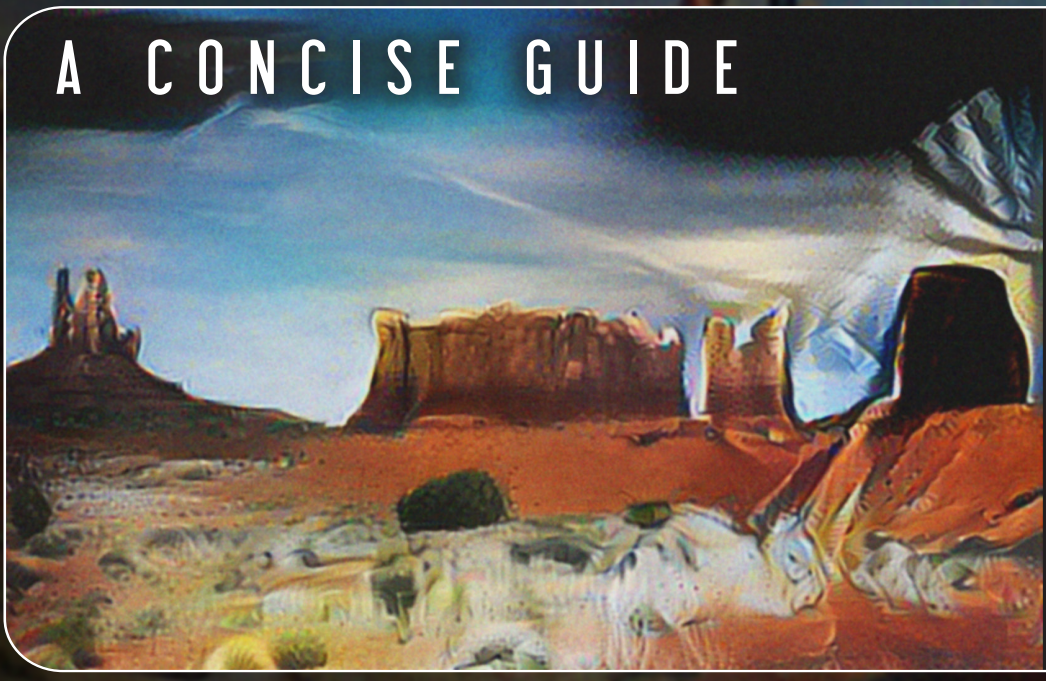
Data-Enabled Engineering Series

Ragav Venkatesan • Baoxin Li

Convolutional Neural Networks in Visual Computing

A CONCISE GUIDE



CRC Press
Taylor & Francis Group

Convolutional Neural Networks in Visual Computing

DATA-ENABLED ENGINEERING

SERIES EDITOR

Nong Ye
Arizona State University, Phoenix, USA

PUBLISHED TITLES

Convolutional Neural Networks in Visual Computing: A Concise Guide
Ragav Venkatesan and Baoxin Li

Convolutional Neural Networks in Visual Computing
A Concise Guide

By
Ragav Venkatesan and Baoxin Li

CRC Press
Taylor & Francis Group
Boca Raton London New York

CRC Press is an imprint of the
Taylor & Francis Group, an **informa** business

CRC Press
Taylor & Francis Group
6000 Broken Sound Parkway NW, Suite 300
Boca Raton, FL 33487-2742

Printed on acid-free paper

International Standard Book Number-13: 978-1-4987-7039-2 (Hardback); 978-1-138-74795-1 (Paperback)

Library of Congress Cataloging-in-Publication Data
Names: Venkatesan, Ragav, author. | Li, Baoxin, author.
Title: Convolutional neural networks in visual computing : a concise guide /
Ragav Venkatesan, Baoxin Li.
Description: Boca Raton ; London : Taylor & Francis, CRC Press, 2017. |
Includes bibliographical references and index.
Identifiers: LCCN 2017029154| ISBN 9781498770392 (hardback : alk. paper) |
ISBN 9781315154282 (ebook)
Subjects: LCSH: Computer vision. | Neural networks (Computer science)
Classification: LCC TA1634 .V37 2017 | DDC 006.3/2--dc23
LC record available at https://lccn.loc.gov/2017029154

Visit the Taylor & Francis Web site at
http://www.taylorandfrancis.com

and the CRC Press Web site at
http://www.crcpress.com

To Jaikrishna Mohan, for growing up with me;
you are a fierce friend, and my brother.
and to Prof. Ravi Naganathan for helping me grow up;
my better angels have always been your philosophy and principles.

—Ragav Venkatesan

To my wife, Julie,
for all your unwavering support over the years.

—Baoxin Li

Taylor & Francis
Taylor & Francis Group
http://taylorandfrancis.com

Contents

Taylor & Francis
Taylor & Francis Group
http://taylorandfrancis.com

Preface

Deep learning architectures have attained incredible popularity in recent years due to their phenomenal success in, among other applications, computer vision tasks. Particularly, convolutional neural networks (CNNs) have been a significant force contributing to state-of-the-art results. The jargon surrounding deep learning and CNNs can often lead to the opinion that it is too labyrinthine for a beginner to study and master. Having this in mind, this book covers the fundamentals of deep learning for computer vision, designing and deploying CNNs, and deep computer vision architecture. This concise book was intended to serve as a beginner's guide for engineers, undergraduate seniors, and graduate students who seek a quick start on learning and/or building deep learning systems of their own. Written in an easy-to-read, mathematically nonabstruse tone, this book aims to provide a gentle introduction to deep learning for computer vision, while still covering the basics in ample depth.

The core of this book is divided into five chapters. Chapter 1 provides a succinct introduction to image representations and some computer vision models that are contemporarily referred to as *hand-carved*. The chapter provides the reader with a fundamental understanding of image representations and an introduction to some linear and nonlinear feature extractors or representations and to properties of these representations. Onwards, this chapter also demonstrates detection

of some basic image entities such as edges. It also covers some basic machine learning tasks that can be performed using these representations. The chapter concludes with a study of two popular non-neural computer vision modeling techniques.

Chapter 2 introduces the concepts of regression, learning machines, and optimization. This chapter begins with an introduction to supervised learning. The first learning machine introduced is the linear regressor. The first solution covered is the analytical solution for least squares. This analytical solution is studied alongside its maximum-likelihood interpretation. The chapter moves on to nonlinear models through basis function expansion. The problem of overfitting and generalization through cross-validation and regularization is further introduced. The latter part of the chapter introduces optimization through gradient descent for both convex and nonconvex error surfaces. Further expanding our study with various types of gradient descent methods and the study of geometries of various regularizers, some modifications to the basic gradient descent method, including second-order loss minimization techniques and learning with momentum, are also presented.

Chapters 3 and 4 are the crux of this book. Chapter 3 builds on Chapter 2 by providing an introduction to the Rosenblatt perceptron and the perceptron learning algorithm. The chapter then introduces a logistic neuron and its activation. The single neuron model is studied in both a two-class and a multiclass setting. The advantages and drawbacks of this neuron are studied, and the XOR problem is introduced. The idea of a multilayer neural network is proposed as a solution to the XOR problem, and the backpropagation algorithm, introduced along with several improvements, provides some pragmatic tips that help in engineering a better, more stable implementation. Chapter 4 introduces the *convpool* layer and the CNN. It studies various properties of this layer and analyzes the features that are extracted for a typical digit recognition dataset. This chapter also introduces four of the most popular contemporary CNNs, AlexNet, VGG, GoogLeNet, and ResNet, and compares their architecture and philosophy.

Chapter 5 further expands and enriches the discussion of deep architectures by studying some modern, novel, and pragmatic uses of CNNs. The chapter is broadly divided into two contiguous sections. The first part deals with the nifty philosophy of using downloadable, pretrained, and off-the-shelf networks. Pretrained networks are essentially trained on a wholesome dataset and made available for the

public-at-large to *fine-tune* for a novel task. These are studied under the scope of generality and transferability. Chapter 5 also studies the compression of these networks and alternative methods of learning a new task given a pretrained network in the form of mentee networks. The second part of the chapter deals with the idea of CNNs that are not used in supervised learning but as generative networks. The section briefly studies autoencoders and the newest novelty in deep computer vision: generative adversarial networks (GANs).

The book comes with a website (convolution.network) which is a supplement and contains code and implementations, color illustrations of some figures, errata and additional materials. This book also led to a graduate level course that was taught in the Spring of 2017 at Arizona State University, lectures and materials for which are also available at the book website.

Figure 1 in Chapter 1 of the book is an original image (original.jpg), that I shot and for which I hold the rights. It is a picture of the monument valley, which as far as imagery goes is representative of the southwest, where ASU is. The art in memory.png was painted in the style of Salvador Dali, particularly of his painting "the persistence of memory" which deals in abstract about the concept of the mind hallucinating and picturing and processing objects in shapeless forms, much like what some representations of the neural networks we study in the book are.

The art in memory.png is not painted by a human but by a neural network similar to the ones we discuss in the book. Ergo the connection to the book. Below is the citation reference.

```
@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}
}
```

This book is also accompanied by a CNN toolbox based on Python and Theano, which was developed by the authors, and a webpage containing color figures, errata, and other accompaniments. The toolbox, named *yann* for "Yet Another Neural Network" toolbox, is available under MIT License at the URL http://www.yann.network. Having in mind the intention of making the material in this book easily accessible for a beginner to build upon, the authors have developed a set of tutorials using yann. The tutorial and the toolbox cover the different architectures and machines discussed in this book with examples and sample code and application programming interface (API) documentation. The *yann* toolbox is under active development at the time of writing this book, and its customer support is provided through GitHub. The book's webpage is hosted at http://guide2cnn.com. While most figures in this book were created as grayscale illustrations, there are some figures that were originally created in color and converted to grayscale during production. The color versions of these figures as well as additional notes, information on related courses, and FAQs are also found on the website.

This toolbox and this book are also intended to be reading material for a semester-long graduate-level course on Deep Learning for Visual Computing offered by the authors at Arizona State University. The course, including recorded lectures, course materials and homework assignments, are available for the public at large at http://www.course.convolution.network. The authors are available via e-mail for both queries regarding the material and supporting code, and for humbly accepting any criticisms or comments on the content of the book. The authors also gladly encourage requests for reproduction of figures, results, and materials described in this book, as long as they conform to the copyright policies of the publisher. The authors hope that readers enjoy this concise guide to convolutional neural networks for computer vision and that a beginner will be able to quickly build his/her own learning machines with the help of this book and its toolbox. We encourage readers to use the knowledge they may gain from this material for the good of humanity while sincerely discouraging them from building "Skynet" or any other apocalyptic artificial intelligence machines.

Acknowledgments

It is a pleasure to acknowledge many colleagues who have made this time-consuming book project possible and enjoyable. Many current and past members of the Visual Representation and Processing Group and the Center for Cognitive and Ubiquitous Computing at Arizona State University have worked on various aspects of deep learning and its applications in visual computing. Their efforts have supplied ingredients for insightful discussion related to the writing of this book, and thus are greatly appreciated. Particularly, we would like to thank Parag Sridhar Chandakkar for providing comments on Chapters 4 and 5, as well as Yuzhen Ding, Yikang Li, Vijetha Gattupalli, and Hemanth Venkateswara for always being available for discussions.

This work stemmed from efforts in several projects sponsored by the National Science Foundation, the Office of Naval Research, the Army Research Office, and Nokia, whose support is greatly appreciated, although any views/conclusions in this book are solely of the authors and do not necessarily reflect those of the sponsors. We also gratefully acknowledge the support of NVIDIA Corporation with the donation of the Tesla K40 GPU, which has been used in our research.

We are grateful to CRC Press, Taylor and Francis Publications, and in particular, to Cindy Carelli, Executive Editor, and Renee Nakash, for their patience and incredible support throughout the writing of this book. We would also like to thank Dr. Alex Krizhevsky for

gracefully giving us permission to use figures from the AlexNet paper. We would further like to acknowledge the developers of Theano and other Python libraries that are used by the yann toolbox and are used in the production of some of the figures in this book. In particular, we would like to thank Frédéric Bastien and Pascal Lamblin from the Theano users group and Montreal Institute of Machine Learning Algorithms of the Université de Montréal for the incredible customer support. We would also like to thank GitHub and Read the Docs for free online hosting of data, code, documentation, and tutorials.

Last, but foremost, we thank our friends and families for their unwavering support during this fun project and for their understanding and tolerance of many weekends and long nights spent on this book by the authors. We dedicate this book to them, with love.

Ragav Venkatesan and Baoxin Li

Authors

Ragav Venkatesan is currently completing his PhD study in computer science in the School of Computing, Informatics and Decision Systems Engineering at Arizona State University (ASU), Tempe, Arizona. He has been a research associate with the Visual Representation and Processing Group at ASU and has worked as a teaching assistant for several graduate-level courses in machine learning, pattern recognition, video processing, and computer vision. Prior to this, he was a research assistant with the Image Processing and Applications Lab in the School of Electrical & Computer Engineering at ASU, where he obtained an MS degree in 2012. From 2013 to 2014, Venkatesan was with the Intel Corporation as a computer vision research intern working on technologies for autonomous vehicles. Venkatesan regularly serves as a reviewer for several peer-reviewed journals and conferences in machine learning and computer vision.

Baoxin Li received his PhD in electrical engineering from the University of Maryland, College Park, in 2000. He is currently a professor and chair of the Computer Science and Engineering program and a graduate faculty in the Electrical Engineering and Computer Engineering programs at Arizona State University, Tempe, Arizona. From 2000 to 2004, he was a senior researcher with SHARP Laboratories of America, Camas, Washington, where he was a

technical lead in developing SHARP's trademarked HiMPACT Sports technologies. From 2003 to 2004, he was also an adjunct professor with Portland State University, Oregon. He holds 18 issued US patents and his current research interests include computer vision and pattern recognition, multimedia, social computing, machine learning, and assistive technologies. He won SHARP Laboratories' President's Award in 2001 and 2004. He also received the SHARP Laboratories' Inventor of the Year Award in 2002. He is a recipient of the National Science Foundation's CAREER Award.