3 Maschinenkunst

In diesem Kapitel stellen wir einige der Konzepte vor, die es Deep-Learning-Modellen scheinbar erlauben, Kunst zu erschaffen – ein Gedanke, den einige von uns vermutlich paradox finden. Der Philosoph Alva Noë von der University of Berkeley in Kalifornien meinte jedenfalls: »Kunst kann uns helfen, ein besseres Bild von unserer menschlichen Natur zu formen.«[1] Falls das stimmt, wie können Maschinen dann Kunst erschaffen? Oder anders formuliert: Sind die Kreationen, die von diesen Maschinen stammen, als Kunst anzusehen? Eine andere Interpretation – die uns übrigens am besten gefällt – lautet, dass diese Kreationen tatsächlich Kunst sind und dass die Programmierer Künstler sind, die ihre Deep-Learning-Modelle wie Pinsel handhaben. Wir sind nicht die einzigen, die diese Werke als wahre Kunst betrachten: Von GAN-Algorithmen (*Generative Adversarial Networks*, zu Deutsch etwa: *erzeugende gegnerische Netzwerke*) geschaffene Gemälde sind teils für mehr als 400.000 US-Dollar über den Tisch gegangen.[2]

In diesem Kapitel werden wir uns die hochentwickelten Konzepte hinter GANs anschauen. Sie werden Beispiele der neuartigen visuellen Werke sehen, die sie produzieren können. Wir werden eine Verbindung zwischen den latenten Räumen, die mit GANs verknüpft sind, und den Wortvektorräumen aus Kapitel 2 ziehen. Und wir werden ein Deep-Learning-Modell behandeln, das als automatisiertes Werkzeug dienen kann, um die Qualität von Fotos drastisch zu verbessern. Aber bevor es losgeht, schnappen Sie sich einen Drink …

3.1 Eine feuchtfröhliche Nacht

Unter den Google-Büros in Montreal gibt es eine Bar namens »Les 3 Brasseurs«, zu Deutsch also »Die 3 Brauer«. Dort dachte sich Ian Goodfellow, der damals, im Jahre 2014, als PhD-Student in Yoshua Bengios renommiertem Labor (Abbildung 1–10) arbeitete, einen Algorithmus zum Herstellen realistisch aussehender

1. Noë, A. (5. Oktober 2015). What art unveils. *New York Times*.
2. Cohn, G. (25. Oktober 2018). AI art at Christie's sells for $432.500. *New York Times*.

Bilder aus[3] – eine Technik, die Yann LeCun (Abbildung 1–9) als »wichtigsten« aktuellen Durchbruch auf dem Gebiet des Deep Learning bejubelte.[4]

Goodfellows Freunde beschrieben ihm ein *generatives Modell*, an dem sie arbeiteten, das heißt, ein Computermodell, das darauf abzielt, etwas Neues zu erschaffen, sei es ein Zitat im Stil von Shakespeare, eine Melodie oder ein abstraktes Kunstwerk. In ihrem speziellen Fall versuchten die Freunde, ein Modell zu entwerfen, das fotorealistische Bilder generieren konnte, wie etwa Porträts menschlicher Gesichter. Damit dies mit dem traditionellen Machine-Learning-Ansatz einigermaßen gut funktioniert (Abbildung 1–12), müssten die Ingenieure, die das Modell entwarfen, nicht nur die entscheidenden Merkmale von Gesichtern katalogisieren und approximieren, wie Augen, Nasen und Münder, sondern auch exakt abschätzen, wie diese Merkmale relativ zueinander angeordnet werden müssten. Bislang waren die Ergebnisse wenig beeindruckend. Die generierten Gesichter waren entweder sehr unscharf oder ihnen fehlten wichtige Elemente wie die Nase oder die Ohren.

Goodfellow, dessen Kreativität möglicherweise durch das eine oder andere Bier angeregt wurde,[5] hatte eine revolutionäre Idee: ein Deep-Learning-Modell, in dem zwei künstliche neuronale Netze (*Artificial Neural Network*, ANN) quasi im Wettstreit gegeneinander antreten. Wie in Abbildung 3–1 dargestellt wird, würde eines dieser ANN darauf programmiert werden, Fälschungen herzustellen, während das andere so programmiert würde, dass es als Detektiv agiert und die Fälschungen von den echten Bildern unterscheidet (diese würden separat angeboten werden). Diese gegnerischen Deep-Learning-Netze würden einander anstacheln: Wenn der *Generator* beim Herstellen der Fälschungen besser wird, muss der *Diskriminator* besser dabei werden, sie zu identifizieren, und so müsste der Generator noch überzeugendere Nachahmungen produzieren und so weiter. Dieser wunderbare Trainingszyklus würde schließlich zu überwältigenden neuen Bildern im Stil der echten Trainingsbilder führen, ob nun von Gesichtern oder anderen Dingen. Und das Beste an der ganzen Sache wäre, dass Goodfellows Ansatz uns der Notwendigkeit entheben würde, manuell Features in das generative Modell zu programmieren. Wie wir schon im Zusammenhang mit dem maschinellen Sehen (Kapitel 1) und der Verarbeitung natürlicher Sprache (Kapitel 2) ausgeführt haben, kümmert sich das Deep Learning automatisch um die Features der Modelle.

---

3. Giles, M. (21. Februar 2018). The GANfather: The man who's given machines the gift of imagination. *MIT Technology Review.*
4. LeCun, Y. (28. Juli 2016). *Quora. bit.ly/DLbreakthru*
5. Jarosz, A., et al. (2012). Uncorking the muse: Alcohol intoxication facilitates creative problem solving. *Consciousness and Cognition. 21*, 487–93.

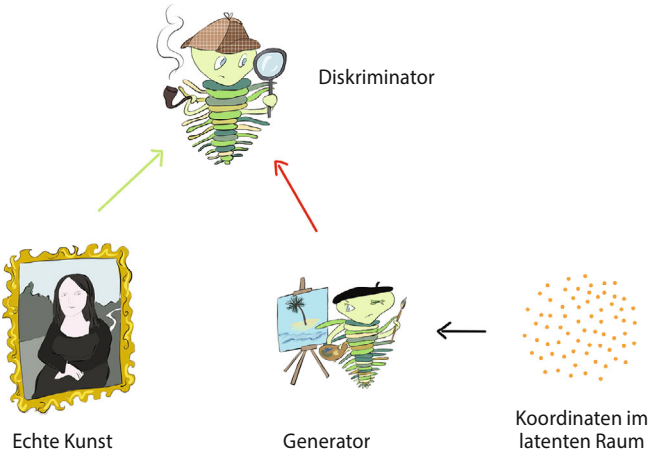

Abb. 3–1 *Schematische Darstellung eines Generative Adversarial Network (GAN). Dem Diskriminator werden sowohl echte Bilder als auch Nachahmungen vorgelegt. Er hat die Aufgabe, die echten Bilder zu identifizieren. Die orange Wolke repräsentiert die Orientierungshilfe durch den latenten Raum (Abbildung 3–4), die dem Fälscher angeboten wird. Diese Lenkung kann entweder zufällig sein (wie das beim Netzwerktraining im Allgemeinen der Fall ist; siehe Kapitel 12) oder kann selektiv (während einer Erkundung nach dem Training; siehe Abbildung 3–3) erfolgen.*

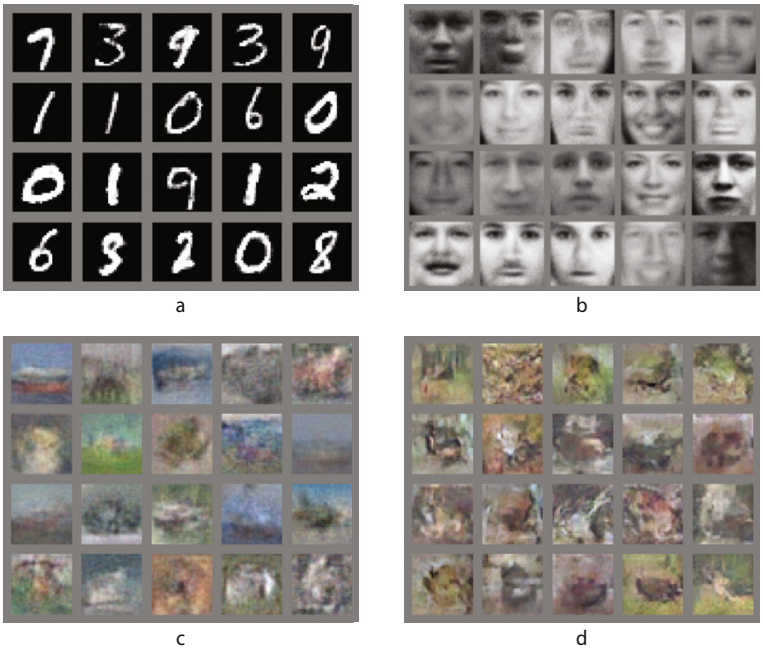

Abb. 3–2 *Ergebnisse aus dem GAN-Artikel von Goodfellow und seinen Kollegen aus dem Jahre 2014*

Goodfellows Freunde bezweifelten, dass sein fantasievolles Vorgehen funktionieren würde. Als er nach Hause kam und seine Freundin schlafend vorfand, arbeitete er bis in die Nacht daran, seinen Entwurf mit den zwei ANN umzusetzen. Das Ganze funktionierte beim ersten Versuch, und die erstaunliche Familie der Generative Adversarial Networks war geboren!

Im selben Jahr präsentierten Goodfellow und seine Kollegen GANs auf der angesehenen *Neural Information Processing Systems (NIPS)*[6] *Conference* der Öffentlichkeit. Einige ihrer Ergebnisse sind in Abbildung 3–2 zu sehen. Ihr GAN erzeugte diese neuartigen Bilder, nachdem es mit (a) handgeschriebenen Ziffern[7], (b) Fotos von menschlichen Gesichtern[8] sowie mit (c) und (d) Fotos aus zehn unterschiedlichen Klassen (z.B. Flugzeuge, Autos, Hunde)[9] trainiert worden war. Die Ergebnisse in (c) sind merklich weniger knackig als die in (d), weil das GAN, das die Bilder aus (d) produziert hatte, speziell für das maschinelle Sehen ausgerichtete Neuronenschichten besaß, sogenannte *Convolutional Layers* oder *Konvolutionsschichten*[10], während das GAN aus (c) nur einen allgemeineren Typ verwendete.[11]

3.2 Berechnungen auf nachgemachten menschlichen Gesichtern

Nach Goodfellows Einstieg bestimmte ein Forschungsteam unter der Leitung des amerikanischen Machine-Learning-Forschers Alec Radford die architektonischen Bedingungen für GANs, die zu einer deutlich realistischeren Bildherstellung führten. Einige der Beispiele für Porträts von nachgemachten Menschen, die von ihren *tiefen convolutional* GANs erzeugt wurden, sind in Abbildung 3–3 zu sehen. In ihrem Artikel demonstrierten Radford und seine Kollegen ziemlich geschickt die Interpolation durch den und die Berechnungen mit dem *latenten Raum*, der mit GANs[12] assoziiert wird. Untersuchen wir zunächst einmal, was latenter Raum ist, bevor wir uns mit Interpolationen und Berechnungen mit dem latenten Raum befassen.

---

6. Goodfellow, I., et al. (2014). Generative adversarial networks. *arXiv:1406.2661*.
7. Aus LeCuns klassischem MNIST-Datensatz, den wir selbst in Teil II benutzen werden.
8. Aus der *Toronto Face*-Datenbank der Forschungsgruppe von Hinton (Abbildung 1–15).
9. Der CIFAR-10-Datensatz, benannt nach dem *Canadian Institute for Advanced Research*, das dessen Erschaffung unterstützt hat.
10. Wir werden in Kapitel 10 ausführlicher auf diesen Schichttyp eingehen.
11. Vollständig verbundene Schichten (*Dense Layers*), die in Kapitel 4 eingeführt und in Kapitel 7 ausführlicher behandelt werden.
12. Radford et al. (2016). Unsupervised representation learning with deep convolutional generative adversarial networks. *arXiv:1511.06434v2*.

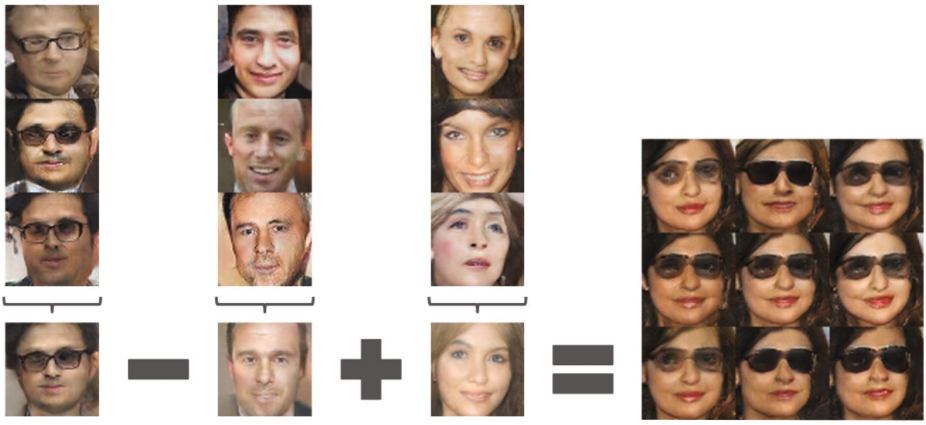

Abb. 3–3 *Ein Beispiel für die Berechnungen mithilfe des latenten Raums aus Radford et al. (2016)*

Die Skizze aus Abbildung 3–4 erinnert vielleicht an den Wortvektorraum aus Abbildung 2–6. Tatsächlich gibt es drei große Ähnlichkeiten zwischen latenten Räumen und Vektorräumen:

1. Während die Skizze aus Gründen der Einfachheit und Verständlichkeit nur einen dreidimensionalen Raum darstellt, sind latente Räume n-dimensional mit meist mehreren Hundert Dimensionen. Der latente Raum des GAN, das Sie in Kapitel 12 selbst herstellen werden, wird zum Beispiel $n = 100$ Dimensionen besitzen.
2. Je dichter zwei Punkte im latenten Raum zueinander liegen, umso ähnlicher sind die Bilder, die diese Punkte repräsentieren.
3. Eine Bewegung durch den latenten Raum in eine bestimmte Richtung kann einer schrittweisen Änderung in einem dargestellten Konzept entsprechen. Im Fall von fotorealistischen Gesichtern könnte dies das Alter oder das Geschlecht sein.

Indem wir zwei Punkte auswählen, die weit voneinander entfernt auf einer n-dimensionalen Achse liegen, die das Alter repräsentiert, könnten wir bei einer Interpolation zwischen ihnen und einer Auswahl beliebiger Punkte von dieser interpolierten Linie einen (nachgeahmten) Mann finden, der schrittweise immer älter zu werden scheint.[13] In unserer Skizze des latenten Raums (Abbildung 3–4) ist eine solche Achse für das »Alter« lila dargestellt. Wenn Sie die Interpolation durch einen authentischen latenten Raum eines GAN beobachten wollen, sollten

13. Ein technischer Hinweis: Wie es bei Vektorräumen der Fall ist, kann diese »Alter«-Achse (oder eine andere Richtung im latenten Raum, die ein sinnvolles Attribut repräsentiert) orthogonal zu allen anderen n Dimensionen liegen, die die Achsen des n-dimensionalen Raums bilden. Wir werden dies in Kapitel 11 näher ausführen.

Sie einen Blick in den Artikel von Radford und seinen Kollegen werfen und dort zum Beispiel nach den sanften Drehungen des »Photo Angle« synthetischer Schlafzimmer suchen. Zum Zeitpunkt der Entstehung dieses Buches konnte man die neuesten Entwicklungen im Bereich der GANs unter *bit.ly/InterpCeleb* anschauen. Dieses Video, das von Forschern des Grafikkartenherstellers Nvidia produziert wurde, bietet eine atemberaubende Interpolation durch hochwertige Porträt-»Fotografien« von B- und C-Promis.[14, 15]

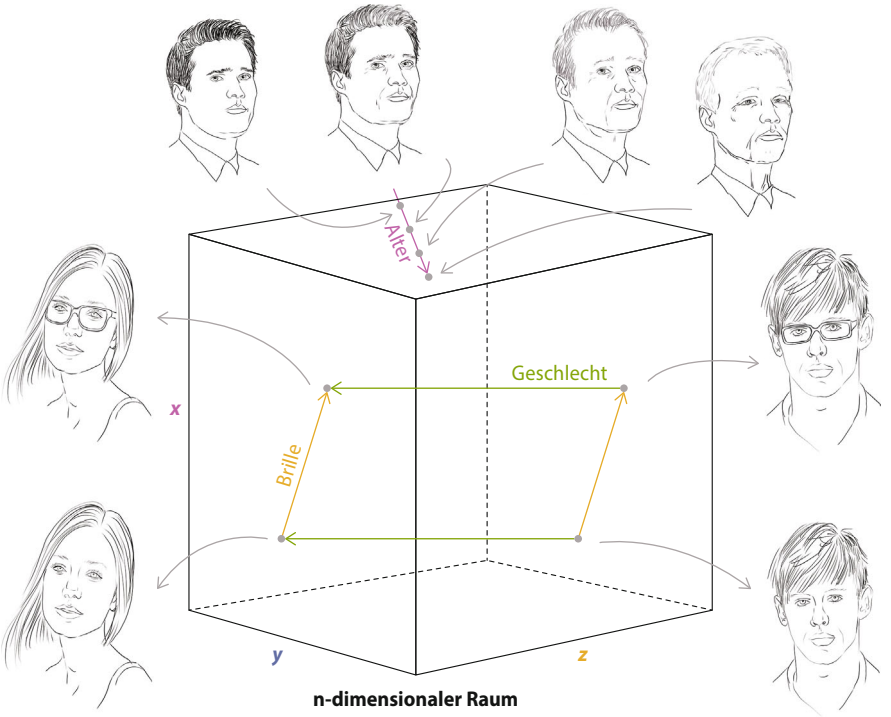

Abb. 3–4 *Eine Skizze des latenten Raums von GANs (Generative Adversarial Networks). Wenn man sich an dem lila Pfeil entlangbewegt, erhält man Bilder, die die Alterung einer Person zeigen, die ansonsten annähernd gleich aussieht. Der grüne Pfeil repräsentiert das Geschlecht und der orange Pfeil das Vorhandensein einer Brille in dem Gesicht.*

Wenn Sie nun mit dem Gelernten einen Schritt weiter gehen wollen, könnten Sie Berechnungen mit den Bildern anstellen, die Sie dem latenten Raum des GAN entnommen haben. Ein Punkt innerhalb des latenten Raums kann durch die Koordinaten seines Standortes dargestellt werden – der daraus resultierende Vek-

14. Karras, t. et al. (2018). Progressive growing of GANs for improved quality, stability and variation. *Proceedings of the International Conference on Learning Representations*.
15. Wenn Sie selbst einmal versuchen wollen, zwischen echten und GAN-generierten Gesichtern zu unterscheiden, besuchen Sie *whichfaceisreal.com*.

tor ist analog den Wortvektoren, die in Kapitel 2 beschrieben wurden. Wie bei den Wortvektoren können Sie Berechnungen mit diesen Vektoren durchführen und den latenten Raum auf semantische Weise durchlaufen. Abbildung 3–3 zeigt ein Beispiel für Berechnungen im latenten Raum aus der Arbeit von Radford und seinen Kollegen. Wir beginnen mit einem Punkt im latenten Raum ihres GANs, der einen Mann mit einer Brille repräsentiert, subtrahieren einen Punkt, der einen Mann *ohne* Brille darstellt, und addieren dann einen Punkt, der eine *Frau* ohne Brille repräsentiert. Der Punkt, den wir als Ergebnis erhalten, existiert im latenten Raum in der Nähe der Bilder, die eine Frau *mit* einer Brille darstellen. Unsere Skizze in Abbildung 3–4 verdeutlicht, wie die Beziehungen zwischen den Bedeutungen im latenten Raum gespeichert werden (auch hier wieder vergleichbar zum Wortvektorraum) und damit die Berechnungen im latenten Raum umsetzen.

3.3 Stiltransfer: Fotos in einen Monet verwandeln (und umgekehrt)

Eine besonders bezaubernde Anwendung von GANs ist der Stiltransfer. Zhu, Park und ihre Mitarbeiter aus dem *Berkeley Artificial Intelligence Research (BAIR) Lab* stellten eine neue Richtung von GAN[16] vor, die erstaunliche Beispiele dafür liefert, wie Abbildung 3–5 zeigt. Alexei Efros, einer der Koautoren des Artikels, nahm im Frankreich-Urlaub Fotos auf, und die Forscher verwandelten diese Fotos dann mithilfe ihres *CycleGAN* in Bilder, die dem jeweiligen Stil des impressionistischen Malers Claude Monet, des Niederländers Vincent van Gogh, des japanischen Genres Ukiyo-E und anderer nachempfunden sind.

Unter *bit.ly/cycleGAN* können Sie Beispiele für den umgekehrten Fall entdecken (Monet-Gemälde, die in fotorealistische Bilder umgewandelt wurden) sowie:

- Sommerszenen, die in Winterszenen verwandelt wurden, und umgekehrt
- Körbe mit Äpfeln, aus denen Körbe mit Orangen wurden, und umgekehrt
- Flache Fotos niedriger Qualität, die auf einmal aussehen, als wären sie mit High-End-(Spiegelreflex-)Kameras aufgenommen worden
- ein Video eines Pferdes, das sich in ein Zebra verwandelt
- ein Video einer am Tage aufgenommenen Fahrt, die in eine Nachtfahrt konvertiert wird

16. Sie werden auch »CycleGANs« genannt, weil sie über viele Zyklen des Netzwerktrainings die Konsistenz des Bildes erhalten. Zhu, J.-Y., et al. (2017). Unpaired image-to-image translation using cycle-consistent adversarial networks. *arXiv:1703.10593*.