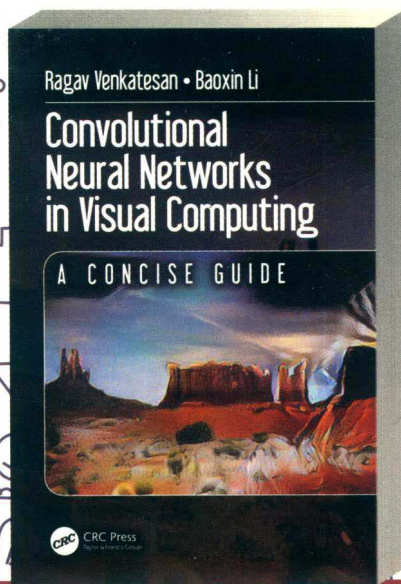


# Convolutional Neural Networks in Visual Computing

## 卷积神经网络 与视觉计算

[美] 拉加夫·维凯特森 (Ragav Venkatesan) 著  
李宝新 (Baixin Li)

钱亚冠 王滨 等译



Convolutional Neural  
Networks in Visual Computing

# 卷积神经网络 与视觉计算

[美] 拉加夫·维凯特森 (Ragav Venkatesan)

李宝新 (Baoxin Li)

◎ 著

钱亚冠 王滨 ◎ 等译



机械工业出版社  
China Machine Press

## 图书在版编目 ( CIP ) 数据

卷积神经网络与视觉计算 / (美) 拉加夫·维凯特森 (Ragav Venkatesan), (美) 李宝新著; 钱亚冠等译. —北京: 机械工业出版社, 2018.11

(智能科学与技术丛书)

书名原文: Convolutional Neural Networks in Visual Computing

ISBN 978-7-111-61239-1

I. 卷… II. ①拉… ②李… ③钱… III. 计算机视觉 - 人工神经网络 - 计算 IV. TP302.7

中国版本图书馆 CIP 数据核字 (2018) 第 250420 号

本书版权登记号: 图字 01-2018-6729

Convolutional Neural Networks in Visual Computing by Ragav Venkatesan and Baoxin Li (9781138747951).  
Copyright © 2018 by Taylor & Francis Group, LLC.

Authorized translation from the English language edition published by CRC Press, part of Taylor & Francis Group LLC. All rights reserved.

China Machine Press is authorized to publish and distribute exclusively the Chinese (Simplified Characters) language edition. This edition is authorized for sale in the People's Republic of China only (excluding Hong Kong, Macao SAR and Taiwan). No part of this publication may be reproduced or distributed in any form or by any means, or stored in a database or retrieval system, without the prior written permission of the publisher.

Copies of this book sold without a Taylor & Francis sticker on the cover are unauthorized and illegal.

本书原版由 Taylor & Francis 出版集团旗下 CRC 出版公司出版, 并授权翻译出版。版权所有, 侵权必究。

本书中文简体字翻译版授权由机械工业出版社独家出版并仅限在中华人民共和国境内 (不包括香港、澳门特别行政区及台湾地区) 销售。未经出版者书面许可, 不得以任何方式复制或抄袭本书的任何内容。

本书封面贴有 Taylor & Francis 公司防伪标签, 无标签者不得销售。

## 卷积神经网络与视觉计算

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: 陈佳媛

责任校对: 殷虹

印刷: 北京文昌阁彩色印刷有限责任公司

版次: 2019 年 1 月第 1 版第 1 次印刷

开本: 186mm × 240mm 1/16

印张: 11.5

书号: ISBN 978-7-111-61239-1

定价: 59.00 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) 88378991 88361066

投稿热线: (010) 88379604

购书热线: (010) 68326294 88379649 68995259

读者信箱: hzit@hzbook.com

版权所有·侵权必究

封底无防伪标均为盗版

本书法律顾问: 北京大成律师事务所 韩光 / 邹晓东

华章科技  
HZBOOKS | Science & Technology



感谢一同成长的 Jaikrishna Mohan，你是我忠实的朋友与兄弟；还有帮助我成长的 Ravi Naganathan 教授。你们的智慧哲学和人生信条使我更为出彩。

——Ragav Venkatesan

感谢我的爱人 Julie，感谢你对我长年坚定不移的支持。

——李宝新

## 译者序

Convolutional Neural Networks in Visual Computing

随着深度学习的兴起，卷积神经网络重新焕发出光彩，它被成功地应用于计算机视觉领域，并获得了前所未有的进步。不管是业界还是教育界，都希望有一本能快速学习这一技术的资料。但由于这一领域的发展很快，大多内容都散布在各种学术论文中，初学者要通读很多论文后才能掌握这些技术，这无疑是个令人痛苦的过程。在为实验室的研究生新生寻找一本合适的教材时，本人也同样感到万分苦恼。一个偶然的的机会，我在亚马逊上发现了本书的英文版。本书从基本的图像知识讲起，到机器学习的核心理论，再到神经网络和卷积神经网络及在计算机视觉中的最新应用，可以说形成了一个非常完整的知识体系。尽管这些内容不管哪一部分单独都可以成为一本教材，但本书的优势就在于用浅显的语言把这些貌似深奥的内容，通过合理的裁剪，围绕视觉计算把逻辑上相关的内容整合到了一起，免去了初学者到处翻阅相关书籍的麻烦，也有助于读者在图像处理、机器学习、深度学习、视觉计算上形成完整的认知。

感谢机械工业出版社华章公司给予翻译这本前沿著作的机会。本书的翻译也得到了浙江省自然科学基金项目（No. LY17F020011）的资助。参加本书翻译的有浙江科技学院大数据学院的钱亚冠、杭州海康威视网络与信息安全实验室的王滨，以及人工智能与机器学习实验室的研究生：卢红波、马骏、马丹峰、郭艳凯、刘新伟、张锡敏等。由于本书内容较新，翻译时间仓促，难免有不足之处，恳请读者批评指正！联系邮箱 QianYaGuan@zust.edu.cn。

钱亚冠

2018年6月于杭州

## Ragav Venkatesan

目前正在亚利桑那州立大学（ASU）的计算、信息学和决策系统工程学院攻读计算机科学博士学位。他是 ASU 的视觉表示和处理组的一名研究助理，曾担任机器学习、模式识别、视频处理和计算机视觉等多个研究生课程的助教。在此之前，他是 ASU 电子计算机工程学院图像处理与应用实验室的研究助理，2012 年获得了硕士学位。从 2013 年到 2014 年，他是英特尔公司的一名研究实习生，研究自动驾驶中的计算机视觉。他是机器学习和计算机视觉领域的几个期刊的同行评审和会议审稿人。

## Baoxin Li

2000 年获马里兰大学帕克学院电气工程博士学位，目前是亚利桑那州立大学计算机科学与工程专业的教授和主任，以及电气工程和计算机工程专业的研究生导师。从 2000 年到 2004 年，他任美国夏普实验室的高级研究员，在那里他领导研发了夏普的 HiIMPACT 体育技术。2003 年至 2004 年，他还是俄勒冈州波特兰州立大学的副教授。他拥有 18 项美国专利，目前的研究兴趣包括计算机视觉和模式识别、多媒体、社会计算、机器学习和辅助技术。2001 年和 2004 年，他获得夏普实验室总裁奖。2002 年，他获得了夏普实验室年度发明者奖，也是国家自然科学基金会职业奖的获得者。



近年来，深度学习体系结构由于在计算机视觉等应用中的极大成功而开始流行起来。特别是卷积神经网络（CNN）已经成为深度学习中最重要的一种网络结构。现在普遍认为深度学习和 CNN 的术语对于初学者来说，太难学习和掌握了。为此，本书涵盖了计算机视觉中的深度学习、设计和部署 CNN，以及深度计算机视觉体系结构的基础知识。本书旨在为工程师、本科生和研究生提供入门指南，帮助他们快速学习和建立自己的深度学习系统。本书以通俗易懂的方式写成，目的是为计算机视觉相关的深度学习技术提供一个易学的教材，同时仍然涵盖了大量必要的基础知识。

本书分为 5 章。第 1 章简要介绍了图像表示和一些计算机视觉模型，这些模型现在被称为人工方式建模。本章为读者提供了对图像表示的基本理解，并介绍了一些线性和非线性的特征提取或表示方法，以及这些表示的特性。此外，本章还介绍了一些基本图像元素（如边缘）的检测方法，还包括用这些表示来完成一些基本的机器学习任务。本章最后对两种流行的非神经网络计算机视觉建模技术进行了介绍。

第 2 章介绍了回归、机器学习和优化的概念。本章首先介绍监督学习，介绍的第一个学习模型是线性回归，它的第一种求解方法是利用最小二乘法得到解析解，并用极大似然来进行解释。本章通过基函数展开的方式实现线性模型向非线性模型的转变，并进一步介绍了交叉验证和正则化的方法解决过拟合和泛化问题。本章的后半部分介绍了凸误差面和非凸误差面的梯度下降最优化方法，进一步介绍了各种梯度下降法和正则化几何，并对基本梯度下降法进行了改进，包括二阶最小化损失技术和动量学习。



第 3 章和第 4 章是本书的核心。第 3 章在第 2 章的基础上，介绍了 Rosenblatt 感知器和感知器学习算法。本章接着介绍了 logistic 神经元及其激活函数，研究了两类和多类问题的单神经元模型，以及这种神经元的优缺点，并介绍了异或问题。针对异或问题，提出了多层神经网络的思想，介绍了反向传播算法，并对其进行了改进，给出了一些实用技巧，有助于工程上更好、更稳定的实现。第 4 章介绍了卷积池化层和 CNN。其中研究了该层的各种属性，并分析了典型的数字手写体数据集提取的特征。本章还介绍了目前最流行的四种 CNN：AlexeNet、VGG、GoogLeNet 和 ResNet，并对它们的架构和思想进行了比较。

第 5 章通过介绍 CNN 的一些当前新颖实用的用法，进一步拓展和丰富了深度神经网络的结构。本章大致分为两个连续的部分。第一部分讨论了使用网上可下载，并已预训练好的现成深度网络的原理。经过预训练的网络是在一个完整的数据集上进行训练得到的，公开提供给研究人员用于新的机器学习任务。这些都是在满足通用、可迁移的条件下研究的。第 5 章还研究了给定一个预训练网络的情况下，压缩网络的和学习新方法。第二部分论述了 CNN 不用于监督学习，而是用于生成网络的思想，简要研究了自编码器和最新的计算机视觉技术：生成对抗网络（GAN）。

本书的相关网站（convolution.network）包含代码及实现（也可登录 [www.hzbook.com](http://www.hzbook.com) 下载）、部分彩色插图、勘误表和补充资料。2017 年春季，本书作为亚利桑那州立大学的一门研究生课程的教材进行讲授，讲座和材料也可以在本书网站上找到。

图 1-1 是我拍摄的原始图像（original.jpg）。这是 monument 山谷的一幅图片，它是美国西南地区的代表，亚利桑那州也在那里。memory.png 是以 Salvador Dali 的风格绘制的，特别是他的绘画《记忆的持久性》，它抽象地描述了心灵幻觉的概念，以无形的形式描绘和处理物体，很像我们在书中研究的神经网络的一些表现

形式。

memory.png 不是人类所画，而是由本书介绍的神经网络绘制的。如果引用了本书，请使用以下的引文参考方式。

```
@article{DBLP:journals/corr/GatysEB15a,
  author    = {Leon A. Gatys and
              Alexander S. Ecker and
              Matthias Bethge},
  title     = {A Neural Algorithm of Artistic Style},
  journal   = {CoRR},
  volume    = {abs/1508.06576},
  year      = {2015},
  url       = {http://arxiv.org/abs/1508.06576},
  timestamp = {Wed, 07 Jun 2017 14:41:58 +0200},
  biburl    = {http://dblp.unitrier.de/rec/bib/
              journals/corr/GatysEB15a},
  bibsource = {dblp computer science bibliography,
              http://dblp.org}
}
```

本书配有一个基于 Python 和 Theano 的 CNN 工具箱（该工具箱是由本书作者开发的），还有一个网页包含彩色图片、勘误表和其他补充资料。这个神经网络工具箱被命名为 yann，可以在 MIT 许可证下使用，网址为 <http://www.yann.network>。为了让初学者更容易地阅读书中的内容，作者开发了一套使用 yann 的教程。该教程和工具箱涵盖了本书讨论的不同体系结构的计算机，并提供了示例代码和应用程序编程接口（API）文档。在撰写本书的时候，yann 工具箱正在积极开发中，它的客户支持是通过 GitHub 提供的。本书的网页在 <http://guide2cnn.com> 上。书中的大多数图片都是灰度图，但有些图片最初是彩色的，是在制作过程中被转换成了灰度图。这些图片的彩色版本以及附加说明、相关课程的信息以及问答也可在网站上找到。

工具箱和本书将作为亚利桑那州立大学一学期的“深入学习视觉计算”研究生课程的阅读材料。本课程包括录制的讲座、课程内容和作业，可供大家使用，

网址为 <http://www.course.convolution.network>。作者可以通过电子邮件回答大家对有关内容和代码的问题，虚心接受大家对本书内容的批评或评论。只要符合出版商的版权政策，作者允许复制书中的图片、结果和内容。作者希望读者喜欢这本书，并希望初学者能够在本书及工具箱的帮助下快速构建自己的学习机器。鼓励读者利用本书中的知识做有益于人类的事，但同时真诚地劝告他们不要建造“Skynet”或任何其他末日人工智能机器。

## 致 谢

Convolutional Neural Networks in Visual Computing

感谢许多同事，是他们使这个耗时的写书项目的完成变得可能并使我从中获得快乐。亚利桑那州立大学的视觉表示与处理组和认知与普适计算中心的许多当前和过去的研究成员都致力于深度学习的各个方面及其在视觉计算方面的应用。他们的努力为本书的写作提供了有见解的内容，对此我们表示非常感激。特别要感谢 Parag Sridhar Chandakkar 对第 4 章和第 5 章提出的建议，以及 Yuzhen Ding、Yikang Li、Vijetha Gattupalli 和 Hemanth Venkateswara，他们总能提供极为有益的观点。

这项工作源于美国国家自然科学基金、海军研究办公室、陆军研究办公室和诺基亚赞助的几个项目。本书的任何观点或结论仅代表作者本人，与赞助者无关。我们也感谢 NVIDIA 公司的支持，他们捐赠的 Tesla K40 GPU 为我们的研究工作提供了帮助。

感谢 Taylor & Francis 出版集团、CRC 出版公司，特别是执行编辑 Cindy Carelli 和 Renee Nakash，感谢他们在本书的写作过程中所给予的耐心和极大支持。我们还要感谢 Alex Krizhevsky 博士，他大度地允许我们使用 AlexNet 论文中的数据。我们还要进一步感谢 Theano 和其他 Python 库的开发人员，这些库被 yann 工具箱所使用，并用于制作本书中的一些图片。特别地，我们要感谢来自 Theano 用户小组的 Frédéric Bastien 和 Pascal Lamblin，以及蒙特利尔大学机器学习算法学院的支持。我们还想感谢 GitHub、Read the Docs 为我们提供免费的数据、代码、文档和教程的在线托管。

最后，也是最重要的，感谢我们的朋友和家人在这个项目中所给予的坚定支持，以及他们对我们占用许多周末和漫漫长夜的理解和宽容。我们把爱和这本书献给他们。

Ragav Venkatesan 和 Baoxin Li

# 目 录

Convolutional Neural Networks in Visual Computing

译者序

作者简介

前言

致谢

## 第 1 章 视觉计算简介 ..... 1

### 1.1 图像表示基础 ..... 3

#### 1.1.1 变换域表示 ..... 6

#### 1.1.2 图像的直方图 ..... 7

#### 1.1.3 图像梯度和边缘 ..... 10

#### 1.1.4 超越图像梯度 ..... 15

### 1.2 基于 Hough 变换的直线

#### 检测 ..... 16

### 1.3 Harris 角点 ..... 16

### 1.4 尺度不变的特征变换 ..... 17

### 1.5 方向梯度直方图 ..... 18

#### 1.5.1 人工设计特征空间中的

#### 决策制定 ..... 19

#### 1.5.2 贝叶斯决策 ..... 21

#### 1.5.3 线性决策边界 ..... 23

### 1.6 可变形零件模型的实例

#### 研究 ..... 25

### 1.7 计算机视觉向神经网络

#### 转变 ..... 27

#### 本章小结 ..... 30

#### 参考文献 ..... 30

## 第 2 章 回归问题中的机器

### 学习 ..... 34

#### 2.1 监督学习 ..... 34

#### 2.2 线性模型 ..... 38

#### 2.3 最小二乘法 ..... 40

#### 2.4 极大似然估计的解释 ..... 43

#### 2.5 扩展到非线性模型 ..... 45

#### 2.6 正则化 ..... 47

#### 2.7 交叉验证 ..... 50

#### 2.8 梯度下降 ..... 52

#### 2.9 几何正则化 ..... 58

#### 2.10 非凸误差面 ..... 60

#### 2.11 随机梯度、批梯度及在线

#### 梯度下降 ..... 61

#### 2.12 其他自适应学习率的更新

#### 规则 ..... 62

#### 2.13 动量 ..... 64

本章小结 .....	65	第 5 章 卷积神经网络的新	
参考文献 .....	66	进展 .....	121
<b>第 3 章 人工神经网络</b> .....	69	5.1 预训练网络 .....	122
3.1 感知器 .....	70	5.1.1 通用性和可传递性 .....	125
3.2 多层神经网络 .....	78	5.1.2 利用预训练网络的模型	
3.3 反向传播算法 .....	82	压缩 .....	130
3.4 改进的反向传播算法 .....	85	5.1.3 Mentee 网络与 FitNet ...	134
3.4.1 激活函数 .....	86	5.1.4 使用预训练网络的应用:	
3.4.2 权重剪枝 .....	88	使用 CNN 的图像美学 ...	136
3.4.3 批量标准化 .....	89	5.2 生成网络 .....	138
本章小结 .....	90	5.2.1 自动编码器 .....	139
参考文献 .....	90	5.2.2 生成对抗网络 .....	142
<b>第 4 章 卷积神经网络</b> .....	93	本章小结 .....	147
4.1 卷积与池化层 .....	94	参考文献 .....	148
4.2 卷积神经网络 .....	102	<b>附录 A Yann</b> .....	152
本章小结 .....	118	<b>后记</b> .....	164
参考文献 .....	119		



## 视觉计算简介

人们进行科学探索的目的是增强人类的能力。我们用火来烹饪食物，减轻了胃来加工食物的负担。这使得人类可消化的食物量增加，加速了文明的发展——这种文明水平是其他物种未达到的。轮子和交通工具的发明，可使我们旅行的速度大大加快。飞机的发明实现了人类遨游蓝天的梦想。人类发明和技术发展就是人类无止境地增强自己能力和探索未来的过程。

这些进步归功于人类的神经系统。人类错综复杂的神经系统具有思维、情感、推理、想象和哲学的能力。计算机视觉研究者偏向于人类视觉系统 (human vision)，但对我们来说，认知视觉系统 (cognitive-visual system) 是人类能力 (包括智力) 中最重要的部分。虽然人类视觉系统及其认知决策过程是很快速的，但并不是所有生物中最强大的，如敏锐性或夜视能力 (Thorpe 等人, 1996; Watamaniuk 和 Duchon, 1992)。人类观察世界局限于一段很窄的电磁波谱，而许多其他生物的视觉感知范围比我们广得多。同时，人类也更容易产生许多视觉缺陷，如近视。基于上述考虑，我们很自然地想要改善我们的视觉能力，弥补缺陷。

多个世纪以来，我们发明了很多工具去超越人类肉眼的观察极限，如望远镜、双筒镜、显微镜和放大镜，去观察更远或更小的物体。无线电、

红外线和 X 射线设备可使我们观察到可见光谱以外的范围。最近，利用干涉仪发现了引力波，使得人类可用引力天文学来观察世界。虽然这些设备增强了人类的视觉能力，但学者和哲学家们早已意识到眼睛只是一种成像工具，真正“看到”的是大脑。

从柏拉图、亚里士多德、查拉卡、欧几里德到达·芬奇都研究了眼睛是如何观察世界的问题，但只有 Hermann von Helmholtz 在 1867 年所著的《生理光学》中首次提出：眼睛只捕捉图像，而大脑才真正“看到”和识别图像中的物体。在他的书中，他基于达·芬奇早期的工作，提出了关于深度和色彩感知、运动感知的新理论。尽管一直以来人们通过各种方式进行研究，Helmholtz 则首次提出了无意识推理（unconscious inference）的概念：不是所有的想法、思维和决定都是大脑有意识地完成的。Helmholtz 注意到人类很容易受到视觉错觉的影响，引用了太阳围绕地球旋转这个著名的误解案例，而事实上是地平线在移动。使用这种类比，Helmholtz 提出在人类无意识的情况下，是大脑在理解眼睛看到的图像，并推断和理解正在观察的物体。这可能是第一次深入洞察神经视觉的研究。一些近代早期的科学家（如 Blakemore 和 Campbell）开始争论一些现在已知的事实：大脑中有神经元负责估计物体的大小和方位感（Blakemore 和 Campbell, 1969）。后来同时期的研究发现了人类视觉系统感知和识别颜色、形状、方向、深度，甚至物体的更复杂原理（Field 等人, 1993; McCollough, 1965; Campbell 和 Kulikowski, 1966; Burton, 1973）。

上述简短的历史叙述说明计算机视觉领域在丰富的人类技术发展中占有一席之地。本书重在简洁介绍现代计算机视觉技术，因为这项技术大多