Biological and Computer Vision

Imagine a world where machines can see and understand the world the way humans do. Rapid progress in artificial intelligence has led to smartphones that recognize faces, cars that detect pedestrians, and algorithms that suggest diagnoses from clinical images, among many other applications. The success of computer vision is founded on a deep understanding of the neural circuits in the brain responsible for visual processing. This book introduces the neuroscientific study of neuronal computations in the visual cortex alongside of the psychological understanding of visual cognition and the burgeoning field of biologically inspired artificial intelligence. Topics include the neurophysiological investigation of the visual cortex, visual illusions, visual disorders, deep convolutional neural networks, machine learning, and generative adversarial networks, among others. It is an ideal resource for students and researchers looking to build bridges across different approaches to studying and developing visual systems.

Gabriel Kreiman is a professor at the Center for Brains, Minds, and Machines and the Children's Hospital at Harvard Medical School. He is the recipient of the New Innovator Award from the National Institute of Health (NIH), National Science Foundation (NSF) Career Award, Pisart Award for Vision Research, and the McKnight Foundation Research Award.

Biological and Computer Vision

GABRIEL KREIMAN
Harvard University

CAMBRIDGE
UNIVERSITY PRESS

CAMBRIDGE
UNIVERSITY PRESS

University Printing House, Cambridge CB2 8BS, United Kingdom

One Liberty Plaza, 20th Floor, New York, NY 10006, USA

477 Williamstown Road, Port Melbourne, VIC 3207, Australia

314–321, 3rd Floor, Plot 3, Splendor Forum, Jasola District Centre, New Delhi – 110025, India

79 Anson Road, #06–04/06, Singapore 079906

Cambridge University Press is part of the University of Cambridge.

It furthers the University's mission by disseminating knowledge in the pursuit of education, learning, and research at the highest international levels of excellence.

www.cambridge.org
Information on this title: www.cambridge.org/9781108483438
DOI: 10.1017/9781108649995

First published 2021

A catalogue record for this publication is available from the British Library.

ISBN 978-1-108-48343-8 Hardback
ISBN 978-1-108-70500-4 Paperback

To all my teachers – current, past, and future

Contents

Figures

Preface

I have enjoyed the privilege of teaching a class on biological and computer vision for undergraduate and graduate students for almost a decade now. The class consists of 10 lectures and a vigorous discussion of seminal or state-of-the-art literature in the field. During this time, a large cadre of extraordinary students took the class; juggled academic, social, and athletic commitments; won fellowships, broke up with their boyfriends and girlfriends or married their sweethearts; joined laboratories for research; pursued internships in computer vision industries; got their first academic rejections; published their first papers, and eventually graduated. After graduation, many of the students continued carrying the torch of research in vision; others became philosophers; some created start-up companies; some became stars in the field; some decided to go after other career options.

Throughout the years, interacting with the students invigorated me, got me thinking, made me revisit some of my preconceptions, and encouraged me to explain things in different ways. This book is a consequence of all of those interactions, all of those iterations, and all of those discussions. It is a story that I write carrying the voices, hopes, puzzles, and questions of all those students.

I had a broad audience in mind while writing these pages. Students have joined our class from a wide spectrum of majors (called concentrations in our jargon). The two modes of the distribution are neurobiology and computer science. However, we also had philosophers, physicists, mathematicians, molecular biologists, psychologists, economists, historians, electrical engineers, and statisticians. We also had a student from public health and a talented film concentrator, either because they were lost the first day of classes or perhaps on purpose. It is my hope that computer scientists, mathematicians, and physicists will be excited to learn about advances in our understanding of visual brains and behavior. It is also my hope that neuroscientists, biologists, and psychologists will be excited to learn about how to train computers to interpret the visual world.

The purpose of this book is not to provide an exhaustive discussion of every possible aspect of vision. The number of scholarly publications on vision is enormous. There have already been more vision studies reported during the last two decades than the total cumulative number in the previous two millennia. The goal of this book is to whet the student's appetite for research in vision.

I am afraid that I cannot do justice to all the exciting work in the field of vision. Each of the topics covered here could well deserve a book of its own. Indeed, there are entire books devoted to the topics in Chapter 2, like John Dowling's *The Retina*. Entire books discuss the visual behavior topics in Chapter 3, such as Dale Purves's *Why We See What We Do*. Even a single one of the visual deficit conditions introduced in Chapter 4 has been expanded into a whole book (Martha Farah's *Visual Agnosia*). Tomaso Poggio teaches an excellent class covering the topics in Chapter 8. I could go on listing other great specialized books that expand on the material of the other chapters as well.

Instead of discussing the different approaches to vision in isolation, my objective in this book is to build bridges, to connect the exquisite recent advances in the study of the neural circuits underlying vision, and the fast-paced developments in computer vision.

The last decade has seen the beginning of a revolution in the field of vision. We now have tools that allow us to examine brains at unprecedented resolution. We can begin to build detailed connectomes describing at high resolution who talks to whom in neuronal hardware. We can simultaneously listen to the activity of hundreds or even thousands of individual neurons. We can turn specific circuits on and off in a reversible manner through impressive new techniques introduced by Ed Boyden and Karl Deisseroth. At the same time, for the first time in the short but exciting history of computer vision, we have algorithms that work quite well in a variety of pattern-recognition tasks. Computational models of vision are rapidly becoming a standard tool for experimentalists. Furthermore, the neurobiological insights capture the imagination of many computational experts to build ever more sophisticated models that provide a reasonable initial approximation to visual behavior and visual neurophysiology.

As a consequence of the vast literature and my modest didactic goal of making links across different areas of vision, many important and exciting topics are left out of this book. I apologize in advance for the major omissions. One enormous topic that is mostly ignored here is vision across the animal kingdom. I have focused on work in rodents, cats, and especially in monkeys and humans at the expense of elegant work in many other species such as flies, salamanders, and locusts. There is a bewildering and fascinating complexity in the diversity of visual systems, and I barely scratched the surface of this richness in Chapter 3. I am convinced that we must study many different species to understand how neural circuits for visual processing work, and we have much to learn from those other species.

Another topic that occupies a substantial literature that has been left out involves noninvasive studies of the human brain: this body of work is so vast that one could probably write entire books about those studies. However, the techniques are evolving so rapidly that I suspect that my notes would be obsolete pretty soon, and new and better measurements will replace the old ones in the near future. Another aspect of human vision omitted here is the clinical practice in ophthalmology. When I tell laypeople that I study vision, they immediately think about the eyes, corrective lenses, cataracts, and glaucoma. Yet another branch of human vision deals with aesthetics and art. My amazing colleague, Margaret Livingstone, has written a delightful book connecting what we know about visual cortex (here discussed in Chapters 5 and 6) to the depiction and interpretation of art (*Vision and Art: The Biology of Seeing*). I would like to encourage readers to venture into all of those other aspects of vision. The work described here provides a foundation to dig deeper into more specialized work in the field.

I have deliberately tried to establish as many links between the different chapters as possible. I am particularly excited about connecting biological and computational circuits. Biological vision is the product of millions of years of evolution. There is no reason to reinvent the wheel when developing computational models. We can learn from how biology solves vision problems and use the solutions as inspiration to build better algorithms. The converse is also true. Ingenious developments in computer vision can help guide us toward what to look for in neural circuits and how to model the complex interactions between neurons.

The sequence of chapters follows the trajectory that I have led students through in my class. This narrative starts with a discussion of the primary constraints in vision, and a definition of the types of problems that the visual system needs to solve (Chapters 1–3), moving onto how neurobiology tackles these challenges (Chapters 2–6), and finally putting it all together through computational models (Chapters 7–9). Although I find this order to be didactically effective, readers and teachers may prefer to create their own curriculum and follow alternative routes.

I conclude the book with a chapter that attempts to connect biology, computer science, and philosophy, discussing a particularly mysterious aspect of vision: conscious experience. I inherited the passion for this problem from my Ph.D. mentor, Christof Koch. Readers interested in consciousness are encouraged to study Prof. Koch's multiple treatises on the topic, particularly *The Quest for Consciousness* and *Confessions of a Romantic Reductionist*.

Another consequence of the rich literature in the field is that one could easily fill out the pages with a long list of references. I struggled with deciding whether to include citations for every statement or not. As a scientist, I am accustomed to justifying every assertion, either by showing data or citing the relevant sources. As a teacher, I was concerned that the references would disrupt the flow in the text and scare students away. While I have not studied this phenomenon carefully, it is my impression that the number of cited works that readers consult is inversely proportional to the number of references. Therefore, I have strictly cut the list of references to only five per chapter. These five references are not meant to represent an exhaustive discussion of work in the field by any means. I tried to mix in historic seminal papers with other interesting and relevant recent work. In the web material accompanying this book, I provide a more extensive list of references for each chapter: (http://klab.tch.harvard.edu/publications/Books/BiologicalAndComputerVision/TableOfContents.html). Even though the web provides a more extensive reference list, it is also by no means intended to be comprehensive in any way. Instead, the references are intended as an invitation to dig deeper into the topics presented in each chapter.

The web material also includes links to relevant video material for each chapter. There are many cases where videos can convey material in ways that are not possible on the printed page. Furthermore, the younger generations are particularly fond of learning from videos. If readers find any educational material that is relevant to the content of this book, or have other comments about the book, I would appreciate receiving feedback to add it potentially. I cannot guarantee that I will add all the material, in part, because of the purported inverse relationship between the number of links and whether people follow through or not. However, I can certainly guarantee that I will take the suggestions very seriously. In this way, these pages open the doors to a discussion on vision, an invitation to commence a dialogue with the readers.

In this dialogue, I did not shy away from discussing controversial topics in the field or outlining speculative ideas about topics that are far from settled. Science is not merely a collection of theorems and facts. It took extensive back and forth between theories and experiments to understand the nature of light in physics. Similarly, research in vision is undoubtedly full of themes that remain the topic of intense debate. I find it useful to engage students and readers in how experiments were designed, what the key ideas and hypotheses were, and how scientists can be wrong and correct their thinking based on the empirical results. I hope that including such interpretation of the findings can help convey the dynamic and vibrant nature of discussions in the field, as opposed to merely pouring a list of facts.

In opening up a dialogue with the readers, highlighting the process of discovery, what is known and what is not known, I would be particularly happy if this book inspires new generations of scholars to take up the challenge and help us figure out how vision works. I would be happy if these pages inspire courageous young scholars to prove that some of the speculations in here are wrong. I hope that these pages will motivate others to figure out how to solve the many open questions in the field.

The ability to bridge neurobiological investigations of the underlying neural circuits and computational models is still in its infancy. Despite recent claims in prominent newspapers, the problem of vision is not solved yet. There has been exhilarating progress in computer vision lately, and we now have machines that can solve certain visual recognition problems at the same level as humans and, in some cases, even algorithms that surpass human performance. However, we are still quite far from solving vision. There is no impediment, no fundamental limit in the laws of physics imposing that we cannot build machines that see and interpret the visual world. I am convinced that we *will* solve vision. The best is yet to come.

Acknowledgments

Some movies depict a scientist as a lonely creature, often evil and with white hair, locked in a faraway room for decades until the apple falls from the tree, which is when she suddenly realizes how everything works. My life as a scientist has little to do with this Hollywood style of rendering the practice of research. Science is particularly appealing to me because of its objective and rigorous endeavor to discover Truth, with capital T. Science does not occur in a vacuum, but, instead, it is part of a social construct, and I have yet to experience an apple hitting me in the head. I am incredibly indebted to a great number of people who have helped in one way or another put these pages together. This book would not have been possible without the insights and help from a very large number of individuals.

First, I would like to thank all the students in the biological and computer vision class. Their perennial curiosity, their seemingly naïve yet astute ways to make me rethink what I was trying to explain, and their constant questioning of our assumptions and statements have kept me on my toes and propelled this work forward. I cannot quite do justice to all of them, but I would like to mention a few students who have stood out – including Kristen Fang, Grigori Guitchounts, Katherine Harrison, Javier Masis, Olivia McGinnis, Daniel Rothchild, Sean Sullivan, Kenneth Shinozuka, Duncan Stothers, Annabelle Tao, Saloni Vishwakarm, and Will Xiao.

The teaching assistants in the class – Joseph Olson, Will Xiao, and Yuchen Xiao – have also played an essential role in this work. We split each class into a lecture and a discussion of primary literature in the field. The teaching assistants always bring provocative questions, new ways of thinking critically about the literature, and they engage the students in exciting discussions about the why and how of a particular study.

Several colleagues have also given guest lectures in this class, and their way of presenting the material was also extremely influential. I would like to thank Frederico Azevedo, Andrei Barbu, Xavier Boix, David Cox, Camille Gomez-Laberge, Till Hartmann, Leyla Isik, Kohitij Kar, Jiye Kim, Bill Lotter, Diego Mendoza-Halliday, Ken Nakayama, Carlos Ponce, Sarit Szpiro, Hanlin Tang, Kasper Vinken, and Will Xiao.

Another constant source of inspiration has been the students and postdocs in my lab. I especially want to mention Yigal Agam, Arjun Bansal, Xavier Boix, Calin Buia, Kendra Burbank, Stephen Casper, Prabaha Gangpadhyay, Camille Gomez-Laberge, Stephan Grzekowski, Shashi Gupta, Walter Hardesty, Eleonora Iaselli, Leyla Isik, Jiye Kim, Phil Kuhnke, Garrett Lam, Daniel Lopez Martinez, Bill Lotter, Radhika

Madhavan, Thomas Miconi, Pranav Misra, Charlotte Moermann, Candace Ross, Martin Schrimpf, Nimrod Shaham, Jedediah Singer, Hanlin Tang, Ben Tsuda, Kasper Vinken, Jerry Wang, Farahnaz Wick, Eric and Kevin Wu, Will Xiao, Ege Yumusak, and Mengmi Zhang. Many of them have provided useful comments and criticism to the pages here. They have also made outstanding scientific contributions that are reflected in the work discussed here.

In the ever more demanding modern days, I see my colleagues continually rushing to meet deadlines and juggling their myriad contributions to the scientific enterprise. I am therefore extremely honored and certainly in debt to many of my friends and colleagues who have been kind enough to read these pages, correct my mistakes, suggest additional material, and provide critical comments. The following people have been particularly kind: Arash Afraz, Katarina Bendtz, Xavier Boix, Alfonso Caramazza, Stephen Casper, Wenqiang Chen, Michael Do, John Dowling, Kohitij Kar, Talia Konkle, Bill Lotter, Richard Masland, David Mazumder, Liad Mudrik, Carlos Ponce, Nao Tsuchiya, Will Xiao, Yaoda Xu, and Mengmi Zhang. I am especially indebted to Ellen Hildreth. She has made major contributions to computer vision herself and she leads the education initiatives at the Center for Brains, Minds and Machines. She has read the entire book and has helped make it better through her detailed corrections, astute questions, and encouraging comments.

Finally, I am eternally thankful to the Center for Brains, Minds and Machines (CBMM), a center sponsored by the National Science Foundation. CBMM has provided an inspiring and influential home to brainstorm about new ideas and synergistically push the frontiers of neuroscience and artificial intelligence research by bringing together a large cadre of talented scholars from multiple fields.

Thank you all!

Abbreviations

We defined all abbreviations in the text. Here are a few that are used throughout.

AI	artificial intelligence (Chapters 1 and 9)
Alexnet	prominent deep convolutional neural network for visual recognition (Krizhevsky et al. 2012); other deep convolutional neural network architectures like VGG, ResNet, and Inception are also mentioned in Chapters 8 and 9
CLEVR	compositional language and elementary visual reasoning dataset (Johnson et al. 2016, Chapter 9)
DCNN	deep convolutional neural network (Chapter 8)
DeepDream	a technique to visualize the type of images that lead to high activation for units in a neural network (Chapter 9)
GAN	generative adversarial network (Goodfellow et al. 2014, Chapter 9)
HMAX	computational model of visual recognition (Riesenhuber and Poggio 1999, Chapter 8)
ImageNet	large dataset of images used to train and test computer vision algorithms (Russakovsky et al. 2014, Chapters 8 and 9)
IOR	inhibition-of-return mechanism to prevent eye movements from always landing on the most salient part of the image (illustrated in Figure 8.9)
ITC	inferior temporal cortex (Chapters 4 and 6)
IVSN	Invariant Visual Search Network (Chapter 8, Figure 8.9)
L2 norm	Euclidian distance between two vectors, typically used to define the error in machine learning (as opposed to the L1 norm, which is the sum of the absolute values of each component of the distance vector) (Chapter 8)
LGN	lateral geniculate nucleus, part of the thalamus that receives input from the retina and projects to primary visual cortex (Chapter 2)
LSTM	Long Short-Term Memory recurrent neural network (Horchreiter and Schmidhuber 1997, Chapter 8)
MNIST	database consisting of images of handwritten digits (Chapter 8)
MSCOCO	segmented object dataset (Lin et al. 2015, Chapter 9)
Neocognitron	computational model of visual recognition (Fukushima 1980, Chapter 8)

PredNet	predictive coding deep convolutional neural network (Lotter et al. 2017, Chapter 8)
QR code	quick response code, matrix barcode that can easily be read by smartphones (Chapter 9)
ReLU	rectifying linear unit (Chapter 7)
ResNet	deep convolutional neural network architecture (He et al. 2015, Chapter 8)
RGC	retinal ganglion cells, output neurons of the retina (Chapter 2)
ROC	receiver operating characteristic curve (Green and Swets 1966, Chapter 9)
Softmax	a function that converts a vector of input numbers into a probability distribution that adds up to 1, where each output value is proportional to the exponential of the input numbers (Chapter 8)
SVM	Support Vector Machine (Chapter 6)
tSNE	t-distributed stochastic network embedding (van der Maaten and Hinton 2008, Chapter 8)
UCF101	large dataset of videos used to train and test action recognition algorithms (Soomro et al. 2012, Chapter 9)
V1	primary visual cortex; other visual areas are defined, including V2, V3, V4, and V5, also known as area MT (Chapters 4 and 5)
WTA	winner-take-all mechanism (for example, the one used in Figure 8.9)
XDream	eXtending DeepDream with real-time evolution for activation maximization, algorithm to probe neuronal tuning in an unbiased manner (Ponce et al. 2019, Chapter 8)