

TURING 图灵原创

彭靖田 林健 白小龙◎著



深入理解 TensorFlow

架构设计与实现原理



中国工信出版集团



人民邮电出版社
POSTS & TELECOM PRESS

TURING

图灵原创

彭靖田 林健 白小龙◎著

深入理解 TensorFlow

架构设计与实现原理

人民邮电出版社
北 京

图书在版编目 (CIP) 数据

深入理解TensorFlow：架构设计与实现原理 / 彭靖田, 林健, 白小龙著. — 北京：人民邮电出版社, 2018.5

(图灵原创)

ISBN 978-7-115-48094-1

I. ①深… II. ①彭… ②林… ③白… III. ①人工智能—算法—研究 IV. ①TP18

中国版本图书馆CIP数据核字(2018)第049742号

内 容 提 要

本书以 TensorFlow 1.2 为基础, 从基本概念、内部实现和最佳实践等方面深入剖析了 TensorFlow。书中首先介绍了 TensorFlow 设计目标、基本架构、环境准备和基础概念, 接着重点介绍了以数据流图为核心的机器学习编程框架的设计原则与核心实现, 紧接着还将 TensorFlow 与深度学习相结合, 从理论基础和程序实现这两个方面系统介绍了 CNN、GAN 和 RNN 等经典模型, 然后深入剖析了 TensorFlow 运行时核心、通信原理和数据流图计算的原理与实现, 最后全面介绍了 TensorFlow 生态系统的发展。

本书适合想深入了解 TensorFlow 的程序员。

-
- ◆ 著 彭靖田 林 健 白小龙
责任编辑 王军花
责任印制 周昇亮
 - ◆ 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>
北京鑫正大印刷有限公司印刷
 - ◆ 开本: 800×1000 1/16
印张: 23 彩插: 2
字数: 544千字 2018年5月第1版
印数: 1—4 000册 2018年5月北京第1次印刷
-

定价: 79.00元

读者服务热线: (010)51095186转600 印装质量热线: (010)81055316

反盗版热线: (010)81055315

广告经营许可证: 京东工商广登字 20170147 号

序

人工智能迎来了继 20 世纪 90 年代以来的又一次大发展，人工智能和深度学习无疑是近年来最受追捧的热点。2016 年和 2017 年 AlphaGo 两次分别战胜韩国和中国围棋国手，更是让人工智能成为街头巷尾的热谈。为什么人工智能在经历了两起两落后再一次迎来了复兴？我认为，这一次的爆发主要由理论、应用、硬件和软件四个方面的原因促成的，貌似偶然实则必然。

Geoffrey Hinton 等人在 2006 年发现了训练高层神经网络的有效算法，为深度学习的研究打开了新局面，这也是人工智能得以重燃的导火线。经过后续研究人员的努力，尤其是 CNN 和 RNN 的出现，深度学习、神经网络方法在图像和语音识别方面显示出非常好的效果，大大超越之前的理论和方法，甚至能够突破人类极限。与计算机视觉、机器人、自然语言理解、信息检索等技术相结合，深度学习的应用也从单纯的图像和语音识别扩展到自动驾驶、图像增强与风格替换、文本语音间转换和推荐系统等。使用深度学习技术的应用和初创企业如雨后春笋般冒了出来。尤其需要指出的是，互联网和大数据的广泛应用是深度学习发展的必要条件。一般而言，数据量越大且数据质量越高，由深度学习训练出来的模型精度也就越好。计算机硬件的发展也直接推动了深度学习的发展。GPU 和其他专用加速器件的出现，大大提高了深度学习的计算效率。在语音识别场景下，GPU 可将数十亿样本的训练时间从数年缩短到数天。而专用的 ASIC 加速芯片比 GPU 的能效更有一个数量级以上的提升，也让深度学习从服务器端走向手机端，进一步拓展了其应用范围。

开放的软件生态和易用的软件形态是形成人工智能和深度学习产业链至关重要的两个方面。没有软件的支撑，理论很难与应用相结合，新硬件也很难为应用提速。从大数据软件的发展历程可以想见，如果没有开源的 Hadoop 生态系统，以及受其设计思想影响的新型大规模并行处理数据库系统，我们现在可能还在为如何管理和处理 PB 乃至 EB 级的数据发愁。开源 TensorFlow 的出现解决了类似的问题，一下子拉近了深度学习理论与实际应用的距离。同时，TensorFlow 也具备迈向成功生态系统的必要条件，即差异化的软件功能、刚需的典型应用和活跃的社区支持，发展前景可期。但是，原生 TensorFlow 的软件形态尚不足以支持深度学习的全流程生产化应用，欠缺诸如数据管理和预处理，模型训练、管理和运行，资源管理、任务调度和运行时监控等能力，导致最终用户形成生产力的成本过高。深度学习是计算密集型重资产类应用，如果有能够提供异构高性能计算资源并能够集成上述平台化功能的深度学习公有云服务，可降低 TensorFlow 的使用门槛并提升用户体验，客观上会与开源效应叠加，起到加速产业发展的作用。

很高兴能够在这个时候看到一本讲述如何使用 TensorFlow 的专业图书。作者是深谙计算机系统之道的一线工程师,带给读者的是基于实战经验的理解。非常难得的是,本书除了讲解如何使用 TensorFlow,还加入了对系统设计原理方面的剖析,有助于读者做针对性的应用和系统优化。相信本书对从事深度学习方面研究和开发的读者定会有所裨益。

查礼

中国科学院计算技术研究所副研究员
中国大数据技术大会(BDTC)发起人

序 二

如果说要评选出一个在过去一年中学术界和产业界最热门的词汇，无疑非“人工智能”莫属。作为新一轮产业变革的核心驱动力，人工智能正重构生产、分配、交换、消费等经济活动各环节，形成从宏观到微观各领域的智能化新需求、新产品、新技术、新业态，引发经济结构重大演化，实现社会生产力的整体跃升。

人工智能系统的智能有三种来源。一种是依靠人类设计者的知识输入，为系统建立人工特征、知识库和推理机制，传统意义上的专家系统属于这一范畴。二是通过数据驱动的归纳式学习，近年来大火的深度学习即属于这一类。与依赖于人工经验、通过手工构建的知识特征不同，深度学习以端到端的方式进行特征学习，其基本动机在于构建多层网络来学习隐含在数据内部的关系，从而使学习得到的特征具有更强、更泛化的表达能力。三是智能体通过与环境交互，学习经验和知识并更新知识表示。近年来，以深度学习为代表的驱动方法在图像识别、语音识别、机器翻译、自然语言理解等任务中取得了一系列突破。人脸识别、自然语言理解在一系列国际评测中展示了超越人类能力的水平，语言识别和机器翻译也达到了一个前所未有的高度。

在我看来，各行各业还会在相当长的一段时间内享受到基于大数据的深度学习红利。将深度学习红利释放到读者所在的行业是提升行业智能水平的一条捷径。而要做到这一点，核心关键在于要降低大数据深度学习技术的使用门槛。要使看起来高深的深度学习技术，早日达到“旧时王谢堂前燕，飞入寻常百姓家”的程度，开源软件生态社区会起到巨大的推动作用，TensorFlow 则无疑正是在这方面的佼佼者。深度学习对于张量计算性能、算子灵活性、自动微分能力、分布式训练、自动调参数、可视化和端侧部署等都有很强的诉求，而 TensorFlow 的设计也充分考虑到了这些因素，这使得它成为了当前业界很流行的一个深度学习引擎。

TensorFlow 还是一个较新的技术，但是发展极为迅猛，这时候出现一本深入浅出讲解 TensorFlow 理论与应用的图书，对于希望学习和应用大数据深度学习技术的广大读者而言，诚“如大旱之望云霓”。本书理论与实践并重，理论上讲清楚了一些本质的东西，并加入了作者对系统设计原理方面的深刻理解，进而通过实际案例，引导读者掌握针对性的系统优化的技能。

本书第一作者是我的学生，2012 年入学时进入了浙大计算机专业的尖子班“求是科学班”，我担任了他们这个班的班主任。他不仅品学兼优，而且作为班上的团支书，帮我这个不太称职的班主任做了很多班级工作。在我心目中，他依然是入学时的青涩模样，转眼间却已成为开源软件

界的技术翘楚，作为老师，最欣慰的莫过于此了吧。是为序。

陈刚

教育部“长江学者”特聘教授

浙江大学计算机学院院长

前言

缘起

2016 年的某个中午，我在知乎上回答了名为“如何高效学习 TensorFlow 代码？”的问题，其中简单介绍了我在 TensorFlow 开源社区的贡献，以及 TensorFlow 的学习路线和方法。此回答引起了一些圈内人的共鸣。人民邮电出版社图灵公司的编辑也在第一时间找到我，希望我能够写一本 TensorFlow 相关的图书。于是，便有了你手中的这本书。

为什么写本书

在 ImageNet 的带动下，深度学习的研究热潮已席卷全球。随着 AlphaGo 的横空出世，资本市场对人工智能的产业化也表现出了空前的兴趣。2017 年 7 月，《国务院关于印发新一代人工智能发展规划的通知》的出台，标志着国家层面对人工智能发展的高度重视，明确了我国新一代人工智能发展的战略目标。未来 10 年，我们将见证人工智能全面升级改造传统行业。在这场深刻变革中，TensorFlow 将迸发出巨大的能量。

TensorFlow 推出短短一个月，就成为了机器学习和深度学习项目中最受欢迎的开源框架。究其原因，离不开 Google 在人工智能与数据处理领域的深厚积淀及其在业界的强大号召力。TensorFlow 自 2015 年 11 月开源以来，已经发布了 30 多个版本。尽管 TensorFlow 整个生态系统是开源的，但由于它版本升级过快，且算子种类众多，大部分公司，尤其是一些中小型或创业公司，难以在有限的时间内快速掌握 TensorFlow 的设计思想和基本原理。TensorFlow 开源的 PS-worker 分布式架构也在快速迭代，与其内部基于 Borg 调度的分布式架构并不相同。自 TensorFlow 开源以来，不断有人撰写图书或博客解释其各组件的实现与原理。但遗憾的是，能够深入剖析 TensorFlow 内部实现细节与设计思想的资料少之又少，而本书则尝试弥补这一缺憾。

本书以 TensorFlow 1.2 为基础，从基本概念、内部实现和最佳实践等方面深入剖析了 TensorFlow。书中重点阐述了以数据流图为核心的机器学习编程框架的设计原则与核心实现，并且介绍了 TensorFlow 生态系统中的两大重要工具：TensorBoard 可视化工具与 TensorFlow Serving 模型托管工具。同时，本书还将 TensorFlow 与深度学习相结合，从理论基础和程序实现这两个方面系统介绍了卷积神经网络（CNN）、生成对抗网络（GAN）和循环神经网络（RNN）等经典模型。本书不仅由浅入深地全面介绍了 TensorFlow 的使用方法，而且结合源代码进行了深入

剖析,使读者可以快速、系统地学习 TensorFlow 的架构设计与实现原理。

读者对象

本书的读者主要包含以下人员。

- **TensorFlow 二次开发人员。**由于在高效性、多平台、多语言、稳定性等方面的诸多优点, TensorFlow 已被国内外越来越多的公司采用并部署到生产环境。而为了解决特定场景下的特定问题,大部分公司选择在开源 TensorFlow 的基础上进行二次开发。通过这本书,这部分人员可以深入而又全面地了解 TensorFlow 的设计原则和实现细节,这是修改 TensorFlow 内核的前提。
- **数据科学家和算法工程师。**如果要使用 TensorFlow 解决生产和生活中的实际问题,仅掌握 TensorFlow 基本使用方法是远远不够的,还必须对 TensorFlow 的设计理念、架构和运作机制有一定了解。尤其是对于分布式训练任务,更需要深入了解 TensorFlow 分布式的架构设计与多种并行模式的实现原理。对于这部分读者来说,本书将带领他们走入 TensorFlow 架构师的内心世界,使其系统、深入地理解 TensorFlow 和数据流图,提高开发水平,从而编写出更加高效的深度学习和机器学习模型。
- **人工智能方向的研究生。**对于一名人工智能专业的研究生来说,除了需要具备扎实的人工智能理论功底外,还应当熟练掌握一种算法模型编程框架,才能将研究课题中的问题快速落实到实际的代码上来。而 TensorFlow 便是当下最受欢迎的机器学习和深度学习框架。通过阅读本书,人工智能方向的研究生可以全面提升复现论文实验结果和开发全新模型的效率,并深入理解 TensorFlow 的设计思想和实现细节。
- **开源软件爱好者。**TensorFlow 是全世界最受欢迎的开源机器学习和深度学习框架,它在设计和实现过程中参考了 Google 第一代分布式机器学习框架 DistBelief 的实践经验,同时又加入了很多值得学习的创新。本书分析 TensorFlow 架构设计和实现原理的方式也许值得许多开源软件爱好者学习和借鉴,这部分读者不仅能够领略到开源软件的优秀设计,还可以掌握分析开源软件源代码的方法和技巧,从而进一步提高使用开源软件的效率和质量。

如何阅读本书

本书分为五大部分(不包括附录)。

第一部分为基础篇(第 1~3 章),简单介绍了 TensorFlow 设计目标、基本架构、环境准备和基础概念,包括数据流图的设计与使用,以及 TensorFlow 运行环境和训练机制,帮助读者快速入门 TensorFlow,迅速上手使用。

第二部分为关键模块篇(第 4~7 章),着重讲解了使用 TensorFlow 端到端解决人工智能问题涉及的关键模块,包括数据处理、编程框架、可视化工具和模型托管工具,帮助读者进一步提

升开发效率，快速落地模型应用。

第三部分为算法模型篇（第 8 ~ 11 章），在读者熟练掌握 TensorFlow 后，该部分将深度学习与 TensorFlow 有机结合，系统介绍了深度学习的发展历史与应用场景，并结合理论与代码实现深入讲解了 CNN、GAN 和 RNN 等经典模型。

第四部分为核心揭秘篇（第 12 ~ 14 章），深入剖析了 TensorFlow 运行时核心、通信原理和数据流图计算的原理与实现，聚焦 C++ 核心层的揭秘，帮助读者进一步理解 TensorFlow 底层的设计思想与实现细节，TensorFlow 二次开发人员需重点关注这部分内容。

第五部分为生态发展篇（第 15 章），全面介绍了 TensorFlow 生态系统发展，并重点介绍了 Keras 深度学习算法库，以及 TensorFlow 与云原生社区 Kubernetes 生态的结合、与大数据社区 Spark 生态的结合，并介绍了 TensorFlow 通信优化技术、TPU 及 NNVM 模块化深度学习技术，帮助读者进一步全面了解深度学习生态发展的现状。

勘误和支持

感谢您在茫茫书海中选择了我们的作品。尽管我们在写作过程中力求以精确的语言传达正确的信息，然而由于专业水平与表达能力的限制，书中难免仍有不周之处。深度学习相关理论与技术演进飞速，书中部分内容在呈现于读者之时或许已不合时宜。我们诚挚地希望你能够指正本书存在的问题，并给我们提出宝贵的意见和建议。

读者可以通过图灵社区的本书主页（<http://www.ituring.com.cn/book/2397>）提交或查看勘误，亦可通过本书专用邮箱 i@use.tf 反馈你的意见和建议。同时，为了帮助读者和众多 TensorFlow 爱好者能够更方便地进行交流和讨论，我们筹建的 TensorFlow 中文社区（<http://use.tf>）也将上线。

致谢

感谢我的合作者林健博士，他在计算机系统方面学识渊博，使我在合作著书的过程中不断进步。同时，林健博士对本书的审稿和校稿工作也做出了重要贡献。

感谢我的另一位合作者白小龙博士，他对深度学习理论和落地的丰富经验使我受益匪浅。

感谢人民邮电出版社图灵公司的王军花编辑在这一年多的时间中始终支持我们的写作，她的鼓励和帮助使我们顺利完成了本书。

感谢丁泽震、辛现银、张震宇、李鹏飞、尉建等朋友在写书和校稿过程中给我们提供的修改建议与各种帮助。

谨以此书献给我们最亲爱的家人，以及众多热爱 TensorFlow 和深度学习的朋友们！

彭靖田
于杭州

作者简介

彭靖田

才云科技技术总监，才云人工智能云平台 Clever 负责人。谷歌机器学习开发专家 (ML Google Developer Experts)。Kubeflow Core Maintainer，作为创始团队与 Google 一起打造云原生机器学习社区——Kubeflow。TensorFlow Contributor，曾一度成为 TensorFlow 社区全球前 40 的贡献者。加州大学圣迭戈分校访问学者，毕业于浙江大学竺可桢学院求是科学班。曾为华为公司深度学习团队核心成员，主要参与华为深度学习平台的设计和研发工作。从高中开始研究算法，并通过 OI 获奖保送至浙江大学，兼具算法工程师、系统工程师和云计算工程师的多重视角，拥有整合人工智能平台、框架和算法模型的丰富经验。

林健

华为公司高级软件工程师。在中国科学院计算技术研究所取得博士学位，并在美国俄亥俄州立大学做过博士后研究。长期从事系统层软件研发，工作方向涉及高性能计算、分布式系统与大数据处理，爱好开源软件与人工智能技术。曾参与开发国家科学计算基础设施——中国国家网格 CNGrid 的核心系统软件 GOS，以及工业界广泛使用的高性能 MPI 通信库——MVAPICH2。以核心开发者身份合作创建开源云服务系统 LingCloud 和开源数据计算库 DataMPI。设计实现教学型中国象棋智能算法引擎 BitStronger，并多次参加计算机棋类博弈赛事。参与翻译《C++ API 设计》《C++探秘：68 讲贯通 C++》等技术图书。

白小龙

华为公司深度学习云服务的技术负责人，主要负责深度学习平台、模型和算法的研发。于 2015 年 6 月毕业于浙江大学并取得工学博士学位，曾获教育部博士生学术新人奖。长期致力于研究面向工业应用的信号处理、图像处理和机器学习算法。曾提出基于图像显著性的大规模缺陷目标快速检测算法用于工业智能检测，并与新加坡国立大学合作研发基于医疗图像的青光眼特征选择和多标签学习算法及软件。相关论文发表于 *IEEE Transactions on Industrial Informatics*、*Computer Methods and Programs in Biomedicine* 等 SCI 期刊。

目 录

第一部分 基础篇

第 1 章 TensorFlow 系统概述	2	2.2.3 Eigen 线性代数计算库	27
1.1 简介	2	2.2.4 CUDA 统一计算设备架构	28
1.1.1 产生背景	2	2.3 源代码结构	29
1.1.2 独特价值	3	2.3.1 根目录	29
1.1.3 版本变迁	4	2.3.2 tensorflow 目录	30
1.1.4 与其他主流深度学习框架的对比	6	2.3.3 tensorflow/core 目录	31
1.2 设计目标	7	2.3.4 tensorflow/python 目录	32
1.2.1 灵活通用的深度学习库	8	2.3.5 安装目录	33
1.2.2 端云结合的人工智能引擎	9	2.4 小结	33
1.2.3 高性能的基础平台软件	10	第 3 章 TensorFlow 基础概念	34
1.3 基本架构	12	3.1 编程范式：数据流图	34
1.3.1 工作形态	12	3.1.1 声明式编程与命令式编程	34
1.3.2 组件结构	13	3.1.2 声明式编程在深度学习应用上的优势	35
1.4 小结	14	3.1.3 TensorFlow 数据流图的基本概念	38
第 2 章 TensorFlow 环境准备	15	3.2 数据载体：张量	40
2.1 安装	15	3.2.1 张量：Tensor	40
2.1.1 TensorFlow 安装概述	15	3.2.2 稀疏张量：SparseTensor	44
2.1.2 使用 Anaconda 安装	17	3.3 模型载体：操作	46
2.1.3 使用原生 pip 安装	17	3.3.1 计算节点：Operation	46
2.1.4 使用 virtualenv 安装	18	3.3.2 存储节点：Variable	49
2.1.5 使用 Docker 安装	19	3.3.3 数据节点：Placeholder	53
2.1.6 使用源代码编译安装	20	3.4 运行环境：会话	55
2.1.7 Hello TensorFlow	22	3.4.1 普通会话：Session	55
2.2 依赖项	23	3.4.2 交互式会话：InteractiveSession	59
2.2.1 Bazel 软件构建工具	24	3.4.3 扩展阅读：会话实现原理	59
2.2.2 Protocol Buffers 数据结构序列化工具	25	3.5 训练工具：优化器	61
		3.5.1 损失函数与优化算法	61
		3.5.2 优化器概述	64

3.5.3 使用 minimize 方法训练模型	66
3.5.4 扩展阅读: 模型训练方法进阶	68
3.6 一元线性回归模型的最佳实践	72
3.7 小结	76

第二部分 关键模块篇

第4章 TensorFlow 数据处理方法 78

4.1 输入数据集	78
4.1.1 使用输入流水线并行读取数据	78
4.1.2 创建批样例数据的方法	86
4.1.3 填充数据节点的方法	87
4.1.4 处理 CIFAR-10 数据集的最佳实践	88
4.1.5 扩展阅读: MNIST 数据集	91
4.2 模型参数	92
4.2.1 模型参数的典型使用流程	92
4.2.2 使用 tf.Variable 创建、初始化和更新模型参数	92
4.2.3 使用 tf.train.Saver 保存和恢复模型参数	98
4.2.4 使用变量作用域处理复杂模型	100
4.3 命令行参数	103
4.3.1 使用 argparse 解析命令行参数	103
4.3.2 使用 tf.app.flags 解析命令行参数	108
4.4 小结	111

第5章 TensorFlow 编程框架 112

5.1 单机程序编程框架	112
5.1.1 概述	112
5.1.2 创建单机数据流图	114
5.1.3 创建并运行单机会话	116
5.2 分布式程序编程框架	118
5.2.1 PS-worker 架构概述	118
5.2.2 分布式程序编程框架概述	120
5.2.3 创建 TensorFlow 集群	121
5.2.4 将操作放置到目标设备	124
5.2.5 数据并行模式	124

5.2.6 同步训练机制	125
5.2.7 异步训练机制	130
5.2.8 使用 Supervisor 管理模型训练	131
5.2.9 分布式同步训练的最佳实践	133
5.3 小结	137

第6章 TensorBoard 可视化工具 138

6.1 概述	138
6.2 可视化数据流图	142
6.2.1 名字作用域与抽象节点	142
6.2.2 可视化数据流图的最佳实践	144
6.2.3 扩展阅读: 汇总数据和事件数据	145
6.2.4 扩展阅读: 揭秘 tf.summary.FileWriter 工作原理	147
6.3 可视化学习过程	149
6.3.1 汇总操作概述	149
6.3.2 使用 tf.summary.scalar 生成折线图	150
6.3.3 使用 tf.summary.histogram 生成数据分布图	152
6.3.4 使用 tf.summary.image 生成图像	154
6.3.5 使用 tf.summary.audio 生成音频	155
6.3.6 可视化 MNIST softmax 模型学习过程的最佳实践	156
6.4 可视化高维数据	158
6.4.1 使用 TensorBoard 可视化高维数据	158
6.4.2 可视化 MNIST 数据集的最佳实践	160
6.5 小结	163

第7章 模型托管工具: TensorFlow Serving 164

7.1 概述	164
7.2 系统架构	165
7.3 安装	167
7.3.1 使用 APT 安装 ModelServer	168

7.3.2 使用源码编译安装 ModelServer	169
7.4 最佳实践	170
7.4.1 导出模型	170
7.4.2 发布模型服务	173
7.4.3 更新线上模型服务	174
7.5 小结	175

第三部分 算法模型篇

第8章 深度学习概述

8.1 深度学习的历史	178
8.1.1 感知机模型与神经网络	178
8.1.2 神经网络的寒冬与复苏	179
8.1.3 神经网络的发展与第二次 寒冬	181
8.1.4 深度学习时代的到来	183
8.2 深度学习的主要应用	184
8.2.1 计算机视觉	185
8.2.2 自然语言处理	186
8.2.3 强化学习	188
8.3 深度学习与 TensorFlow	190
8.4 小结	191

第9章 CNN 模型

9.1 CNN	192
9.1.1 CNN 简介	192
9.1.2 卷积层	193
9.1.3 激活层	195
9.1.4 池化层	195
9.1.5 全连接层	196
9.1.6 Dropout 层	196
9.1.7 BN 层	197
9.1.8 常用的 CNN 图像分类模型	197
9.2 TensorFlow-Slim	204
9.2.1 TensorFlow-Slim 总体结构	204
9.2.2 datasets 包和 data 包	205
9.2.3 preprocessing 包	207
9.2.4 deployment 包	207
9.2.5 nets 包	209

9.2.6 TensorFlow-Slim 最佳实践	212
9.3 应用	216
9.3.1 物体检测	216
9.3.2 图像分割	221
9.4 小结	222

第10章 GAN 模型

10.1 原理、特点及应用	223
10.1.1 原理	224
10.1.2 特点	225
10.1.3 应用	226
10.2 GAN 模型的改进	228
10.2.1 CGAN 模型	228
10.2.2 LAPGAN 模型	229
10.2.3 DCGAN 模型	230
10.2.4 InfoGAN 模型	230
10.2.5 LSGAN 模型	231
10.2.6 WGAN 模型	232
10.3 最佳实践	233
10.4 小结	238

第11章 RNN 模型

11.1 基本 RNN 单元及其变种	239
11.1.1 RNN 模型简介	239
11.1.2 基本 RNN 单元	240
11.1.3 LSTM 单元	242
11.1.4 GRU 单元	243
11.1.5 双向 RNN 单元	244
11.1.6 带有其他特性的 RNN 单元	245
11.2 RNN 模型	247
11.2.1 PTB-LSTM 语言模型	247
11.2.2 Seq2Seq 模型	251
11.3 小结	254

第四部分 核心揭秘篇

第12章 TensorFlow 运行时核心 设计与实现

12.1 运行时框架概述	256
12.2 关键数据结构	257

第 13 章 通信原理与实现.....27714.4 分布式会话运行..... 315

第五部分 生态发展篇

第 15 章 TensorFlow 生态环境 330附录 A 354附录 A 354

第一部分

基础篇

- 第 1 章 TensorFlow 系统概述
- 第 2 章 TensorFlow 环境准备
- 第 3 章 TensorFlow 基础概念



人工智能和深度学习的热潮将 TensorFlow 推向了很高的地位，媒体的追捧和业界的宣传也为这一源自 Google 的开源软件增添了传奇的色彩。对于技术从业者或爱好者而言，我们初识 TensorFlow 时有必要拨开表象看本质。本章作为引子，首先从技术视角概括性地介绍 TensorFlow 的产生背景、独特价值、版本变迁，以及它与其他主流深度学习框架的异同。同时，本章从灵活通用性、异构支持性和性能高效性三个视角解析 TensorFlow 的设计目标，展示 TensorFlow 作为一款兼具深度学习库、人工智能引擎和基础平台软件身份的开源产品的优势所在。最后，我们将简单介绍 TensorFlow 的工作形态和组件结构，帮助读者快速建立对 TensorFlow 软件架构的第一印象。

1.1 简介

当今，人工智能领域最受欢迎的深度学习和机器学习框架非 Google 开源的 TensorFlow 莫属。本节我们将依次介绍 TensorFlow 的产生背景、独特价值和版本变迁，并横向对比目前主流机器学习和深度学习框架各自的特点和优劣。

1.1.1 产生背景

随着近年来深度学习模型在图像、视觉和语音领域不断取得突破，相关研究热潮持续高涨，开源深度学习框架出现百花齐放之势，其中具有代表性的框架包括 XGBoost、Theano、Torch、Caffe 和 MXNet 等。它们有的计算速度快，有的可移植性好，有的内存占用少，有的易于上手。在 Google 推出 TensorFlow 之前，大家仍处于“你方唱罢我登场”的百家争鸣状态。然而，Google 推出深度学习框架 TensorFlow 之后，江湖巨变。

2015 年 10 月，Google 旗下的 DeepMind 公司研发的 AlphaGo 击败樊麾，成为第一个无须让子即可在 19 路棋盘上击败职业围棋选手的电脑程序。这一壮举不仅打破了人工智能无法在围棋领域战胜人类顶尖棋手的诅咒，而且吸引了全球各界对人工智能研究的高度关注。2016 年 1 月，AlphaGo 的研究成果发表在知名学术期刊《自然》上，这一事件也将人工智能热潮推向了新的高度。2017 年 5 月，强化后的 AlphaGo 与世界围棋三甲之一的柯洁对弈，获得 3 : 0 全胜战绩。这