由 krizhevsky 等人提出的 8 层卷积神经网络。其在当年的 ImageNet 比赛中脱颖而出,将图像识别的正确率提高了一倍多,可见其杰出的分类功力。但 AlexNet 对医学图像没有很好的针对性,本文将针对 AD 的医学意义对原有模型作出改进,这也是实验的重点。改进后的 AlexNet 提高了预测的正确率。

对 AD 的早期诊断主要有以下几个亮点:

(1)本文提出的针对医学图像的增强的 AlexNet 属于首创。增强的 AlexNet 模型是通过实验多种算法,针对 AD 特点比较优劣后得出的。实验基于深度学习方法,分别使用 AlexNet 和增强的 AlexNet 模型对实验数据集进行训练。实验针对这两个模型各获得了相应的 4 个分类:AD vs. HC(AH),AD vs. MCI(AM),MCI vs. HC(MH),AD vs. MCI vs. HC(3-WAY)。本文提出的增强的 AlexNet 模型更适合 AD 的早期诊断并且具有一定的医学意义。与原始模型相比,使用改进后的模型进行测试的正确率更高。

(2)Gupta 等人提出的无监督的学习方法在进行 MCI 预测 AD 方面取得了不错的效果。他们使用 Sparse AutoEncoders(SAE)方法进行特征提取,使用简单网络模型进行训练。本文提出使用 CNN 模型进行特征提取和训练,让机器全面地学习核磁共振图像。

(3)深度学习需要大量的训练样本才能达到准确分类的要求。在大数据面前,时间成本不可忽视。本文采用多 GPU 并行计算,使用曙光 W780-G20,利用 8 块 NVIDIA Tesla K80,使 4 个分类器同时进行训练,大大节约了训练时间。

(4)本文的数据库来源于网上公开的数据库 ADNI。另外,首都医科大学附属北京天坛医院也提供了中国的被试数据。首都医科大学附属北京天坛医院是北京脑重大疾病研究院,神经医学的顶尖专业临床与研究单位,拥有多套核磁共振扫描系统设备。ADNI 的数据量十分巨大,数据集包含了多种患病类型,这奠定了 AD 早期诊断的研究基础。实验得到首都医科大学附属北京天坛医院的临床数据支持,这对研究中国的阿尔茨海默病起到很重要的作用。众所周知,环境、天气、饮食等对疾病有很大的影响。而 ADNI 提供的数据为全世界的,不具有针对性。双管齐下,这使得对 AD 的研究更为透彻。在老年退行性疾病的预测平台建成之后,将使用首都医科大学附属北京天坛医院的被试数据进行实验并邀请医生进行评测。

此外,本文还将两种模型与现有的研究结果进行了评测和比较,对目前的研究现状及存在的问题进行了阐述。最后,提出了今后的研究方向。

2 基本概念

2.1 MRI 图像的张量表示

2.1.1 相关概念

张量^[9]是高维数据的表示形式,是矩阵的高维推广。标量可以看作是零阶张量,矢量是一阶张量,矩阵是二阶张量。张量是矩阵的推广,因此它也满足一些矩阵的性质。

Tucker 分解是张量分解的一种,它可以得到一个核心张量和一系列正交矩阵。Tucker 分解的概念最初是在研究心理学^[1-11]时提出的,然后逐渐被应用在许多学科中^[12]。Tucker 分解的优势在于减少像素之间的信息损失。在高维图像中,核心张量保证了原张量的重要信息不会丢失。Tucker 分解在数据挖掘、数据压缩中有很多成功应用。其函数可以表示为:

$$W = \{G; A_1, A_2, \dots, A_n\} = G \times A_1 \times A_2 \times \dots \times A_n \quad (1)$$

其中, W 是要被 Tucker 分解的张量, G 是核心张量, A_1, A_2, \dots, A_n 是一系列正交矩阵。

2.1.2 MRI 图像的预处理

MRI 图像为三维图像,但深度学习算法的输入要求为二

维图像。为了降低维度造成的图像信息的丢失,实验应用 Tucker 分解的概念将三维图像降维。Tucker 分解被广泛应用于信号过滤^[13]、图片去噪^[14]。图 1 给出了针对三阶张量的 Tucker 分解。待分解的张量是核心张量和每一个分量的正交矩阵的乘积。

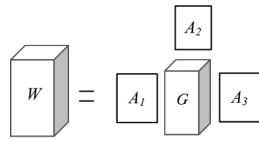


图 1 三维张量的 Tucker 分解

Tucker 分解应用于彩色图像,宽、高和 RGB 作为它的阶。但实验的 MRI 图像为医学图像,呈现黑白动态图,不同于 Tucker 分解之前处理的三维图像,所以 MRI 图像的预处理采用 Tucker 分解的概念,但又与之前的图像处理方式不完全一致。那么除了宽和高的另一阶设置成什么合适呢?因为 MRI 是通过射频脉冲信号实现对机体的检测的,所以经过一个固定的时间就会产生相应的横截面图。实验将选用时间作为它的另一阶。图 2 的 MRI 用张量表示,令它的阶为 I_1, I_2, I_3 ,分别表示图像的宽、高和时间。因为 MRI 图像和平常的彩色图像有一定的区别,它是多切面形成的动态脑图,实验为了减少对有效信息的压缩,只根据核磁共振射频参数(即 I_3)进行了张量分解。这个时间的长短主要参考核磁共振射频参数。如果时间过短,可能出现相同信息的多次采集,造成信息的冗余;如果时间过长,采集的信息可能会出现缺损,造成结果的误差。切片图像如图 2 所示。

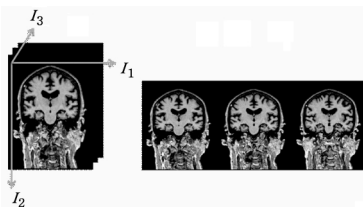


图 2 MRI 的切片

2.2 深度学习

1958 年,David Hubel 和 Torsten Wiesel 利用猫的瞳孔来研究瞳孔区域与大脑皮层神经元的对应关系^[15]。他们在猫的头骨中插入电极,让猫的神经元感受各种形状、亮度的物体,最终发现猫的不同的视觉神经元对刺激的反应不同,但存在某种对应关系。物体的边缘对猫的视觉神经元的刺激最大。具体的刺激过程是:猫的瞳孔接收到某种信息,视觉神经元将信息抽象化,大脑对抽象的信息不断细化最终得出结果。这说明生物的视觉系统会逐级抽象分层,抽象的迭代次数越多,得到的结果越精确。深度学习借鉴于此,研究了人脑的判定过程。通俗来说,抽象的过程就是寻找特征的过程。

深度学习的思想是构建多层次神经网络,下一层从上一层提取特征进行组合。因为物体的边缘信息的组成部分是相近的,所以不同物体的第一层的基本结构是相似的。深度学习让机器提取特征,避免了手工选择的劳累和不可避免的失误。但是每一层的特征提取次数需要根据经验来完成。太多的特征使计算量变大,太少的特征不能完成正确的表示。假设一个深度学习的模型共有 n 层,将每一层定义为 $P[1], P[2], \dots, P[n]$,输入为 I ,输出为 O 。

$$I \Rightarrow P[1] \Rightarrow P[2] \dots P[n] \Rightarrow O \quad (2)$$

其中, \Rightarrow 表示从上一层提取特征,但不能避免该过程的信息丢失。如果令 $O=I$,那么得到的特征最能表达原图。但这几乎是不可实现的任务,所以使 $O-I$ 的误差最小即可。图 3 是深度学习的多层结构,它具有多个隐藏层,与传统的神经网络不同。

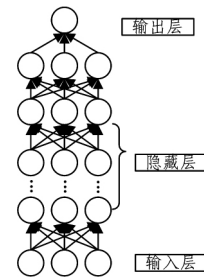


图 3 神经网络的多层结构

传统的神经网络应用反向传播 BP 算法,它是基于统计模型的机器学习算法,在早些年受到热捧^[16]。因为其可以从大量数据样本中基于统计原理得到特征并对其分类,所以解放了手工分类。但它是只有一层隐层结构的模型,如图 4 所示。而深度学习希望建立多层隐层结构,逐层提取特征,找到最适合的基底。深度学习更强调模型的层次、深度,通过逐层筛选特征达到准确预测的目的,如图 3 所示。与大数据结合后,其更能提高预测的准确性。

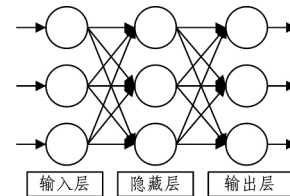


图 4 传统神经网络的层次结构

BP 算法可以表示为:

$$U_k = \sum_j W_{ij} X_j^{k-1} \quad (3)$$

其中, k 表示第 k 层, i 表示在第 i 层的神经元 i ,输入总和表示为 U_k , X 表示样本, X_j^{k-1} 表示第 $k-1$ 层的第 j 个神经元, W_{ij} 为神经元的权重系数。BP 算法利用输出与实际输出的最小误差来调整参数,最主要采用非线性规划方法中的最速下降方法。 W_{ij} 根据误差函数来调整。误差函数可以表示为:

$$e = \frac{1}{2} \sum_i (T_i^m - V_i)^2 \quad (4)$$

其中, e 为误差, T_i^m 为输出层 m 的实际期望, V_i 为输出期望。

BP 算法的激活函数为 S 型激活函数,公式如下:

$$C = c_1 w_1 + c_2 w_2 + \dots + c_n w_n \quad (5)$$

其中, C 代表输入, c_n 代表输入向量,权值为 w_n 。

$$f(C) = \frac{1}{1 + e^{-C}} \quad (6)$$

其中, $f(C)$ 为输出函数。

熟悉了传统的神经网络模型后可以更好地理解深度学习模型。这两种模型之间有相似之处。它们的共同点是都采用前向传播输出计算值,利用反向传播调整权重和偏置。不同点在于深度学习是用多层次的神经网络模拟人脑的神经元达到分类、诊断的作用,深度学习通过将低层次的特征抽象为高层次完成特征的学习;而传统的神经网络只有一层的隐藏层,结构比深度学习的卷积神经网络简单许多。但深度学习也有缺点,即它的参数难以调整,容易产生过拟合现象,所以在发

展初期沉寂了很长时间。

大脑的神经元众多,深度学习在输入层和输出层之间建立了多层隐层结构,每一层都与它的前一层相连,跨层之间没有联系,这是比较接近大脑中的神经元分布的。

从图 3 中可以看出,深度学习更强调深层的结构。在传统的神经网络中应用 BP 算法做训练。随机设定初值,根据输出和最终结果的误差来调整参数,直到收敛。由于 BP 算法只有一个隐藏层,所以可以根据误差调整参数。但在多层的结构中,这点不适用。因为这时误差逐层往下传值,到达底部的误差已经变得太小,没办法达到调整参数的作用。而且 BP 算法作为有监督的深度学习算法,需要提供结果的标签(label),但人脑可以识别没有标签的东西,这就是深度学习可以完成无监督的学习的原因。有监督的学习在训练时每个输入对应一个输出;无监督的学习只有输入值,让机器识别输出是什么。

2006 年,Hinton 提出 DL 的训练过程^[17]:

1)自底向上的非监督学习。逐层训练,令训练前的特征与训练后的特征最相近。

2)自顶向下的监督学习。训练完所有层后,使用 Wake-Sleep 算法进行调优。

Wake-Sleep 算法分为醒(Wake)和睡(Sleep)两个部分。

1)Wake 阶段:对样本进行抽象表示。通过输出和权重产生每一层的抽象特征。从底层到顶层进行抽象。使用梯度下降算法修改权重,使得抽象的结果最能代表本层。

2)Sleep 阶段:从最终输出和权重值完成底层状态的描述。在表示过程中,从顶层到底层,通过误差不断修正层间的权重。

需说明的是,大脑学习的特征并不是图片的像素值,而是一个结构。这个结构的线性组合能描述出原图。Bruno Olshause 和 David Field 提出了这个想法,并付诸了实践。他们收集了很多灰度的风景图片,从这些图片中提取了 400 个尺寸为 16×16 的碎片,用碎片的线性组合表示另一个尺寸相同的碎片。将 400 个碎片记作 $S[1], S[2], \dots, S[400]$,将待表示的碎片记作 T 。

$$\sum a[k] * S[k] \rightarrow Q \quad (7)$$

其中, $a[k]$ 是叠加碎片 $S[k]$ 时的权重系数。在能准确表示 Q 的情况下, $S[k]$ 取最小值。这就是后来提出的稀疏编码(Sparse Coding)算法^[18]的核心思想。

1)随机选择一组 $S[k]$,令 $\sum a[k] * S[k]$ 和 Q 最接近,得到 $a[k]$ 。

2)选择其它更合适的碎片 $S'[k]$ 替代原先的 $S[k]$,使得 $\sum a[k] * S'[k]$ 和 Q 最接近。

3)重复步骤 1)一步骤 2)。

经过几次迭代后得到 $S[k]$,这时 $S[k]$ 的线性组合与 Q 最接近。研究表明,这时的 $S[k]$ 大多为原图片的边缘曲线。这些边缘曲线构成特征的基本结构,可以完整地表示一幅图片。因为有效表示 Q 的 $S[k]$ 的个数最少,所以大多数不表示 Q 的碎片的 $a[k]$ 为 0,故取名为稀疏编码。

稀疏编码分为训练和编码两个阶段。在训练阶段,输入的图片通过学习得到字典 $S[k]$ 。学习的过程是一个迭代的过程。先固定 $S[k]$,得到使目标函数和输出误差最小的 $a[k]$ 。然后固定 $a[k]$,得到使目标函数和输出误差最小的

$S[k]$ 。迭代的次数取决于目标函数是否收敛。当目标函数收敛时,得到最适合的字典 $S[k]$ 。得到 $S[k]$ 后,根据输入的测试图片,原方程可解,得到 $a[k]$,这就是编码阶段的求解。式(8)为求解字典的公式。

$$\min \sum_{i=1}^m \| P_i - \sum_{j=1}^k a_{ij} S_j \|^2 + \gamma \sum_{i=1}^m \sum_{j=1}^k |a_{ij}| \quad (8)$$

其中, P_i 为输入, $\sum_{j=1}^k a_{ij} S_j$ 求得输出, γ 为约束系数。

2.3 卷积神经网络

卷积神经网络(Convolutional Neural Network,CNN)^[19]是一种多层网络模型,是深度学习的神经网络模型。训练好的过滤器和局部邻域池化操作交替作用在输入的图片上,然后输出一系列越来越复杂的特征^[20]。它是对 BP (Back Propagation)神经网络的一种改进。它们的共同点是都采用前向传播输出计算值,利用反向传播调整权重和偏置。但 CNN 在层与层之间采用局部连接,而 BP 神经网络采用的是全连接。这个局部连接的概念被命名为局部感受野。CNN 有一个输入层,每一层都接收上一层的输入,直到输出,但在整个网络中没有反馈,由一个有向环图表示。CNN 在原始输入中重复使用卷积操作和局部池(pooling)操作得到越来越复杂的特征图(feature map)。CNN 中的层(layer)由多个图(map)组成,每个 map 由多个神经元构成。map 起到过滤器的作用,同一个 map 的所有神经元共用一个特征,这个特征叫做卷积核,在 CNN 中被命名为权重。针对这个特征计算输入图片的卷积值,其中数值最大的区域与特征匹配,得到合适的特征,并将其输入到下一层结构中。但计算卷积值会耗费许多的时间成本,其实在实际操作中只是用固定大小的权重矩阵与输入图片比对,而不真正地计算。

CNN 的历史可以追溯到 1962 年,Hubel 和 Wiesel 通过猫的视觉皮层细胞提出了感受野(Receptive Field,RF)的概念^[21],这在上文中已经详细阐述。RF 的经典定义是影响神经元工作的一个视觉空间^[22]。通俗来说,RF 就是特征图与输入图像的某个区域匹配,这个区域就叫做感受野。之后,日本学者 Fukushima 基于感受野提出了神经认知机,这可以认为是最早的 CNN。神经认知机在理论上可以识别文字,但需要大量的神经元,在实现上还存在困难,而且对大小改变、平移等数学变换束手无策。随着 CNN 发展的深入,多层次高适应性的网络结构出现。现在已经可以实现手写文字的识别^[23]、物体识别^[19]等图像分类问题。

因为 CNN 在图像识别上的优势,所以实验决定使用多层 CNN 网络作为训练模型。图 5 展示了切片后的 MRI 经过卷积操作和池化操作的过程。

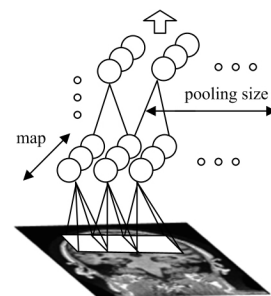


图 5 脑图的卷积、池化

图 5 的卷积过程由 3 个带有偏置的过滤器完成,每个过滤器带有一个卷积核。切片后的脑图进行卷积操作,然后增

加一个偏置得到卷积层。接下来,进行池化过程,通过求和、增加权重与偏置、激活函数计算等操作得到新的特征图。池化操作是对特征进行聚合统计的过程,这些特征可能分布在不同区域。池化只需要计算某个部分的某个特征的平均值或者最大值,而不需要关联所有特征。图5的脑图在经过池化操作后还需要经过卷积等过程,这是因为核磁共振图像是信息量很大的图片,需要经过多次卷积和池化才能保证特征被完全识别。在经历几个周期后,图片上的像素值被栅格化(rasterize),连接成一个长向量输入到传统的神经网络中,得到输出。图6示出了一个完整的CNN的框架^[24]。需要注意的是,大多CNN都不止经过一个周期的卷积、池化过程,而且卷积、池化操作也不一定非要组合出现。

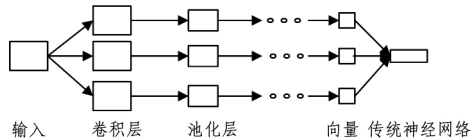


图6 CNN的框架

1)特征提取:特征提取作用在卷积层。神经元连接上一层的局部感受野,然后进行特征提取。特征的位置取决于现有的其它特征^[4]。

2)特征映射:特征映射作用在池化层,也叫作子采样层。它具有平移不变性,在特征经过移动后还可以做出正确的判断,在图像识别中很有优势。另外,在神经元之间进行权值共享可以达到减少自由参数的目的。

一般情况下,经过卷积层后map的数量会出现变化,这是由于卷积核不同造成的。但是在最大池化(max-pooling)操作后的map数量是不变的,因为池化并不改变特征的数量,只是聚合统计的一种方法。

CNN优于其他网络结构的地方在于局部感受野和权值共享。举例:如图7所示,在一张 256×256 的脑图中,假设隐层神经元为 256×256 ,如果使用传统神经网络做全连接,那么将产生 $256^4 = 4294967296$ 个连接。如果引入局部感受野,每个神经元不需要连接所有像素点,也即不需要做全连接,那么结果会是怎样呢?假设局部感受野为 10×10 ,现在 256×256 个神经元只需要与局部感受野连接,那么只需要 $256 \times 256 \times 10 \times 10 = 6553600$ 个连接,是之前连接的 $1/65$ 。而且CNN的卷积层如果用同一个卷积核对图像操作,那么不论隐层神经元多大,这之间的权值共享只有100,这是其他传统神经网络望尘莫及的。

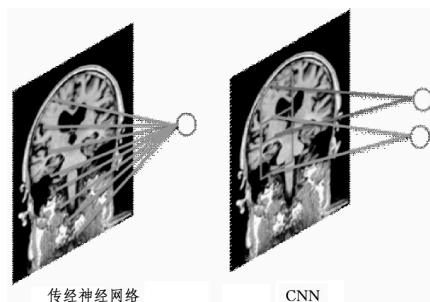


图7 CNN的优势

由于实验要对AD的MRI图像进行诊断,而MRI图像与日常的静态图像不同,它有自己的医学意义,因此设计针对医学图像的CNN是对AD进行深度学习的关键。

3 数据集和预处理

3.1 MRI数据集

本文的实验数据是从阿尔茨海默病神经影像学网站得到的(ADNI)。这是一个由国家生物医学成像和生物工程研究所(NIBIB)、美国食品和药物管理局(FDA)、民营医药企业与非营利组织组成的研究协会,协会创建于2014年。虽然它看上去很年轻,但网站上的数据集十分庞大。ADNI收集了超过800名来自全球59个地区的测试者。被跟踪时间最长的测试者提供了3年的MRI、正电子发射断层扫描。他们在每6个月或12个月时接受针对AD的心理测试。更多的信息可以在<http://www.adni-info.org>网站上搜索。针对AD的心理测试主要包括临床痴呆评定量表(Clinical dementia rating,CDR)^[25]和简易精神状态检查表(mini-mental state examination,MMSE)^[26],这两类表可以作为AD的心理学诊断标准。表1和表2分别描述了痴呆程度和受教育程度。ADNI为AD和MCI的研究提供了丰富的数据支持和诊断决策。表1—表3所列心理学测试和AD的诊断标准。

表1 CDR的诊断标准

CDR			
CDR 0.5	CDR 1.0	CDR 2.0	CDR 3.0
疑似痴呆	轻度痴呆	中度痴呆	严重痴呆
总分为3			

表2 MMSE的诊断标准

MMSE		
≤ 17	≤ 20	≤ 24
文盲	小学程度	中学程度
总分为30		

表3 一种判定AD的诊断标准

诊断标准		
HC	MCI	AD
CDR $0 \sim 0.5$ &	CDR $0.5 \sim 1$ &	CDR $1 \sim 3$ &
MMSE ≥ 24	MMSE ≥ 24	MMSE < 24

实验选择的ADNI网站上的数据集是使用快速梯度回波成像技术(MPRAGE)采集并跟踪2年的1.5T T1相,详细参数为:TR为2400ms;层厚为1.2mm;图片尺寸为 $192 \times 192 \times 160$ 。为了研究出符合中国国情的诊断系统,实验得到了首都医科大学附属北京天坛医院的数据支持,该数据将用于老年退行性疾病的预测平台建成后的测评工作,但本文的测试数据中不包含此数据。本研究通过首都医科大学附属北京天坛医院伦理委员会审批。表4列出甄选出来的数据的详细信息。

表4 ADNI数据集的详细信息

类别	类别人数	男/女	年龄	切片数量
			(平均值±偏差)	
AD	43	24/19	76.86 ± 6.85	6768
MCI	113	69/44	76.40 ± 7.21	10115
HC	148	69/79	76.70 ± 5.17	10164

3.2 MRI预处理

如果用深度学习来诊断疾病,那么数据样本量必须足够大。实验没有借鉴国外的标准化处理过程,因为当样本量足够大时,各种各样的情况都会包括进去,实验也力求降低信息的缺损。但是,不做标准化处理不代表对数据不做处理,实验需要的是没有信息缺损、脑动微小的MRI图像。在这一方

面,实验得到了医生的医学帮助,使用医学软件 MRIcron 进行了数据集的甄选。然后,将经过分解的 MRI 图像的大小调整为 256×256 。实验选用了大脑的冠状位,这是因为医生在确诊 AD 时更多是根据冠状位图,实验力求可以与医生的临床诊断过程一致。图 8 为 AD,MCI 和 HC 的冠状位图。

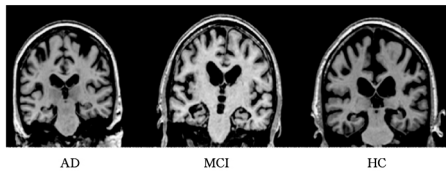


图 8 AD,MCI 和 HC 的冠状位图

4 实验

4.1 训练集和测试集

实验建立了 4 组数据集,分别是 AD vs. HC(AH),AD vs. MCI(AM),MCI vs. HC(MH)和 AD vs. MCI vs. HC(3-WAY)。因为要对 AD 的早期进行诊断,所以需要了解 MCI 和 AD,MCI 和 HC 的关系。实验数据被分为训练集和测试集,约占被试数量的 90%和 10%。其中,训练集的 25%作为验证集。

4.2 卷积神经网络 AlexNet

数据集准备完毕后,再通过实验选择卷积神经网络 AlexNet 作为训练的网络模型原型。在阅读文献中发现 AlexNet 在脑图像识别上有自身的优势^[27]。AlexNet 拥有 8 层结构,在图片识别上十分出色。它由 Krizhevsky 等人^[28]发明并夺得了 2012 年 ImageNet 比赛的冠军。它将 ImageNet 的正确识别率提高了一倍多。AlexNet 的具体结构如图 9 所示。

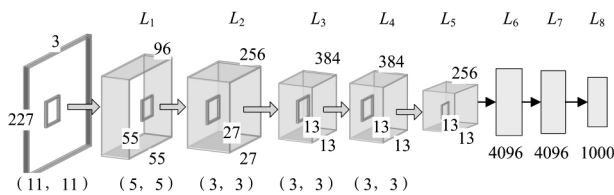


图 9 AlexNet 网络结构图

输入为 $227 \times 227 \times 3$,227 为数据源的宽和高,3 代表彩色图片 RGB。因为 MRI 为灰度图,在训练之前还需要对它进行彩色化。第一层和第二层经历了相同的计算过程,即卷积(conv)、ReLU(rectified linear units)、最大池化(max-pooling)和归一化(norm)处理。以第一层为例,具体过程如图 10 所示。

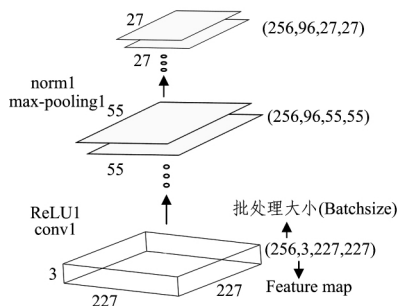


图 10 AlexNet 第一层的详细解释

数据源经过 96 个卷积核(num_output),大小(kernel_

size)为 11,步长(stride)为 4。卷积后的数据大小为 $55 \times 55 \times 96(55=(227-11)/4+1)$,经过 ReLU 激活函数,数据源大小不变。然后做 max-pooling,其中 kernel_size 为 3, stride 为 2,得到 map 大小为 $27 \times 27 \times 96(27=(55-3)/2+1)$ 。最后,经过 norm,参数 local_size 设置为 5, alpha 为 0.0001, beta 为 0.75,得到输入下一层的数据大小为 $27 \times 27 \times 96$ 。

第二层的计算过程与第一层一样。第一层输出的结果经过 256 个 kernel_size 为 5, stride 为 1 的 feature map 做卷积操作,其他参数与第一层一样。需要注意的是在第二层处理图像时需要加边(pad),数值为 2。卷积完的大小为 $27 \times 27 \times 256(27=(27+2 \times 2-5)/1+1)$,其中加边是双向的,所以需要 2×2 。第三层和第四层只做了 conv 和 ReLU。第五层与第一层的处理过程类似,只是没经过 norm。在得到第五层的结果后,将结果变成一个长向量,输入到传统的神经网络中,这里使用了 3 层的全连接网络。传统神经网络与之前的卷积网络的区别是没有共享权重。前两个全连接层的 kernel 数为 4096。最终层产生 1000 个节点,应用 Softmax 回归函数得到分类(label)值^[29]。Softmax 回归函数可以解决多分类问题,实验要做 AD 的早期诊断,必定会对 3-WAY 做分类。经典的 Sigmoid 回归只能解决二分类问题。Softmax 公式如式(9)所示。

$$E(d)_j = \frac{e_j^{d_j}}{\sum_g e_g^{d_g}} \quad (9)$$

其中, g 是种类个数, d 是 g 的训练参数,每个 d 与一个向量对应。

4.3 增强的 AlexNet 网络模型

虽然 AlexNet 在图像识别上十分出色,但它并不是针对医学图像的网络模型。针对 AD 问题,AlexNet 没有显示出带有医学意义的框架设计,但它在诊断 AD 上已经有了很好的表现,下文将介绍实验结果。虽然它已完成了对 AD 进行诊断的要求,但改进原有模型使它在医学诊断上更有意义仍然是实验工作的重点。下面给出原始 AlexNet 的算法流程,如图 11 所示。

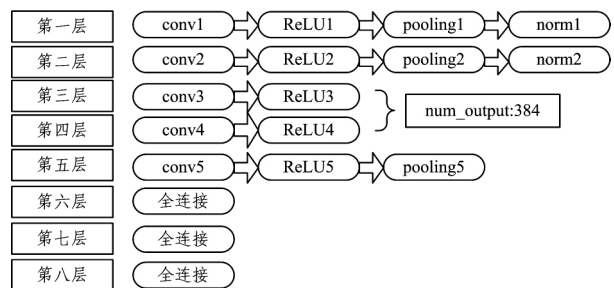


图 11 AlexNet 的算法流程图

AlexNet 有八层结构,前五层构建的是卷积神经网络,第六层到第八层为传统神经网络。第一层和第二层都使用了卷积、ReLU、池化、标准化操作。第三层和第四层参数相同,都使用了卷积和 ReLU 操作。第五层使用了卷积、ReLU、池化操作。操作后的结果输入到传统的神经网络中,最后三层使用全连接的网络结构。具体算法如下。

算法 1 AlexNet

将 input 标准化为 $227 \times 227 \times 3$;
进入第一层:

```
conv1:
    num_output=96;
    kernel_size=1;
    stride=4;
ReLU1;
pooling1:
    kernel_size=3;
    stride=2;
norm1:
    local_size=5;
进入第二层:
conv2:
    num_output=256;
    pad=2;
    kernel_size=5;
    group=2;
ReLU2;
pooling2:
    kernel_size=3;
    stride=2;
norm2:
    local_size=5;
进入第三层:
conv3:
    num_output=384;
    pad=1;
    kernel_size=3;
ReLU3;
进入第四层:
conv4:
    num_output=384;
    pad=1;
    kernel_size=3;
ReLU4;
进入第五层:
conv5:
    num_output=256;
    pad=1;
    kernel_size=3;
ReLU5;
pooling5:
    kernel_size=3;
    stride=2;
进入第六层到第八层:
全连接操作;
```

使用算法 1 后对测试集的诊断精确度如表 5 所列。

表 5 算法 1 的诊断精确度/%

	AH	AM	MH	3-WAY
精确度	95.44	89.66	86.80	84.39

MRI 图像中,海马体、颞叶、扣带回等区域是诊断 AD 的关键部分。因为 AD 患者的海马体已重度萎缩,所以根据疾病特征增加卷积核个数或者改变卷积层可以更好地进行诊断。经过阅读文献发现在原 AlexNet 模型的第三层和第四层对 AD, MCI 和 HC 提取特征的能力最强^[30]。所以,对模型的

改进重点集中在这两层。实验尝试在第四层之后增加一层,这一层与原来的第四层结构一致,力求对特征进行更精确的提取。算法流程图如图 12 所示。

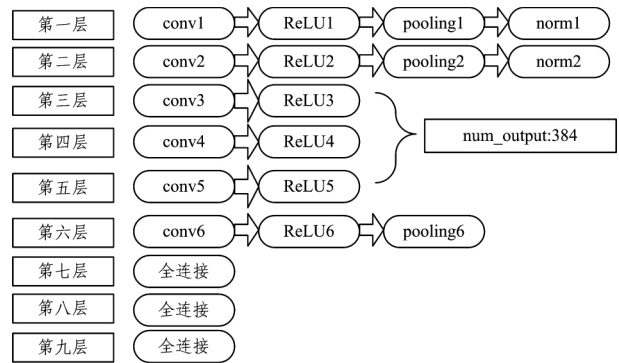


图 12 算法 2 的算法流程图

在该算法中,增加了与第四层相同参数的第五层。因为文献资料表中,第三层和第四层的分类效果好于其他层^[30]。第三层和第四层的参数一致,但是第四层的分类效果略优于第三层。因此,实验尝试在第四层的基础上增加与其参数一致的第五层,实现九层结构的卷积神经网络。具体算法如下。

算法 2 AlexNet-4same5

将 input 标准化为 $227 * 227 * 3$;

第一层到第四层与算法 1 一致。

进入第五层:

```
conv5:
    num_output=384;
    pad=1;
    kernel_size=3;
ReLU5;
```

第六层到第九层的参数与算法 1 的第五层到第八层的参数一致,这里不再赘述。

使用算法 2 后对测试集的诊断精确度如表 6 所列。

表 6 算法 2 的诊断精确度/%

	AH	AM	MH	3-WAY
精确度	96.49	67.24	76.01	76.74

修改后的第五层与第四层有相同的结构。经过实验发现,修改的算法在 AH 这个分类上比原先的 AlexNet 更好。其他的分类器不如算法 1 的分类效果好。

然后,实验尝试让第五层的结构与算法 1 第一层的结构相同,完成九层结构的卷积神经网络。算法流程图如图 13 所示。

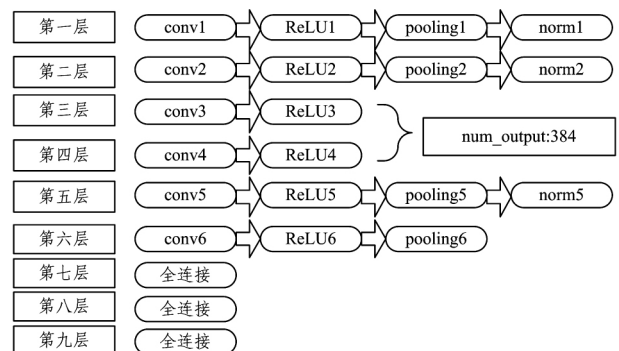


图 13 算法 3 的算法流程图

在该算法中,增加了与算法 1 第一层相同结构的第五层,实现九层结构的卷积神经网络。这时的第五层包括了卷积、ReLU、池化、标准化操作。实验试图增强原网络模型 AlexNet 的提取特征强度来提高诊断的精确性。具体算法如下。

算法 3 AlexNet-sIsames5

将 input 标准化为 $227 * 227 * 3$;

第一层到第四层与算法 1 一致。

进入第五层:

```
conv5:
    num_output=256;
    pad=1;
    kernel_size=3;
```

ReLU5;

pooling5:

```
    kernel_size=3
    stride=2
```

norm5;

第六层到第九层的参数与算法 1 的第五层到第八层的参数一致,这里不赘述。

使用算法 3 后对测试集的诊断精确度如表 7 所列。

表 7 算法 3 的诊断精确度/%

	AH	AM	MH	3-WAY
精确度	85.26	74.14	82.09	77.74

经过实验发现,修改的算法不如算法 1 的诊断效果好。

实验开始尝试对卷积核的参数进行更改。修改参数的目的是对海马体、颞叶部分更准确地进行特征提取,减少信息丢失。所以,选择合适的卷积核个数十分重要。MRI 为三维图片,海马体、颞叶等诊断区域随着数轴进行变化,实验尝试找到患病部位的最合适的卷积核个数。

算法流程如图 14 所示。

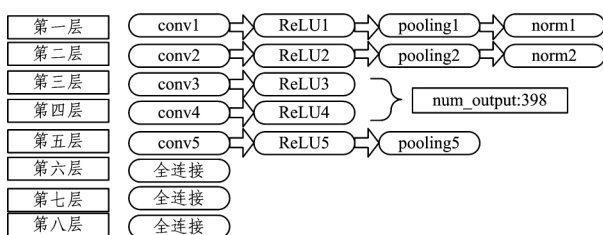


图 14 算法 4 的算法流程图

文献显示第三层和第四层的特征提取能力最强。因此,实验更改了第三层和第四层的 num_output,即 kernel 的数量,从 370 开始做公差为 2 的批处理。具体算法如下。

算法 4 AlexNet-l3l4_398

将 input 标准化为 $227 * 227 * 3$;

FOR n LESS THAN N

前两层与算法 1 的前两层一致。

进入第三层:

```
conv3:
    num_output=n;
    pad=1;
    kernel_size=3;
```

ReLU3;

进入第四层:

```
conv4:
```

```
    num_output=n;
```

```
    pad=1;
```

```
    kernel_size=3;
```

ReLU4;

第五层到第八层与算法 1 一致。

n=n+2;

END FOR

n 从 370 取值,N 为 410。在 num_output 为 398 的 kernel 上达到测试结果的峰值。

使用算法 4 后对测试集的诊断精确度如表 8 所列。

表 8 算法 4 的诊断精确度/%

	AH	AM	MH	3-WAY
精确度	96.14	90.52	84.80	86.05

从表 8 中看出,算法 4 在 AH,AM,3-WAY 这 3 个数据集上的分类效果比算法 1 出色,但对 MH 的分类效果不如算法 1。算法 4 的 AH 分类效果比算法 2 稍差一些。算法 1 在图像识别上已经非常出众,但算法 4 针对 AD 的诊断有其医学意义并且正确率有所提高。

综合来看,算法 4 的精确度比之前设计的算法有一定提高。如果在算法 4 的基础上再增加一层的卷积、ReLU 操作,结果会不会更加好?带着这样的猜测,本文设计了算法 5。算法 5 增加了一层具有卷积、ReLU 操作的结构,并将算法 4 更改的参数应用在算法 5 中。算法流程如图 15 所示。

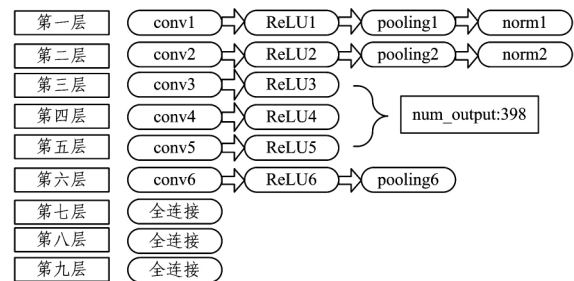


图 15 算法 5 的算法流程图

算法 5 增加的第五层的参数和结构与自身的三层和第四层一致。具体算法如下。

算法 5 AlexNet-i5_398

将 input 标准化为 $227 * 227 * 3$;

前四层与算法 4 的前四层一致。

进入第五层:

```
conv5:
    num_output=398;
    pad=1;
    kernel_size=3;
```

ReLU5;

第六层至第九层与算法 4 的第五层至第八层一致。

使用算法 5 后对测试集的诊断精确度如表 9 所列。

表 9 算法 5 的诊断精确度/%

	AH	AM	MH	3-WAY
精确度	94.39	87.06	79.73	74.75

从表 9 中看出,算法 5 的精确度不如算法 1。

最终,使用柱状图表示各个算法的精确度以更好地比较各个算法。各算法的精确度对比图如图 16 所示。

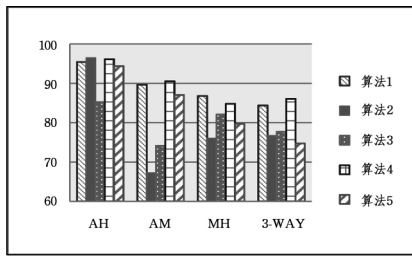


图 16 各算法精确度对比

综合上述 5 种算法可知,算法 2 AlexNet-4same5 对 AH 的分类效果最好。这是因为它增加了卷积操作,过滤出了更多的有效特征。算法 4 AlexNet-l3l4_398 对 AM 和 3-WAY 的分类效果最好,这是因为它根据 AD 患病部位的特点进行了卷积操作。算法 1 对 MH 的分类效果最好。虽然算法 4 对 AH 的分类效果不如算法 2,但差距极小,而且它更多地针对了 AD 的特点。所以,下文将使用算法 4 的网络模型作为增强的 AlexNet 网络模型与原始 AlexNet 网络模型进行对比。算法 4 被命名为增强的 AlexNet 网络模型,在下文实验对比中将用此名称进行结果的评测。

4.4 实验过程

MRI 图像进行张量分解等预处理后输入到深度神经网络中。实验使用的深度学习软件为 DIGHTS2.0,它是基于 Caffe 的一种可视化软件。Caffe^[31] 是基于深度学习算法开发的一个开源库,由 C++ 和 CUDA 语言实现。用户只需调用其中的深度学习算法,而不需要了解其中复杂的底层代码,具有方便快捷的功能。实验建立了 4 个数据库,分别是 AH, AM, MH 和 3-WAY。在建立数据库之后会产生相应的统计信息。图 17 是 AH 数据库的平均值(mean)。之后,将训练集分别放入 AlexNet 和增强的 AlexNet 网络模型中进行训练。将测试集放入训练好的模型中进行测试,并对实验结果进行比较和评价。

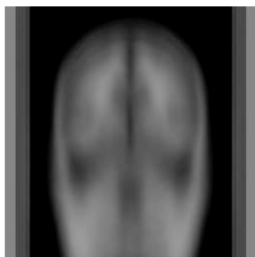


图 17 AH 的 mean 图

5 结果和评价

现在已经拥有了两种网络模型: AlexNet 模型(算法 1)和增强的 AlexNet 模型(算法 4)。分别对这两种模型进行实验,每个实验分为 4 组: AH, AM, MH 和 3-WAY。使用的 GPU 设备为曙光 W780-G20,4U 机架式,配备 8 块 NVIDIA Tesla K80,2 块 Intel Xeon E5-2620 v3 6 核处理器(2.4GHz)。4 组数据集并行执行,每组使用 2 块 K80,4 个 GPU 运算。在这 4 组中,单个训练集最大达到 26000 多张训练图片。完成 4 组实验不超过 30min。

将卷积神经网络应用在核磁共振图像上。为了更清楚地看到卷积神经网络的工作过程,本文对 AH 组进行测试。使用一张 AD 测试图片通过 AlexNet 的第一层,将其中的计算

过程可视化。因为增强的 AlexNet 模型的前两层与原 AlexNet 模型相同,所以图 18 的结构可以对两种模型通用。这也是针对上文 CNN 结构的应用示例。虽然实验输入的训练集为灰度图,但因为 AlexNet 网络结构输入为 $227 \times 227 \times 3$ (3 代表着 RGB 三通道),所以服务器程序将灰度图转化成了彩色图。

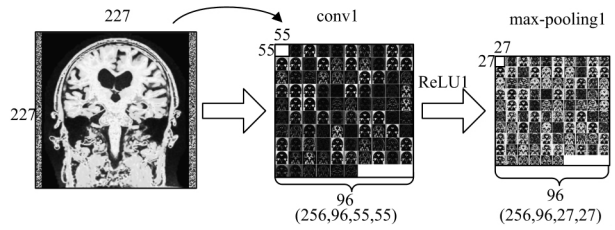


图 18 CNN 第一层计算过程的可视化展示(AH)

增强的 AlexNet 模型的测试过程如图 19 所示。因为第八层将第七层的输入变成 1000 个节点,并通过激活函数输出 label 值,所以在图中并没有做出可视化显示。

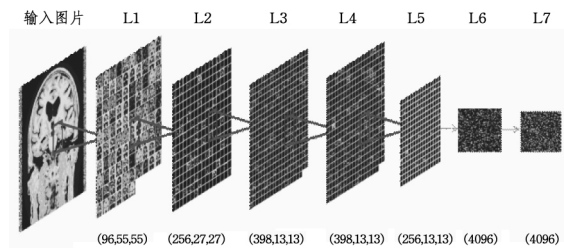


图 19 使用增强的 AlexNet 模型的测试过程

实验分别对两种模型进行了训练,经过 30 次迭代后产生训练结果。表 10 显示两种模型的验证精确度均超过 95%,说明训练出的模型为有效模型。

表 10 验证集的验证结果

	AH	AM	MH	3-WAY
精确度 (AlexNet)/%	98.00	97.32	95.65	96.14
精确度 (增强的 AlexNet)/%	98.10	96.93	95.63	96.95
损失(lose) (AlexNet)	0.058	0.076	0.120	0.105
损失(lose) (增强的 AlexNet)	0.056	0.082	0.117	0.091

图 20 示出了增强的 AlexNet 模型对 AH 数据集进行训练的迭代过程。因为已经对两种模型的验证结果做了表格展示,所以这里只展示了增强的 AlexNet 模型的一组实验的迭代过程。从图 20 可以看出,实验的曲线符合深度学习曲线的规律,迭代 30 次的曲线已经趋于平稳。

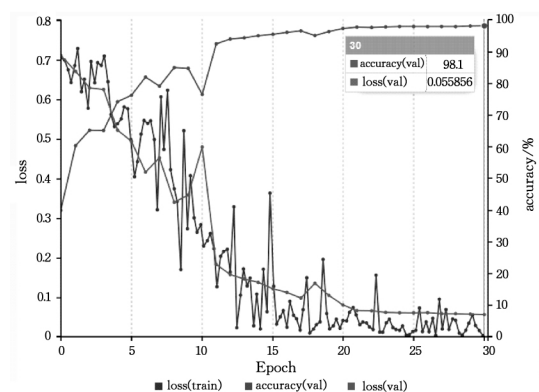


图 20 增强的 AlexNet 模型对 AH 数据集训练的迭代过程

经过大量的研究和测试,实验得到了较好的结果。由于 MRI 图像的动态化,使得一部分的数据对疾病来说是无用的。如果大量地使用无用数据,则对诊断结果的影响很大。这里的无用是指不包含病变的敏感区。如果大量测试这样的数据,是无法对疾病进行正确诊断的。所以测试集的图片经过了医学的筛选,将海马体、颞叶等敏感区明显的被试图片作为测试样本,这样做也是力求与医生诊断 AD 和 MCI 的过程一致。实验对测试集进行了多次测试,表 11 列出了使用原始的 AlexNet 网络模型和增强的 AlexNet 网络模型产生的分类器的精确度,总体结果较好。

表 11 测试集的验证结果/%

方法	AH	AM	MH	3-WAY
AlexNet	95.44	89.66	86.80	84.39
增强的 AlexNet	96.14	90.52	84.80	86.05

增强的 AlexNet 模型在诊断 AD、辨别 AD 和 MCI 以及三分类问题上均好于原始 AlexNet 模型测试出的结果。为了分析两个网络的优缺点,实验基于测试结果绘制受试者工作特征曲线(ROC),计算灵敏度、特异度。ROC 曲线广泛用于医学诊断范围,可以通过统计数据确定疾病的最佳诊断点。因为 ROC 曲线只针对二分类问题,所以图 21 是针对 AH、AM 和 MH 做的 ROC 曲线。

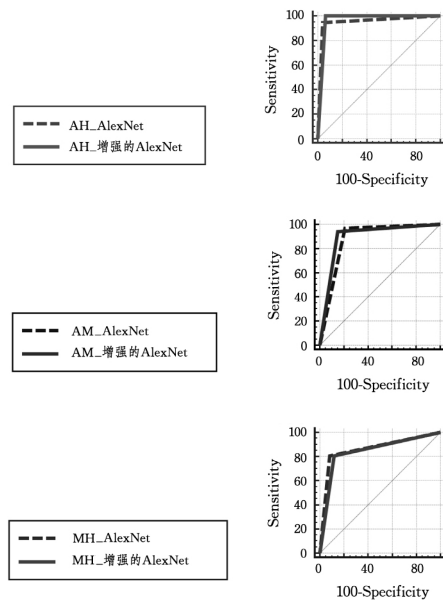


图 21 两种模型针对不同的分类问题绘制的 ROC 曲线

在 ROC 曲线中有两个重要指标:灵敏度(Sensitivity)和特异度(Specificity)。灵敏度又叫做真阳性率(TPR),是正样本预测结果数和正样本实际数的比值。100-特异度又叫做假阳性率(FPR),指被错误地预测为正样本但其实是负样本的结果数和负样本实际数的比值。不难推测,特异度就是负样本预测的结果数与负样本实际数的比值。在医生看来,找出所有患者是当务之急,所以 TPR 越高越好。如果把没病的误诊为有病的也可能也会引起严重的后果,所以 FPR 越低越好。如果医生对病情很敏感,稍微一点症状都当做患病,那么他的统计数据中 TPR 会很高,相应的 FPR 也会很高。所以这两个指标也是相互制约的。理想化的情况是 $TPR=1, FPR=0$,即把所有患者都正确诊断出来,并且没有误诊任何一个没

有患病的检测者。在 ROC 曲线上,如果拐角越接近左上方,说明结果越理想。ROC 曲线下方的面积代表诊断的精确度(AUC)。表 12 列出了原始 AlexNet 和增强的 AlexNet 的统计数据。

表 12 AH,AM 和 MH 组在 ROC 上的统计数据/%

	灵敏度	特异度	精确度
AH(AlexNet)	94.74	95.91	95.44
AH(增强的 AlexNet)	100.00	93.57	96.14
AM(AlexNet)	97.10	78.72	89.66
AM(增强的 AlexNet)	94.20	85.11	90.52
MH(AlexNet)	80.62	91.62	86.80
MH(增强的 AlexNet)	80.62	88.02	84.80

从统计数据上不难看出,增强的 AlexNet 网络在诊断 AD 上更有优势,它的灵敏度达到 100%,说明该模型诊断出了所有的 AD 患者。但它的特异度小于原始模型,说明增强的 AlexNet 的误诊率要大于原有模型。这不难理解,因为这两个指标之间存在制约关系,对病情的诊断越敏感,误诊率就越高。从前两个指标上看,这两个模型都有各自的优势。但在精确度上,增强的 AlexNet 模型优于原始模型。综合来看,改进的模型还是略优于原始模型的。不得不说,原始模型的正确率已经到达了一定的高度,在这个上面做提升确实存在一定困难。增强的 AlexNet 模型能在 TPR 和 AUC 上略胜一筹,还是难能可贵的。

在分辨 AD 和 MCI 上,增强的 AlexNet 模型在正确诊断 AD 上的效果不如原始模型。MCI 是比较复杂的疾病,它表现为两种类型:1)可以转化为 AD;2)终生为稳定性 MCI。所以,这给分辨 AD 和 MCI 提供了一定的阻碍。检测的数据集的追踪时间只有 2 年,类型尚不明确。但增强的 AlexNet 模型的误诊率却远远小于原始 AlexNet 模型,这说明改进的模型识别 MCI 的能力比原始模型强。而且改进模型的精确度高于原始模型。若要达到早期诊断 AD 的目的,则需对 AD 做预测。虽然在实验结果上改进模型在 AD 的诊断率上不如原始模型好,但会不会是因为这其中的一些 MCI 将要转化为 AD,这个推测需要之后继续验证。

在诊断 MCI 上,无论从哪个指标来看,原始的 AlexNet 模型都优于增强的 AlexNet。显而易见,增强的 AlexNet 在诊断 MCI 上并没有发挥功效。在对 AD、MCI、HC 3 种情况做诊断时,从精确度上来看,改进模型的效果优于原始模型。

虽然两种模型各有优势,但通过对 ROC 曲线分析发现了很多问题。这些评测结果将为后续进一步改进网络模型提供指导。另外,测试集较大地依赖医学的指导,这将是今后要解决的关键问题。机器学习算法既然可以通过训练集做诊断,那么它也应该对测试集数据做选择,这将是之后研究的方向。

结束语 本文实验设计了 4 种基于卷积神经网络 AlexNet 的改进算法。改进的 AlexNet 网络模型是根据 AD 特点设计的,对疾病的诊断具有一定的医学意义。改进的 AlexNet 网络模型通过与原始 AlexNet 网络模型做对比和分析,展示出了较好的分类效果。在经过统计学评测后,发现算法仍有不足之处,也提供了改进的思路。在今后的研究中,将尝试使用深度学习对大脑中的患病部位进行定位。这无论对训练模型还是测试来说都实现了全自动化,可避免人工筛选造成的误差。总而言之,实验为 AD 的研究提供了较好的研究

思路和算法,证明了深度学习对 AD 和 MCI 的诊断有很好的效果。

致谢 十分感谢首都医科大学附属北京天坛医院的大力支持,以及 ADNI(<https://ida.loni.usc.edu>)和首都医科大学附属北京天坛医院提供的实验数据。

参考文献

- [1] World Alzheimer Report 2009[OL]. <http://www.alz.co.uk/research/files/WorldAlzheimerReport.pdf>.
- [2] PETERSEN R C, SMITH G E, WARING S C, et al. Mild cognitive impairment: clinical characterization and outcome[J]. *Archives of Neurology*, 1999, 56(3): 303-308.
- [3] LIU S, CAI W, et al. Early diagnosis of Alzheimer's disease with deep learning[C]//Proceedings of the 2014 IEEE 11th International Symposium on Biomedical Imaging (ISBI). Beijing, China: 2014: 1015-1018.
- [4] GUPTA A, AYHAN M, MAIDA A. Natural image bases to represent neuroimaging data[C]//Proceedings of the 30th International Conference on Machine Learning (ICML-13). Atlanta, USA, 2013: 987-994.
- [5] KLOPPPEL S, STONNINGTON C M, BARNES J, et al. Accuracy of dementia diagnosis—a direct comparison between radiologists and a computerized method[J]. *Brain*, 2008, 131(11): 2969-2974.
- [6] VAN LEEMPUT K, MAES F, VANDERMEULEN D, et al. Automated model-based tissue classification of MR images of the brain[J]. *IEEE Transactions on Medical Imaging*, 1999, 18(10): 897-908.
- [7] TRIPOLITI E E. A supervised method to assist the diagnosis and monitor progression of Alzheimer's disease using data from an fMRI experiment[J]. *Artificial Intelligence in Medicine*, 2011, 53(1): 35-45.
- [8] CHRIS H, VIKAS S, LOPAMUDRA M, et al. Spatially augmented LP boosting for AD classification with evaluations on the ADNI dataset[J]. *Neuroimage*, 2009, 48(1): 138-149.
- [9] BADER B W, KOLDA T G. Tensor decompositions and applications. Preprint of article[J]. *Siam Review*, 2009, 51(3): 455-500.
- [10] TUCKER L R. Some mathematical notes on three-mode factor analysis[J]. *Psychometrika*, 1966, 31(3): 279-311.
- [11] TUCKER L R. The extension of factor analysis to three dimensional matrices[C]//Proceedings of the Contributions to Mathematical Psychology. New York, USA, 1964: 110-127.
- [12] PHAN A H, CICHOCKI A S, TICHAUSKY P. On Fast algorithms for orthogonal Tucker decomposition[C]//Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing. Florence, Italy, 2014: 6766-6770.
- [13] MUTI D, BOURENNANE S. Survey on tensor signal algebraic filtering[J]. *Signal Processing*, 2007, 87(2): 237-249.
- [14] LETEXIER D, BOURENNANE S. Adaptive Flattening for Multidimensional Image Restoration[J]. *Signal Processing Letters IEEE*, 2008, 15: 229-232.
- [15] KANDEL E R. An introduction to the work of David Hubel and Torsten Wiesel[J]. *Journal of Physiology*, 2009, 587(12): 2733-2741.
- [16] 王青海. BP 神经网络算法的一种改进[J]. *青海大学学报(自然版)*, 2004, 22(3): 82-84.
- [17] HINTON G E, OSINDERO S, YW T. A Fast Learning Algorithm for Deep Belief Nets[J]. *Neural Computation*, 2006, 18(7): 1527-1554.
- [18] BENGIO Y S. Practical Recommendations for Gradient-Based Training of Deep Architectures[M]//*Neural Networks: Tricks of the Trade*. Springer Berlin Heidelberg, 2012: 437-478.
- [19] 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.
- [20] SHUIWANG J, MING Y, KAI Y. 3D Convolutional Neural Networks for Human Action Recognition[J]. *IEEE Transactions on Pattern Analysis & Machine Intelligence*, 2013, 35(1): 221-231.
- [21] 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): 106-154.
- [22] HARTLINE H K. The receptive fields of optic nerve fibers[J]. *American Journal of Physiology—Legacy Content*, 1940, 130(4): 690-699.
- [23] SIMARD P Y, DAVE, STEINKRAUS, PLATT J C. Best practices for convolutional neural networks applied to visual document analysis[C]//Proceedings of the Seventh International Conference on Document Analysis and Recognition. Edinburgh, Scotland, 2003: 958-963.
- [24] LECUN Y, BENGIO Y. Convolutional networks for images, speech, and time series[M]. *The Handbook of Brain Theory & Neural*, 1995.
- [25] MORRIS J C. The Clinical Dementia Rating (CDR): current version and scoring rules[J]. *Neurology*, 1993, 43(11): 2412-2414.
- [26] TOMBAUGH T N, MCLNTYRE N J. The mini-mental state examination; a comprehensive review[J]. *Journal of the American Geriatrics Society*, 1992, 40(9): 922-935.
- [27] PAYAN A, MONTANA G. Predicting Alzheimer's disease: a neuroimaging study with 3D convolutional neural networks [OL]. <http://arxiv.org/pdf/1502.02506.pdf>.
- [28] KRIZHEVSKY A, SUTSKEVER I, HINTON G E. ImageNet classification with deep convolutional neural networks[C]//Proceedings of the Advances in Neural Information Processing Systems. South Lake Tahoe, US, 2012: 1097-1105.
- [29] HINTON G, DENG L, YU D, et al. Deep Neural Networks for Acoustic Modeling in Speech Recognition[J]. *IEEE Signal Processing Magazine*, 2012, 29(6): 82-97.
- [30] LIU F, SHEN C. Learning Deep Convolutional Features for MRI Based Alzheimer's Disease Classification [OL]. <http://arxiv-web.arxiv.org/pdf/1404.3366v1>.
- [31] JIA Y, SHELHAMER E, DONAHUE J, et al. Caffe: Convolutional Architecture for Fast Feature Embedding[J]. *EprintArxiv*, 2014: 675-678.