

移动平台深度神经网络实战：原理、架构与优化



[移动平台深度神经网络实战：原理、架构与优化 下载链接1](#)

著者:卢誉声

出版者:机械工业出版社

出版时间:2019-12-1

装帧:平装

isbn:9787111641001

本书精讲移动平台深度学习系统所需核心算法、硬件级指令集、系统设计与编程实战、海量数据处理、业界流行框架裁剪与产品级性能优化策略等，深入、翔实。

深度学习基础（第1~4章），介绍开发机器学习系统所需重要知识点，以及开发移动平台机器学习系统算法基石，诸如人工神经网络、稀疏自编码器、深度网络、卷积神经网络等。

移动平台深度学习基础（第5~6章），介绍移动平台开发环境搭建、移动平台开发基础、ARM指令集加速技术，以及轻量级网络的实现原理与实战。

深入理解深度学习（第7~8章），剖析数据预处理原理与方法，高性能实时处理系统开发，以及基于深度神经网络的物体检测与识别。本篇是下一篇文章的前导与“基石”。

深入理解移动平台深度学习（第9~12章），本篇应用前述章节的框架与技术，实现移动平台深度学习系统的实现与集成，具体涵盖：①

移动平台性能优化，数据采集与训练，为开发移动平台图像分类系统建立基础；②

深入剖析TensorFlow

Lite代码体系、构建原理、集成方法以及核心代码与裁剪分析，模型处理工具，并完成移动平台系统集成；③

结合实战分析主流移动平台机器学习框架、接口，并展望未来。

作者介绍:

卢誉声

Autodesk数据平台和计算平台资深工程师，负责平台架构研发工作。工作内容涵盖大规模分布式系统的服务器后端、前端以及SDK的设计与研发，在数据处理、实时计算、分布式系统设计与实现、性能调优、高可用性和自动化等方面积累了丰富的经验。擅长C/C++、JavaScript开发，此外对Scala、Java以及移动平台等也有一定研究。著有《移动平台深度神经网络实战：原理、架构与优化》、《分布式实时处理系统：原理架构与实现》，并译有《高级C/C++编译技术》和《JavaScript编程精解（原书第2版）》等。

目录: 序一

序二

前言

第一篇 深度学习基础

第1章 向未来问好 2

1.1 机器学习即正义 2

1.1.1 照本宣科 3

1.1.2 关键概念概述 4

1.1.3 数学之美 5

1.2 机器学习的场景和任务 6

1.3 机器学习算法 8

1.3.1 分类算法 8

1.3.2 回归算法 8

1.3.3 聚类算法 8

1.3.4 关联分析算法 9

1.3.5 集成算法 9

1.3.6 强化算法 10

1.4 如何掌握机器学习 10

1.4.1 学习曲线 10

1.4.2 技术栈 11

1.5 深度学习 12

1.5.1 深度学习的贡献 12

1.5.2 深度学习框架简介 13

1.5.3 安装使用深度学习框架 16

1.5.4 深度学习进展 22

1.6 走进移动世界的深度学习 25

1.6.1 移动平台机器学习概述 25

1.6.2 难度和挑战 26

1.7 本书框架 26

1.8 本章小结 27

第2章 机器学习基础	28
2.1 机器学习的主要任务	28
2.2 贝叶斯模型	29
2.3 Logistic回归	33
2.3.1 线性回归	33
2.3.2 几率与Logit	35
2.3.3 Logistic回归	38
2.3.4 背景溯源	39
2.3.5 实现Logistic回归	42
2.4 本章小结	44
第3章 人工神经网络	45
3.1 人工神经网络简介	45
3.2 基本结构与前向传播	46
3.2.1 神经元	46
3.2.2 连接与网络	47
3.2.3 神经网络向量化	48
3.2.4 前向传播	50
3.3 反向传播算法	50
3.4 实现前向神经网络	53
3.4.1 神经网络与前向传播实现	53
3.4.2 Softmax回归	60
3.5 稀疏自编码器	61
3.5.1 引子	61
3.5.2 自编码器简介	61
3.5.3 稀疏自编码算法	63
3.6 神经网络数据预处理	64
3.6.1 去均值	64
3.6.2 归一化	65
3.7 本章小结	65
第4章 深度网络与卷积神经网络	66
4.1 深度网络	66
4.1.1 自我学习	66
4.1.2 特征学习	67
4.1.3 深度神经网络	68
4.1.4 逐层贪婪训练方法	69
4.2 卷积神经网络	70
4.2.1 全连接与部分连接网络	70
4.2.2 卷积	70
4.2.3 池化	72
4.2.4 卷积神经网络	73
4.3 卷积神经网络实现	73
4.3.1 Layer实现	74
4.3.2 Net实现	79
4.3.3 InnerProduct实现	92
4.3.4 Convolution实现	95
4.3.5 Pooling实现	101
4.3.6 定义注册头文件	109
4.4 本章小结	110
第二篇 移动平台深度学习基础	
第5章 移动平台深度学习框架设计与实现	112
5.1 移动平台深度学习系统开发简介	112
5.2 ARM Linux基础开发环境	113
5.2.1 通用ARM工具链安装	114
5.2.2 Android NDK安装	114

5.2.3 树莓派工具链安装	115
5.3 TensorFlow Lite介绍	115
5.3.1 TensorFlow Lite特性	115
5.3.2 TensorFlow Lite架构	116
5.3.3 TensorFlow Lite代码结构	117
5.4 移动平台性能优化基础	118
5.4.1 ARM v8体系结构	119
5.4.2 ARM v8数据类型与寄存器	120
5.4.3 Neon指令集介绍	122
5.4.4 ARM v8内存模型	124
5.4.5 Neon指令集加速实例	127
5.5 本章小结	140
第6章 移动平台轻量级网络实战	141
6.1 适用于移动平台的轻量级网络	141
6.2 SqueezeNet	142
6.2.1 微观结构	142
6.2.2 宏观结构	142
6.2.3 核心思路	143
6.2.4 实战：用PyTorch实现SqueezeNet	144
6.3 MobileNet	153
6.4 ShuffleNet	154
6.5 MobileNet V2	155
6.5.1 MobileNet的缺陷	155
6.5.2 MobileNet V2的改进	155
6.5.3 网络结构	156
6.5.4 实战：用PyTorch实现MobileNet V2	157
6.6 本章小结	161
第三篇 深入理解深度学习	
第7章 高性能数据预处理实战	164
7.1 数据预处理任务	164
7.1.1 数据清理	165
7.1.2 数据集成	165
7.1.3 数据归约	165
7.1.4 数据变换	166
7.2 数据标准化	166
7.3 PCA	167
7.3.1 PCA的现实问题	167
7.3.2 PCA的计算方法	167
7.3.3 PCA的数学理论基础	169
7.4 在Hurricane之上实现PCA	170
7.4.1 Hurricane实时处理系统	171
7.4.2 实现Hurricane Topology	172
7.4.3 实现PCA	178
7.5 本章小结	192
第8章 基于深度神经网络的物体检测与识别	193
8.1 模式识别与物体识别	193
8.1.1 模式识别	193
8.1.2 模式识别系统	194
8.1.3 传统模式识别方法	194
8.1.4 深度学习模式识别方法	197
8.2 图像分类	197
8.2.1 LeNet	197
8.2.2 AlexNet	200
8.2.3 数据抓取整理	203

- 8.2.4 数据预处理 204
- 8.2.5 数据训练 206
- 8.3 目标识别与物体检测 207
 - 8.3.1 目标识别简介 207
 - 8.3.2 R-CNN 208
 - 8.3.3 SPP-Net 209
 - 8.3.4 Fast R-CNN 211
 - 8.3.5 Faster R-CNN 211
 - 8.3.6 RetinaNet 213
- 8.4 检测识别实战 213
 - 8.4.1 Faster R-CNN 214
 - 8.4.2 RetinaNet 230
- 8.5 移动平台检测识别实战 237
 - 8.5.1 移动平台系统开发思路 237
 - 8.5.2 基于RetinaNet的检测定位实现 237
 - 8.5.3 基于AlexNet的识别分类实现 244
 - 8.5.4 接口设计封装 247
- 8.6 本章小结 258
- 第四篇 深入理解移动平台深度学习
- 第9章 深入移动平台性能优化 260
 - 9.1 模型压缩 260
 - 9.2 权重稀疏化 262
 - 9.2.1 Structured Sparsity Learning 262
 - 9.2.2 Dynamic Network Surgery 262
 - 9.2.3 Dynamic Network Surgery实现 264
 - 9.3 模型加速 275
 - 9.3.1 半精度与权重量化 275
 - 9.3.2 深度压缩 276
 - 9.3.3 二值化网络 278
 - 9.3.4 三值化网络 280
 - 9.3.5 DoReFa-Net 282
 - 9.3.6 编程实战 283
 - 9.4 嵌入式优化 287
 - 9.4.1 算法局限与改进 287
 - 9.4.2 理论改进 287
 - 9.4.3 编程实战 288
 - 9.5 嵌入式优化代码实现 290
 - 9.5.1 量化分析实现 290
 - 9.5.2 层实现 302
 - 9.5.3 量化矩阵计算 309
 - 9.6 本章小结 313
- 第10章 数据采集与模型训练实战 314
 - 10.1 收集海量数据 314
 - 10.1.1 搜索引擎工作原理 315
 - 10.1.2 HTTP会话 316
 - 10.1.3 解决JavaScript渲染问题 316
 - 10.2 图片数据爬虫实现 317
 - 10.2.1 获取任务 318
 - 10.2.2 解析图片 320
 - 10.2.3 图片存储 326
 - 10.2.4 图片去重 327
 - 10.2.5 完成Topology 328
 - 10.3 训练与测试 330
 - 10.3.1 模型定义 330

10.3.2 训练	334
10.3.3 测试	342
10.3.4 封装	344
10.4 本章小结	345
第11章 移动和嵌入式平台引擎与工具实战	346
11.1 TensorFlow Lite构建	346
11.2 集成TensorFlow Lite	357
11.3 核心实现分析	358
11.3.1 解释器代码分析	358
11.3.2 图代码分析	373
11.3.3 操作符注册	381
11.3.4 操作符扩展实现	384
11.3.5 计算与优化模块	399
11.4 模型处理工具	407
11.5 本章小结	425
第12章 移动平台框架与接口实战	426
12.1 Core ML	426
12.1.1 准备数据和生成模型	427
12.1.2 App实战：引入Core ML实现	430
12.2 Android Neural Networks API	437
12.2.1 等等，Google还有一个ML Kit	437
12.2.2 NNAPI编程模型	437
12.2.3 创建网络与计算	439
12.2.4 JNI封装与调用	451
12.2.5 App实战：集成NNAPI	454
12.3 实战：实现Android图像分类器App	459
12.3.1 JNI封装	459
12.3.2 Java调用	474
12.4 未来之路	479
12.5 本章小结	480
• • • • •	(收起)

[移动平台深度神经网络实战：原理、架构与优化_下载链接1](#)

标签

好书，值得一读

深度学习

AI

计算机-机器学习

编程

程序设计

专业书

计算机

评论

公司的产品准备在移动设备上落地，发现了这本书，实在是及时雨，对知识的覆盖和编程实用性方面都可圈可点，对我帮助很大。

很实用的工具书，少了很多理论

现在移动端应用深度学习技术的趋势太明显了，大势所趋啊。买来这本书读了一下，对我帮助比较大，主要是因为对知识体系的梳理方面恰到好处，而且结合实例代码，这一点我很喜欢。

一直在找结合移动和神经网络的书，终于有了，初步翻了下，由浅入深，层层递减，有理论有实践，是我要的书。

别的不说，光是对ARM指令集的解读以及应用，就值得购买，市面上这类书籍比较少，新出这一本值得购买。

这本书对模型优化主题进行了深入探讨，我正好需要这方面的指南，正好看到就买了。

水军太对，书内容一般，移动端优化原理介绍的太浅

从采集数据、处理数据到使用数据，并结合成熟的移动平台机器学习框架来实现一个可以工作的机器学习产品

一直在找结合移动和神经网络的书，终于有了，初步翻了下，由浅入深，层层递减，有理论有实践，是我要的书。

[移动平台深度神经网络实战：原理、架构与优化_下载链接1](#)

书评

[移动平台深度神经网络实战：原理、架构与优化_下载链接1](#)