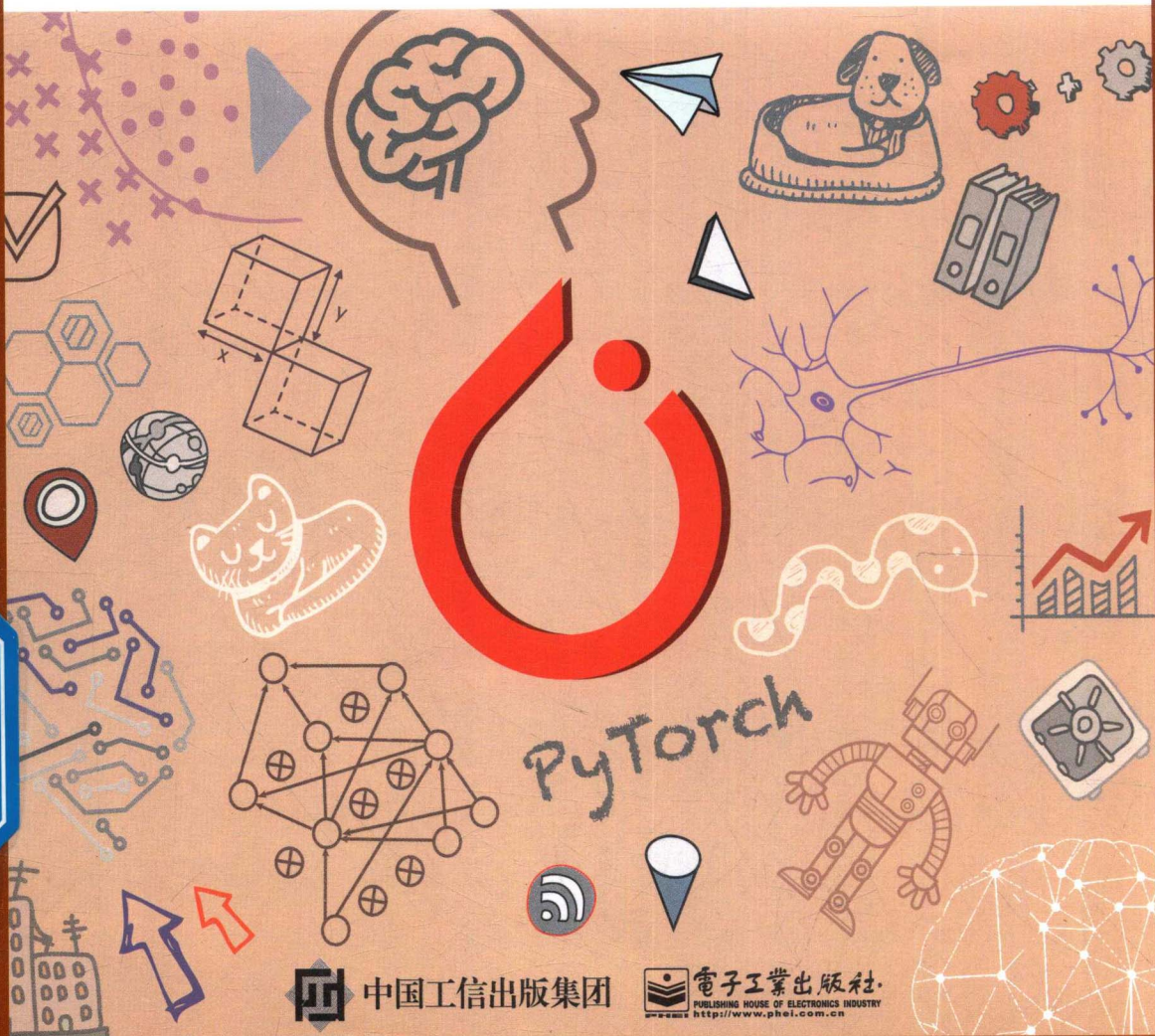




# 深度学习之 PyTorch实战计算机视觉

唐进民 编著



# 深度学习之 PyTorch实战计算机视觉

唐进民 编著



电子工业出版社

Publishing House of Electronics Industry

北京·BEIJING

## 内 容 简 介

计算机视觉、自然语言处理和语音识别是目前深度学习领域很热门的三大应用方向，本书旨在帮助零基础或基础较为薄弱的读者入门深度学习，达到能够独立使用深度学习知识处理计算机视觉问题的水平。通过阅读本书，读者将学到人工智能的基础概念及 Python 编程技能，掌握 PyTorch 的使用方法，学到深度学习相关的理论知识，比如卷积神经网络、循环神经网络、自动编码器，等等。在掌握深度学习理论和编程技能之后，读者还会学到如何基于 PyTorch 深度学习框架实战计算机视觉。本书中的大量实例可让读者在循序渐进地学习的同时，不断地获得成就感。

本书面向对深度学习技术感兴趣、但是相关基础知识较为薄弱或者零基础的读者。

未经许可，不得以任何方式复制或抄袭本书之部分或全部内容。  
版权所有，侵权必究。

### 图书在版编目 (CIP) 数据

深度学习之 PyTorch 实战计算机视觉 / 唐进民编著. —北京: 电子工业出版社, 2018.6  
(博文视点 AI 系列)  
ISBN 978-7-121-34144-1

I. ①深… II. ①唐… III. ①机器学习 IV.①TP181

中国版本图书馆 CIP 数据核字 (2018) 第 088388 号

策划编辑: 张国霞

责任编辑: 徐津平

印 刷: 三河市良远印务有限公司

装 订: 三河市良远印务有限公司

出版发行: 电子工业出版社

北京市海淀区万寿路 173 信箱 邮编: 100036

开 本: 787×980 1/16 印张: 17.75 字数: 370 千字

版 次: 2018 年 6 月第 1 版

印 次: 2018 年 6 月第 2 次印刷

印 数: 2501~3500 册 定价: 79.00 元

凡所购买电子工业出版社图书有缺损问题, 请向购买书店调换。若书店售缺, 请与本社发行部联系, 联系及邮购电话: (010) 88254888, 88258888。

质量投诉请发邮件至 [zlts@phei.com.cn](mailto:zlts@phei.com.cn), 盗版侵权举报请发邮件至 [dbqq@phei.com.cn](mailto:dbqq@phei.com.cn)。

本书咨询联系方式: (010) 51260888-819, [faq@phei.com.cn](mailto:faq@phei.com.cn)。



# 前言

“人工智能”（Artificial Intelligence，简称 AI）一词在很久以前就有了，被大众津津乐道却是近几年的事情，这和机器学习（Machine Learning）、深度学习（Deep Learning）等技术的崛起有着千丝万缕的联系，而这一切又得益于大数据的发展和计算机处理性能的不不断提升。

本书将带领读者了解人工智能的相关技术和发展近况，通过一些实例来掌握必备的技能，并能够独立使用相关技术完成对计算机视觉问题的分析和处理。本书各个章节的知识要点如下。

第 1 章主要介绍人工智能、神经网络和计算机视觉的发展历史，让读者对这一领域有一个全面的认识。

第 2 章主要介绍在理解和掌握后面章节的内容时需要用到的数学知识，以及在实战操作的过程中进行环境搭建及安装相关软件的方法。本书中数学相关的大部分知识都集中在本章中，其主要目的是让读者先对这个领域的知识产生兴趣，这样才能更好地深入学习和研究。在本章中不会插入大量的数学公式，这样做会让初学者望而却步，在不断消化公式的过程中丧失学习兴趣和动力。通过不断实战来学习，可以累积成就感，这种自顶向下的方式不失为一种更好的学习方法。

第 3 章主要介绍在学习神经网络的过程中会经常遇到的一些概念和定义。比如后向传播（Back Propagation）、激活函数（Activation Function）、监督学习（Supervised Learning）、无监督学习（Unsupervised Learning），等等，这也是为之后学习深度神经网络做准备。在搭建一个完整的深度神经网络模型时，就需要不断地用到本章的内容了。

第 4 章主要介绍深度神经网络中的卷积神经网络（Convolutional Neural Network，简称 CNN）。首先介绍卷积层、全连接层、池化层等相关内容，之后又列举了目前主流的一些卷积神经网络架构，并对比它们之间的相同点和不同点，以便于掌握不同的卷积神经网络



络的结构和技术细节。

第 5 章主要介绍 Python 编程语言的相关知识，目的是让读者掌握 Python 语言的语法定义和使用方式，并使用 Python 语言进行功能代码的编写；还会介绍在处理计算机视觉问题时需要用到的两个重要的 Python 包：NumPy 和 Matplotlib。本章内容丰富，而且 Python 语言自身就很简单且易上手，读者很快就能掌握 Python 这门编程语言。

第 6 章主要介绍如何使用 PyTorch 深度学习框架。PyTorch 非常简单易用，能够根据我们的需求快速搭建出我们想要的深度神经网络模型，这在很大程度上归功于 PyTorch 基于动态图计算的特性，它与基于静态图计算的深度学习框架相比，有更多的优势，比如 PyTorch 不仅速度快，还有许多功能强大的包可供调用。本章先介绍 PyTorch 中常用的包和类的使用方法；然后介绍如何使用 PyTorch 中的一些自动化方法来提升代码的执行效率和简洁度；最后会通过一个综合实例，使用本章的内容解决一个实际的计算机视觉问题。

第 7 章一开始就是一个关于计算机视觉问题的实战，介绍了一种非常实用的深度神经网络复用方法，即迁移学习（Transfer Learning）。在掌握迁移学习的原理之后，会基于 PyTorch 对迁移学习进行实战，并解决比之前更复杂的计算机视觉问题。对实战代码的解析会贯穿本章，让读者更深刻地理解代码。

第 8 章讲解如何基于 PyTorch 实战图像风格迁移（Neural Style）。通过对本章的学习，读者会发现，利用卷积神经网络不仅能处理图片分类问题，只要有想法和创意，还能做更多、更有趣的事情。

第 9 章介绍一种多模型融合方法，在现有的模型遭遇性能提升瓶颈时，可通过搭建一种经过科学融合的新模型达到超过预期的泛化能力。本章依然会基于 PyTorch 对多模型融合方法进行实战。

第 10 章介绍一种区别于卷积神经网络的新神经网络结构，即循环神经网络（Recurrent Neural Network，简称 RNN）。不同于卷积神经网络强大的图像特征提取能力，循环神经网络主要用于处理有序输入的数据。为了方便读者理解模型如何对有序数据进行处理，本章会基于 PyTorch 使用循环神经网络来处理一个计算机视觉问题。

第 11 章讲解自动编码器，它是一种使用非监督学习方法的神经网络。自动编码器能够实现很多功能，本章会选取一个图像去噪问题来进行自动编码器实战。

本书前 6 章的内容可作为后 5 章的铺垫，前 6 章的知识偏向基础和理论，不过，只有掌握了这些内容，才能从容应对后 5 章的实战。这个循序渐进的过程会让读者对知识的理

解更深刻，技能提升更迅速。

人工智能在近几年大热，网络上的相关资料良莠不齐且没有体系，即使有优秀的干货，对于基础薄弱的初学者来说起点也太高。本书也是出于对这一现状的考虑，通过从基础到实战、由浅入深的过程，让读者基于 PyTorch 来使用深度学习方法实际解决一些计算机视觉相关的问题，这样，读者在获取知识的过程中会更有成就感，学起来也会更积极、主动。

感谢家人的鼓励和支持，也感谢张国霞编辑的帮助和付出，笔者才能以更好的方式将这部作品呈现在读者的面前。希望读者能遵从敏捷学习的思路，多实战、多思考并不断进步。在本书中会有很多实例，读者可以举一反三、不断实践，在发现问题时要多思考，毕竟本书涉及的内容有限，若想让自己的能力得到更高层次的提升，则需要获取更多的资料来充实自己。

唐进民

2018年5月

---

轻松注册成为博文视点社区用户 ([www.broadview.com.cn](http://www.broadview.com.cn))，扫码直达本书页面。

- **提交勘误：**您对书中内容的修改意见可在 [提交勘误](#) 处提交，若被采纳，将获赠博文视点社区积分（在您购买电子书时，积分可用来抵扣相应金额）。
- **交流互动：**在页面下方 [读者评论](#) 处留下您的疑问或观点，与我们和其他读者一同学习交流。

页面入口：<http://www.broadview.com.cn/34144>



# 目 录

第 1 章 浅谈人工智能、神经网络和计算机视觉	1
1.1 人工还是智能	1
1.2 人工智能的三起两落	2
1.2.1 两起两落	2
1.2.2 卷土重来	3
1.3 神经网络简史	5
1.3.1 生物神经网络和人工神经网络	5
1.3.2 M-P 模型	6
1.3.3 感知机的诞生	9
1.3.4 你好，深度学习	10
1.4 计算机视觉	11
1.5 深度学习+	12
1.5.1 图片分类	12
1.5.2 图像的目标识别和语义分割	13
1.5.3 自动驾驶	13
1.5.4 图像风格迁移	14
第 2 章 相关的数学知识	15
2.1 矩阵运算入门	15
2.1.1 标量、向量、矩阵和张量	15
2.1.2 矩阵的转置	17
2.1.3 矩阵的基本运算	18
2.2 导数求解	22
2.2.1 一阶导数的几何意义	23

2.2.2	初等函数的求导公式	24
2.2.3	初等函数的和、差、积、商求导	26
2.2.4	复合函数的链式法则	27
<b>第 3 章</b>	<b>神经网络基础</b>	<b>29</b>
3.1	监督学习和无监督学习	29
3.1.1	监督学习	30
3.1.2	无监督学习	32
3.1.3	小结	33
3.2	欠拟合和过拟合	34
3.2.1	欠拟合	34
3.2.2	过拟合	35
3.3	后向传播	36
3.4	损失和优化	38
3.4.1	损失函数	38
3.4.2	优化函数	39
3.5	激活函数	42
3.5.1	Sigmoid	44
3.5.2	tanh	45
3.5.3	ReLU	46
3.6	本地深度学习工作站	47
3.6.1	GPU 和 CPU	47
3.6.2	配置建议	49
<b>第 4 章</b>	<b>卷积神经网络</b>	<b>51</b>
4.1	卷积神经网络基础	51
4.1.1	卷积层	51
4.1.2	池化层	54
4.1.3	全连接层	56
4.2	LeNet 模型	57
4.3	AlexNet 模型	59
4.4	VGGNet 模型	61
4.5	GoogLeNet	65



4.6	ResNet	69
<b>第 5 章 Python 基础</b> ..... 72		
5.1	Python 简介	72
5.2	Jupyter Notebook	73
5.2.1	Anaconda 的安装与使用	73
5.2.2	环境管理	76
5.2.3	环境包管理	77
5.2.4	Jupyter Notebook 的安装	79
5.2.5	Jupyter Notebook 的使用	80
5.2.6	Jupyter Notebook 常用的快捷键	86
5.3	Python 入门	88
5.3.1	Python 的基本语法	88
5.3.2	Python 变量	92
5.3.3	常用的数据类型	94
5.3.4	Python 运算	99
5.3.5	Python 条件判断语句	107
5.3.6	Python 循环语句	109
5.3.7	Python 中的函数	113
5.3.8	Python 中的类	116
5.4	Python 中的 NumPy	119
5.4.1	NumPy 的安装	119
5.4.2	多维数组	119
5.4.3	多维数组的基本操作	125
5.5	Python 中的 Matplotlib	133
5.5.1	Matplotlib 的安装	133
5.5.2	创建图	133
<b>第 6 章 PyTorch 基础</b> ..... 142		
6.1	PyTorch 中的 Tensor	142
6.1.1	Tensor 的数据类型	143
6.1.2	Tensor 的运算	146
6.1.3	搭建一个简易神经网络	153

6.2	自动梯度	156
6.2.1	torch.autograd 和 Variable	156
6.2.2	自定义传播函数	159
6.3	模型搭建和参数优化	162
6.3.1	PyTorch 之 torch.nn	162
6.3.2	PyTorch 之 torch.optim	167
6.4	实战手写数字识别	169
6.4.1	torch 和 torchvision	170
6.4.2	PyTorch 之 torch.transforms	171
6.4.3	数据预览和数据装载	173
6.4.4	模型搭建和参数优化	174
<b>第 7 章</b>	<b>迁移学习</b>	<b>180</b>
7.1	迁移学习入门	180
7.2	数据集处理	181
7.2.1	验证数据集和测试数据集	182
7.2.2	数据预览	182
7.3	模型搭建和参数优化	185
7.3.1	自定义 VGGNet	185
7.3.2	迁移 VGG16	196
7.3.3	迁移 ResNet50	203
7.4	小结	219
<b>第 8 章</b>	<b>图像风格迁移实战</b>	<b>220</b>
8.1	风格迁移入门	220
8.2	PyTorch 图像风格迁移实战	222
8.2.1	图像的内容损失	222
8.2.2	图像的风格损失	223
8.2.3	模型搭建和参数优化	224
8.2.4	训练新定义的卷积神经网络	226
8.3	小结	232

<b>第 9 章 多模型融合</b> .....	233
9.1 多模型融合入门.....	233
9.1.1 结果多数表决.....	234
9.1.2 结果直接平均.....	236
9.1.3 结果加权平均.....	237
9.2 PyTorch 之多模型融合实战.....	239
9.3 小结.....	246
<b>第 10 章 循环神经网络</b> .....	247
10.1 循环神经网络入门.....	247
10.2 PyTorch 之循环神经网络实战.....	249
10.3 小结.....	257
<b>第 11 章 自动编码器</b> .....	258
11.1 自动编码器入门.....	258
11.2 PyTorch 之自动编码实战.....	259
11.2.1 通过线性变换实现自动编码器模型.....	260
11.2.2 通过卷积变换实现自动编码器模型.....	267
11.3 小结.....	273

# 第 1 章

## 浅谈人工智能、神经网络和计算机视觉

---

目前，关于人工智能，有不少值得我们期待的新技术涌现，比如自动驾驶、智能图像识别、智能医疗、智能金融等，这些技术正不断被应用到真实的场景中，而且真实的场景还在不断丰富，这势必在无形之中改变各行各业现状。本章将对人工智能及其相关知识进行简单介绍，带领读者对人工智能及其相关知识进行初步了解。

### 1.1 人工还是智能

---

人工智能可以细分为强人工智能和弱人工智能，弱人工智能更注重“人工”的重要性，强人工智能更注重“智能”的重要性。

通俗地讲，在弱人工智能机器对某个问题进行决策时，人仍然需要积极参与其中，所以弱人工智能也被称作限制领域的人工智能或应用型人工智能。弱人工智能只能在特定的领域解决特定的问题，而且其中的一些问题已经有了明确的答案，比如作为人的智能助手，

在某方面代替人的日常重复劳动。弱人工智能技术已经在某些特定领域落地，并对互联网、金融、制造业、医疗等各个领域产生了不小的冲击，而且会持续下去。

而我们在使用强人工智能机器对问题进行决策时，就不再需要人参与其中，因为强人工智能机器能够“思考”，进而不断优化和拓展自己解决问题的能力，甚至能够创造出全新的技能，所以强人工智能也被称作通用人工智能或完全人工智能，即已经具备了能够完全替代人在各领域工作的能力。

人工智能的舞台是巨大的，改变世界的机会无处不在，相信通过我们的不断努力，人工智能技术会发展得更好，应用场景会更丰富，人们的生活、工作方式也将因此发生翻天覆地的变化。

## 1.2 人工智能的三起两落

人工智能技术发展至今，已经是第 3 次受到公众的高度关注了，这得益于计算机和大数据的发展，更重要的是人们看到了能够真正落地应用的产品。下面让我们回顾一下人工智能三起两落的历史。

### 1.2.1 两起两落

1956 年夏季，John McCarthy、Marvin Minsky、Claude Shannon 等人在美国举办的达特茅斯会议（Dartmouth Conference）上首次提出了“人工智能”的概念。这是人类历史上第 1 个有真正意义的关于人工智能的研讨会，也是人工智能学科诞生的标志，具有十分重要的意义。人工智能概念一经提出，便收获了空前的反响，人工智能历史上的第 1 股浪潮就这样顺理成章地形成了，该浪潮随即席卷全球。当时，普通大众和研究人工智能的科学家都极为乐观，相信人工智能技术在几年内必将取得重大突破和快速进展，甚至预言在 20 年内智能机器能完全取代人在各个领域的工作。这种乐观情绪持续高涨，直到 1973 年《莱特希尔报告》的出现将其终结，该报告用翔实的数据明确指出人工智能的任何部分都没有达到科学家一开始承诺达到的影响力水平，至此人工智能泡沫被无情地戳破，在人们幡然醒悟的同时，人工智能历史上的第 1 个寒冬到来，人们对人工智能的热情逐渐消退，社会各界的关注度和资金投入也逐年减少。

20 世纪 80 年代，专家系统（Expert System）出现又让企业家和科学家看到了人工智



能学科的新希望，继而形成人工智能历史上的第 2 股浪潮。专家系统是指解决特定领域问题的能力已达到该领域的专家能力水平，其核心是通过运用专家多年积累的丰富经验和专业知识，不断模拟专家解决问题的思维，处理只有专家才能处理的问题。专家系统的出现实现了人工智能学科从理论走向专业知识领域的应用，各种应用场景不断丰富，在人工智能历史上是一次重大突破和转折，具有深远的意义。真正意义上的计算机视觉、机器人、自然语言处理、语音识别等专业领域也诞生于这个阶段。但是随着时间的推移，专家系统的缺点也暴露无遗，最为致命的就是专家系统的应用领域相对狭窄，在很多方面缺乏常识性知识和专业理论的支撑，这直接将第 2 股人工智能浪潮推向了寒冬。

20 世纪 90 年代后期，机器学习（Machine Learning）、深度学习（Deep Learning）等技术成为人工智能的主流，再加上大数据和计算机硬件的快速发展，使人工智能再次卷土重来，这一次，以语音识别、计算机视觉、自然语言处理为代表的专业领域均取得了巨大突破和进展。

## 1.2.2 卷土重来

第 3 股人工智能浪潮的兴起与机器学习、深度学习技术被广泛应用和研究有着千丝万缕的联系。但是深度学习和机器学习不是什么新兴技术，并且深度学习还是机器学习的一个分支，人工智能、机器学习和深度学习之间的包含关系如图 1-1 所示。

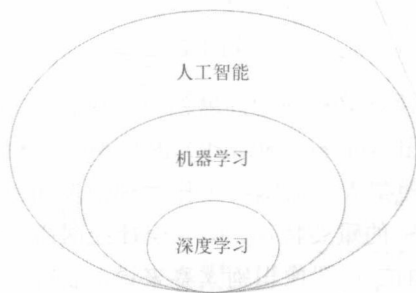


图 1-1

机器学习也被称为统计学习方法，顾名思义，机器学习中的大部分学习算法都是基于统计学原理的，所以机器学习和深度学习技术具备一个共同的特点：它们都需要使用尽可能多的数据来完成对自身模型的训练，这样才能让最终输出的模型拥有强大的泛化能力。

因为本书涉及的是计算机视觉相关的内容，所以我们重点看看计算机视觉领域因为第 3 股人工智能浪潮的冲击又发生了哪些变化。计算机视觉是人工智能学科中最能体现智能

成分的技术，如果计算机视觉问题得到了完美解决，就可以说人类在人工智能领域又迈进了一大步。当前在计算机视觉领域应用得最好的技术是深度学习方法，也可以说深度学习之所以有如此大的影响力，和它在计算机视觉领域取得的突出成绩是分不开的，所以计算机视觉和深度学习成就了彼此。

使用深度学习方法处理计算机视觉问题的过程类似于人类的学习过程：我们搭建的深度学习模型通过对现有图片的不断学习总结出各类图片的特征，最后输出一个理想的模型，该模型能够准确预测新图片所属的类别。图 1-2 展示了两个不同的学习过程，上半部分是通过使用深度学习模型解决图片分类问题，下半部分是人通过学习总结的方式解决物体识别分类问题，它们之间的工作机理非常相似，这也让深度学习技术拥有了更浓郁的智能色彩。

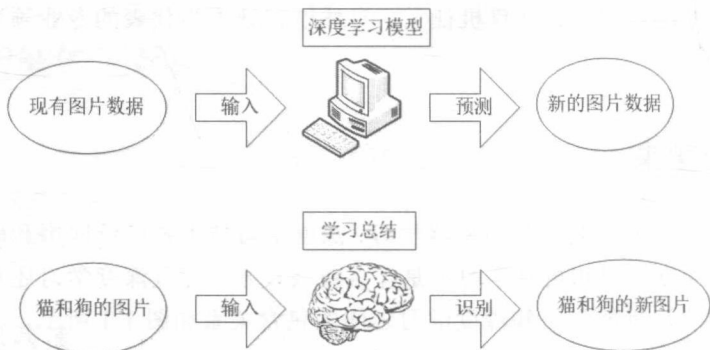


图 1-2

在使用深度学习方法解决计算机视觉问题的过程中，用得最多的网络架构是一种叫作卷积神经网络（Convolutional Neural Network）的模型。卷积神经网络是人工神经网络的变化和升级，是科学家通过模拟人类大脑的工作原理而发明的。人们发现人工神经网络模型能够很好地提取输入图像中的重要特征，而卷积神经网络在图像特征提取方面具有更明显的优势。就拿近几年举办的专业图像识别大赛来说，取得优异成绩的参赛队伍基本上都使用了卷积神经网络模型，这也证明了深度学习方法在图像识别、图像处理、图像特征的提取上要比目前的一些主流的传统机器学习方法效果更好。

人工智能技术是未来各行各业的生产力，国外的 Google、Facebook、Microsoft 等，以及国内的百度、腾讯和阿里巴巴等，都在大量招揽人工智能方向的人才，许多国家也 already 将人工智能技术的发展提升到国家战略高度，人工智能相关技术领域的薪酬也是水涨船高，大家都在为通过人工智能学科改变相关领域和行业时刻准备着。

## 1.3 神经网络简史

神经网络的概念来自生物学科，人脑中错综复杂的生物神经网络承担着对人自身庞大的生物信息的处理工作。之后出现的人工神经网络其实是科学家根据人脑中生物神经网络的工作原理而抽象出的一种可以用数学进行定义的模式，但这个抽象的过程仅限于认知领域，因为在实际情况下生物神经网络的工作机理会比用数学定义的人工神经网络的表达式复杂许多。即便如此，人工神经网络在处理问题时效果却非常出众。如图 1-3 所示是我们对机械脑和人脑的一个臆想。

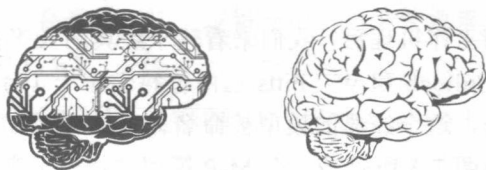


图 1-3

### 1.3.1 生物神经网络和人工神经网络

构成生物神经网络系统和功能的基本单位是生物神经细胞，也叫作生物神经元，一个生物神经元由细胞核、树突、轴突、突触等组织构成，而一个完整的生物神经网络系统由成千上万个生物神经元构造而成。如图 1-4 所示就是一个生物神经元的简单图例。

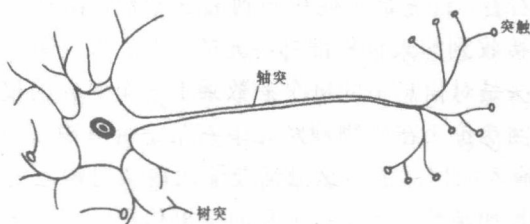


图 1-4

在图 1-4 中左边类似于树枝的是树突，树突是神经元的信息输入端，一个神经元通过树突就可以接收来自外部神经元的信息传入；图中中间的长条是轴突，轴突是神经元的信

息输出端，神经元通过轴突可以把神经元中已处理完成的信息传递到突触；图中右边同样类似于树枝但是带有小圆点的是突触，突触是本神经元和外部神经元之间的连接接口。大量的神经元通过树突和突触互相连接，最后构造出一个复杂的神经网络。生物神经元的信  
息处理流程简单来说是先通过本神经元的树突接收外部神经元传入本神经元的信  
息，这个信息会根据神经元内定义的激活阈值选择是否激活信息，如果输入的信息最终被神经元激活，那么会通过本神经元的轴突将信息输送到突触，最后通过突触传递至与本神经元连接的其他神经元。

### 1.3.2 M-P 模型

在知道生物神经元的工作机理后，我们来看一个经典的人工神经元模型。人工神经元的基础模型是由 W.S.McCulloch 和 W. Pitts 这两位科学家于 1943 年根据生物神经元的生物特性和运行机理发明的，这个经典的模型被命名为 M-P 模型，M 和 P 分别是这两位科学家的名字的首字母。如图 1-5 所示是一个 M-P 模型的结构示意图。

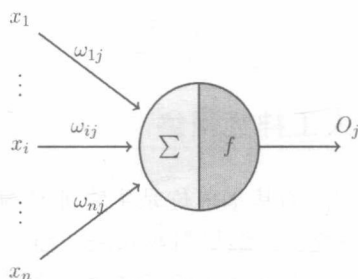


图 1-5

在图 1-5 中从左向右看，首先是一列从  $x_1$  到  $x_n$  的参数，我们可以将从  $x_1$  到  $x_n$  看作类似于生物神经元中的树突接收到的来自外部神经元的信  
息，不过还需要对这些输入的信息进行相应的处理，处理方法是  
对信息中的每个参数乘上一个对应的权重值，权重值的范围是从  $w_{1j}$  到  $w_{nj}$ ；图中的圆圈等价于在生物神经元中判断是否对输入的信息进行激活、输出的部分，M-P 模型在判断输入的信息能否被激活及输出前会对输入的信息使用  $\Sigma$  来完成求和处理， $\Sigma$  就是数学中的累加函数，然后将求和的结果传送给函数  $f$ ，函数  $f$  是一个定义了目标阈值的激活函数，这个激活函数只有在满足目标阈值时才能将信息激活及输出； $O_j$  类似于生物神经元中的轴突，用于承载输出的信息。

M-P 模型的数学表达式如下：