HZ BOOKS
华章IT

智能系统与技术丛书

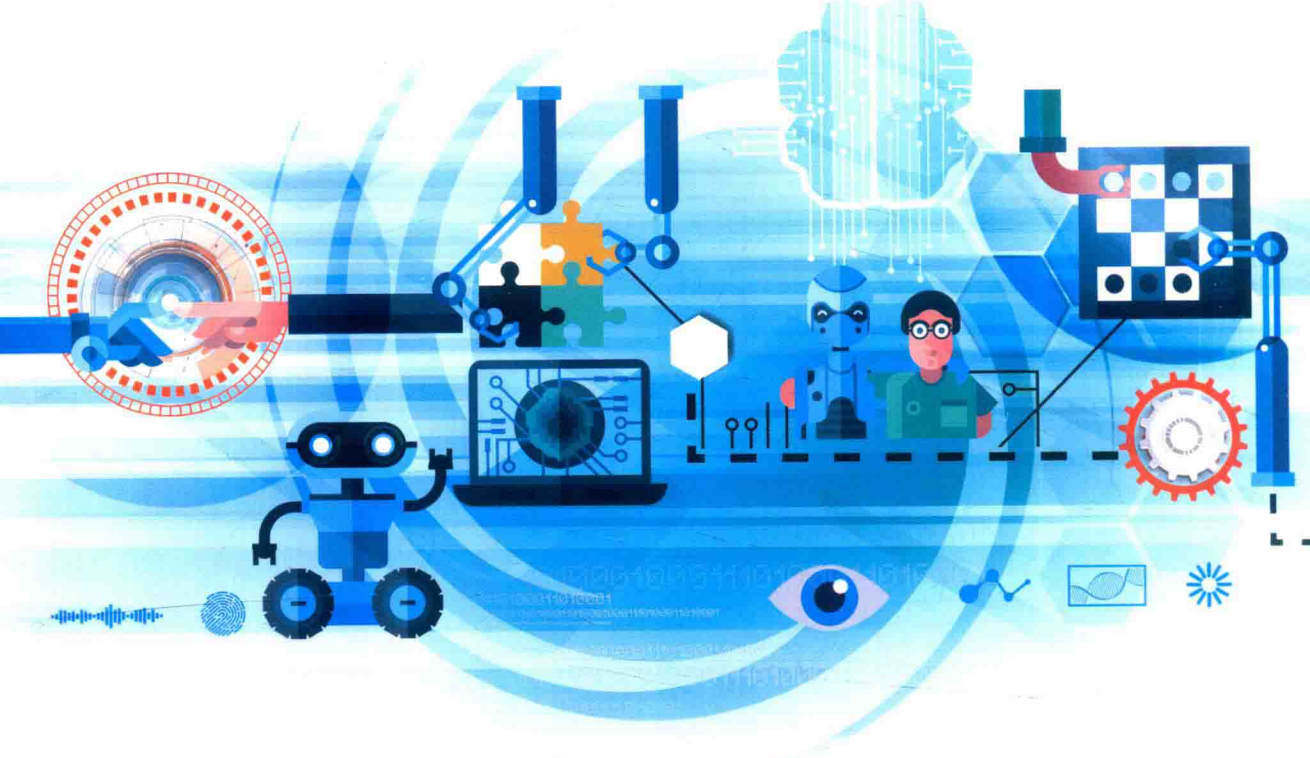
从技术原理、算法和工程实践3个维度系统讲解图像识别
阿里巴巴达摩院算法专家、阿里巴巴技术发展专家、阿里巴巴数据架构师联合撰写

彩色印刷

Deep Learning and Image Recognition
Principle and Practice

深度学习与图像识别
原理与实践

魏溪含 涂铭 张修鹏 著



机械工业出版社
China Machine Press

智能系统与技术丛书

Deep Learning and Image Recognition
Principle and Practice

深度学习与图像识别

原理与实践

魏溪含 涂铭 张修鹏 著

机械工业出版社
China Machine Press

图书在版编目（CIP）数据

深度学习与图像识别：原理与实践 / 魏溪含，涂铭，张修鹏著 . —北京：机械工业出版社，2019.6
（智能系统与技术丛书）

ISBN 978-7-111-63003-6

I. 深… II. ①魏… ②涂… ③张… III. 人工智能 - 算法 - 应用 - 图像识别 - 教材 IV. TP391.413

中国版本图书馆 CIP 数据核字（2019）第 122633 号

深度学习与图像识别：原理与实践

出版发行：机械工业出版社（北京市西城区百万庄大街 22 号　邮政编码：100037）
责任编辑：张锡鹏　　责任校对：殷　虹
印　　刷：中国电影出版社印刷厂　　版　　次：2019 年 7 月第 1 版第 1 次印刷
开　　本：186mm × 240mm　1/16　　印　　张：17.25
书　　号：ISBN 978-7-111-63003-6　　定　　价：129.00 元

凡购本书，如有缺页、倒页、脱页，由本社发行部调换
客服热线：（010）88379426　88361066　　投稿热线：（010）88379604
购书热线：（010）68326294　　读者信箱：hzit@hzbook.com

内容简介

这是一部从技术原理、算法和工程实践3个维度系统讲解图像识别的著作，由阿里巴巴达摩院算法专家、阿里巴巴技术发展专家、阿里巴巴数据架构师联合撰写。

在知识点的选择上，本书广度和深度兼顾，既能让完全没有基础的读者迅速入门，又能让有基础的读者深入掌握图像识别的核心技术；在写作方式上，本书避开了复杂的数学公式及其推导，从问题的前因后果 、创造者的思考过程角度展开，利用简单的数学计算来做模型分析和讲解，通俗易懂。更重要的是，本书不仅聚焦于技术，更是将重点放在了如何用技术解决实际的业务问题。

全书一共13章：

第1~2章主要介绍了图像识别的应用场景、工具和工作环境的搭建；

第3~6章详细讲解了图像分类算法、机器学习、神经网络、误差反向传播等图像识别的基础技术及其原理；

第7章讲解了如何利用PyTorch来实现神经网络的图像分类，专注于实操，是从基础向高阶的过渡；

第8~12章深入讲解了图像识别的核心技术及其原理，包括卷积神经网络、目标检测、分割、产生式模型、神经网络可视化等主题；

第13章从工程实践的角度讲解了图像识别算法的部署模式。

作者简介

魏溪含 爱丁堡大学人工智能硕士，阿里巴巴达摩院算法专家，在计算机视觉、大数据领域有8年以上的算法架构和研发经验。

在大数据领域，曾带领团队对阿里巴巴个性化推荐系统进行升级；在计算机视觉领域，主导并攻克了光伏EL全自动瑕疵识别的世界难题，并在行为识别领域带领团队参赛打破世界纪录等。

涂铭 资深阿里巴巴数据架构师，对大数据、自然语言处理、图像识别、Python、Java相关技术有深入的研究，积累了丰富的实践经验。在工业领域曾参与燃煤优化、设备故障诊断项目，以及正泰光伏电池片和组件EL图像检测项目；在自然语言处理方面，担任导购机器人项目的架构师，主导开发机器人的语义理解、短文本相似度匹配、上下文理解，以及通过自然语言检索产品库，在项目中构建了NoSQL+文本检索等大数据架构，也同时负责问答对的整理和商品属性的提取，带领NLP团队构建语义解析层。

张修鹏 毕业于中南大学，阿里巴巴技术发展专家，长期从事云计算、大数据、人工智能与物联网技术的商业化应用，在阿里巴巴首次将图像识别技术引入工业，并推动图像识别产品化、平台化，擅于整合前沿技术解决产业问题，主导多个大数据和AI为核心的数字化转型项目成功实施，对技术和商业结合有着深刻的理解。

投稿通道
联系人：杨福川 邮箱：yfc@hzbook.com 微信：15693352

封面设计 锡彬

华章IT
HZBOOKS | Information Technology

PREFACE

前　言

为什么要写这本书

随着深度学习技术的发展、计算能力的提升和视觉数据的增长，视觉智能计算技术在许多应用领域如拍照搜索、智能相册、人脸闸机、城市智能交通管理、智慧医疗等都取得了令人瞩目的成绩。因此越来越多的人开始对机器视觉感兴趣，并开始从事这个行业。就图像识别领域来说，运行一个开源的代码并不是什么难事，但搞懂其中的原理确实会稍有些难度。因此本书在每章中都会用相对通俗的语言来介绍算法的背景和原理，并会在读者“似懂非懂”时给出实战案例。实战案例的代码已全部在线下运行通过，代码并不复杂，可以很好地帮助读者理解其中的细节，希望读者在学习理论之后可以亲自动手实践。图像识别的理论和实践是相辅相成的，希望本书可以带领读者走进图像识别的世界。

本书从章节规划到具体的讲述方式，具有以下两个特点：

第一个特点是本书的主要目标读者定位为高校相关专业的本科生（统计学、计算机技术）、图像识别爱好者，以及不具备专业数学知识的人群。图像识别是一系列学科的集合体，它以机器学习、模式识别等知识为基础，因此依赖很多数学知识。本书尽量绕开复杂的数学证明和推导，从问题的前因后果、创造者思考的过程和简单的数学计算的角度来做模型的分析和讲解，目的是以更通俗易懂的方式带领读者入门。另外，在第 8～12 章的后面都附有参考文献，想要深入了解的读者可以继续阅读。

第二个特点是本书在每章后面都附有实战案例，读者可以结合案例学习，通过实践验证自己想法的价值。在本书的内容编排上，遵循知识点背景介绍——原理剖析——实战案例的介绍方式，同时所有的代码会在书中详细列出或者上传到 GitHub，以方便读者下载与调试，帮助读者快速掌握知识点，快速上手，而且这些代码也可以应用到后续实际的开发项目中。在实际项目章节中，选取目前在图像识别领域中比较热门的项目，对之前的知识点进行汇总，帮助读者巩固与提升。

读者对象

❑ 统计学或相关 IT 专业学生

本书的初衷是面向相关专业的学生——拥有大量基于理论知识的认知却缺乏实战经验的人员，让其在理论的基础上深入了解。通过本书，学生可以跟随本书的教程一起操作学习，达到对自己使用的人工智能工具、算法和技术知其然亦知其所以然的目的。

❑ 信息科学和计算机科学爱好者

本书是一本近现代科技的历史书，也是一本科普书，还是一本人工智能思想和技术的教科书。通过本书可以了解人工智能领域的前辈们在探索的道路上做出的努力和思考，理解他们不同的观点和思路，有助于开拓自己的思维和视野。

❑ 人工智能相关专业的研究人员

本书详细介绍了图像识别的相关知识。通过本书可以了解其理论知识，了解哪些才是项目所需的内容以及如何在项目中实现，能够快速上手。

如何阅读本书

本书从以下几个方面阐述图像识别：

第 1 章介绍图像识别的一些应用场景，让读者对图像识别有个初步的认识。

第 2 章主要对图像识别的工程背景做简单介绍，同时介绍了本书后续章节实战案例中会用到的环境，因此该章是实战的基础。

第 3～6 章是图像识别的技术基础，包括机器学习、神经网络等。该部分的代码主要使用 Python 实现。没有机器学习基础的同学需要理解这几章之后再往下看，有机器学习基础的同学可以有选择地学习。

第 7 章是一个过渡章节，虽然第 6 章中手动用 Python 实现了神经网络，但由于本书后面的图像识别部分主要使用 PyTorch 实现，因此使用该章作为过渡，介绍如何使用 PyTorch 来搭建神经网络。

第 8～12 章为图像识别的核心。第 8 章首先介绍了图像中的卷积神经网络与普通神经网络的异同，并给出了常见的卷积神经网络结构。接下来的第 9～12 章分别介绍了图像识别中的检测、分割、产生式模型以及可视化的问题，并在每章后面给出相应的实战案例。

第 13 章简单介绍了图像识别的工业部署模式，以帮助读者构建一个更完整的知识体系。

第 8～12 章包含参考文献，主要是本书中介绍的一些方法，或者本书中提到但是没有深入说明的方法，感兴趣的读者可以自行查询学习。

关于附件的使用方法：除了第 1 章外，本书的每一章都有对应的源数据和完整代码，这些内容可在本书中直接找到，有些代码需要从 GitHub 中下载，地址为 https://github.com/

image_recognition/learning-recognition。需要注意的是，为了让读者更好地了解每行代码的含义，在注释信息中使用了中文标注，每个程序文件的编码格式都是 UTF-8。

勘误和支持

由于本书的作者水平及撰稿时间有限，书中难免会出现一些错误或者不准确的地方，恳请读者批评指正。读者可通过发送电子邮件到 weixihan1@163.com 和 kenny_tm@hotmail.com 联系并反馈建议或意见。

致谢

首先非常感谢我的家人，由于业余时间常常被工作挤占，本书的撰写又用了所剩不多的业余时间，因此少了很多陪伴家人的时间，感谢他们的理解、支持和鼓励。

撰写一本书，将自己的知识重新梳理后分享给读者，在技术发展的道路上帮助到其他人，这件事情是非常有价值的，因此也非常感谢两位合著者涂铭、张修鹏。

感谢机械工业出版社华章公司的杨福川老师，以及全程参与审核、校验等工作的张锡鹏、孙海亮老师等出版工作者，是他们的辛勤付出才能保证本书顺利面世。

感谢我身边的朋友、同事、同学，感谢一路走来你们的支持、鼓励和帮助。

谨以此书献给热爱算法并为之奋斗的朋友们，愿大家身体健康、生活美满、事业有成！

魏溪含

书籍初成，感慨良多。

在接受邀请撰写该书时，从未想到过程如此艰辛与波折。这里需要感谢一路陪我走来的所有人。

感谢我的家人的理解和支持，陪伴我度过写作本书的漫长岁月。

感谢我的合写者——魏溪含和张修鹏，与他们合作轻松愉快，他们给予我很多的理解和包容。

感谢参与审阅、校验等工作的杨福川老师以及其他老师，是他们在幕后的辛勤付出保证了本书的成功出版。

另外在本书的写作期间，有很多专业领域的内容都得到了各个领域专家的指导甚至亲笔编著。这里需要特别感谢阿里云计算公司产品方面的专家李骏，编写了第 13 章全部内容，感谢他在产品和技术上利用其丰富的行业经验为本书留下的宝贵财富。

再次感谢大家！

涂　铭

首先要感谢我的妻子金晖，我能在工作繁忙的情况下参与此书的编写，离不开她的付出和支持，感谢我的宝贝张正延，给了我无穷的动力，感谢我的父亲、母亲，永远深爱你们。

感谢魏溪含和涂铭！魏溪含在书中贡献了她图像识别领域多年的经验，涂铭为此书的出版付出了最多的心血。

这本书是友谊和工作成果的结晶，本书作为我们并肩奋斗的见证，希望能将我们实践经验沉淀成的知识，帮助到更多希望了解和学习深度学习与图像识别的读者。

感谢杨福川等机械工业出版社的老师们，他们在幕后的付出和支持，是本书得以出版的保障。

最后感谢这些年一路走来帮助过我的亲人、老师、朋友、同事、同学，始终满怀感恩！

张修鹏

目　录

01 CHAPTER

第 1 章
机器视觉在行业中的应用

本章将介绍机器视觉的发展背景，而后针对机器视觉的主要应用场景做一个简单的介绍，带领读者了解机器视觉都能应用在哪些领域、解决哪些问题。

1.1 机器视觉的发展背景

1.1.1 人工智能

人工智能（Artificial Intelligence，AI）是计算机科学的一个分支，其意在了解智能的实质，并生产出一种新的能以人类智能相似的方式做出反应的智能机器。该领域的研究包括机器人、语言识别、机器视觉、自然语言处理和专家系统等。

那么，人们常说的人工智能、机器学习、深度学习的关系是什么呢。如图 1-1 所示，人工智能是一个比较大的领域，其中包括机器学习、深度学习、模式识别等，而神经网络是机器学习中的一种方法，深度学习又是神经网络方法中的一个子集。

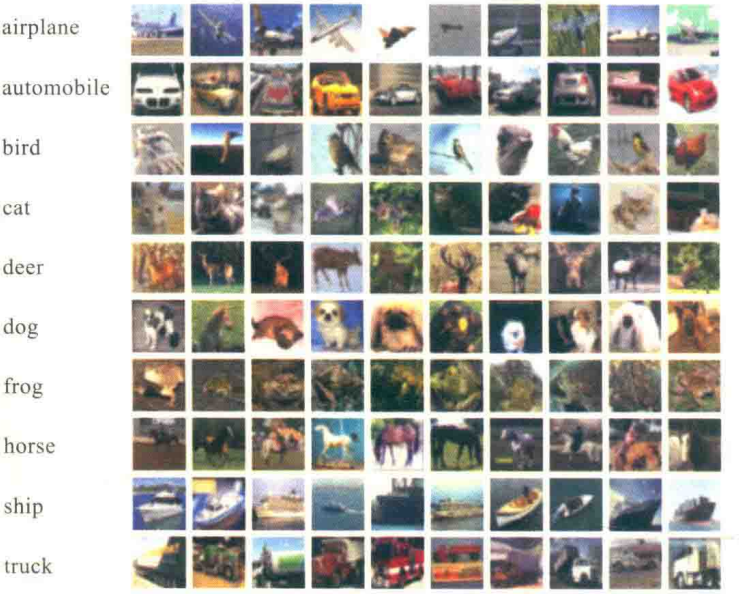
图 1-1 人工智能相关领域关系图

历史上人工智能经历了三次“春天”。人工智能的概念于20世纪50年代被首次提出，当时人们觉得人工智能在20年之内会改变世界，所有的工作都会被人工智能颠覆。直到1973年的《莱特希尔报告》明确指出当时人工智能的任何部分都没有达到人们想象的水平，第一个“春天”随之结束。第二个“春天”是20世纪80年代，神经网络和反向传播算法的提出，以及专家系统的初步结果，让科学家和企业家再次看到了希望。但因为普通神经网络不可避免的问题以及专家系统的局限，第二次热浪也逐渐冷却。现在，随着深度学习技术的崛起，人工智能正迎来第三个“春天”。

1.1.2　机器视觉

机器视觉是人工智能的一个重要分支，其核心是使用“机器眼”来代替人眼。机器视觉系统通过图像/视频采集装置，将采集到的图像/视频输入到视觉算法中进行计算，最终得到人类需要的信息。这里提到的视觉算法有很多种，例如，传统的图像处理方法以及近些年的深度学习方法等。

对于人工智能的一个重要研究方向——机器视觉来说，这个春天与以往有什么不同呢，我们来看图1-2。图1-2a展示了一个由彩色图像组成的、分类的数据集Cifar10（第3章有详细介绍），其中有飞机、汽车、鸟、猫、鹿、狗、青蛙、马、船、卡车10个类别，且每个类别中都有1000张32×32的彩色图片。图1-2b展示的是不同算法在Cifar10数据集上的分类效果。从中我们可以看出，在深度学习出现以前，传统的图像处理和机器学习方法并不能很好地完成这样一个简单的分类任务，而深度学习的出现使得机器有了达到人类水平的可能。事实上，AlphaGo的出现已经证明了在一些领域，机器有了超越人类的能力。



a）Cifar10数据集展示

图1-2　人工智能的第三个“春天”

b）传统图像处理方法与深度学习方法在 Cifar10 数据集上的效果对比

图 1-2 （续）

1.2　机器视觉的主要应用场景

由于深度学习技术的发展、计算能力的提升和视觉数据的增长，视觉智能计算技术在不少应用当中都取得了令人瞩目的成绩。图像视频的识别、检测、分割、生成、超分辨、captioning、搜索等经典和新生的问题纷纷取得了不小的突破。这些技术正广泛应用于城市治理、金融、工业、互联网等领域。本节将以 9 个场景为例，对一些常见的应用场景进行介绍，让读者直观地理解机器视觉都能解决哪些问题。

1.2.1　人脸识别

人脸识别（Face Recognition）是基于人的面部特征信息进行身份识别的一种生物识别技术。它通过采集含有人脸的图片或视频流，并在图片中自动检测和跟踪人脸，进而对检测到的人脸进行面部识别。人脸识别可提供图像或视频中的人脸检测定位、人脸属性识别、人脸比对、活体检测等功能。

人脸识别是机器视觉最成熟、最热门的领域，近几年，人脸识别已经逐步超过指纹识别成为生物识别的主导技术。人脸识别分为 4 个处理过程——人脸图像采集及检测、人脸图像预处理、人脸图像特征提取以及匹配与识别，其主要应用场景如表 1-1 所示。

表 1-1　人脸识别的主要应用场景

应 用 场 景	说　　明
人脸支付	将人脸与用户的支付渠道绑定，支付阶段即可刷脸付款，无须出示银行卡、手机等，提高支付效率（如图 1-3）
人脸开卡	客户在银行等部门开卡时，可通过身份证和人脸识别进行身份校验，以防止借用身份证进行开卡