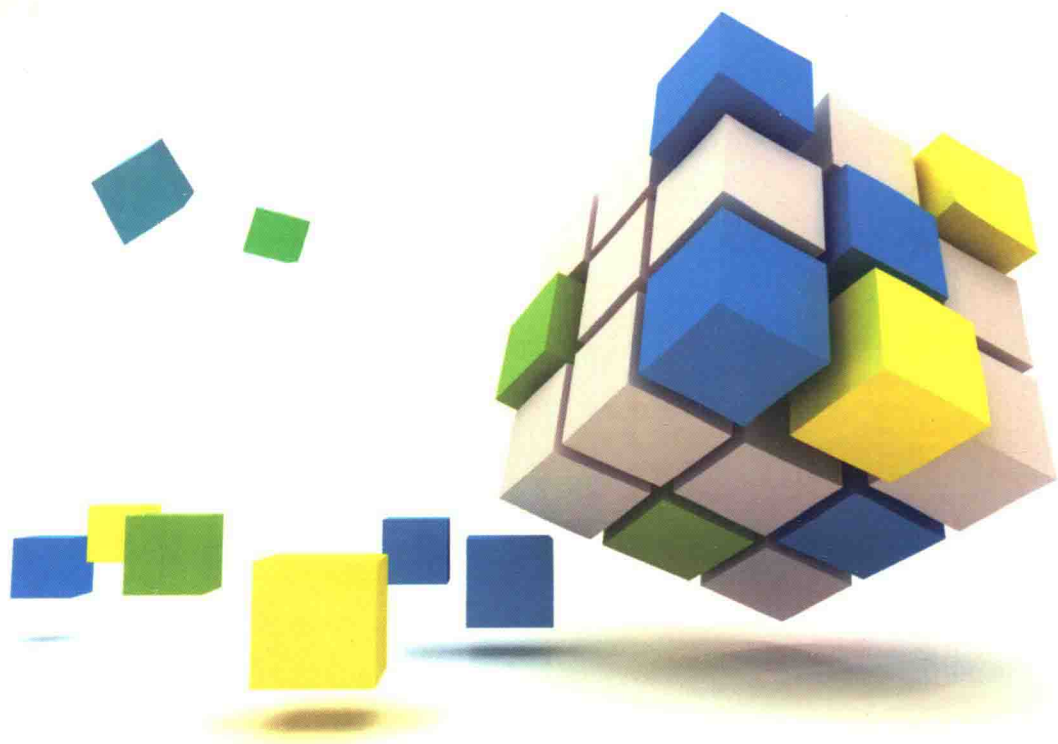




APPLICATION AND PRACTICE
OF DEEP LEARNING IN DYNAMIC MEDIA

深度学习在动态媒体中的 应用与实践

唐宏 陈麒 庄一嵘◎编著

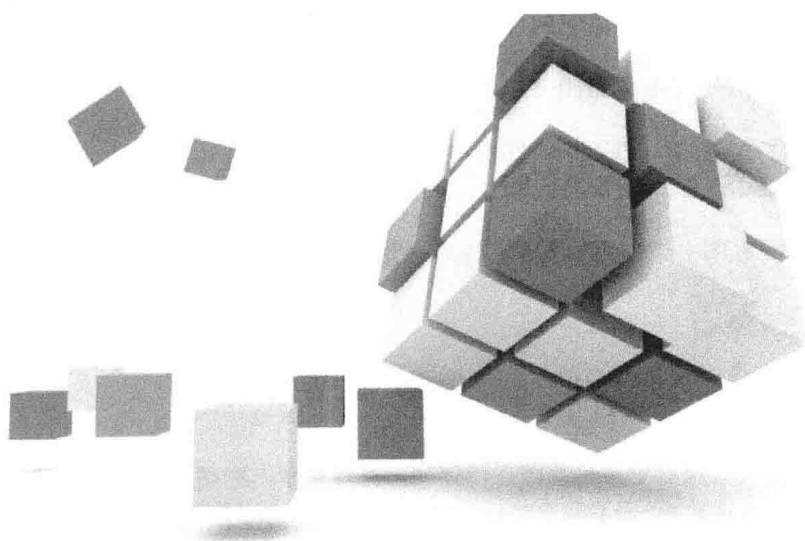
对深度学习有浓厚兴趣的读者，希望用深度学习完成设计的计算机专业
或者电子信息专业的高校毕业生以及想从实战项目入手的深度学习研发工程师或者算法工程师



中国工信出版集团



人民邮电出版社
POSTS & TELECOM PRESS



APPLICATION AND PRACTICE
OF DEEP LEARNING IN DYNAMIC MEDIA

深度学习在动态媒体中的 应用与实践

唐宏 陈麒 庄一嵘◎编著

人民邮电出版社
北京

图书在版编目 (C I P) 数据

深度学习在动态媒体中的应用与实践 / 唐宏, 陈麒,
庄一嵘编著. — 北京: 人民邮电出版社, 2018.3 (2018.6 重印)
ISBN 978-7-115-48010-1

I. ①深… II. ①唐… ②陈… ③庄… III. ①
Linux操作系统 IV. ①TP316.85

中国版本图书馆CIP数据核字(2018)第053368号

内 容 提 要

本书是一本深度学习的基础入门读物, 对深度学习的基本理论进行了介绍, 主要以 Ubuntu 系统为例搭建了三大主流框架——Caffe、TensorFlow、Torch, 然后分别在 3 个框架下, 通过 3 个实战项目掌握了框架的使用方法, 并详细描述了生产流程, 最后讲述了通过集群部署深度学习的项目以及运营维护的注意事项。

本书适合对深度学习有浓厚兴趣的读者、希望用深度学习完成设计的计算机专业或电子信息专业的高校毕业生以及想从实战项目入手的深度学习研发工程师或算法工程师。

◆ 编 著 唐 宏 陈 麒 庄一嵘

责任编辑 李慧恬

责任印制 彭志环

◆ 人民邮电出版社出版发行 北京市丰台区成寿寺路 11 号

邮编 100164 电子邮件 315@ptpress.com.cn

网址 <http://www.ptpress.com.cn>

北京虎彩文化传播有限公司印刷

◆ 开本: 700×1000 1/16

印张: 9.25

2018 年 3 月第 1 版

字数: 100 千字

2018 年 6 月北京第 2 次印刷

定价: 59.00 元

读者服务热线: (010)81055488 印装质量热线: (010)81055316

反盗版热线: (010)81055315

前 言

深度学习在被《MIT 技术评论》评选为 2013 年世界十大突破性技术之首后，保持着迅猛发展的趋势，现已应用在人脸识别、自然语言处理、图像审核、车型识别、图像搜索等前沿技术中。作为目前热门的技术，很多计算机从业者、科研爱好者都对深度学习充满了好奇，但是由于深度学习涉及很多高深的数学原理，所以对于初学者而言，深度学习的入门门槛较高，急需一本以项目实战为主的书籍指引初学者登入深度学习的殿堂。市场中现有的关于深度学习的书籍大多偏重理论、公式的推导，缺乏实战的讲解，对读者从“入门到放弃”有很大的影响。本书注重实战项目的讲解，让读者从实战快速入门深度学习。

本书只涉及深度学习的基本原理，不过多纠结数学公式推导，能让读者快速上手书中的实战项目，用于实际生产。本书非常适合以下读者：对深度学习有浓厚兴趣的读者、希望通过深度学习完成设计的计算机专业或电子信息专业的高校毕业生、想从实战项目入手的深度学习研发工程师或算法工程师。

第一部分是概要，由两章构成。第 1 章主要从深度学习的发展、深度学习的应用、主流的几款深度学习框架的对比这 3 个方面，对深度学习进行了介绍。第 2 章主要从深度学习的基本概念、训练过程以及模型 3 个部分来阐述深度学习的基本理论。

第二部分是深度学习三大主流框架以及各个框架下的生产实例的详细介绍。第 3 章介绍了 3 种主流框架，分别是 Caffe、TensorFlow、Torch，

主要以 Ubuntu 系统为主，介绍了三大框架的安装过程，并记录了一些安装过程中的常见问题和解决方法，在每一个深度学习框架的最后，都利用该框架解决了一个实际生产问题。第 4、5、6 章分别是这 3 个框架应用于人脸识别、车辆识别、不良视频识别的生产实例，首先对项目进行了概述和需求分析，然后设计了项目系统，对功能和模块进行了描述，紧接着在 Ubuntu 系统上部署了生产环境，并对生产环境进行了验证，最后通过脚本实现了生产流程。

第三部分是集群部署和运营维护，也即第 7 章。首先阐述了由于深度学习框架部署的复杂性，引出了 Docker 容器，介绍了 Docker 的主要构成组件及其优点，然后以 TensorFlow 为例，详细介绍了基于 Docker 的 TensorFlow 实验环境搭建步骤。最后简单介绍了运营维护需要注意的要点。

中国电信股份有限公司广州研究院的农德华也参与了各章节的编写工作，卢琳完成了漫画配图的创作工作，在此对他们的辛勤付出表示衷心的感谢！此外，由于作者经验有限，书中难免有不妥之处，敬请读者批评指正。

编者

2018 年 1 月于广州

目 录

第 1 章 深度学习简介	1
1.1 深度学习的发展	1
1.2 深度学习的应用及研究方向	3
1.3 深度学习工具介绍和对比	4
1.3.1 Caffe	4
1.3.2 TensorFlow	5
1.3.3 Torch	6
1.4 小结	7
第 2 章 深度学习基本理论	9
2.1 深度学习的基本概念	9
2.2 深度学习的训练过程	13
2.3 深度学习的常用模型和方法	14
2.4 小结	20
第 3 章 深度学习环境搭建	23
3.1 Caffe 安装	23

3.1.1	安装 Caffe 的相关依赖项	24
3.1.2	安装 NVIDIA 驱动	24
3.1.3	安装 CUDA	27
3.1.4	配置 cuDNN	30
3.1.5	源代码编译安装 OpenCV	32
3.1.6	编译 Caffe, 并配置 Python 接口	34
3.2	Caffe 框架下的 MNIST 数字识别问题	41
3.3	TensorFlow 安装	42
3.3.1	基于 pip 安装	42
3.3.2	基于 Anaconda 安装	46
3.3.3	基于源代码安装	51
3.3.4	常见安装问题	56
3.4	TensorFlow 框架下的 CIFAR 图像识别问题	59
3.5	Torch 安装	61
3.5.1	无 CUDA 的 Torch 7 安装	61
3.5.2	CUDA 的 Torch 7 安装	61
3.6	Torch 框架下 neural-style 图像合成问题	62
3.7	小结	74

第 4 章 人脸识别 75

4.1	人脸识别概述	75
4.2	人脸识别系统设计	76
4.2.1	需求分析	76
4.2.2	功能设计	77
4.2.3	模块设计	78
4.3	系统生产环境部署及验证	81
4.3.1	抽帧环境部署	81

4.3.2 抽帧功能验证	82
4.3.3 OpenFace 环境部署	82
4.3.4 OpenFace 环境验证	84
4.4 批量生产	90
4.5 小结	102
第 5 章 车辆识别	103
5.1 概述	103
5.2 系统设计	104
5.2.1 需求分析	104
5.2.2 功能设计	104
5.2.3 模块设计	105
5.3 系统生产环境部署及验证	106
5.3.1 生产环境部署	106
5.3.2 项目部署	107
5.3.3 环境验证	108
5.4 批量生产	109
5.5 小结	117
第 6 章 不良视频识别	119
6.1 概述	119
6.2 不良图片模型简介	120
6.3 系统设计	122
6.4 系统部署及系统测试验证	123
6.5 批量生产	125
6.5.1 批量节目元数据信息检索与筛选	125

6.5.2 基于 FFmpeg 的 SDK 抽取视频 I 帧	126
6.5.3 基于肤色比例检测的快速筛查	128
6.5.4 基于 Caffe 框架的不良图片检测	128
6.6 小结	129

第 7 章 集群部署与运营维护 131

7.1 认识 Docker	131
7.2 基于 Docker 的 TensorFlow 实验环境	134
7.3 运营维护	137
7.4 小结	138

参考文献 139



第 1 章

深度学习简介

1

最近几年，人工智能的概念已经深入人心，各类基于人工智能的技术已经开始进入普通人的生活，国内外各大知名企业纷纷将大量的人力物力投入其中。对于计算机相关行业来说，下一个战略要点是人工智能已经确凿无疑。当然，人工智能是一个很广义的概念，它有很多分支。本书所要讲述的深度学习就是人工智能旗下的其中一个分支，它也是机器学习的一种主要实现方式。为了更详细地了解深度学习，作者将从深度学习的发展、深度学习的应用以及主流的几款深度学习工具的对比等方面逐一进行介绍。



1.1 深度学习的发展

说到深度学习，就不得不说神经网络，它们两个应该说是同一概念在

不同时代背景下的表现形式。神经网络诞生的发展总体上经历了3个阶段。

第一阶段，沃伦·麦卡洛克和沃尔特·皮茨在1943年发表的一篇名为“A logical calculus of the ideas immanent in nervous activity”的论文中提出了神经网络数学模型，这个模型的实现方式是通过从某种物体（无论是实物还是虚拟物体）中提取尽可能多的特征值，我们假设为 n ， n 越大，对物体的分析就越具体，越趋近正确答案，然后通过上面的神经网络数学模型结构，对这些特征值加权并放入一些特定的阈值函数中，这就构成了一个最基础的神经网络算法，当我们输入一些特征值后，通过该函数处理就可以获得输出结果，该数据结果为0或1，对应于逻辑判断的是和否，根据输出结果我们就能判断某个物体是否为这种指定物体。举个例子，我们收集了猫的尽可能多的特征值，并放入阈值函数中构成算法公式，这时我们再输入一个猫或其他的特征值，就能区分这个动物是否为猫，当然这是最基础的实现。

第二阶段，20世纪80年代，反向传播算法和分布式的提出推动了新一轮的神经网络的发展，但是受限于当时的计算水平，神经网络的发展还是一度受阻。

第三阶段，到了21世纪初，CPU的计算水平和GPU的图形处理能力大幅提升，计算水平已经不是神经网络发展的阻碍，并且由于大数据的发展，获取海量特征数据也不是难事，神经网络迎来全新的发展契机，从2012年至今，深度学习的搜索热度逐年上升，人类迎来了人工智能时代！



1.2 深度学习的应用及研究方向

深度学习的应用涉及机器学习的方方面面，除了人们所熟知的图像识别、语音识别等，还包括诸如自动驾驶、机器人、搜索排名、医疗诊断以及游戏等各个方面。

(1) 图像识别

深度学习最开始运用在图像识别领域，像目前比较热门的无人驾驶、人脸识别等功能都是基于图像识别技术。图像识别的实现方式很多，但是基于深度学习的图像识别准确度更高。图像识别技术历久弥新，不管是现在还是未来，其在深度学习的推动下必将走向一个更辉煌的发展道路！

(2) 语音识别

说到语音识别，不得不说到科大讯飞，它是中国最大的智能语音技术服务商，得益于人工智能的发展，科大讯飞的语音识别技术已经达到世界顶级水准，国内多家互联网巨头纷纷与其合作，语音识别的前景不言而喻，相信在未来的发展也无可限量。

(3) 游戏对战

在游戏中加入 AI（人工智能）早已不是新鲜事，AI 的出现不仅让游戏变得不再那么简单乏味，而且添加了人与电脑的互动，这个就要拿典型案例英雄联盟来说说了。过去打人机都非常简单，玩家可以很轻易

地预判 AI 的活动方式，但是 AI 将不再不堪一击，它们将会变得异常难以击败，对玩家的考验也非常大，在游戏行业，对 AI 的需求只会越来越大！



1.3 深度学习工具介绍和对比

深度学习的框架很多，作者选取了几个最流行的深度学习框架来分别讲述它们的优缺点，读者可以根据自己的实际情况选取适合自己的学习框架。

4

1.3.1 Caffe

Caffe 是一款老牌的深度学习框架，它主要运用在视频和图像处理等方面。对其他深度学习应用，比如文本、声音和时序数据等，Caffe 并不是一个好的选择。Caffe 实现基于 C/C++，它使用 MATLAB 和 Python 作为接口语言。

优点：

- 使用 Python 进行开发；
- 在前馈网络和图像处理上较好；
- 在微调已有网络方面较好；
- 不写任何代码就可训练模型。

缺点：

- 不擅长循环网络；

- 需要为新的 GPU 层编写 C++/CUDA ；
- 面对大型网络（GoogLeNet、ResNet）有点吃力 ；
- 不可扩展 ；
- 无商业化支持。

1.3.2 TensorFlow

谷歌作为科技界的龙头老大，也有自己的深度学习框架——TensorFlow，TensorFlow 的出现基本上取代了 Theano 框架，当然读者可以发现这两个框架的众多雷同之处。

和大部分深度学习框架的编程语言类似，TensorFlow 框架语言使用 C/C++，使用 C 语言的引擎机制可以加速其运行，接口语言使用 Python。对于有众多开发者的 Java 社区来说这并不友好，当然 Java 受制于其虚拟机的原因并不是很适合于深度学习。

TensorFlow 的框架不仅解决深度学习方面的问题，也涉及强化学习等其他算法。

优点：

- 计算图抽象，如同 Theano ；
- 比 Theano 的编译速度更快 ；
- 进行可视化的 TensorBoard ；
- 数据和模型并行 ；
- GitHub 社区非常活跃，远超其他深度学习框架的活跃度。

缺点：

- 比其他框架慢；
- 预训练模型不多；
- 计算图是纯 Python 的，因此更慢；
- 无商业化支持；
- 需要退出到 Python 才能加载每个新的训练批处理（Batch）；
- 不能进行太大的调整；
- 在大型软件项目上，动态键容易出错。

1.3.3 Torch

Torch 是用 Lua 语言编写的面向机器学习算法的计算框架。这个框架深受大型科技公司如 Facebook、Twitter 等的喜爱，也广泛地运用在它们的内部项目中，它们拥有专门的内部团队开发自己的深度学习平台。

优点：

- 很多容易结合的模块碎片；
- 易于编写自己的层类型和在 GPU 上运行；
- Lua（大部分库代码是 Lua 语言，易于读取）；
- 大量的预训练模型。

缺点：

- Lua 是小众语言；
- 需要编写自己的训练代码；



- 对循环神经网络不太好；
- 没有商业化支持；
- 糟糕的文档支持；
- 不能即插即用。

其他的深度学习框架如 Theano、CNTK、DSSTNE 和 MXNet 等在这里就不做介绍，读者可以自行搜索相关资料。这里值得一提的是国内的百度公司也推出了自己的深度学习框架——PaddlePaddle，作者没有深入地研究这个框架，不过相比于国外的深度学习框架，这款国内的框架应该会更易上手，毕竟没有语言障碍。当然其官网提供的模型和案例也很丰富，唯一的缺点是只支持在 Docker 中使用，所以还需要使用者有 Docker 的基础，算是一点小小的遗憾吧。



1.4 小结

本章从深度学习的发展、应用以及主流的几款深度学习工具的对比 3 个方面，对深度学习进行了阐述。其中介绍主流的深度学习工具时，以目前最流行的 Caffe、TensorFlow 和 Torch 这三大框架为例，介绍了它们的基本概念并比较了其优缺点。

