**深度学习卷积神经网络原理与应用研究**

龚皓靖 120191140216

**摘要** 深度学习作为机器学习领域中重要的技术手段, 有着广阔的应用前景. 文中简述了深度学习的发展历程, 介绍了卷积神经网络及其衍生的系列方法模型, 以及 Caffe, TensorFlow, Torch 等6种主流深度框架; 论述了深度学习在图像、 语音、 视频、 文本、 数据分析方面的应用情况, 分析了深度学习现阶段存在的问题以及未来的发展趋势

**关键词** 深度学习，框架，神经网络，原理

**Abstract** As an important technology in the field of machine learning, deep learning has broad application prospects. In this paper, the development process of deep learning is briefly introduced, including convolutional neural network and its derived series of method models, as well as Caffe, tensorflow, torch and other six mainstream depth frameworks; This paper discusses the application of deep learning in image, voice, video, text and data analysis, and analyzes the current problems and future development trend of deep learning

**Keywords:** deep learning, framework, neural network, principle

**1 引言**

深度学习(deep learning) 作为机器学习领域的重大分支,不仅深刻影响着机器学习领域的走向, 更是实现人工智能的一条有力的途径. 深度学习本质上是层次特征提取学习的过程, 它通过构建多层隐含神经网络模型, 利用海量数据训练出模型特征来提取最有利的参数, 将简单的特征组合抽象成高层次的特征, 以实现对数据或实际对象的抽象表达.鉴于近几年深度学习的火热程度, 本文将从发展历程、 常用方法、 主流框架、 应用情况这 几个方面对深度学习卷积神经网络展开系统的介绍.

从深度学习诞生以来，就吸引了很多公司和个人加入到这个领域的研究。最近几年关于深度学习方向的著作和应用如雨后春笋般涌现出来。阅读深度学习领域的著作，我们能够快速掌握该领域的热点和趋势。2015年发表在《自然》杂志上一篇名为《Deep Learning》的文章正式将深度学习推向高潮。随后一大批高校和科研单位投入进深度学习的研究中，并且硕果累累，新的观点层出不穷。

国际上有很多大公司在做深度学习方面的研究。谷歌、微软、苹果、亚马逊等公司都积极投入到深度学习的研究中，大多一方面做研究项目，如“谷歌大脑”，另一方面做实际应用，如微软公司推出的聊天机器人“小冰”、苹果的智能语音助手“siri”等。

我国关于深度学习的研究属于后起之秀，并且大有追赶国际的势头。一些大型科技公司都开始涉足这个领域。国内的阿里、腾讯、百度、华为等公司也加紧了这方面的研究。另外还有很多初创公司，将这项技术应用于生物、医疗、广告等各行各业。

过去几年来，计算机视觉研究主要集中在卷积神经网络（常简称为 ConvNet 或 CNN）上。这些工作已经在广泛的分类和回归任务上实现了新的当前最佳表现。相对而言，尽管这些方法的历史可以追溯到多年前，但对这些系统得到出色结果的方式的理论理解还很滞后。事实上，当前计算机视觉领域的很多成果都是将 CNN 当作黑箱使用，这种做法是有效的，但其有效的原因却非常模糊不清，这严重满足不了科学研究的要求。尤其是这两个可以互补的问题：（1）在被学习的方面（比如卷积核），究竟被学习的是什么？（2）在架构设计方面（比如层的数量、核的数量、池化策略、非线性的选择），为什么某些选择优于另一些选择？这些问题的答案不仅有利于提升我们对 CNN 的科学理解，而且还能提升它们的实用性。

此外，目前实现 CNN 的方法需要大量训练数据，而且设计决策对结果表现有很大的影响。更深度的理论理解应该能减轻对数据驱动的设计的依赖。尽管已有实证研究调查了所实现的网络的运行方式，但到目前为止，这些结果很大程度上还局限在内部处理过程的可视化上，目的是为了理解 CNN 中不同层中发生的情况。

**2 深度学习的发展历程**

深度学习是神经网络发展到一定时期的产物. 最早的神经网络模型可以追溯到 1943 年 McCulloch 等提出的 McCulloch-Pitts 计算结构, 它大致模拟了人类神经元的工作原理, 但需要手动设置权重, 十分不便.1958年,Rosenblatt教授提出了感知机模型( perceptron) , 尽管相比前者, 该模型能更自动合理地设置权重, 但同样存在较大的局限, 难以展开更多的研究. 特 别 地, Minsky 教 授 于 和 Paper 教 授 于 1969年证明了感知机模型只能解决线性可分问题, 并且否定了多层神经网络训练的可能性, 甚至提出了“基于感知机的研究终会失败”的观点, 此后十多年的时间内, 神经网络领域的研究基本处于停滞状态

20 世纪80年代, 计算机飞速发展, 计算能力相较以前也有了质的飞跃. 1986 年, Rumelhart教授团队在 Nature 上发表文章, 提出了反向传播算法(Back Propagation, BP) . BP算法的提出不仅有力地回击了 Minsky 教授等人的观点, 更引领了神经网络研究的第二次高潮. 随后, 玻尔兹曼机、 卷积神经网络、 循环神经网络等神经网络结构模型均在这一时期得到了较好的发展.

2006年, 机器学习领域泰斗 Hinton及其团队在Science上发表了关于神经网络理念突破性的文章, 首次提出了深度学习的概念, 并指明可以通过逐层初始化来解决深度神经网络在训练上的难题. 该理论的提出再次激起了神经网络领域研究的浪潮. 此后, 随着大数据时代的到来、 互联网+模式的发展, 以及计算机性能的飞速提高, 深度学习开始渗透到各行各业, 神经网络也迎来了发展的春天. 2011年, 吴恩达领导Google 科学家们用 16000台电脑成功模拟了一个人脑神经网络; 2012年, Hinton 团队将深度学习应用于 ImageNet 上,并取得了惊人的识别效果; 2013年, 欧洲委员会发起模仿人脑的超级计算机项目, 同年1月, 百度宣布成立深度学习研究院. 2014年, 深度学习模型 Top-5 在 ImageNet 2014 计算机识别竞赛上拔得头筹, 同年, 腾讯和京东也分别成立了自己的深度学习研究室. 2016 年, AlphaGo 击败围棋世界冠军李在石, 同年9 月, 中国科学院计算技术研究所发布“寒武纪 1A”深度神经元网络处理器. 这一切都显著地表明了一个事实:深度学习正在有条不紊地发展着, 其影响力不断扩大.

**3 深度学习原理方法**

**3.1卷积神经网络**

## 卷积神经网络(Convolutional Neural Networks,CNN)是在模式识别、图像处理领域的一种高效且稳定的方法,它通过局部感知、共享权值、空间或时间上的池采样来充分利用数据本身包含的局部特性,以优化网络结构,保证一定程度上的位移和变形的不变性. 所谓局部感知,是指在卷积神经网络结构中每个神经元只与部分图像产生特征映射关系,不用感知全局图像.这一特性启发于 １ ９ ６ ２ 年 Hubel 和 Wiesel 通过对猫视觉皮层细胞的研究而提出的感受野理论. 共享权值则是指:当用卷积神经网络提取 １ 种特征时,神经元间共享一套权值,且用这一个相同的卷积核对图像做卷积;当提取 n种特征时,神经元间共享 n  套权值,并用这 n  个相同的卷积对图像做卷积.局部感知和共享权值这两大特性大大地减少了卷积神经网络结构中的参数数目,使网络结构变得更加简单清晰.一个完整的卷积神经网络基本上由输入层、卷积层、池化层、全连接层和 SoftMax 层这5种结构组成.输入层是整个卷积神经网络的输入,一般代表一张图片的像素矩阵;卷积层是整个神经网络的核心,一般将前一层神经网络上的子节点矩阵卷积转化为下一层神经网络上的节点矩阵,并增加节点矩阵的深度,以达到更深层次抽象特征表达的目的;池化层的实质是数理统计矩阵块不重叠区域的聚合特征,一般有平均池采样和最大池采样两种方法,池采样层不会改变特征矩阵的深度,但可以缩小矩阵的大小,简化神经网络的结构;全连接层一般位于多次卷积池化处理后,用以给出最后的分类结果. 卷积神经网络的训练实际上是监督学习的过程,主要包括向前传播和反向传播两个阶段,在训练的过程中,参数能被不断地优化,从而达到最好的模拟效果. 随着卷积神经网络探究的深入,以卷积神经网络为基础,出现了一系列的优化改进结构模型,如全卷积神经网络等。

## **3.2主流深度学习框架**

深度学习框架是专为深度学习领域开发的具有一套独立的体系结构、统一的风格模板、可以复用的解决方案.它一般具有高内聚、严规范、可扩展、可维护、高通用的特点,且拥有统一的代码风格、模板化的结构,能减少大量重复代码的编写.随着深度学习的日益火热,越来越多的深度学习框架被开发出来.目前主流的深度学习框架主要有5 种

**4 深度学习的应用**

随着深度学习在 MNIST 数据集数字识别上的成功应用,越来越多的人开始注意到并尝试运用深度学习这种新方法,同时,深度学习涉及的行业也越来越广.通过对深度学习最新相关文献的阅读,本文将从图像、语音、视频、文本、数据分析这 5个方面对深度学习的应用现状展开论述.

**4.1 图像领域**

卷积神经网络最早应用在图像识别上,它在人脸识别、车牌识别方面都有较好的应用,如 LeNet 模型、GoogLeNet 模型、FCN 模型等都是典型的卷积神经网络模型.魏英姿等通过构建深度卷积神经网络模型,在 MATLAB 平台上成功地对玉米籽粒的完整性进行了识别,并选用了 BP 算法作比较,验证了深度卷积神经网络模型的适用性;李楠则借助深度学习框架 Caffe,采用 LeNet-5模型对破损路面进行了识别,准确率达 92％ ;杜 敬构建了深度卷积神经网络模型

(DCNN)来进行遥感水体的识别,实验采用无人机获取的遥感图像,并将支持向量机法和面向对象法作为对比方法,结果显示 DCNN 模型的识别效果最佳;寇广等结合 Caffe框架对云服务中僵尸云的检测做了详尽的研究,实验中不仅改进了 LeNet-5模型,建立了基于 LeNet-5模型的 6种不同网络结构的卷积神经网络模型,更设置了不同的池采样方案,以对比决策树法和支持向量机法,最终结果显示了深度学习方法具有无可比拟的优越性;付罡等则提出了基于改进全卷积网络模型的高分辨率遥感影像准确分类的方法,并选用高分彩色遥感图像进行实验,结果所提方法获得了较高的精度.除此之外,还有许多成功应用 CNN 模型的研究,在此不一一赘述.

除卷积神经网络外,深度置信网络(DBN)、层叠去噪自编码器(SDAE)这两年也在图像处理方面有较好的应用. 刘大伟等构建了合适的 DBN 模型,完成了对高分辨率遥感影像的分类研究;樊恒等利用 KTH 数据库来训练 DBN 模型,以识别人体行为;李婷构建了 SDAE 模型来学习盲文,解决了盲文识别中特征自动提取和降维的问题;阚希等采用多光谱卫星遥感数据,以无监督训练结合有监督训练来进行全局微调,训练出合适的 SDAE 模型,并将其成功运用到青藏高原地区的积雪识别中,最终绘制出青藏高原的积雪分布图.

**4.2  语音领域**

随着信息科技的发展,海量数据已成为各个领域发展中的必然趋势,如何利用大数据也是各个行业面临的必然挑战.传统的语音识别方法在大语音数据中难以获得较好的效果,而深度学习作为新方法,恰好能提供新的思路,目前在语音识别领域已有较好的应用. Hinton 等改进了传统的混合高斯- 隐马尔科夫模型 GMM- HMM,将模型中的 GMM 替换为深度神经网络 (DNN),从 而生成了新的模型 DNN- HMM.相比于原模型,改进后的模型的词汇错误率大大降低. 王山海等则尝试将深度自动编码器(DAE)应用到孤立词语音识别中,通过在 MATLAB上搭建模型,与 BP 算法的结果进行对比,最终验证了 DAE 模型在语音识别上的适用性.王强等引入 CNN 模型和 DBN 模型,并将其运用到水声识别中,验证了这两种模型在声音识别上的可行性.语音的高质量和可读度是语音数据里最重要的部分,也是语音增强中的核心内容. 近几年,深度学习在语音增强方面同样有亮眼的表现,如深度循环神经网络(DRNN)、堆叠式降噪自动编码(SCAE)和LSMT 神经网络等都能很好地提高语音质量.

**4.3  视频领域**

深度学习在视频领域的应用主要集中在交通监管和目标跟踪上.杨红红等构建了基于稀疏约束的 DAE 模型,以非监督训练监督式微调的方法来训练模型,将其运用到交通目标跟踪上.实验选取了一组视频,将 IVT,MIL,OAB3种算法进行比较,最终发现,基于系数约束的 DAE 模型在不同的场景下都有较高的准确率.司朋举等在 GooLeNet 模型的基础上改进并构建了一个三分类的卷积神经网络模型,并将其成功运用在仿生眼监控系统中,以对人与车进行识别.王慧燕等提出了基于卷积神经网络的辅助跟踪方法——TaNet,并验证了该方法在长距离、多目标、目标遮挡、目标消失重现等多种场景下具有极高的准确性. 文孟飞等建立了一种异构多模态深度学习结构,将视频数据描述为视与听模式,分别利用 CNN 模型和 RBM 模型进行处理,通过共享特征建立关联,最终实现对视频信息目标的识别.

**4.4文本领域**

近年来,对文本领域的研究也都基本集中在深度学习上.张立民等构建了基于浅层结构 RSM 的 DBN 模型,在MATLAB 平台上,利用newgroups 文档集进行训练和测试,证明了该模型在大规模文本 分析上的可行性. 周盈盈等以 topicvec 词向量模型作为模型的输入特征,改进了卷积神经网络,并将改进后的模型运用到文本分类问题的研究上.基于知乎实验数据的测试结果表明,该模型对于短文本的分类准确度能达到 ９ ３ ．２ ７ ％ ,对于长文本的分类准确度高达 ９ ８ ．０ ６ ％.李阳辉等则利用降噪自动编码器模型在深度学习框架 theano 下展开对文本的情感分析研究,证实了降噪自动编码器模型在文本信息分析上的优越性.

**4.5数据分析领域**

在大数据时代,有深度的数据分析不仅是日常的业务要求,更是准确把握行业现状及趋势的必要条件. 深度学习作为信息爆炸时代下的新方法,在大数据的处理上有着不可替

代的优势,目前在数据诊断、参数估计、建模预测等方面均有所应用。

**5  深度学习的现状分析**

目前,深度学习尽管已经取得了很多成果,在图像、语音等领域的表现也十分亮眼,但是该领域依然存在很多需要解决的问题.

1. 训练模式的优化问题. 目前,深度学习模型训练在大多数情况下还是以监督学习为主,而监督学习的开展需要大量的人工标注训练样本,这个过程往往会耗费较长的时间,也需要投入大量的人力,极其繁杂. 其次,训练样本的多少、训练规模的大小会直接影响模型应用的精度,因此如何平衡训练规模、训练时间、训练精度和研究基础条件间的关系也是每项研究开展时所必须面临的问题。
2. 理论创新问题.大部分深度学习模型都是基于卷积神经网络、受限玻尔兹曼机、自动编码器等最简单的神经网络结构叠加或混合组成的,在解决一些数据类型多样、数据结构复杂、数据关系繁复的问题时,往往难以获得较好的效果,因此,在深度学习模型结构的创新上还需要更深层次的研究与探索.
3. 应用规模化问题. 目前,大多数深度学习研究都还只是处于科研阶段,即使在语音和图像等比较成熟的领域,也少有规模化投入生产生活的应用产品.
4. 中文资料较少的问题. 深度学习这项技术起源于国外,能被及时译制的中文基础资料还较少,市面上大部分也都是一些炒概念的书籍,真正实用的工具资料较少.因此,文献资料的普及、核心技术的推广也是国内研究需要努力的方向.

随着 TensorFlow Windows 版本的推出以及 Caffe ２ 在

Facebook 平台的落户,未来深度学习将会有更多的人参与,上述问题与挑战也终将被解决和完成;其次,行业大数据化也会为深度学习提供更宽广的舞台.未来深度学习发展将会呈现出多平台、深层次、全网络的趋势.

**6 总结**

本论文主要研究了深度神经网络，尤其是卷积神经网络在图像物体检测领域的研究与应用。论文首先阐述了深度学习的背景和物体识别领域的国内外研究现状，以及物体识别技术的新旧交替。然后，论文详细介绍了深度网络的相关概念和技术要点。虽然目前基于卷积神经网络图像分类系统虽然很多，而且在识别效果上非常不错。但其中一些基本问题仍然没有得到很好的解决，主要表现在两个方面：第一、尚未形成一套完整的通用理论。现在许多识别系统都是根据特定的数据库进行特别的设计网络的深度和层次，通过不断的摸索发现最佳的参数和优化算法，人为因素比较大，也没有较系统的理论阐述影响卷积神经网络识别效果的因素。第二、现有的方法尚存在一些缺陷。特别是对自然图像进行分类识别时，对卷积神经网络的初始状态参数以及寻优算法的选取，会对网络训练造成很大影响，选择不好会造成网络的不工作，或者有可能陷入局部极小、欠拟合、过拟合等诸多问题。

**参考文献**

[1] 傅文博，孙涛 深度学习原理与应用综述[J].计算机科学,2018,45(z1)

[2] 杜敬.基于深度学习的无人机遥感影像水体识别[J].江西科学,2017,(1).158-161,170

[3] 刘毅铭.深度学习研究与应用综述[J].绿色科技,2019,(11):281-283.

[4] 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.

[5] 张荣,李伟平,莫同.深度学习研究综述[J].信息与控制,2018,47(4):385-397,410.

[6] 尹宝才,王文通,王立春.深度学习研究综述[J].北京工业大学学报,2015,(1):48-59.

[7] 王瀚文.深度学习在嵌入式设备上的应用综述[J].应用能源技术,2018,(7):54-56.