

Jon Krohn • Grant Beyleveld • Aglaé Bassens

Deep Learning illustriert

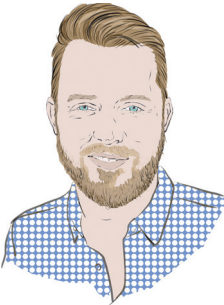
Eine anschauliche Einführung
in Machine Vision,
Natural Language Processing
und Bilderzeugung
für Programmierer und Datenanalysten



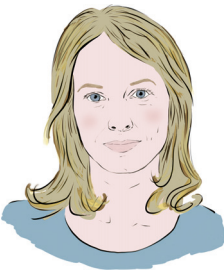
dpunkt.verlag



Jon Krohn ist Chief Data Scientist beim Machine-Learning-Unternehmen *untapt*. Er präsentiert eine viel gerühmte Serie aus Tutorials, die von Addison-Wesley herausgebracht wurden, darunter *Deep Learning with TensorFlow LiveLessons* und *Deep Learning for Natural Language Processing LiveLessons*. Jon unterrichtet Deep Learning an der *New York City Data Science Academy* und als Gastdozent an der *Columbia University*. Er besitzt einen Dokortitel in Neurowissenschaften von der Universität Oxford und veröffentlicht seit 2010 Artikel zum Thema Machine Learning in führenden Fachzeitschriften, darunter *Advances in Neural Information Processing Systems*.



Grant Beyleveld ist Data Scientist bei *untapt*, wo er auf dem Gebiet der Verarbeitung natürlicher Sprache mittels Deep Learning arbeitet. Er besitzt einen Dokortitel in biomedizinischer Wissenschaft von der *Icahn School of Medicine* am *Mount Sinai Hospital* in New York City, wo er die Beziehung zwischen Viren und ihren Wirten untersuchte. Er ist Gründungsmitglied von *deeplearning-studygroup.org*.



Aglaé Bassens ist eine in Paris lebende belgische Künstlerin. Sie studierte bildende Kunst an *The Ruskin School of Drawing and Fine Art* der Universität Oxford und an der *Slade School of Fine Arts* des *University College London*. Neben ihrer Arbeit als Illustratorin malt sie Stillleben und Wandbilder.

Jon Krohn · Grant Beyleveld · Aglaé Bassens

Deep Learning illustriert

**Eine anschauliche Einführung in Machine Vision,
Natural Language Processing und Bilderzeugung
für Programmierer und Datenanalysten**

Aus dem Englischen von Kathrin Lichtenberg



dpunkt.verlag

Jon Krohn · Grant Beyleveld · Aglaé Bassens

Lektorat: Gabriel Neumann

Übersetzung: Kathrin Lichtenberg, Ilmenau

Copy-Editing: Friederike Daenecke, Züllich

Terminologie-Beratung: Marcus Fraaß

Satz: Birgit Bäuerlein

Herstellung: Stefanie Weidner

Umschlaggestaltung: Helmut Kraus, www.exclam.de

Druck und Bindung: mediaprint solutions GmbH, 33100 Paderborn

Bibliografische Information der Deutschen Nationalbibliothek

Die Deutsche Nationalbibliothek verzeichnet diese Publikation in der Deutschen Nationalbibliografie; detaillierte bibliografische Daten sind im Internet über <http://dnb.d-nb.de> abrufbar.

ISBN:

Print 978-3-86490-663-3

PDF 978-3-96088-751-5

ePub 978-3-96088-752-2

mobi 978-3-96088-753-9

Translation Copyright für die deutschsprachige Ausgabe © 2020 dpunkt.verlag GmbH

Wieblingen Weg 17 · 69123 Heidelberg

Authorized German translation of the English original »Deep Learning Illustrated«, 1st edition by

Jon Krohn, Beyleveld Grant, Bassens Aglae, published by Pearson Education, Inc, publishing as

Addison-Wesley Professional, Copyright © 2019 Pearson Education, Inc

All rights reserved. No part of this book may be reproduced or transmitted in any form or by any means, electronic or mechanical, including photocopying, recording or by any information storage retrieval system, without permission from Pearson Education, Inc.

Hinweis:

Dieses Buch wurde auf PEFC-zertifiziertem Papier aus nachhaltiger Waldwirtschaft gedruckt. Der Umwelt zuliebe verzichten wir zusätzlich auf die Einschweißfolie.

Schreiben Sie uns:

Falls Sie Anregungen, Wünsche und Kommentare haben, lassen Sie es uns wissen: hallo@dpunkt.de.



Die vorliegende Publikation ist urheberrechtlich geschützt. Alle Rechte vorbehalten. Die Verwendung der Texte und Abbildungen, auch auszugsweise, ist ohne die schriftliche Zustimmung des Verlags urheberrechtswidrig und daher strafbar. Dies gilt insbesondere für die Vervielfältigung, Übersetzung oder die Verwendung in elektronischen Systemen.

Es wird darauf hingewiesen, dass die im Buch verwendeten Soft- und Hardware-Bezeichnungen sowie Markennamen und Produktbezeichnungen der jeweiligen Firmen im Allgemeinen warenzeichen-, marken- oder patentrechtlichem Schutz unterliegen.

Alle Angaben und Programme in diesem Buch wurden mit größter Sorgfalt kontrolliert. Weder Autor noch Verlag noch Übersetzer können jedoch für Schäden haftbar gemacht werden, die in Zusammenhang mit der Verwendung dieses Buches stehen.

❖ *Für Gigi* ❖

Vorwort

Machine Learning gilt vielen Menschen als die Zukunft der Statistik und Computertechnik, da es völlig neue Akzente in Kundendienst, Design, Bankwesen, Medizin, Produktion und in vielen anderen Bereichen und Branchen setzt. Es ist kaum möglich, seinen Einfluss auf die Welt und jene Veränderungen, die Machine Learning in den kommenden Jahren und Jahrzehnten bringen wird, überzubewerten. Von der Vielzahl an Machine-Learning-Methoden, die von Experten eingesetzt werden, etwa *Penalized Regression*, *Random Forest* und *Boosted Trees*, ist *Deep Learning* vermutlich die aufregendste.

Deep Learning hat die Gebiete *Computer Vision* (maschinelles Sehen) und *Natural Language Processing* (Verarbeitung natürlicher Sprache) revolutioniert, und Forscher finden immer neue Bereiche, die sie mit der Macht neuronaler Netze verwandeln wollen. Seine größte und beeindruckendste Wirkung zeigt Deep Learning oft bei den Versuchen, das menschliche Erleben nachzuahmen, wie bei der erwähnten Seh- und Sprachverarbeitung sowie bei der Audiosynthese und bei Übersetzungen. Die Berechnungen und Konzepte, die dem Deep Learning zugrunde liegen, wirken möglicherweise abschreckend und hindern Menschen unnötigerweise daran, sich damit zu befassen.

Die Autoren von *Deep Learning illustriert* gehen diese traditionell wahrgenommenen Hürden an und vermitteln ihr Wissen ruhig und gelassen – und das entstandene Buch ist eine wahre Freude. Wie die anderen Bücher aus dieser Reihe – *R for Everyone*, *Pandas for Everyone*, *Programming Skills for Data Science* und *Machine Learning with Python for Everyone* – wendet sich dieses Buch an ein breites Publikum mit ganz unterschiedlichem Wissen und Können. Die mathematischen Notationen sind auf ein Minimum beschränkt, und falls dennoch Gleichungen erforderlich sind, werden sie von verständlichem Text begleitet. Die meisten Erkenntnisse werden durch Grafiken, Illustrationen und Keras-Code ergänzt, der in Form leicht nachzuvollziehender Jupyter-Notebooks zur Verfügung steht.

Jon Krohn unterrichtet schon seit vielen Jahren Deep Learning. Besonders denkwürdig war eine Präsentation beim *Open Statistical Programming Meetup* in New York – bei derselben Vereinigung, in der er seine *Deep Learning Study Group* startete. Seine Brillanz in diesem Thema zeigt sich an seinen Texten, die Lesern Bildung vermitteln und ihnen gleichzeitig zeigen, wie spannend und aufregend das Material ist. Für dieses Buch arbeitet er mit Grant Beyleveld und Aglaé Bassens zusammen, die ihr Wissen bei der Anwendung von Deep-Learning-Algorithmen und ihre gekonnten und witzigen Zeichnungen beisteuern.

Deep Learning illustriert kombiniert Theorie, Mathematik (dort, wo es nötig ist), Code und Visualisierungen zu einer umfassenden Behandlung des Themas Deep Learning. Das Buch behandelt die volle Breite des Themas, einschließlich vollständig verbundener Netzwerke, Convolutional Neural Networks, Recurrent Neural Networks, Generative Adversarial Networks und Reinforcement Learning sowie deren Anwendungen. Dadurch ist dieses Buch die ideale Wahl für jemanden, der neuronale Netze kennenlernen und gleichzeitig praktische Hinweise für deren Implementierung haben möchte. Jeder kann und sollte davon profitieren und außerdem seine Zeit beim Lesen mit Jon, Grant und Aglaé genießen.

Jared Lander

Herausgeber der Reihe

Einführung

Milliarden miteinander verbundener Neuronen, gemeinhin als Gehirn bezeichnet, bilden Ihr Nervensystem und erlauben es Ihnen, zu spüren, zu denken und zu handeln. Durch akribisches Einfärben und Untersuchen dünner Scheiben von Gehirnmasse konnte der spanische Arzt Santiago Cajal (Abbildung 1) als erster¹ Neuronen identifizieren (Abbildung 2). In der ersten Hälfte des 20. Jahrhunderts begannen Forscher zu verstehen, wie diese Zellen arbeiten. In den 1950er-Jahren experimentierten Wissenschaftler, die von unserem zunehmenden Verständnis für das Gehirn inspiriert waren, mit computerbasierten künstlichen Neuronen und verknüpften diese zu künstlichen neuronalen Netzen, die versuchten, die Funktionsweise ihres natürlichen Namensvetters nachzuahmen.

Gewappnet mit dieser kurzen Geschichte der Neuronen, können wir den Begriff Deep Learning täuschend leicht definieren: Deep Learning beinhaltet ein Netzwerk, in dem künstliche Neuronen – üblicherweise Tausende, Millionen oder noch mehr davon – wenigstens mehrere Schichten tief gestapelt sind. Die künstlichen Neuronen in der ersten Schicht übergeben Informationen an die zweite, die zweite Schicht reicht sie an die dritte und so weiter, bis die letzte Schicht irgendwelche Werte ausgibt. Wie wir allerdings im Laufe dieses Buches zeigen werden, kann diese simple Definition die bemerkenswerte Breite der Funktionalität des Deep Learning sowie seine außerordentlichen Zwischentöne nicht annähernd erfassen.

Wie wir in Kapitel 1 genauer ausführen werden, war die erste Welle des Deep-Learning-Tsunami, die metaphorisch gesprochen ans Ufer brandete, eine herausragende Leistung in einem wichtigen Machine-Vision-Wettbewerb im Jahre 2012. Sie wurde getrieben und unterstützt durch das Vorhandensein einigermaßen preiswerter Rechenleistung, ausreichend großer Datensätze und einer Handvoll wesentlicher theoretischer Fortschritte. Akademiker und Techniker merkten auf, und in den turbulenten Jahren seither hat das Deep Learning zahlreiche, mittlerweile alltägliche Anwendungen gefunden. Von Teslas Autopilot bis zur Stimmerkennung von Alexa, von Echtzeitübersetzungen zwischen Sprachen bis hin zu seiner Integration in Hunderte von Google-Produkten hat Deep Lear-

1. Cajal, S.-R. (1894). *Les Nouvelles Idées sur la Structure du Système Nerveux chez l'Homme et chez les Vertébrés*. Paris: C. Reinwald & Companie.

ning die Genauigkeit vieler durch Computer erledigter Aufgaben von 95 Prozent auf teils mehr als 99 Prozent verbessert – die entscheidenden Prozentpunkte, die dafür sorgen, dass ein automatisierter Dienst sich tatsächlich anfühlt, als würde er von Zauberhand ausgeführt werden. Auch wenn die in diesem Buch gelieferten interaktiven Codebeispiele die vorgebliche Magie entzaubern, verschafft das Deep Learning den Maschinen eine übermenschliche Fähigkeit bei komplexen Aufgaben, die so verschieden sind wie das Erkennen von Gesichtern, das Zusammenfassen von Texten und das Spielen schwieriger Brettspiele.² Angesichts dieser markanten Fortschritte überrascht es kaum, dass »Deep Learning« gleichgesetzt wird mit »künstlicher Intelligenz« – in der Presse, am Arbeitsplatz und zu Hause.



Abb. 1 Santiago Cajal (1852–1934)

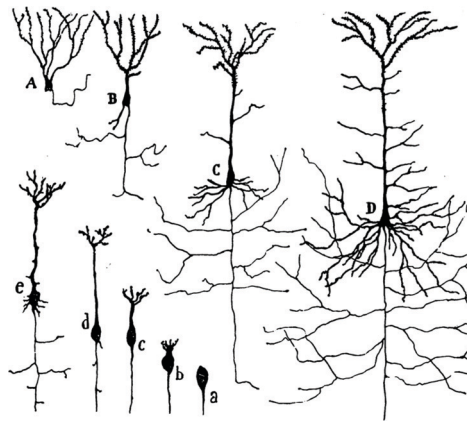


Abb. 2 Ein handgezeichnetes Diagramm aus Cajals Veröffentlichung (1894) zeigt das Wachstum eines Neurons (a–e) und verschiedenartige Neuronen eines Frosches (A), einer Eidechse (B), einer Ratte (C) und eines Menschen (D)

2. Unter bit.ly/aiindex18 finden Sie einen Vergleich zwischen menschlicher und maschineller Leistungsfähigkeit.

Es sind aufregende Zeiten, weil – wie Sie in diesem Buch entdecken werden – vermutlich nur einmal im Leben ein einziges Konzept in so kurzer Zeit so umfassende Umstürze mit sich bringt. Wir sind hocherfreut, dass auch Sie Interesse an Deep Learning gefunden haben, und können es kaum erwarten, unseren Enthusiasmus für diese beispiellose Technik mit Ihnen zu teilen.

Wie Sie dieses Buch lesen sollten

Dieses Buch besteht aus vier Teilen. Teil I, »Deep Learning vorgestellt«, eignet sich für alle interessierten Leserinnen und Leser. Es ist ein allgemeiner Überblick, der uns verrät, was Deep Learning eigentlich ist, wie es sich entwickelt hat und wie es mit Konzepten wie KI, Machine Learning und Reinforcement Learning verwandt ist. Voller eigens geschaffener Illustrationen, eingängiger Analogien und auf das Wesentliche konzentrierter Beschreibungen, sollte Teil I für alle erhellend sein, also auch für diejenigen, die keine besondere Programmiererfahrung mitbringen.

Die Teile II bis IV wenden sich hingegen an Softwareentwickler, Data Scientists, Forscher, Analysten und andere, die gern lernen möchten, wie sich Deep-Learning-Techniken auf ihrem Gebiet einsetzen lassen. In diesen Teilen unseres Buches wird die wesentliche zugrunde liegende Theorie behandelt. Hierbei wird der Einsatz mathematischer Formeln auf das Mindestmaß reduziert und stattdessen auf intuitive visuelle Darstellungen und praktische Beispiele in Python gesetzt. Neben dieser Theorie vermitteln funktionierende Codeausschnitte, die in den begleitenden Jupyter-Notebooks³ zur Verfügung stehen, ein praktisches Verständnis für die wichtigsten Familien der Deep-Learning-Ansätze und -Anwendungen: Maschinelles Sehen (Machine Vision) (Kapitel 10), Verarbeitung natürlicher Sprache (Natural Language Processing) (Kapitel 11), Bildherstellung (Kapitel 12) und Spiele (Kapitel 13). Damit er besser zu erkennen ist, geben wir Code immer in einer solchen Nichtproportionalschrift (also in einer Schrift mit fester Breite) an. Außerdem verwenden wir in den Codeausschnitten den üblichen Jupyter-Stil (Zahlen in Grün, Strings in Rot usw.).

Falls Sie sich nach detaillierteren Erklärungen der mathematischen und statistischen Grundlagen des Deep Learning sehnen, als wir in diesem Buch anbieten, könnten Sie sich unsere Tipps für weitere Studien anschauen:

1. Michael Nielsens E-Book *Neural Networks and Deep Learning*⁴, das kurz ist, Konzepte mithilfe netter interaktiver Applets demonstriert und eine ähnliche mathematische Notation verwendet wie wir

3. github.com/the-deep-learners/deep-learning-illustrated

4. Nielsen, M. (2015). *Neural Networks and Deep Learning*. Determination Press.
Kostenlos verfügbar unter: neuralnetworksanddeeplearning.com

2. Das Buch *Deep Learning*⁵ von Ian Goodfellow (vorgestellt in Kapitel 3), Yoshua Bengio (Abbildung 1–10) und Aaron Courville, das ausführlich die mathematischen Grundlagen neuronaler Netzwerktechniken behandelt

Überall im Buch finden Sie freundliche Trilobiten, die Ihnen gern kleine Schnipsel nicht ganz so notwendiger Informationen anbieten möchten, die für Sie vielleicht dennoch interessant oder hilfreich sein könnten. Der *lesende Trilobit* (wie in Abbildung 3) ist ein Bücherwurm, der Freude daran hat, Ihr Wissen zu erweitern. Der Trilobit, der um Ihre Aufmerksamkeit bittet (wie in Abbildung 4), hat eine Textpassage bemerkt, die möglicherweise problematisch für Sie ist, und würde in dieser Situation gern helfen. Zusätzlich zu den Trilobiten, die die Kästen bevölkern, haben wir reichlich Gebrauch von Fußnoten gemacht. Diese müssen Sie nicht unbedingt lesen, aber sie enthalten kurze Erklärungen neuer Begriffe und Abkürzungen sowie Quellenangaben zu wichtigen Artikeln, Büchern und anderen Referenzen, die Sie bei Interesse bemühen können.

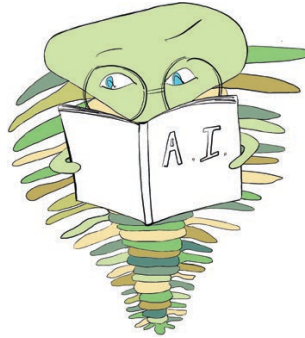


Abb. 3 Der lesende Trilobit hat Freude daran, Ihr Wissen zu erweitern.



Abb. 4 Dieser Trilobit möchte Ihre Aufmerksamkeit auf eine schwierige Textpassage lenken. Achten Sie auf ihn!

-
5. Goodfellow, I., et al. (2016). *Deep Learning*. MIT Press. Kostenlos verfügbar unter: deeplearningbook.org

Für einen Großteil des Inhalts dieses Buches gibt es begleitende Video-Tutorials in englischer Sprache. Dieses Buch bot uns die Möglichkeit, die theoretischen Konzepte gründlicher darzustellen, und die Videos erlauben es Ihnen, sich aus einer anderen Perspektive mit den Jupyter-Notebooks vertraut zu machen: Hier wird die Bedeutung der einzelnen Codezeilen bereits beim Eintippen beschrieben.⁶ Die Serie der Video-Tutorials verteilt sich über drei Titel, die jeweils bestimmte Kapitel dieses Buches begleiten:

1. *Deep Learning with TensorFlow LiveLessons*:⁷
Kapitel 1 und Kapitel 5 bis 10
2. *Deep Learning for Natural Language Processing LiveLessons*:⁸
Kapitel 2 und 11
3. *Deep Reinforcement Learning and GANs LiveLessons*:⁹
Kapitel 3, 4, 12 und 13

-
6. Viele der Jupyter-Notebooks, die in diesem Buch behandelt werden, sind direkt aus den Videos abgeleitet, die alle vor dem Schreiben des Buches aufgezeichnet wurden. An manchen Stellen haben wir entschieden, den Code für das Buch zu aktualisieren, sodass die Video-Version und die Buchversion eines Notebooks einander zwar ähnlich sind, aber nicht unbedingt völlig identisch sein müssen.
 7. Krohn, J. (2017). *Deep Learning with TensorFlow LiveLessons: Applications of Deep Neural Networks to Machine Learning Tasks* (Videokurs). Boston: Addison-Wesley.
 8. Krohn, J. (2017). *Deep Learning for Natural Language Processing LiveLessons: Applications of Deep Neural Networks to Machine Learning Tasks* (Videokurs). Boston: Addison-Wesley.
 9. Krohn, J. (2018). *Deep Reinforcement Learning and GANs LiveLessons: Advanced Topics in Deep Learning* (Videokurs). Boston: Addison-Wesley.

Danksagungen

Wir danken dem Team bei *untapt*, vor allem Andrew Vlahutin, Sam Kenny und Vince Petaccio II, die uns unterstützten, während wir dieses Buch schrieben. Besonders erwähnen wollen wir Ed Donner, der neuronale Netze liebt und uns pausenlos ermutigte, unserer Leidenschaft auf dem Gebiet des Deep Learning zu folgen.

Außerdem danken wir den Mitgliedern der *Deep Learning Study Group*¹, die regelmäßig unsere stimulierenden und lebhaften Treffen im New Yorker Büro von untapt besuchen. Da dieses Buch aufgrund der Diskussionen unserer Study Group entstand, kann man sich kaum vorstellen, wie es ohne diese Treffen zustande gekommen wäre.

Dank geht an unsere technischen Gutachter für ihre wertvollen Ratschläge, die den Inhalt des Buches deutlich verbessert haben: Alex Lipatov, Andrew Vlahutin, Claudia Perlich, Dmitri Nesterenko, Jason Baik, Laura Graesser, Michael Griffiths, Paul Dix und Wah Loon Keng. Danke auch an die Lektoren und Hersteller des Buches – Chris Zahn, Betsy Hardinger, Anna Popick und Julie Nahil –, deren Sorgfalt und Aufmerksamkeit die Qualität, Klarheit und Gestaltung dieses Buches sicherstellten. Dank an Jared Lander, der die New Yorker Open-Statistical-Programming-Gemeinschaft leitet, die sowohl unsere Deep Learning Study Group begründete als auch ein Treffen mit Debra Williams Cauley in die Wege leitete. Ein besonderer Dank gilt Debra selbst, die unsere fantasievollen Publikationsideen von dem Tag an unterstützt hat, an dem wir sie kennenlernten, und die entscheidend an ihrer Umsetzung beteiligt war. Wir danken auch den Wissenschaftlern und Machine-Learning-Experten, die uns akademisch geleitet haben und uns weiterhin inspirieren, vor allem Jonathan Flint, Felix Agakov und Will Valdar.

Und schließlich geht ein unendlicher Dank an unsere Familien und Freunde, die nicht nur ertragen haben, dass wir auch im Urlaub und an den Wochenenden gearbeitet haben, sondern uns auch selbstlos motiviert haben, es zu tun.

1. deeplearningstudygroup.org

Inhaltsübersicht

Teil I	Deep Learning vorgestellt	1
1	Biologisches und maschinelles Sehen	3
2	Menschen- und Maschinensprache	25
3	Maschinenkunst	47
4	Spielende Maschinen	61
Teil II	Die nötige Theorie	91
5	Der (Code-)Karren vor dem (Theorie-)Pferd	93
6	Künstliche Neuronen, die Hotdogs erkennen	105
7	Künstliche neuronale Netze	121
8	Deep Networks trainieren	137
9	Deep Networks verbessern	163
Teil III	Interaktive Anwendungen des Deep Learning	195
10	Maschinelles Sehen	197
11	Natural Language Processing	241
12	Generative Adversarial Networks	315
13	Deep Reinforcement Learning	341

Teil IV	KI und Sie	375
14	Mit Ihren eigenen Deep-Learning-Projekten beginnen	377
Anhang		399
A	Die formale Notation neuronaler Netze	401
B	Backpropagation	403
C	PyTorch	407
D	Bildnachweise	415
	Abbildungsverzeichnis	417
	Tabellenverzeichnis	429
	Beispielverzeichnis	431
	Index	435

Inhaltsverzeichnis

Teil I	Deep Learning vorgestellt	1
1	Biologisches und maschinelles Sehen	3
1.1	Das biologische Sehen	3
1.2	Maschinelles Sehen	10
1.2.1	Das Neocognitron	10
1.2.2	LeNet-5	11
1.2.3	Der traditionelle Machine-Learning-Ansatz	14
1.2.4	ImageNet und die ILSVRC	15
1.2.5	AlexNet	17
1.3	TensorFlow Playground	20
1.4	Quick, Draw!	22
1.5	Zusammenfassung	23
2	Menschen- und Maschinensprache	25
2.1	Deep Learning für Natural Language Processing	26
2.1.1	Deep-Learning-Netze lernen Repräsentationen automatisch	26
2.1.2	Natural Language Processing	28
2.1.3	Eine kurze Geschichte des Deep Learning für NLP	30
2.2	Repräsentationen von Sprache im Computer	31
2.2.1	1-aus-n-Repräsentationen von Wörtern	31
2.2.2	Wortvektoren	32
2.2.3	Wortvektor-Arithmetik	36
2.2.4	word2viz	37
2.2.5	Lokalistische versus verteilte Repräsentationen	39
2.3	Elemente der natürlichen menschlichen Sprache	41
2.4	Google Duplex	44
2.5	Zusammenfassung	46

3	Maschinenkunst	47
3.1	Eine feuchtfröhliche Nacht	47
3.2	Berechnungen auf nachgemachten menschlichen Gesichtern	50
3.3	Stiltransfer: Fotos in einen Monet verwandeln (und umgekehrt)	53
3.4	Machen Sie Ihre eigenen Skizzen fotorealistisch	54
3.5	Fotorealistische Bilder aus Text erzeugen	55
3.6	Bildverarbeitung mittels Deep Learning	56
3.7	Zusammenfassung	58
4	Spielende Maschinen	61
4.1	Deep Learning, KI und andere Monster	61
4.1.1	Künstliche Intelligenz	61
4.1.2	Machine Learning	63
4.1.3	Representation Learning	63
4.1.4	Künstliche neuronale Netze	63
4.1.5	Deep Learning	64
4.1.6	Maschinelles Sehen	65
4.1.7	Natural Language Processing	66
4.2	Drei Arten von Machine-Learning-Problemen	66
4.2.1	Supervised Learning	67
4.2.2	Unsupervised Learning	67
4.2.3	Reinforcement Learning	68
4.3	Deep Reinforcement Learning	70
4.4	Videospiele	72
4.5	Brettspiele	73
4.5.1	AlphaGo	74
4.5.2	AlphaGo Zero	78
4.5.3	AlphaZero	81
4.6	Manipulation von Objekten	83
4.7	Populäre Umgebungen für das Deep-Reinforcement-Learning	85
4.7.1	OpenAI Gym	85
4.7.2	DeepMind Lab	86
4.7.3	UnityML-Agents	88
4.8	Drei Arten von KI	89
4.8.1	Artificial Narrow Intelligence	89
4.8.2	Artificial General Intelligence	89
4.8.3	Artificial Super Intelligence	89
4.8.4	Zusammenfassung	90

Teil II	Die nötige Theorie	91
5	Der (Code-)Karren vor dem (Theorie-)Pferd	93
5.1	Voraussetzungen	93
5.2	Installation	94
5.3	Ein flaches Netzwerk in Keras	94
5.3.1	Der MNIST-Datensatz handgeschriebener Ziffern	95
5.3.2	Ein schematisches Diagramm des Netzwerks	96
5.3.3	Die Daten laden	99
5.3.4	Die Daten umformatieren	101
5.3.5	Die Architektur eines neuronalen Netzes entwerfen	102
5.3.6	Trainieren eines Deep-Learning-Modells	103
5.4	Zusammenfassung	104
6	Künstliche Neuronen, die Hotdogs erkennen	105
6.1	Das Einmaleins der biologischen Neuroanatomie	105
6.2	Das Perzeptron	106
6.2.1	Der Hotdog/Nicht-Hotdog-Detektor	107
6.2.2	Die wichtigste Gleichung in diesem Buch	111
6.3	Moderne Neuronen und Aktivierungsfunktionen	112
6.3.1	Das Sigmoid-Neuron	113
6.3.2	Das Tanh-Neuron	115
6.3.3	ReLU: Rectified Linear Units	116
6.4	Ein Neuron auswählen	118
6.5	Zusammenfassung	119
	Schlüsselkonzepte	119
7	Künstliche neuronale Netze	121
7.1	Die Eingabeschicht	121
7.2	Vollständig verbundene Schichten	122
7.3	Ein vollständig verbundenes Netzwerk zum Erkennen von Hotdogs	124
7.3.1	Forwardpropagation durch die erste verborgene Schicht	125
7.3.2	Forwardpropagation durch nachfolgende Schichten	126
7.4	Die Softmax-Schicht eines Netzwerks zum Klassifizieren von Fastfood	129
7.5	Zurück zu unserem flachen Netzwerk	132
7.6	Zusammenfassung	134
	Schlüsselkonzepte	135

8	Deep Networks trainieren	137
8.1	Kostenfunktionen	137
8.1.1	Quadratische Kosten	138
8.1.2	Gesättigte Neuronen	139
8.1.3	Kreuzentropie-Kosten	140
8.2	Optimierung: Lernen, um die Kosten zu minimieren	143
8.2.1	Der Gradientenabstieg	143
8.2.2	Die Lernrate	146
8.2.3	Batch-Größe und stochastischer Gradientenabstieg	148
8.2.4	Dem lokalen Minimum entkommen	152
8.3	Backpropagation	155
8.4	Die Anzahl der verborgenen Schichten und der Neuronen anpassen	156
8.5	Ein mittleres Netz in Keras	158
8.6	Zusammenfassung	161
	Schlüsselkonzepte	162
9	Deep Networks verbessern	163
9.1	Die Initialisierung der Gewichte	163
9.1.1	Xavier-Glorot-Verteilungen	168
9.2	Instabile Gradienten	171
9.2.1	Verschwindende Gradienten	171
9.2.2	Explodierende Gradienten	172
9.2.3	Batch-Normalisierung	172
9.3	Modellgeneralisierung (Überanpassung vermeiden)	174
9.3.1	L1- und L2-Regularisierung	176
9.3.2	Dropout	177
9.3.3	Datenaugmentation	180
9.4	Intelligente Optimierer	181
9.4.1	Momentum	181
9.4.2	Nesterov-Momentum	182
9.4.3	AdaGrad	182
9.4.4	AdaDelta und RMSProp	183
9.4.5	Adam	183
9.5	Ein tiefes neuronales Netz in Keras	184
9.6	Regression	186
9.7	TensorBoard	189
9.8	Zusammenfassung	192
	Schlüsselkonzepte	193

Teil III	Interaktive Anwendungen des Deep Learning	195
10	Maschinelles Sehen	197
10.1	Convolutional Neural Networks	197
10.1.1	Die zweidimensionale Struktur der visuellen Bilddarstellung	198
10.1.2	Berechnungskomplexität	198
10.1.3	Konvolutionsschichten	199
10.1.4	Mehrere Filter	202
10.1.5	Ein Beispiel für Konvolutionsschichten	203
10.2	Hyperparameter von Konvolutionsfiltern	208
10.2.1	Kernel-Größe	208
10.2.2	Schrittlänge	209
10.2.3	Padding	209
10.3	Pooling-Schichten	210
10.4	LeNet-5 in Keras	212
10.5	AlexNet und VGGNet in Keras	218
10.6	Residualnetzwerke	221
10.6.1	Schwindende Gradienten: Das Grauen der tiefen CNN	221
10.6.2	Residualverbindungen	222
10.6.3	ResNet	225
10.7	Anwendungen des maschinellen Sehens	225
10.7.1	Objekterkennung	226
10.7.2	Bildsegmentierung	230
10.7.3	Transfer-Lernen	233
10.7.4	Capsule Networks	237
10.8	Zusammenfassung	238
	Schlüsselkonzepte	239
11	Natural Language Processing	241
11.1	Natürliche Sprachdaten vorverarbeiten	241
11.1.1	Tokenisierung	244
11.1.2	Alle Zeichen in Kleinbuchstaben umwandeln	247
11.1.3	Stoppwörter und Interpunktionszeichen entfernen	247
11.1.4	Stemming	248
11.1.5	N-Gramme verarbeiten	249
11.1.6	Vorverarbeitung des kompletten Textkorpus	251

11.2	Worteinbettungen mit word2vec erzeugen	254
11.2.1	Die prinzipielle Theorie hinter word2vec	254
11.2.2	Wortvektoren evaluieren	257
11.2.3	word2vec ausführen	258
11.2.4	Wortvektoren plotten	263
11.3	Der Bereich unter der ROC-Kurve	268
11.3.1	Die Wahrheitsmatrix	269
11.3.2	Die ROC-AUC-Metrik berechnen	270
11.4	Klassifikation natürlicher Sprache mit vertrauten Netzwerken	274
11.4.1	Die IMDb-Filmkritiken laden	274
11.4.2	Die IMDb-Daten untersuchen	278
11.4.3	Die Länge der Filmkritiken standardisieren	281
11.4.4	Vollständig verbundenes Netzwerk	281
11.4.5	Convolutional Networks	288
11.5	Netzwerke für die Verarbeitung sequenzieller Daten	293
11.5.1	Recurrent Neural Networks	294
11.5.2	Ein RNN in Keras implementieren	296
11.5.3	Long Short-Term Memory Units	299
11.5.4	Bidirektionale LSTMs	303
11.5.5	Gestapelte rekurrente Modelle	303
11.5.6	Seq2seq und Attention	305
11.5.7	Transfer-Lernen in NLP	307
11.6	Nichtsequenzielle Architekturen: Die funktionale API in Keras	308
11.7	Zusammenfassung	312
	Schlüsselkonzepte	313
12	Generative Adversarial Networks	315
12.1	Die grundlegende GAN-Theorie	315
12.2	Der Quick, Draw!-Datensatz	319
12.3	Das Diskriminator-Netzwerk	323
12.4	Das Generator-Netzwerk	326
12.5	Das Adversarial-Netzwerk	329
12.6	Das GAN-Training	331
12.7	Zusammenfassung	337
	Schlüsselkonzepte	338

13	Deep Reinforcement Learning	341
13.1	Die grundlegende Theorie des Reinforcement Learning	341
13.1.1	Das Cart-Pole-Spiel	342
13.1.2	Markow-Entscheidungsprozesse	344
13.1.3	Die optimale Strategie	347
13.2	Die grundlegende Theorie von Deep-Q-Learning-Netzwerken	349
13.2.1	Value-Funktionen	350
13.2.2	Q-Value-Funktionen	350
13.2.3	Einen optimalen Q-Value schätzen	351
13.3	Einen DQN-Agenten definieren	353
13.3.1	Initialisierungsparameter	355
13.3.2	Das neuronale-Netze-Modell des Agenten bauen	358
13.3.3	Sich an das Spiel erinnern	359
13.3.4	Training über Memory Replay	359
13.3.5	Eine Aktion auswählen	361
13.3.6	Speichern und Laden der Modellparameter	362
13.4	Mit einer OpenAI-Gym-Umgebung interagieren	362
13.4.1	Hyperparameter-Optimierung mit SLM Lab	366
13.5	Agenten jenseits von DQN	369
13.5.1	Policy-Gradienten und der REINFORCE-Algorithmus	370
13.5.2	Der Actor-Critic-Algorithmus	371
13.6	Zusammenfassung	372
	Schlüsselkonzepte	373

Teil IV	KI und Sie	375
----------------	-------------------	------------

14	Mit Ihren eigenen Deep-Learning-Projekten beginnen	377
14.1	Ideen für Deep-Learning-Projekte	377
14.1.1	Machine Vision und GANs	377
14.1.2	Natural Language Processing	380
14.1.3	Deep Reinforcement Learning	381
14.1.4	Ein vorhandenes Machine-Learning-Projekt überführen	381
14.2	Ressourcen für weitere Projekte	382
14.2.1	Gesellschaftlich nützliche Projekte	383
14.3	Der Modellierungsprozess einschließlich der Anpassung der Hyperparameter	384
14.3.1	Automatisierung der Hyperparameter-Suche	387

14.4	Deep-Learning-Bibliotheken	387
14.4.1	Keras und TensorFlow	388
14.4.2	PyTorch	390
14.4.3	MXNet, CNTK, Caffe und so weiter	391
14.5	Software 2.0	391
14.6	Die kommende Artificial General Intelligence	394
14.7	Zusammenfassung	397

Anhang	399
---------------	------------

A	Die formale Notation neuronaler Netze	401
B	Backpropagation	403
C	PyTorch	407
C.1	PyTorch-Eigenschaften	407
C.1.1	Das Autograd System	407
C.1.2	Das Define-by-Run-Framework	407
C.1.3	PyTorch im Vergleich mit TensorFlow	408
C.2	PyTorch in der Praxis	409
C.2.1	Die PyTorch-Installation	409
C.2.2	Die grundlegenden Bausteine in PyTorch	410
C.2.3	Ein tiefes neuronales Netz in PyTorch bauen	411
D	Bildnachweise	415
	Abbildungsverzeichnis	417
	Tabellenverzeichnis	429
	Beispielverzeichnis	431
	Index	435

Teil I

Deep Learning vorgestellt

1	Biologisches und maschinelles Sehen	3
2	Menschen- und Maschinensprache	25
3	Maschinenkunst	47
4	Spielende Maschinen	61

1 Biologisches und maschinelles Sehen

In diesem Kapitel und auch im Laufe dieses Buches dient uns das visuelle System biologischer Organismen als Analogie, um das Deep Learning »zum Leben zu erwecken«. Diese Analogie vermittelt nicht nur ein tiefgreifendes Verständnis für das, was Deep Learning ausmacht, sondern verdeutlicht auch, weshalb Deep-Learning-Ansätze so machtvoll und so überaus vielfältig einsetzbar sind.

1.1 Das biologische Sehen

Vor 550 Millionen Jahren, in der prähistorischen Periode des Kambrium, stieg die Anzahl der Arten auf unserem Planeten schlagartig an (Abbildung 1–1). Aus den Fossilienfunden lässt sich ablesen,¹ dass diese explosionsartige Zunahme (die auch tatsächlich als Kambrische Explosion bezeichnet wird) durch die Entwicklung von Lichtdetektoren bei Trilobiten gefördert wurde, einem kleinen Meereslebewesen, das mit den heutigen Krebsen verwandt ist (Abbildung 1–2). Ein visuelles System, selbst wenn es nur primitiv ausgebildet ist, bringt eine wunderbare Vielfalt neuer Fähigkeiten mit sich. Man kann beispielsweise bereits aus einiger Entfernung Nahrung, Feinde und freundlich aussehende Gefährten ausmachen. Auch andere Sinne, wie der Geruchssinn, erlauben es Tieren, diese Dinge wahrzunehmen, allerdings nicht mit der Genauigkeit und Schnelligkeit des Sehvermögens. Die Hypothese besagt, dass mit dem Sehvermögen der Trilobiten ein Wettrennen einsetzte, dessen Ergebnis die Kambrische Explosion war: Die Beutetiere und auch die Feinde der Trilobiten mussten sich weiterentwickeln, um zu überleben.

1. Parker, A. (2004). *In the Blink of an Eye: How Vision Sparked the Big Bang of Evolution*. New York: Basic Books.

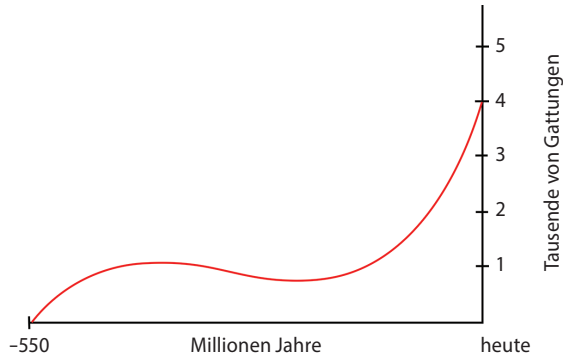


Abb. 1-1 Die Anzahl der Arten auf unserem Planeten begann vor 550 Millionen Jahren, während der Periode des Kambrium, schlagartig anzusteigen. »Gattungen« sind Kategorien miteinander verwandter Arten.

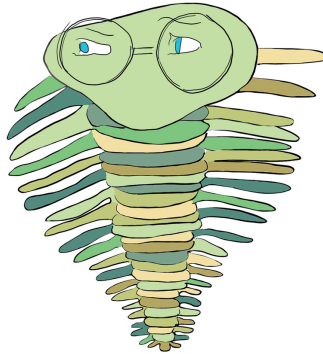


Abb. 1-2 Ein Trilobit mit Brille

In den mehr als eine halbe Milliarde Jahren, seit die Trilobiten das Sehen entwickelten, hat die Komplexität dieses Sinnes ungemein zugenommen. So ist bei heutigen Säugetieren ein Großteil der Großhirnrinde – das ist die äußere graue Masse des Gehirns – der visuellen Wahrnehmung vorbehalten.² Ende der 1950er-Jahre begannen die Physiologen David Hubel und Torsten Wiesel (Abbildung 1-3) an der John Hopkins University mit ihren bahnbrechenden Forschungen darüber, wie visuelle Informationen in der Großhirnrinde von Säugetieren verarbeitet werden,³ für die sie später mit dem Nobelpreis⁴ ausgezeichnet wurden. Wie in Abbildung 1-4 dargestellt wird, führten Hubel und Wiesel ihre Forschungen durch, indem sie narkotisierten Katzen Bilder zeigten, während sie gleichzeitig die Aktivität einzelner Neuronen aus dem primären visuellen Cortex aufzeichneten, also dem ersten Teil der Großhirnrinde, der visuellen Input von den Augen erhält.

Hubel und Wiesel zeigten den Katzen mithilfe von Dias, die sie auf eine Leinwand projizierten, einfache Formen, wie den Punkt aus Abbildung 1-4. Ihre ersten Ergebnisse waren entmutigend: Ihre Bemühungen lösten keine Reaktion der