

# 卷积神经网络的发展综述

包 俊 董亚超 刘宏哲

北京联合大学北京市信息服务工程重点实验室 北京 100101  
(15617976473@163.com)

**摘 要** 深度学习(Deep Learning,DL)是机器学习的技术和研究领域之一。通过建立具有阶层结构的人工神经网络(Artificial Neural Networks,ANNs),在计算系统中实现人工智能,是目前很流行的机器学习方法。与此同时,卷积神经网络可谓是现在深度学习领域中大红大紫的网络框架,在计算机视觉领域更是一枝独秀。对卷积神经网络的发展进行了概述,列举了几种比较经典的卷积神经网络,并提出了卷积神经网络的未来发展趋势。

**关键词:**神经网络;卷积神经网络;结构演化;经典网络;发展综述

**中图法分类号** TP311

## Review of Development of Convolutional Neural Networks

BAO Jun,DONG Ya-chao and LIU Hong-zhe

Beijing Key Laboratory of Information Service Engineering,Beijing Union University,Beijing 100101,China

**Abstract** Deep Learning (DL) is one of the technology and research fields of machine learning. Artificial Neural Networks (ANNs) with hierarchical structure is a popular method of machine learning to realize artificial intelligence in computer systems. At the same time,convolutional neural network is a popular network framework in the field of deep learning,especially in the field of computer vision. This paper summarizes the development of convolutional neural networks,lists several classical convolutional neural networks and puts forward the future development trend of convolutional neural networks.

**Keywords** Neural network, Convolutional neural network, Structural evolution, Classical network, Development review

## 1 引言

近年来,深度学习越来越多地得到大家的研究,并被用于对复杂结构和大样本的高维数据进行学习,在计算机视觉、自然语言处理、生物信息学、自动控制等领域得到了广泛应用,且在人像识别、机器翻

基金项目:国家自然科学基金(61871039,61802019,61906017);北京市属高校高水平教师队伍建设支持计划项目(IDHT20170511);北京市自然科学基金(4184088);北京联合大学领军人才项目(BPHR2019AZ01);北京市教委项目(KM201911417001,KM201711417005);国家科技支撑计划项目(2015BAH55F03);智能驾驶大数据协同创新中心项目(CYXC1902)

This work was supported by the National Natural Science Foundation of China(61871039,61802019,61906017),Supporting Plan for Cultivating High Level Teachers in Colleges and Universities in Beijing(IDHT20170511),Natural Science Foundation of Beijing,China(4184088),Premium Funding Project for Academic Human Resources Development in Beijing Union University(BPHR2019AZ01),Beijing Municipal Commission of Education Project(KM201911417001,KM201711417005),National Key Technology R&D Program of the Ministry of Science and Technology of China(2015BAH55F03) and Big Data Collaborative Innovation Center for Intelligent Driving(CYXC1902).

通信作者:刘宏哲(liuhongzhe@buu.edu.cn)

译、自动驾驶等现实问题中取得了成功。随着研究的深入,目前已出现了许多深度学习技术,比如:监督学习和无监督学习、深度前馈网络、正则化、稀疏编码和 dropout、自适应学习算法、卷积神经网络、循环神经网络、递归神经网络、深度神经网络和深度堆叠网络、LSTM 长短时记忆<sup>[1]</sup>、蒙特卡洛、受限波尔兹曼机、softmax 回归、决策树和聚类算法、KNN 和 SVM、生成对抗网络和有向生成网络等。

卷积神经网络作为深度学习的技术之一,是一类包含卷积计算且具有深度结构的前馈神经网络,是深度学习的代表性算法。对卷积神经网络的研究始于 20 世纪 80 至 90 年代,时间延迟网络和 LeNet-5 是最早出现的卷积神经网络<sup>[2]</sup>;21 世纪以来,随着深度学习理论的提出和数值计算设备的改进,卷积神经网络得到了快速发展,并被应用于计算机视觉、自然语言处理等领域。

了解卷积神经网络的发展过程,知道每种经典卷积神经网络的优化过程,了解它们的优势与劣势,对我们更好地学习与发展卷积神经网络是十分必要的,也是我们快速了解卷积神经网络的有效途径。本文分别从神经网络的发展、卷积神经网络的发展和卷积神经网络的未来发展趋势方面,对这些内容进行详细的说明。

2 神经网络

神经网络是一组大致模仿人类大脑构造设计的算法<sup>[3]</sup>,用于识别模式。神经网络通过机器感知系统解释传感器数据,能够对原始输入进行标记或聚类等操作。神经网络所能识别的模式是数值形式,因此必须将图像、声音、文本、时间序列等一切现实世界的的数据转换为数值形式。

2.1 神经网络的兴起

神经网络首次出现于 1958 年感知机的提出。但感知机模型存在限制,比如不能求解异或问题(原因在于没有引入非线性)。直到 1986 年 Hinton 等将 BP 算法用于 ANN 的训练过程,ANN 才迎来了复苏<sup>[4]</sup>。进入 90 年代,随着 ANN 本身对硬件要求的提高,以及 SVM 等判别式分类模型的有效应用,神经网络的发展再次进入低谷。随着 2006 年 Hinton 等人提出了深度受限玻尔兹曼机(Restricted Boltzmann Machine,RBM)模型,以及基于预训练和反向传播的训练方法,人们的目光才被再次拉回神经网络领域。2012 年,Alex 等人第一次使用 DCNN 模型参加图像分类竞赛,并取得了大幅度的效果提升,至此神经网络正式进入飞速发展阶段<sup>[5]</sup>。图 1 以图示化的形式,对神经网络的兴衰进行了形象化说明。

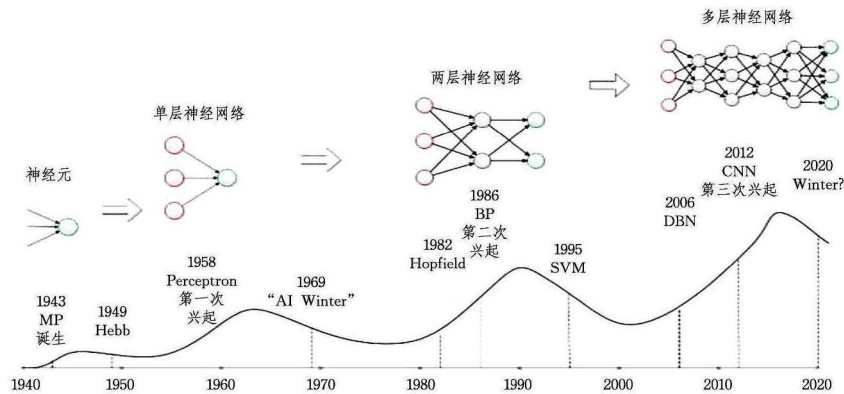


图 1 神经网络的兴衰史

Fig. 1 Development of neural network

2.2 人工神经网络

人工神经网络(Artificial Neural Network, ANN)<sup>[6]</sup>,是 20 世纪 80 年代人工智能领域兴起的研究热点。它从信息处理角度对人脑神经网络进行抽象,建立某种简单模型,并按不同的连接方式组

成不同的网络。在工程与学术界常直接将其简称为神经网络或类神经网络。神经网络是一种运算模型,由大量的节点(或称神经元)相互联接构成。每个节点代表一种特定的输出函数,称为激励函数(activation function-n)。每两个节点间的连接都代

表一个通过该连接信号的加权重,称之为权重,这相当于人工神经网络的记忆。网络的输出则依网络的连接方式、权重值和激励函数的不同而不同。而网络自身通常都是对自然界某种算法或者函数的逼近,也可能是对一种逻辑策略的表达。

最近十多年来,人工神经网络的研究工作不断深入,已经取得了很大的进展,其在模式识别、智能机器人、自动控制、预测估计、生物、医学、经济等领域已成功地解决了许多现代计算机难以解决的实际问题,表现出了良好的智能特性<sup>[7]</sup>。

### 3 卷积神经网络的发展

随着神经网络的飞速发展,其定义也层出不穷。在了解卷积神经网络之前,我们要先知道机器学习、神经网络、深度学习和卷积神经网络的关系<sup>[8]</sup>,如图2所示。

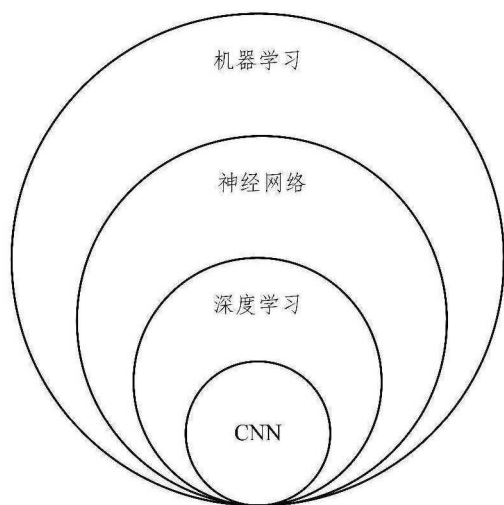


图2 机器学习、神经网络和深度学习之间的关系图

Fig.2 Diagram of relationship between machine learning, neural network and deep learning

#### 3.1 卷积神经网络的提出背景

早在1989年,Yann LeCun和他的同事们就发表了卷积神经网络(Convolution Neural Networks, CNN)的成果<sup>[9]</sup>。

CNN是一种带有卷积结构的深度神经网络,通常至少有两个非线性可训练的卷积层、两个非线性的固定卷积层(又称为 Pooling Layer)和一个全连接层,一共至少5个隐含层。

CNN的结构受到著名的 Hubel-Wiesel 生物视觉模型的启发,尤其是模拟视觉皮层 V1 和 V2 层中

Simple Cell 和 Complex Cell 的行为。

#### 3.2 卷积神经网络的结构演化

CNN 结构演化的历史如图3所示。第一个 CNN 模型诞生于1989年,1998年诞生了 LeNet<sup>[10]</sup>。随着 ReLU 和 dropout 的提出,以及 GPU 和大数据带来的历史机遇,CNN 在2012年迎来了历史突破。2012年之后,CNN 的演化路径可以总结为4条:1)更深的网络;2)增强卷积模的功能,以及上述两种思路的融合;3)从分类到检测;4)增加新的功能模块。

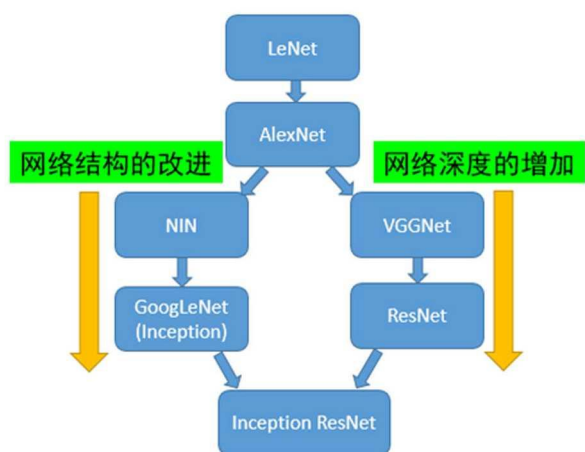


图3 卷积神经网络的演化脉络

Fig.3 Evolutionary of convolutional neural network

#### 3.3 经典的卷积神经网络

经过20多年的发展,几种经典神经网络的网络结构越来越庞大,网络层数越来越深,处理能力也越来越强。在这些基本网络的基础上,也衍生了一系列的深度学习方法。目前,图像分类中的 ResNet,目标检测领域占统治地位的 Faster R-CNN,分割中最牛的 Mask-RCNN,UNet 和经典的 FCN,都是以这几种常见网络为基础<sup>[11]</sup>。

##### 3.3.1 LeNet

最早的深度卷积神经网络模型,被用于字符识别,其具有如下特点。

- 1)卷积神经网络使用3个层作为一个系列:卷积,池化,非线性;
- 2)使用卷积提取空间特征;
- 3)使用映射到空间均值的下采样(subsample);
- 4)双曲线(tanh)或S型(sigmoid)形式的非线性;
- 5)多层神经网络(Multi-Layer Perceptron, MLP)作为最后的分类器。

LeNet 提供了利用卷积层堆叠进行特征提取的框架,开启了深度卷积神经网络的发展。

### 3.3.2 AlexNet

AlexNet 在 LeNet 基础上进行了更宽、更深的网络设计<sup>[12]</sup>,首次在 CNN 中引入了 ReLU,Dropout 和 Local Response Norm (LRN)等技巧。网络的技术特点如下:

1) 使用 ReLU (Rectified Linear Units) 作为 CNN 的激活函数,并验证其在较深网络中的效果超过了 Sigmoid<sup>[13]</sup>,成功解决了 Sigmoid 在网络较深时的梯度弥散问题,提高了网络的训练速率;

2) 为避免过拟合,训练时使用 Dropout 随机忽略一部分神经元;

3) 使用重叠的最大池化(max pooling),最大池化可以避免平均池化的模糊化效果,而采用重叠技巧可以提升特征的丰富性;

4) 提出了 LRN 层(ReLU 后进行归一化处理),对局部神经元的活动创建竞争机制,使得其中响应比较大的值变得相对更大,并抑制其他反馈较小的神经元,增强了模型的泛化能力;

5) 利用 GPU 强大的并行计算能力加速网络训练过程,并采用 GPU 分块训练的方式解决显存对网络规模的限制<sup>[14]</sup>;

6) 数据增强,利用随机裁剪和翻转镜像操作增加训练数据量,降低过拟合。

### 3.3.3 VGGNet

VGGNet 的结构特点如下:

1) 对卷积核和池化大小进行了统一,网络中进行  $3 \times 3$  的卷积操作和  $2 \times 2$  的最大池化操作<sup>[15]</sup>;

2) 采用卷积层堆叠的策略,将多个连续的卷积层构成卷积层组。

其优点如下:

1) 与单个卷积层相比,卷积组可以扩大感受野范围,增强网络的学习能力和特征表达能力;

2) 与具有较大核的卷积层相比,采用多个具有小卷积核的卷积层串联的方式能够减少网络参数;

3) 在每层卷积之后进行 ReLU 非线性操作,可以进一步提升网络的特征学习能力。

### 3.3.4 Network in Network(NIN)

研究目的:

1) 提高网络对非线性特征的描述能力<sup>[16]</sup>;

2) 减少网络参数。

研究策略:

1) 采用多层感知器代替传统的线性卷积层;

2) 用全局均值池化(global average-pooling)代替全连接层,大大减小了网络复杂度,降低了过拟合。

### 3.3.5 GoogLeNet

与 Network in Network 类似,GoogLeNet 采用子网络堆叠的方式搭建,每个子网络为一个 Inception 模块。Inception module 包含如下 4 个分支。

1) Shortcut 连接:将前一层输入通过  $1 \times 1$  卷积;

2) 多尺度滤波:输入通过  $1 \times 1$  卷积之后分别连接卷积核大小为 3 和 5 的卷积;

3) 池化分支:相继连接  $3 \times 3$  pooling 和  $1 \times 1$  卷积。

Inception module 的优点:

1) 减少网络参数,降低运算量;

2) 多尺度、多层次滤波。

### 3.3.6 残差网络 ResNet

简单的通过叠加卷积层的方式来增加网络深度,并不能改善模型效果,甚至还会使模型变得更差,更难训练。因为此时梯度减缓和梯度消失的现象会变得十分严重,梯度反向传播到较早的层,重复乘法可能使梯度无穷大,从而导致随着网络的加深,其性能饱和甚至开始迅速恶化。

为了解决这个问题,MS 尝试搭建一个快捷通道(shortcut connections)来传递梯度。ResNet 并不是第一个使用快捷通道,Highway Network 也有相似的想法,不过 ResNet 的最终效果比较好。

残差网络的提出,本质上还是要解决层次比较深时无法训练的问题。这种借鉴了 Highway Network 思想的网络,相当于旁边专门开个通道使得输入可以直达输出,而优化的目标则由原来的拟合输出  $H(x)$  变成输出和输入的差  $H(x) - x$ ,其中  $H(X)$  是某一层原始的期望映射输出, $x$  是输入。

### 3.3.7 DenseNet

在 DenseNet 之前,就有其他网络进行更多的跨层连接的尝试,但是 DenseNet 更加粗暴,直接将所有的模块连接起来。

DenseNet 作为另一种拥有较深层数的卷积神经网络,具有如下优点:1)拥有比 ResNet 更少的参数数量;2)旁路加强了特征的重用;3)网络更易于训练,并具有一定的正则效果;4)缓解了 gradient vanishing 和 model degradation 的问题。

#### 4 卷积神经网络的未来发展趋势

现在越来越多的 CNN 模型从巨型网络到轻量化网络一步步演变,模型准确率也越来越高<sup>[17]</sup>。现在在工业界追求的重点已从准确率的提升(因为都已经很高了)转变为了速度与准确率的 trade off,都希望模型又快又准。因此,从原来的 AlexNet、VGG-Net,到体积小一点的 Inception 和 ResNet 系列,再到目前能移植到移动端的 MobileNet 和 ShuffleNet (体积能降低到 0.5mb),我们可以看到如下一些趋势。

卷积核方面:

- 1)大卷积核用多个小卷积核代替;
- 2)单一尺寸卷积核用多尺寸卷积核代替;
- 3)固定形状卷积核趋于使用可变形卷积核;
- 4)使用  $1 \times 1$  卷积核。

卷积层通道方面:

- 1)标准卷积用 depthwise 卷积代替;
- 2)使用分组卷积;
- 3)分组卷积前使用 channel shuffle;
- 4)通道加权计算。

卷积层连接方面:

- 1)使用 skip connection,让模型更深;
- 2)Densely connection,使每一层都融合上其他

层的特征输出(DenseNet)。

**结束语** 卷积神经网络依然还在不断地发展创新,但每一次的进步,都是在以前神经网络结构的基础上进行的改进与创新,都是对已有网络不足之处弥补。AlexNet 之后,出现了两个大方向的创新,一个是网络结构的改进,另一个是网络深度的改进。但目前神经网络的参数量庞大,对计算能力的要求还是很高,如何选择合适的超参数,比如学习率、卷积过滤的核大小、层数等,需要大量的技术和经验,这也是卷积神经网络未来的发展方向,需要更深入的研究。

#### 参考文献

- [1] ZHANG J. Gradient Descent based Optimization Algorithms for Deep Learning Models Training[J]. arXiv: 1903.03614, 2019.
- [2] 周志华. 机器学习[M]. 北京:清华大学出版社, 2015: 114-115.
- [3] FARABET C, COUPRIE C, NAJMAN L, et al. Learning Hierarchical Features for Scene Labeling[J]. IEEE Transactions on Pattern Analysis & Machine Intelligence, 2013, 35(8): 1915-1929.
- [4] SUN Y, WANG X, TANG X. Deep Learning Face Representation from Predicting 10,000 Classes[C]// Computer Vision and Pattern Recognition. IEEE, 2014: 1891-1898.
- [5] ABDEL-HAMID O, MOHAMED A R, JIANGH, et al. Convolutional Neural Networks for Speech Recognition [J]. IEEE/ACM Transactions on Audio Speech & Language Processing, 2015, 22(10): 1533-1545.
- [6] ScienceOpen. Visualizing and Comparing Convolutional Neural Networks [J]. Computer Science, 2014.
- [7] GU J, WANG Z, KUENJ, et al. Recent Advances in Convolutional Neural Networks [J]. Computer Science, 2015.
- [8] OORD A V D, DIELEMAN S, SCHRAUWEN B. Deep content-based music recommendation [J]. Advances in Neural Information Processing Systems, 2013, 26: 2643-2651.
- [9] KRIZHEVSKY A, SUTSKEVER I, HINTON G E. ImageNet Classification with Deep Convolutional Neural Networks [J]. Advances in Neural Information Processing Systems, 2012, 25(2): 2012.
- [10] FAN J, XU W, WU Y, et al. Human Tracking Using Convolutional Neural Networks [J]. IEEE Transactions on Neural Networks, 2010, 21(10): 1610-1623.
- [11] TOSHEV A, SZEGEDY C. DeepPose: Human Pose Estimation via Deep Neural Networks [C]// Computer Vision & Pattern Recognition. IEEE, 2014: 1653-1660.
- [12] ZHANG Y, WALLACE B. A Sensitivity Analysis of (and Practitioners' Guide to) Convolutional Neural Networks for Sentence Classification [J]. Computer Science, 2015.
- [13] SIMONYAN K, ZISSERMAN A. Very Deep Convolu-

- tional Networks for Large-Scale Image Recognition[J], Computer Science, 2014.
- [14] SUN Y, WANG X, TANG X. Deep Learning Face Representation by Joint Identification-Verification [J]. Advances in Neural Information Processing Systems, 2014, 27: 1988-1996.
- [15] LIN T Y, DOLLÁR P, GIRSHICK R, et al. Feature pyramid networks for object detection [C] // CVPR. 2017: 4.
- [16] LIN M, CHEN Q, YAN S C. Network in Network [C] // ICLR. 2014.
- [17] 蹇宇澄, 刘昭策. 深度学习的实现与发展——从神经网络到机器学习[J]. 电子技术与软件工程, 2017(11): 30-31.



**BAO Jun**, born in 1995, male, master's degree. His main research interests include deep learning and target detection.



**LIU Hong-zhe**, female, professor. Her main research interests include semantic computing, digital image processing and computer vision.