



AMiner学术头条 原创

2019/05/27 10:54

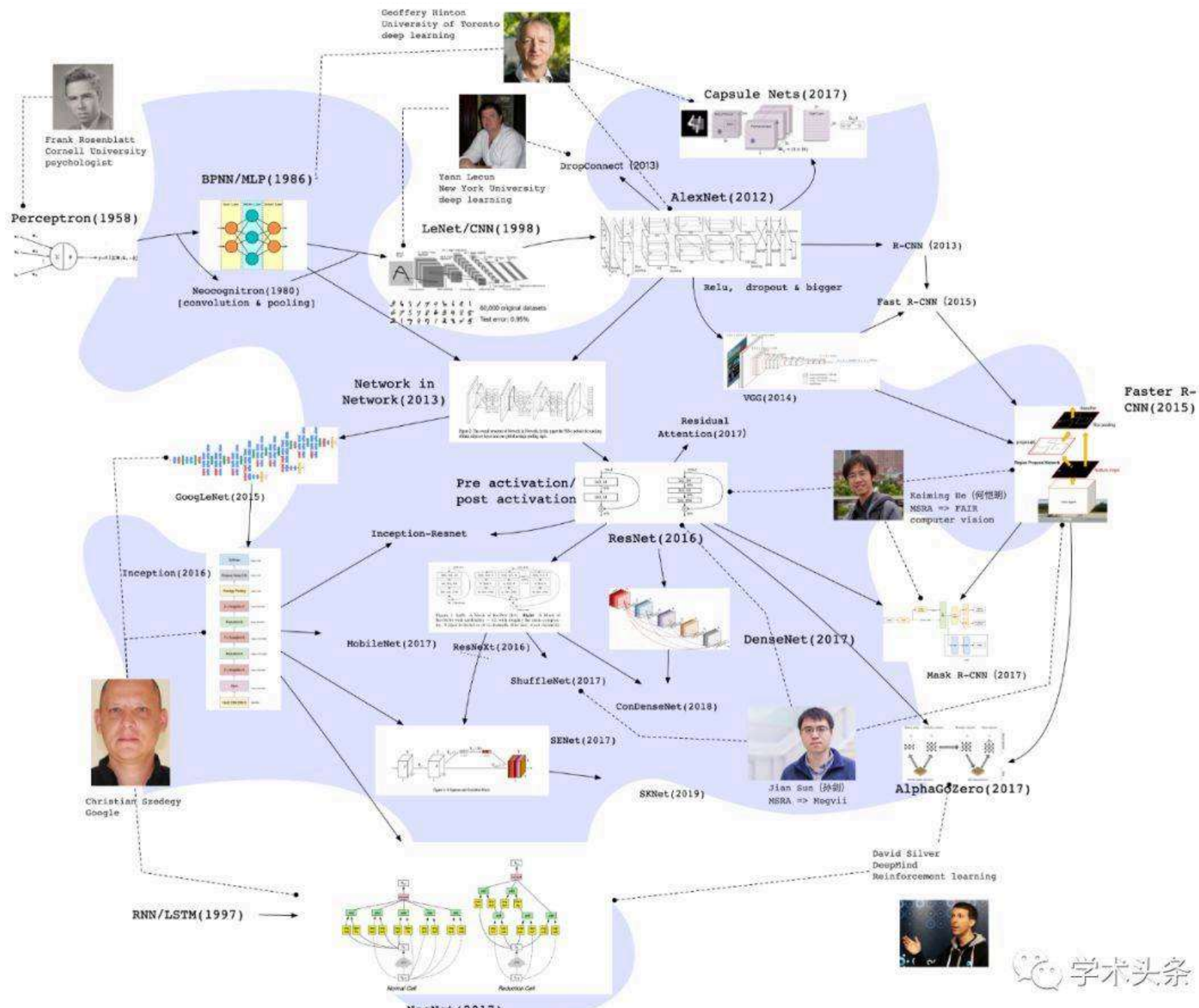
邹旭 唐杰  
作者

# 一文带你了解卷积神经网络CNN的发展史

卷积神经网络 (Convolutional Neural Network, CNN) 是一种前馈神经网络，它的人工神经元可以响应一部分覆盖范围内的周围单元，对于大型图像处理有出色表现。

卷积神经网络由一个或多个卷积层和顶端的全连通层（对应经典的神经网络）组成，同时也包括关联权重和池化层（pooling layer）。这一结构使得卷积神经网络能够利用输入数据的二维结构。与其他深度学习结构相比，卷积神经网络在图像和语音识别方面能够给出更好的结果。这一模型也可以使用反向传播算法进行训练。相比较其他深度、前馈神经网络，卷积神经网络需要考量的参数更少，使之成为一种颇具吸引力的深度学习结构。

本文主要介绍了一些CNN的历史进展。



## 1962年 Hubel和Wiesel

卷积神经网络的发展，最早可以追溯到1962年，Hubel和Wiesel对猫大脑中的视觉系统的研究。



Hubel和Wiesel (图片来源: harvard brain tour)

20世纪60年代初，David Hubel和Torsten Wiesel从约翰霍普金斯大学和Steven Kuffler一起来到哈佛大学，在哈佛医学院建立了神经生物学系。他们在论文《Receptive fields, binocular interaction and functional architecture in the cat's visual cortex》中提出了Receptive fields的概念，因其在视觉系统中信息处理方面的杰出贡献，他们在1981年获得了诺贝尔生理学或医学奖。

Hubel和Wiesel记录了猫脑中各个神经元的电活动。他们使用幻灯机向猫展示特定的模式，并指出特定的模式刺激了大脑特定部位的活动。这种单神经元记录是当时的一项创新，由Hubel早期发明的特殊记录电极实现，他们通过这些实验系统地创建了视觉皮层的地图。

论文地址：<https://www.aminer.cn/archive/receptive-fields-binocular-interaction-and-functional-architecture-in-the-cat-s-visual-cortex/55a5761e612c6b12ab1cc946>

## 1980年 福岛邦彦

1980年，日本科学家福岛邦彦在论文《Neocognitron: A self-organizing neural network model for a mechanism of pattern recognition unaffected by shift in position》提出了一个包含卷积层、池化层的神经网络结构。



老人家现在已经退休了，被誉为“八十多岁仍在奋斗的全球人工智能专家”。除了后来发展出卷积神经网络的Neurocognition（认知控制），现在深度学习中开始热闹起来的Attention（注意力）网络背后也有他的身影，他也在上世纪80年就提出过Attention概念和网络。

## 1998年 Yann Lecun

1998年，在这个基础上，Yann Lecun在论文《Gradient-Based Learning Applied to Document Recognition》中提出了LeNet-5，将BP算法应用到这个神经网络结构的训练上，就形成了当代卷积神经网络的雏形。





(图片来源: Forbes)

原始的CNN效果并不算好, 而且训练也非常困难。虽然也在阅读支票、识别数字之类的任务上很有效果, 但由于在一般的实际任务中表现不如SVM、Boosting等算法好, 一直处于学术界边缘的地位。

论文地址: <https://www.aminer.cn/archive/gradient-based-learning-applied-to-document-recognition/53e9b85bb7602d970441f6c2>

## 2012年 Hinton组

直到2012年，Imagenet图像识别大赛中，Hinton组的论文《ImageNet Classification with Deep Convolutional Neural Networks》中提到的Alexnet引入了全新的深层结构和dropout方法，一下子把error rate从25%以上提升到了15%，颠覆了图像识别领域。



Alexnet有很多创新点，但现在看来是一项非常简陋的工作。他主要是让人们意识到原来那个福岛邦彦提出，Yann Lecun优化的Lenet结构是有很大的改进空间的；只要通过一些方法能够加深这个网络到8层左右，让网络表达能力提升，就能得到出人意料的好结果。

论文地址: <https://www.aminer.cn/archive/imagenet-classification-with-deep-convolutional-neural-networks/53e9a281b7602d9702b88a98>

顺着Alexnet的思想, Lecun组2013年提出一个Dropconnect, 把error rate提升到了11%。而NUS的颜水成组则提出了Network in Network, NIN的思想是CNN原来的结构是完全可变的, 然后加入了一个1\*1conv层, NIN的应用也得到了2014年Imagine另一个挑战——图像检测的冠军。Network in Network的思想是CNN结构可以大胆去变化, 由此, Inception和VGG在2014年把网络加深到了20层左右, 图像识别的error rate也大幅提升到6.7%, 接近人类的5.1%。

2015年, MSRA的任少卿、何凯明、孙剑等人, 尝试把identity加入到神经网络中。最简单的Identity却出人意料的有效, 直接使CNN能够深化到152层、1202层等, error rate也降到了3.6%。后来, ResNeXt, Residual-Attention, DenseNet, SENet等也各有贡献, 各自引入了Group convolution, Attention, Dense connection, channelwise-attention等, 最终Imagenet上error rate降到了2.2%, 完爆人类。现在, 即使手机上的神经网络, 也能达到超过人类的水平。

而另一个挑战——图像检测中, 也是任少卿、何凯明、孙剑等优化了原先的R-CNN, fast R-CNN等通过其他方法提出region proposal, 然后用CNN去判断是否是object的方法, 提出了faster R-CNN。Faster R-CNN的主要贡献是使用和图像识别相同的CNN feature, 发现那个feature不仅可以识别图片是什么东西, 还可以用来识别图片在哪个位置! 也就是说, CNN的feature非常有用, 包含了大量的信息, 可以同时用来做不同的task。这个创新一下子把图像检测的MAP也翻倍了。

在短短的4年中, Imagenet图像检测的MAP从最初的0.22达到了最终的0.73。何凯明后来还提出了Mask R-CNN, 给faster R-CNN又加了一个mask head。即使只在train中使用mask head, 但mask head的信息传递回了原先的CNN feature中, 因此使得原先的feature包含更精细的信息。由此, Mask R-CNN得到了更好的结果。



何凯明在2009年时候就以一个简单有效的去雾算法得到了CVPR best paper，在计算机视觉领域声名鹊起。后来更是提出了Resnet和Faster R-CNN两大创新，直接颠覆了整个计算机视觉/机器学习领域。前些年有很多质疑说高考选拔出的不是人才，几十年几千个状元“没有一个取得成就”。而何凯明正是2003年的广东理科状元，Densenet的共同一作刘壮是2013年安徽省的状元，质疑者对这些却又视而不见了。

CNN结构越来越复杂，于是谷歌提出了Nasnet来自动用Reinforcement Learning 去search一个优化的结构。Nas是目前CV界一个主流的方向，自动寻找出最好的结构，以及给定参数数量/运算量下最好的结构（这样就可以应用于手机），是目前图像识别的发展方向。但何凯明前几天（2019年4月）又发表了一篇论文，表示其实random生成的网络连接结构只要按某些比较好的random方法，都会取得非常好的效果，比标准的好很多。Random和Nas哪个是真的正确的道路，这就有待研究了。

正由于CNN的发展，才引发其他领域很多变革。利用CNN，AlphaGo战胜了李世石，攻破了围棋。但基础版本的AlphaGo其实和人类高手比起来是有胜有负的。但利用了Resnet和Faster-RCNN的思想，一年后的Master则完虐了所有人类围棋高手，达到神一般的境界，人类棋手毫无胜机。后来又有很多复现的开源围棋AI，每一个都能用不大的计算量吊打所有的人类高手。以至于现在人们讲棋的时候，都是按着AI的胜率来讲了。AI的出现也打脸了很多“古今无类之妙手”，人们称颂了几百年的丈和、秀荣妙手，在当下的AI看来，反而是大恶手。而有些默默无闻，人们都认为下的不好的棋，反而在AI分析后大放异彩了。