

# 漫威电影中的深度学习

## Deep learning in Marvel Cinematic Universe

王元卓 中国科学院计算技术研究所

江旭晖 中国科学院计算技术研究所

陆源 竞技世界(北京)网络技术有限公司

在《钢铁侠》中,钢铁侠研发出第一套战甲后回到了恐怖分子基地解救难民。当恐怖分子挟持人质时,钢铁侠通过快速扫描人物特征,分辨出恐怖分子和难民的身份,成功击毙了恐怖分子并解救人质。这个识别的过程就是运用了深度学习的相关技术,通过学习恐怖分子和人质的显著特征,并利用人工神经网络来将特征转化为战甲能够理解的“信息”,使战甲能够自动分辨出恐怖分子和人质,如图1所示。



图1 《钢铁侠》电影中的利用深度学习识别恐怖份子片段

深度学习的核心是人工神经网络,人工神经网络的设计思想源自于人脑中的生物神经网络。大脑中的神经网络由许多神经元组成,每个神经元由多个树突和一个轴突组成。一个神经元可以接收多个刺激信号输入,并产生一个输出。神经元通过突触相互连接的复杂神经网络,实现了人脑对外部信息的分析与学习。

那么什么是深度学习呢?

深度学习是机器学习的分支之一,建立在模拟人脑进行分析与学习的人工神经网络(以下简称神经网络)的基础上,神经网络对于文字、图像和声音等样本数据的内在规律和特征进行学习,并转化成机器能够理解的高维向量,最终让机器能够像人一样具有分析和学习能力,能够识别和分析文字、图像和声音等。

神经网络的研究经历了一段漫长的发展史,可大致分为3个阶段。

20世纪40年代到60年代是人工神经网络的起源阶段,这个时期的数学家和心理学家提出通过数学模型模仿人类大脑,提出了基于神经网络模拟人类神经元的数学模型和对应的学习方法,为以后的神经

网络学习奠定了基础。在此基础上，他们提出了由两层神经元组成的神经网络，称之为“感知机”，这对于神经网络的发展具有里程碑式的意义。但随着研究深入，科学家发现单层感知机无法解决非线性分类，难以实际应用，至此神经网络研究进入了第一个寒冬期。

20世纪80年代到21世纪初是神经网络的发展阶段，这个时期科学家提出了适用于多层感知机的反向传播算法，完美解决了非线性分类问题，使神经网络再次受到关注。但当时有限的运算能力限制了神经网络的规模，神经网络的研究进入了第二个寒冬期。

21世纪初深度学习思想被提出，使得神经网络的研究进入爆发阶段，随着运算能力的大幅提高，神经网络的数据处理能力也随之增强，深度学习的概念正式被提出。在著名的图像识别竞赛中，基于神经网络的深度学习算法脱颖而出，这使得深度学习再次受到关注，并且蓬勃发展至今。如AlphaGo、AlphaFold等基于深度学习的算法甚至在特定领域取得了超越人类的能力。

接下来我们看看深度学习是如何工作的，如图2所示。

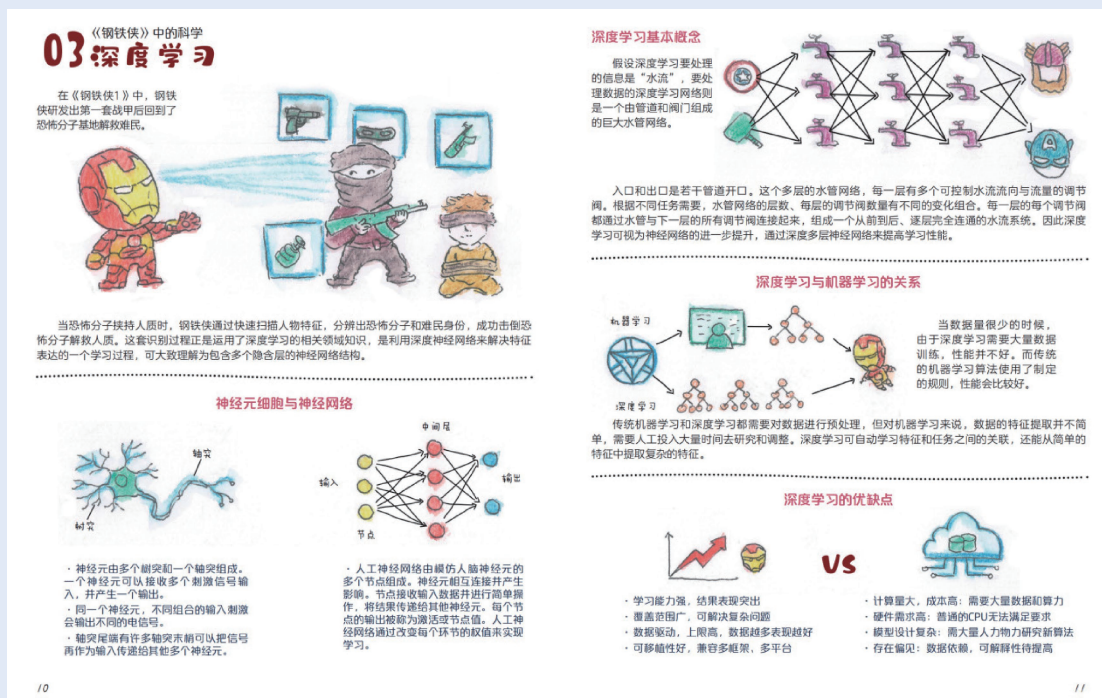


图2 深度学习讲解图

(选自《科幻电影中的科学：科学家奶爸的AI手绘》)

深度学习可以被认为是一种深度神经网络，可大致理解为包含多个隐含层的神经网络结构。该结构由模仿人脑神经元的多个节点组成。神经元相互连接并产生影响。节点接收输入数据并进行简单操作，将结果传递给其他神经元。神经网络通过改变每个环节的权值来实现学习。

打个比方，假设深度学习要处理的信息是“水流”，那么深度学习神经网络则是一个由管道和阀门组成的巨大水管网络。入口和出口是若干管道开口，这个多层的水管网络，每一层有多个可控制水流流向与流量的调节阀。根据不同任务需要，水管网络的层数、每层的调节阀数量有不同的组合变化。每一层的每个调节阀都通过水管与下一层的所有调节阀连接起来，组成一个从前到后、逐层完全连通的水流系统。

那么，计算机该如何使用这个庞大的水管网络来学习呢？

比如我们想让计算机认识汉字。当计算机看到一张写有“国”字的图片时，就将组成这张图片的所有数字信息全都变成信息的“水流”，从入口灌进水管网络。我们预先在水管网络的每个出口都插一块字牌，对应于每一个我们想让计算机认识的汉字。这时，因为输入的是“国”这个汉字，等“水流”流过整个水管网络，计算机就会跑到管道出口位置去看一看，是不是标记有“国”字的管道出口流出来的“水

流”最多。如果是这样，就说明这个管道网络符合要求。如果不是这样，就调节水管网络里的流量调节阀，让“国”字出口流出的水最多。

因此深度学习可被视为神经网络的进一步提升，通过增加神经网络的层数来提高学习的能力。

那么，深度学习与传统机器学习又有怎样的关系呢？

深度学习的过程需要提供大量的数据，因此当数据量很少的时候，深度学习的性能会相对一般。而传统的机器学习算法使用了制定的规则，性能会相对较好。传统的机器学习需要制定规则，而数据的特征提取和规则制定对于机器学习来说并不简单，需要人工投入大量时间去研究和调整。此时，深度学习的优势就体现出来了，深度学习可自动学习特征和任务之间的关联，还能从简单的特征中提取深层次的复杂特征。

整体来讲，深度学习有很强的学习能力，在实际应用中表现亮眼；覆盖范围广，可移植性好，能够兼容多框架、多平台，也适用于多样的数据形式，并解决复杂的问题。在数据规模快速增长的当下，数据越多，基于数据驱动的深度学习的表现也会越好。

但是，深度学习也具有一定的缺陷。深度学习的计算量大、成本高，需要大量的数据和算力的支撑，普通的CPU通常无法满足需求，需要更高性能的GPU甚至是TPU。并且深度学习的模型设计复杂，其黑盒的特性导致模型的结果存在潜在的偏见，可解释性有待提高。

当今，深度学习技术已经应用于各行各业，并潜移默化地影响着人们的生活方式。人脸识别的背后有着深度学习在计算机视觉领域的应用；网购、刷短视频的背后有着深度学习在智能推荐领域的应用；与小冰、小爱和小度等智能助手对话的背后有着深度学习在信息检索、语音识别以及自然语言处理领域的应用。深度学习技术在带给人们生活便利的同时，也在推动着生产力的提升以及社会的进步。

在过去的几十年的发展，深度学习大量借鉴了我们关于人脑、统计学和应用数学的知识。近年来得益于更强大的计算能力、更大的数据集和能够训练更多层神经网络的技术，深度学习的普及性和实用性都有了极大的发展。未来，深度学习在更多领域的普适应用以及可解释性研究将成为发展趋势。

#### 作者简介



**王元卓**（1978- ），男，博士，中国科学院计算技术研究所研究员、博士生导师，中科大数据研究院院长，中国科普作家协会副理事长，中国计算机学会科学普及工作委员会主任，主要研究方向为大数据与人工智能。



**江旭晖**（1994- ），男，中国科学院计算技术研究所博士生，主要研究方向为社交知识图谱。



**陆源**（1990- ），男，现就职于竞技世界（北京）网络技术有限公司，从事数据产品工作，主要研究方向为大数据与社交网络。热心科普创作，科普畅销书《科幻电影中的科学：科学家奶爸的宇宙手绘》《科幻电影中的科学：科学家奶爸的AI手绘》作者之一。