

GAN

实战生成对抗网络

[美] Kuntal Ganguly 著
刘梦馨 译

Learning Generative Adversarial Networks

下一代简化版
深度学习



中国工信出版集团



电子工业出版社
PUBLISHING HOUSE OF ELECTRONICS INDUSTRY
http://www.phei.com.cn

GAN 实战生成对抗网络

Learning Generative Adversarial Networks

北京·BEIJING

内 容 简 介

本书介绍深度学习领域一个十分活跃的分支——生成对抗网络(GAN)。书中覆盖了深度学习的基础、对抗网络背后的原理以及构建方式等内容。同时本书还介绍了多个真实世界中使用对抗网络构建智能应用的案例并提供了具体的代码以及部署方法,旨在帮助读者在真正的生产环境中使用生成对抗网络。

本书适合数据科学家、算法工程师、数据挖掘工程师以及其他机器学习领域相关的从业人员,以帮助他们的工作中应用生成对抗模型;也适合机器学习和深度学习的爱好者、初学者来了解深度学习领域的最新魔力。

Copyright © 2017 Packt Publishing. First published in the English language under the title 'Learning Generative Adversarial Networks'.

本书简体中文版专有出版权由 Packt Publishing 授予电子工业出版社。未经许可,不得以任何方式复制或抄袭本书的任何部分。专有出版权受法律保护。

版权贸易合同登记号 图字:01-2018-3819

图书在版编目(CIP)数据

GAN: 实战生成对抗网络 / (美) 昆塔勒·甘古力(Kuntal Ganguly)著; 刘梦馨译. —北京: 电子工业出版社, 2018.6

书名原文: Learning Generative Adversarial Networks

ISBN 978-7-121-34254-7

I. ①G… II. ①昆… ②刘… III. ①机器学习—研究 IV. ①TP181

中国版本图书馆 CIP 数据核字(2018)第 106112 号

策划编辑: 张春雨

责任编辑: 李云静

印 刷: 三河市华成印务有限公司

装 订: 三河市华成印务有限公司

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: 10

字数: 195 千字

版 次: 2018 年 6 月第 1 版

印 次: 2018 年 6 月第 1 次印刷

定 价: 65.00 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, 盗版侵权举报请发邮件至 dbqq@phei.com.cn。

本书咨询联系方式: 010-51260888-819 faq@phei.com.cn。

译者序

机器学习领域的发展日新月异，并已经深入我们的日常生活。语音识别、图像识别、机器翻译这些技术都在改变着我们的生活方式，而这些都和深度学习紧密相关。尽管机器拥有的能力越来越令人震撼，然而作为人类的我们依然可以安慰自己说，机器拥有的只是计算能力，只有我们人类才拥有创造性的智能。

诚然目前应用最广泛的机器学习方式还是监督学习，需要人类预先对数据进行标记，需要人类先验的经验总结，机器才能更好地工作。由于人类并不能完全将自己的经验符号化、数字化，因此，即使机器可以处理更大量的数据，也无法完全拥有人类的智能。而至于音乐、绘画、文学等艺术创作更被认为是人类的专属，机器永远望尘莫及，机器离真正的智能还有很大的距离。

而生成对抗网络（GAN）的出现，让事情发生了变化，生成对抗网络采用无监督学习的方式，自动从源数据中进行学习，在不需要人工对数据集进行标记的情况下就可以产生令人惊叹的效果。从维基百科的词条学习专业知识，然后写出一个令普通人可以理解的科普文章，模仿知名画家的作品并学习他们的风格进行再创作，这种看似只有人类这样拥有智能和创造性的物种才会做的事情，机器学习现在真的可以做到了！

而这本书的作者 Kuntal Ganguly 就将在本书中向你介绍这个神奇的技术。生成对抗网络也是如今深度学习里最为前沿也是最为令人着迷的领域之一。本书从生成模型的基础讲起，逐步介绍了各种生成对抗网络的变体和使用场景。除了深度学习相关的理论，书中还有大量基于现实世界的使用案例和代码，给读者提供了充足的实践方法。相信通过

本书你可以快速地领略到生成对抗网络的魅力及其无限的可能性，并将其应用到你的实际工作中。

非常荣幸获得电子工业出版社张春雨先生的邀请来负责本书的中文翻译。在翻译的过程中我深感自己的中英文能力十分有限，加之自己在机器学习领域也还是个新手，故译文仍有不足之处，十分期待大家的指正。

刘梦馨

2018年5月19日

作者简介

Kuntal Ganguly 是一位大数据分析工程师，其利用机器学习和大数据框架来搭建大规模数据驱动系统。Kuntal 具有 7 年的大数据和机器学习应用构建经验。

Kuntal 为云上客户提供搭建实时分析系统的解决方案。这其中利用到了托管式的云服务和开源 Hadoop 生态系统工具，诸如 Spark、Kafka、Storm、Solr 以及机器学习和深度学习框架。

Kuntal 也喜欢亲自动手参与软件的开发过程，并且曾经独自一人完成了多个大规模分布式应用从构思、架构、开发一直到部署的整个过程。他也是一位机器学习和深度学习的从业者，十分热衷于构建智能应用。

下面是他的 LinkedIn 主页链接：<https://www.linkedin.com/in/kuntal-ganguly-59564088/>。

十分感谢我的母亲 Chitra Ganguly 以及我的父亲 Gopal Ganguly 对我的爱和支持，以及他们多次教导我要努力工作。从他们身上学到的点点滴滴对我的整个人生都有巨大的帮助。同时，我也想感谢这么多年来陪伴我的朋友、同事以及导师。

审稿人介绍

Max Strakhov 是一位具有 8 年计算机编程工作经验以及 4 年机器学习工作经验的软件工程师和研究者。他曾在 Google 和 Yandex 工作，目前是 AURA Devices LLC 的联合创始人及 CTO。

他对深度学习尤其是生成网络在人工智能领域的应用有着深厚的兴趣。他在自己的博客 <http://monnoroch.github.io/> 中记录了关于深度学习、软件工程以及其他相关技术的分享。

目录

我相信数据科学和人工智能将会赋予我们超能力。

前言

本书中提到的概念和模型可以帮助你快速地建立深度网络，创造性地利用无监督学习方法来完成监督学习领域的任务，例如图像分类。

利用生成网络的基本概念，你会学到如何从无标签的数据中生成逼真的图像，根据文本的描述信息制作图像，以及探索跨域的风格迁移以及域之间的关系。

本书的内容

第 1 章 深度学习概述，以一种简单的没有太多数学和公式的方式介绍了深度学习最新的常用概念和术语。同时也展示了深度学习这些年来的演化，以及深度学习如何利用生成模型逐渐进入无监督学习领域的过程。

第 2 章 无监督学习 GAN，介绍生成对抗网络的工作原理以及如何构建生成对抗网络。同时还会介绍深度学习网络如何应用于半监督学习领域，以及如何应用它们来进行图像的生成和创作。GAN 的训练十分困难，在本章中会介绍一些提升训练/学习效果的技术。

第 3 章 图像风格跨域转换，介绍如何利用简单但强大的 CGAN 和 CycleGAN 模型进行创意制作。本章展示了利用条件 GAN 根据某种特定风格或者条件来制作图像。本章还讨论了如何利用 BEGAN 进行网络固化来避免模型坍塌的问题。最后，展示了如何利用 CycleGAN 来进行跨域转换（苹果变成橘子，马变成斑马）。

第 4 章 从文本构建逼真的图像，展示了利用最新的层叠 GAN 技术将从文本生成图像的问题拆分成两个可控的子问题，并利用 StackGAN 进行处理。本章还会展示如何利用

DiscoGAN 成功地进行图像风格的跨域转换，包括以边框的图像为输入来输出手提包的图像，以及对名人的图像进行性别转换。

第 5 章 利用多种生成模型生成图像，介绍预训练模型的概念并讨论了利用大规模分布式系统 Apache Spark 运行深度学习和生成模型。接下来我们会利用 GAN 和预训练模型来提升低品质图像的清晰度。最后，我们会介绍用 DeepDream 和 VAE 等其他生成模型来生成图像和风格转换。

第 6 章 将机器学习带入生产环境，介绍多种将以机器学习和深度学习为基础的智能应用部署到生产环境的方法，包括基于数据中心的部署方式，以及基于云环境的容器化微服务部署方式，或基于无服务器技术的部署方式。

阅读本书的准备工作

本书中所用到的工具、库以及数据集都是开源的并可免费获取。本书中提到的云环境会提供一些免费试用来给用户进行评测。通过本书以及其他一些机器学习和深度学习的资料，读者将会发现生成对抗网络所拥有的无限创造力。

你需要安装 Python 并用 pip 安装一些相关的 Python 软件包来有效地运行本书中提供的样例代码。

本书面向的读者

本书所面向的读者是那些希望了解生成模型的威力以及从无标记的源数据和噪声中制作逼真图像的机器学习专家以及数据科学家。

拥有足够机器学习和神经网络知识并且能熟练使用 Python 进行编程的读者可以通过本书学会如何从文本生成图像，自动发现相似域之间的关联，并能基于深度学习和生成模型来探索无监督学习领域的问题。

约定惯例

在本书中，你会发现我们将用不同风格的数字和文本来区分不同的信息。下面是一些文本风格的样例，以及对应含义的解释。

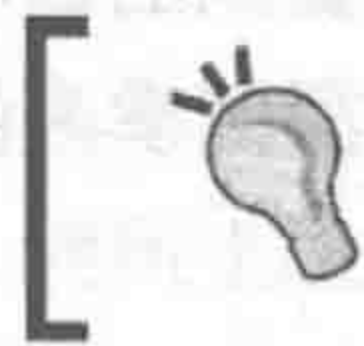
导入相关软件模块的代码块如下所示：

```
<div class="packt_code">
nsamples=6
Z_sample = sample_Z(nsamples, noise_dim)
y_sample = np.zeros(shape=[nsamples, num_labels])
y_sample[:, 7] = 1 # generating image based on label
samples = sess.run(G_sample, feed_dict={Z: Z_sample, Y:y_sample})
</div>
```

新出现的术语和重要词汇会被加粗显示。如果你使用屏幕阅读，屏幕中在菜单和对话框中出现的词汇，在文本中以下面的形式展现，例如：“单击下一页按钮将进入下一页”。



警告和重要的笔记出现在这样的方框中。



提示和技巧出现在这样的方框中。

读者服务

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

- **下载资源：**本书提供示例代码及资源文件，可在[下载资源处](#)下载。
- **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/34254>



目录

前言.....	XII
1 深度学习概述	1
1.1 深度学习的演化	1
1.1.1 sigmoid 激发	3
1.1.2 修正线性单元 (ReLU)	3
1.1.3 指数线性单元 (ELU)	4
1.1.4 随机梯度下降 (SGD)	5
1.1.5 学习速率调优	6
1.1.6 正则化	7
1.1.7 权重分享以及池化	8
1.1.8 局部感受野	10
1.1.9 卷积网络 (ConvNet)	11
1.2 逆卷积/转置卷积	13
1.2.1 递归神经网络和 LSTM	13
1.2.2 深度神经网络	14
1.2.3 判别模型和生成模型的对比	16
1.3 总结	16

2 无监督学习 GAN	17
2.1 利用深度神经网络自动化人类任务	17
2.1.1 GAN 的目的	18
2.1.2 现实世界的一个比喻	19
2.1.3 GAN 的组成	20
2.2 GAN 的实现	22
2.2.1 GAN 的应用	25
2.2.2 在 Keras 上利用 DCGAN 实现图像生成	26
2.2.3 利用 TensorFlow 实现 SSGAN	29
2.3 GAN 模型的挑战	38
2.3.1 启动及初始化的问题	38
2.3.2 模型坍塌	38
2.3.3 计数方面的问题	39
2.3.4 角度方面的问题	39
2.3.5 全局结构方面的问题	40
2.4 提升 GAN 训练效果的方法	41
2.4.1 特征匹配	41
2.4.2 小批量	42
2.4.3 历史平均	42
2.4.4 单侧标签平滑	42
2.4.5 输入规范化	42
2.4.6 批规范化	42
2.4.7 利用 ReLU 和 MaxPool 避免稀疏梯度	43
2.4.8 优化器和噪声	43
2.4.9 不要仅根据统计信息平衡损失	43
2.5 总结	43
3 图像风格跨域转换	45
3.1 弥补监督学习和无监督学习之间的空隙	45
3.2 条件 GAN 介绍	46
3.2.1 利用 CGAN 生成时尚衣柜	47

3.2.2	利用边界均衡固化 GAN 训练.....	51
3.3	BEGAN 的训练过程.....	52
3.3.1	BEGAN 的架构.....	52
3.3.2	利用 TensorFlow 实现 BEGAN.....	53
3.4	利用 CycleGAN 实现图像风格的转换.....	57
3.4.1	CycleGAN 的模型公式.....	58
3.4.2	利用 TensorFlow 将苹果变成橘子.....	58
3.4.3	利用 CycleGAN 将马变为斑马.....	61
3.5	总结.....	63
4	从文本构建逼真的图像.....	65
4.1	StackGAN 介绍.....	65
4.1.1	条件强化.....	66
4.1.2	StackGAN 的架构细节.....	68
4.1.3	利用 TensorFlow 从文本生成图像.....	69
4.2	利用 DiscoGAN 探索跨域的关系.....	72
4.2.1	DiscoGAN 架构以及模型公式.....	73
4.2.2	DiscoGAN 的实现.....	75
4.3	利用 PyTorch 从边框生成手提包.....	78
4.4	利用 PyTorch 进行性别转换.....	80
4.5	DiscoGAN 和 CycleGAN 的对比.....	82
4.6	总结.....	82
5	利用多种生成模型生成图像.....	83
5.1	迁移学习介绍.....	84
5.1.1	迁移学习的目的.....	84
5.1.2	多种利用预训练模型的方法.....	85
5.1.3	利用 Keras 对车、猫、狗和花进行分类.....	86
5.2	利用 Apache Spark 进行大规模深度学习.....	90
5.2.1	利用 Spark 深度学习模块运行预训练模型.....	91
5.2.2	利用 BigDL 运行大规模手写数字识别.....	94

5.2.3	利用 SRGAN 生成高清晰度图像.....	98
5.2.4	SRGAN 的架构.....	99
5.3	利用 DeepDream 生成梦幻的艺术图像.....	105
5.4	在 TensorFlow 上利用 VAE 生成手写数字.....	107
5.5	VAE 在真实世界的比喻.....	108
5.6	GAN 和 VAE 两个生成模型的比较.....	111
5.7	总结.....	111
6	将机器学习带入生产环境.....	113
6.1	利用 DCGAN 构建一个图像矫正系统.....	113
6.1.1	构建图像矫正系统的步骤.....	115
6.1.2	在生产环境部署模型的挑战.....	117
6.2	利用容器的微服务架构.....	118
6.2.1	单体架构的缺陷.....	118
6.2.2	微服务架构的优点.....	118
6.2.3	使用容器的优点.....	119
6.3	部署深度模型的多种方法.....	120
6.3.1	方法 1——离线建模和基于微服务的容器化部署.....	120
6.3.2	方法 2——离线建模和无服务器部署.....	121
6.3.3	方法 3——在线学习.....	121
6.3.4	方法 4——利用托管机器学习服务.....	121
6.4	在 Docker 上运行基于 Keras 的深度模型.....	121
6.5	在 GKE 上部署深度模型.....	124
6.6	利用 AWS Lambda 和 Polly 进行无服务器的图像识别并生成音频.....	127
6.6.1	修改 Lambda 环境下代码和包的步骤.....	137
6.6.2	利用云托管服务进行人脸识别.....	138
6.7	总结.....	145

1

深度学习概述

深度神经网络目前在解决多种问题，在图像识别、语音识别、机器翻译和自然语言处理等多个领域可以提供效果和人工类似的解决方案。

在本章中，我们将会回顾神经网络这种从生物学获得启发的架构这些年来是如何不断演化的。作为之后几章的基础，接下来我们将会进入深度学习相关的一些最重要的概念和术语。最后我们将通过生成模型来了解深度网络创造性特性背后的“原力”。

本章将会包含以下内容：

- 深度学习的演化。
- 随机梯度下降、修正线性单元（ReLU）、学习速率和一些其他概念。
- 条件网络、递归神经网络以及 LSTM。
- 判别模型和生成模型之间的区别。

1.1 深度学习的演化

和神经网络相关的大量工作诞生于 20 世纪八九十年代，然而当时的计算机运行得十分缓慢，数据量也很小，科研人员并没有发现在现实世界中能够使用神经网络的场景。因此，在 21 世纪的头 10 年，神经网络几乎从机器学习的世界中消失了。直到最近几年，最先在 2009 年的语言识别领域，接下来在 2012 年的计算机视觉中，神经网络凭借优异的表现重新归来（同时还带来了 LeNet、AlexNet 等）。这期间什么发生了变化？

大量的数据（Big Data）以及廉价高速的 GPU 改变了这一切。现如今神经网络已经进入了世界的每一个角落。如果你正在做任何和数据相关的工作，则不管是分析还是预测，

深度学习都将会是你必须要了解的技术。

下面看一下图 1-1。

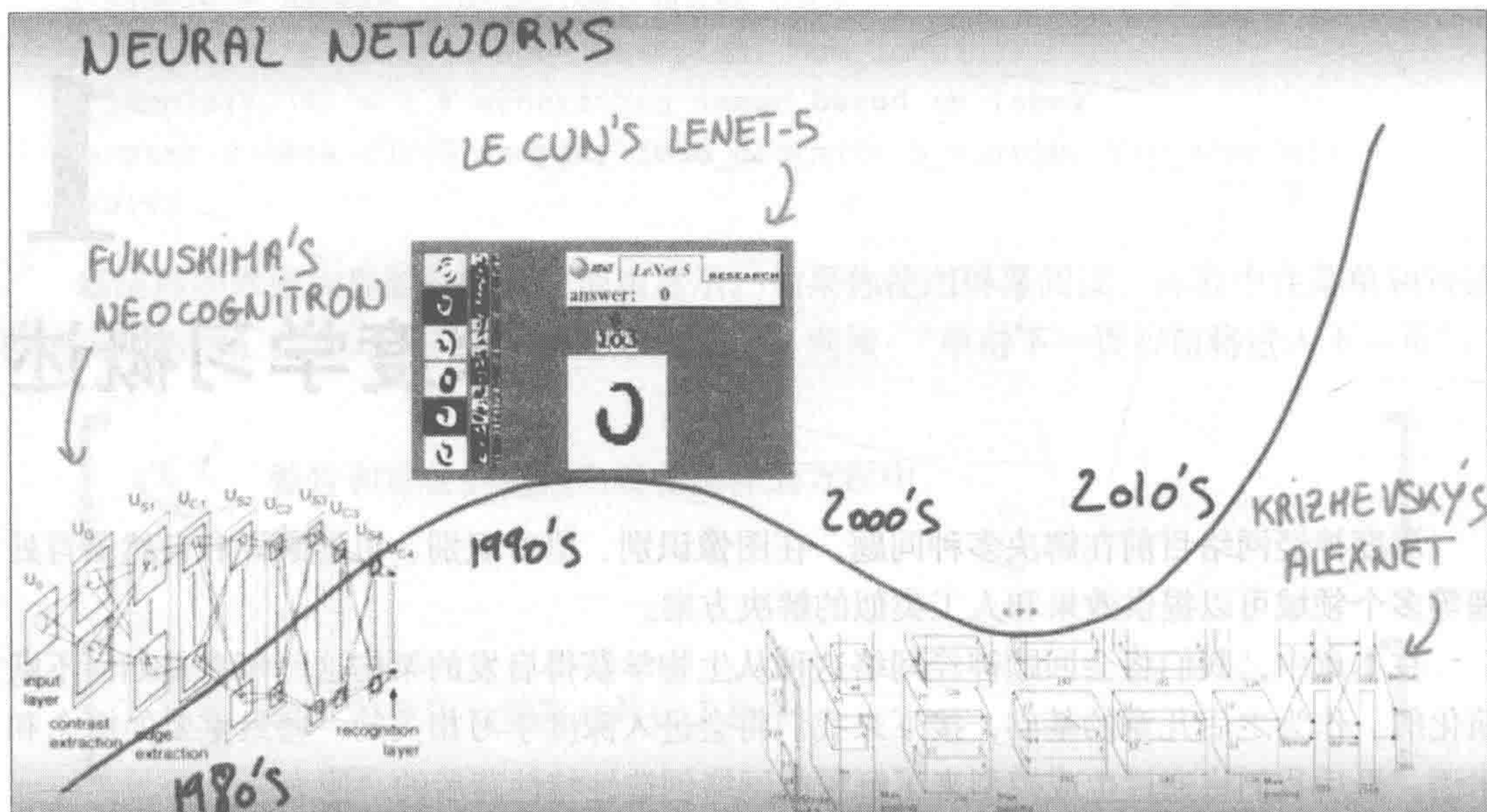


图1-1：深度学习的演化

深度学习是机器学习中十分令人振奋的一个分支，它利用大量的数据来训练计算机去做一些之前只有人类才能做的事情，诸如如何分辨图像中有哪些物体，分辨人们在打电话时对话的内容，将文档翻译成另一种语言，帮助机器人探索世界并对各种事情及时响应，等等。深度学习成为解决机器认知问题最为核心的工具，并且是计算机视觉和语言识别领域当下最为优秀的技术。

现如今大量的工作都将深度学习作为机器学习工具链中最为核心的一部分。由于深度学习在大规模数据量和解决复杂问题中能发挥最大的优势，因此，Facebook、Baidu、Amazon、Microsoft 和 Google 都在产品中使用到了深度学习。

深度学习通常是“深度神经网络”的代称，深度神经网络中包含了多个层，每一层又是由多个节点组成的。计算发生在每个节点，每个节点会将输入以及一系列的参数和权重进行计算，将输入进行增强或者抑制。这些输入-权重组合的结果将会被汇总，它们的总和会被传入一个激发（activation）函数，通过这个函数来确认这个汇总值该如何在整个网络中进行传递，以及会对最终的结果产生怎样的影响。例如在一个分类任务中，这个值会对分类结果产生何种影响。神经网络的每一层包含一排的节点，通过输入数据的不同来决