

在前面的文章中，我们介绍了反向传播(Back Propagation, 简称BP) 算法，在本质上，BP算法是一种全连接神经网络。BP算法也有很多成功的应用，但只能适用于“浅层”网络，因为“肤浅”，所以也就限制了它的特征表征能力，进而也就局限了它的应用范围。

为什么它难以“深刻”呢？在很大程度上问题就出在它的“全连接”上。难道“全连接”不好吗？它更全面啊，难道全面反而是缺陷？

我们暂时不讨论这个问题，等读者朋友阅读完本专题的系列文章之后，答案自然就会了然于胸。在本章，我们讨论一种应用范围更为广泛的网络——卷积神经网络(Convolutional Neural Network, 简称CNN)，它在图像、语音识别等众多任务（比如GoogleNet、微软的ResNet等）上发挥神勇，近几年深度学习大放异彩，CNN可谓是功不可没，拔得头筹。

可为什么CNN能这么生猛呢？答案还得从历史中追寻。著名人类学家费孝通先生曾指出[1]，我们所谓的“当前”，其实包含着从“过去”历史中拔萃出来的投影和时间选择的积累。历史对于我们来说，并不是什么可有可无的点缀之饰物，而是实用的、不可或缺的前行之基础。

下面我们就先聊聊卷积神经网络的历史脉络，希望能从中找到一点启发。在回顾历史之前，我们先尝试思考这样一个“看似题外话而实则不然”的问题：为什么几乎所有低级动物的双眼都是长在头部两侧？

9.1眼在何方？路在何方？

的确，如果你仔细观察，低级动物的双眼大多都长在两侧。从进化论的角度来看，“物竞天择，适者生存”。大自然既然这么安排，自然有它的道理。其中一个解释就是，那些低级动物正是因为这样的“造物安排”，所以能够同时看到上下左右前后等各个方向，从而就不存在视觉盲区。这确实是一种极为安全的配置，有了安全性，它们才能更好地在地球上生存。

可这样的配置又有什么局限呢？相比于低级动物，人的双眼可都是长在面部前方的。这样的配置，肯定不能全方位地观察周遭的一切，这岂不是很糟糕？但事实上是，只有人类进化成为这个地球上最为“高级”的动物。

有人是这样解释的（这个解释的意义可能更多的是来自于意象上的，而非生物学的，所以读者也无需死磕）：低级动物无死角的眼睛配置，虽然能够更全面的关注周围，但副作用却在于，它们没办法把自己的目光集中在某一处，自然也没有办法仔细、长期地观察某个点，于是它们也就不可能进化出深入思考的能力。而人类却因为眼睛的缺陷（接受了视野中的盲区）而能注视前方，从而能给出观察事物的深刻洞察（Insight）。“高级”动物就是这么“练就”出来了。

换句话说，肤浅的全面观察，有时候还不如局部的深入洞察。想一想，那个著名的蝴蝶效应（The Butterfly Effect）是怎么回事，它以“统筹全局”的视角来解释天气的变化：亚马逊雨林一只蝴蝶翅膀偶尔振动，也许两周后就会引起美国得克萨斯州的一场龙卷风。可现实生活中，又有谁真的用“蝴蝶效应”来解决过天气变化问题了呢？我们还不是挥起一把芭蕉扇，利用局部的空气流动，来让自己在酷暑之下获得一丝丝凉意？

可这，和我们今天的主题——卷积神经网络又有什么联系呢？当然有联系，这个联系体现在方法论层面，且听我慢慢道来。

9.2卷积神经网络的历史脉络

我们知道，所谓动物的“高级”特性，体现出来的是行为表象。更深层地，它们也会体现在大脑皮层的进化上。1968年，神经生物学家大卫·休伯尔（David Hunter Hubel）与托斯坦·威泽尔（Torsten N. Wiesel）在研究动物(先后以猫和人类的近亲——猴子为实验对象)视觉信息处理时，有了两个重要而有趣的发现[1]：（1）对于视觉的编码，动物大脑皮层的神经元实际上是存在局部的感受域(receptive field)的，具体说来，它们是局部敏感且具有方向选择性（论文如图9-1所示）。（2）动物大脑皮层是分级、分层处理的。在大脑的初级视觉皮层中存在两种细胞：简单细胞（simple cell）、复杂细胞（complex cell）和超复杂细胞（hyper-complex cell），这些不同类型细胞承担不同抽象层次的视觉感知功能。

RECEPTIVE FIELDS AND FUNCTIONAL ARCHITECTURE
OF MONKEY STRIATE CORTEX

By D. H. HUBEL AND T. N. WIESEL

From the Department of Physiology, Harvard Medical School,
Boston, Mass., U.S.A.

(Received 6 October 1967)

SUMMARY

1. The striate cortex was studied in lightly anaesthetized macaque and spider monkeys by recording extracellularly from single units and stimulating the retinas with spots or patterns of light. Most cells can be categorized as simple, complex, or hypercomplex, with response properties very similar to those previously described in the cat. On the average, however, receptive fields are smaller, and there is a greater sensitivity to changes in stimulus orientation. A small proportion of the cells are colour coded.
2. Evidence is presented for at least two independent systems of columns

图9-1休伯尔与威泽尔的经典论文

正是因为这些重要的生理学发现，使得休伯尔与威泽尔二人获得了1981年的诺贝尔医学奖。而这个科学发现的意义，并不仅仅局限于生理学，它也间接促成了人工智能在五十年后的突破性发展。

休伯尔等人的研究成果对人工智能启发意义在于，人工神经网络的设计可以不必考虑使用神经元的“全连接”模式。如此一来，就可以大大降低神经网络的复杂性。受此理念的启发，1980年日本学者福岛邦彦（Fukushima）提出了神经认知机(neocognitron，亦译为“新识别机”)模型[3]，这是一个使用无监督学习训练的神经网络模型，其实也就是卷积神经网络的雏形（论文如图9-2所示）。从图9-2中可以看到，neocognitron借鉴了休伯尔等人提出的视觉可视区（visual area）分层和高级区关联等理念。

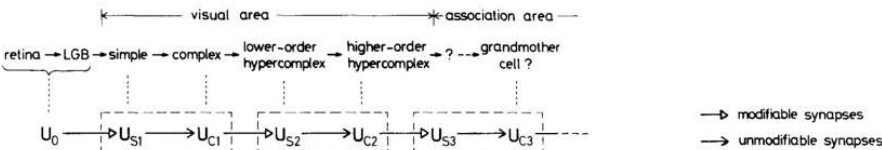


Fig. 1. Correspondence between the hierarchy model by Hubel and Wiesel, and the neural network of the neocognitron

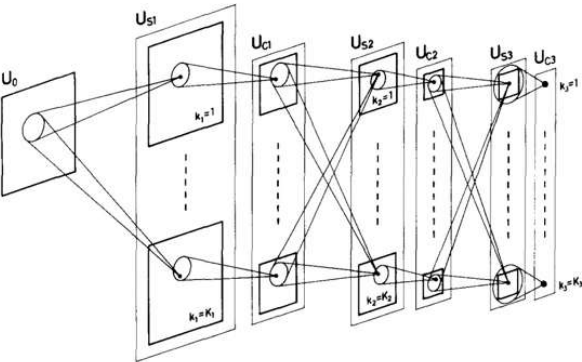


Fig. 2. Schematic diagram illustrating the interconnections between layers in the neocognitron

图9-2 神经认知机的论文

自此之后，很多计算机科学家先后对“神经认知机”做了深入研究和改进，但其效果却不尽人意。直到1990年，在AT&T贝尔实验室工作的Yann LeCun等人，把有监督的反向传播（BP）算法应用于福岛邦彦等人提出的架构，从而奠定了现代CNN的结构[4]。

相比于传统的图像处理算法，LeCun等人提出的CNN，避免了对图像进行复杂的前期处理（即大量的人工图像特征提取工作），也就是说，CNN能够直接从原始图像出发，经过非常少的预处理，就能从图像中找出视觉规律，进而完成识别分类任务，其实这就是端到端（end-end）的含义。在手写邮政编码的识别问题上，LeCun等人把错误率整到5%左右（论文如图9-3所示）。有了严格的理论基础，并有成功的商业案例应用为之站台，一下子就让卷积神经网络学术界和产业界引爆。

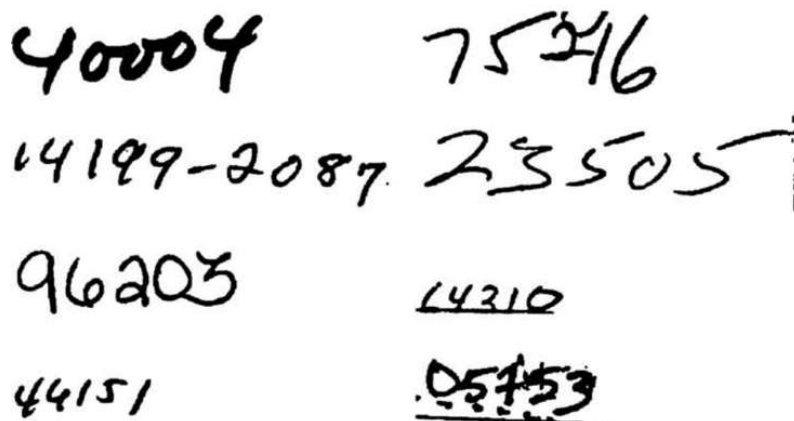


Figure 1: Examples of original zipcodes from the testing set.

图9-3 LeCun等人利用卷积神经网络识别手写邮政编码

LeCun其实也挺“自恋”的，他把自己的研究工作命名为 LeNet，几经版本的更新，最终定格为 LeNet 5。在当时，LeNet 架构可谓是风靡一时，但它的核心业务主要用于字符识别任务，比如前文提到的读取邮政编码、数字等。

可问题来了，从1990年到现在，小30年过去了。为什么30年前的技术，到现在突然又以深度学习的面目重新火爆起来了呢？

对于深度学习，吴恩达先生有个形象的比喻：深度学习的过程就犹如发射火箭。火箭想“发飙”，得依靠两法宝：一是发动机，二是燃料。而对深度学习而言，它的发动机就是“大计算”，它的燃料就是“大数据”。

而在30年前，LeCun等人虽然提出了CNN，但其性能严重受限当时的环境：没有大规模的训练数据，也没有跟得上的计算能力，这导致了当时CNN网络的训练过于耗时且识别效果有上限。

而与LeCun同在一个实验室的“老冤家”Vapnik（万普尼克），提出并发扬光大了支持向量机（Support Vector Machine，简称SVM）。这个SVM很了不起，在1998年，就把LeCun等人同类任务的识别错误率一下子降低至0.8%，远超同期的卷积神经网络。就这样，在SVM的崛起中，神经网络研究沉寂于一个新的低潮！

有人把“Yann LeCun”的中文名称翻译成“严乐春”。的确，“三十年河东，三十年河西”，当“大计算”和“大数据”不再是问题的时候，严乐春用了三十年迎来了自己的又一“春”。当前，严乐春以深度学习大牛的身份出席各大会，做了一场又一样的主题报告（keynote speech），他用自己的亲身经历，生动地演绎了“形势比人强”。

LeCun 提出的LeNet，在推进深度学习的发展上可谓功不可没。后期的卷积神经网络也在LeNet 5 基础上做了很多改良，比如2012年Hinton教授采用新的激活函数（ReLU）等。当前主流的卷积神经网络结构如图9-4所示，其精华大致体现在3个核心操作和3个概念。3个核心是指：卷积（Convolution）、池化（Pooling）和非线性处理（ReLU）。三概念是指：局部感受域（Local receptive field）、权值共享（Weight sharing）和亚采样（Subsampling）。

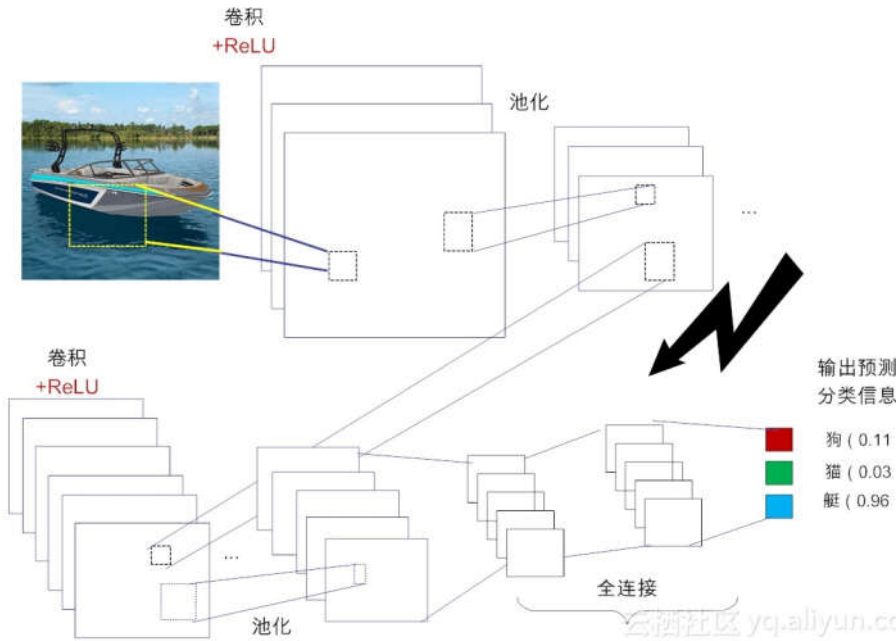


图9-4 CNN的基本结构

在后面的章节中，我们会对这些概念——进行生动形象地介绍，敬请关注。

9.3 小结与思考

在本节中，我们主要回顾了卷积神经网络的发展史，从历史的脉络中，我们可得出一个意象上的结论：深而洞察的专注可能要胜过广而肤浅的观察。这也从方法论上部分解释了“深度学习”的成功所在。

进一步地，请你思考如下问题，人类双眼的布局其实是有缺陷的，因为它带来不安全。李笑来先生认为，人只有放弃了部分的安全感，才能获得进步。为了弥补这方面的缺陷，人们可以通过有效的社交和协作将缺点逆转为优势。

社交进而催生社群，但社群数量超过150人，人们单纯的社交就显得无能为力了。新锐历史学家尤瓦尔·赫拉利在《人类简史》中表明观点，人类需要通过“讲故事”的方式来完成大规模的协作。你认可这个观点吗？为什么？

留言写下你的感悟，祝你每天都有收获。

参考文献

[1] 费孝通.乡土中国.北京大学出版社.2012年10月

[2] Hubel D H, Wiesel T N. Receptive fields and functional architecture of monkey striate cortex[J]. The Journal of physiology, 1968, 195(1): 215-243.

[3] Fukushima K, Miyake S. Neocognitron: A self-organizing neural network model for a mechanism of visual pattern recognition[M]//Competition and cooperation in neural nets. Springer, Berlin, Heidelberg, 1982: 267-285.

[4] LeCun Y, Boser B E, Denker J S, et al. Handwritten digit recognition with a back-propagation network[C]//Advances in neural information processing systems. 1990: 396-404.

文章作者：张玉宏，著有《品味大数据 (<http://product.dangdang.com/24048575.html>)》、本文节选自《深度学习之美 (<https://item.jd.com/12382640.html>)》（最通俗易懂的深度学习入门）2018年7月电子工业出版社出版。

审校：我是主题曲哥哥。

(未完待续)

系列文章：

卷地风来忽吹散，积得飘零美如画（深度学习入门系列之十）(<https://yq.aliyun.com/articles/156269>)