NEURONALE NETZE
REVOLUTION FÜR DIE WISSENSCHAFT?

Ein Konzept aus den 1980er Jahren nimmt durch neue Entwicklungen in der Computer-Hardware einen kometenhaften Aufschwung. Vielschichtige, »tiefe« neuronale Netze revolutionieren nicht nur die Bilderkennung und die Datenanalyse, sondern gewinnen inzwischen auch wissenschaftliche Erkenntnisse.

Christoph Angerer ist Senior Compute Performance Engineer bei der Firma NVIDIA in München. Dort kümmert er sich um Anwendungen der künstlichen Intelligenz zu wissenschaftlichen Zwecken sowie die systemnahe Optimierung von Deep Learning Kernels. 2011 promovierte er an der ETH Zürich in Informatik.

spektrum.de/artikel/1520773

Am 15. März 2016 musste sich Lee Sedol, einer der weltbesten Go-Spieler, einer Maschine geschlagen geben. AlphaGo, eine künstliche Intelligenz aus dem zu Google gehörenden Unternehmen DeepMind, hatte in einem geradezu historischen Turnier vier von fünf Spielen gewonnen. Obendrein ist AlphaGo jüngst selbst geschlagen worden – von seinem verbesserten Nachfolger AlphaGoZero (siehe den Artikel S. 22).

Spätestens seit diesem Ereignis ist offensichtlich, dass die künstliche Intelligenz aus ihrem langen Winterschlaf erwacht ist. Inzwischen überschlagen sich die Meldungen über Anwendungen, die bis vor kurzem noch der Sciencefiction zugeschrieben wurden. Die Technik, deren Verbesserung diesen rasanten Fortschritt möglich gemacht hat, ist unter dem Namen »künstliches neuronales Netz« oder auch einfach »neuronales Netz« bekannt (**Spektrum** November 1992, S. 134).

Bereits wenn man eine Suchmaschine wie Google bittet, Bilder anhand einer Beschreibung in Worten zu finden, leisten neuronale Netze die wesentliche Arbeit. Doch weiterentwickelte Systeme, die generativen neuronalen Netze, gehen noch weit darüber hinaus: Sie erzeugen aus einer textuellen Beschreibung annähernd fotorealistische Bilder, die der Beschreibung entsprechen, bis dahin aber noch nicht existierten (Bilder S. 16/17 und 20). Sie malen sich diese Bilder gewissermaßen selbstständig im Geiste aus.

Andere Netze erzeugen vollständige akustische Rohdaten, die von echter Klaviermusik nicht zu unterscheiden sind, ohne dass jemals eine Taste eines Klaviers gedrückt wurde. Wieder andere haben den Stil eines Künstlers so »verinnerlicht«, dass sie ein beliebiges Foto in diesem Stil verfremden können (Bild S. 19). Neuronale Netze sind auch die Gehirne in selbstfahrenden Autos. Und selbst ein häufig belächeltes Filmklischee machen neuronale Netze inzwischen möglich: Sie verwandeln auf Knopfdruck stark verpixelte Fotos in hochauflösende Bilder (Bilder rechts).

Künstliche neuronale Netze sind heute keine akademische Nischendisziplin mehr, sondern werden in zahllosen kommerziellen Anwendungen erfolgreich eingesetzt.

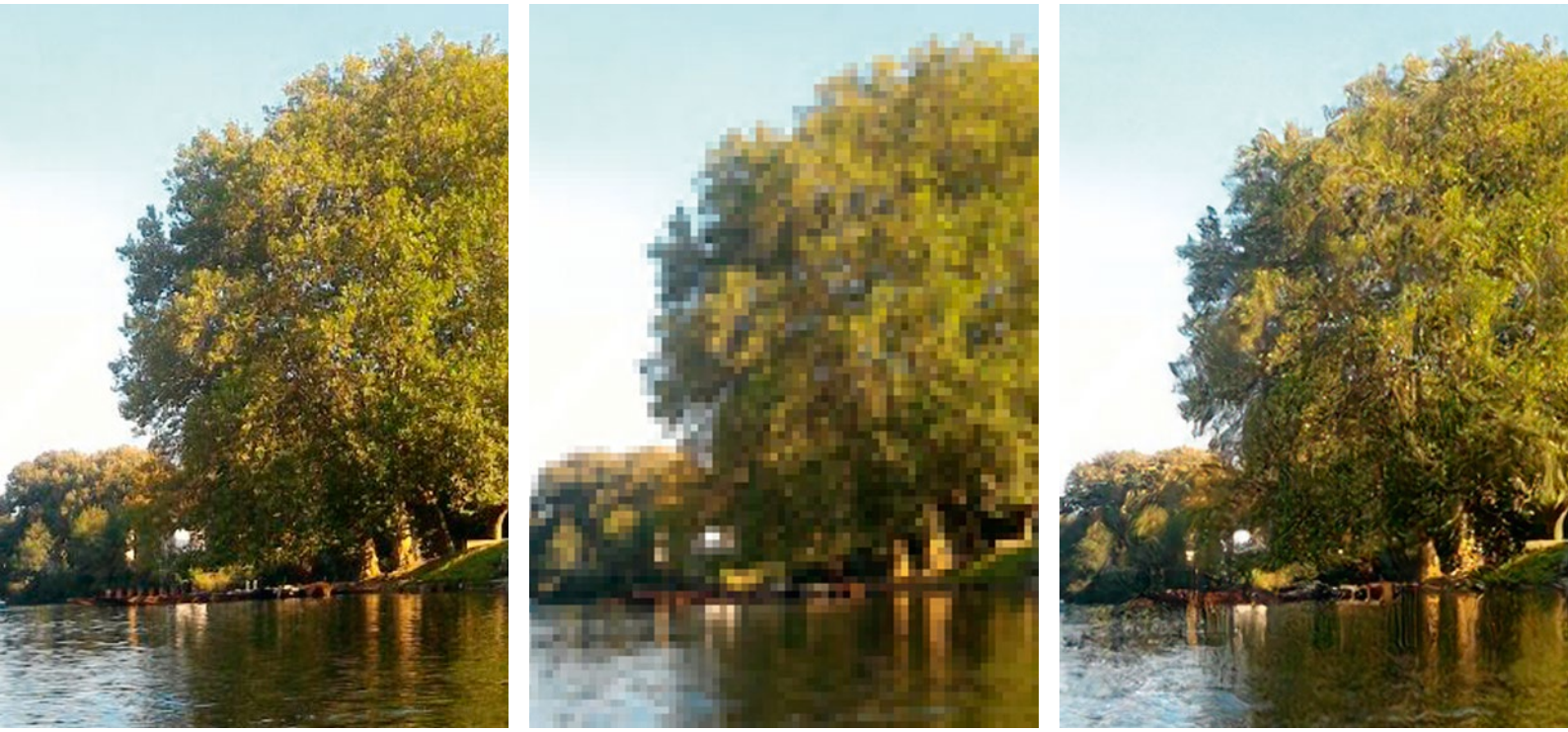
AUF EINEN BLICK
SIEG DER KÜNSTLICHEN INTELLIGENZ

1 Neuronale Netze extrahieren aus einer großen Menge von Daten – zum Beispiel Bildern – charakteristische Eigenschaften durch einen Lernprozess.

2 Dank riesiger verfügbarer Datenmengen und der durch Grafikprozessoren massiv angestiegenen Rechenleistung haben sie in letzter Zeit einen ungeahnten Aufschwung genommen.

3 Mittlerweile erzielen sie Erfolge nicht nur in Bildverarbeitung und Datenanalyse, sondern auch bei wissenschaftlichen Fragen, zum Beispiel bei Simulationen in der Quantenchemie und der Strömungsdynamik.

Aus zwei Bildern (Mitte), die wegen viel zu geringer Pixelzahl die Originale (links) nur mangelhaft wiedergeben, errechnet ein neuronales Netz Rekonstruktionen, die den Originalen nahekommen (rechts). Die dafür erforderlichen Zusatzinformation entstammen nicht dem jeweiligen Originalbild (das dem Netz nicht zur Verfügung stand), sondern allgemeinem Wissen über das Aussehen von Bäumen und Tieren, welches das Netz sich durch Betrachten zahlreicher Bilder zugelegt und in seinen Parametern gespeichert hat. Programmiert wurde das Netz von Mehdi S. M. Sajjadi vom Max-Planck-Institut für intelligente Systeme in Tübingen.



MEHDI S.M. SAJJADI, MAX-PLANCK-INSTITUT FÜR INTELLIGENTE SYSTEME

ORIGINAL: ISTOCK / ZORAN KOLUNDZIJA; BEARBEITUNG: MEHDI S.M. SAJJADI, MAX-PLANCK-INSTITUT FÜR INTELLIGENTE SYSTEME

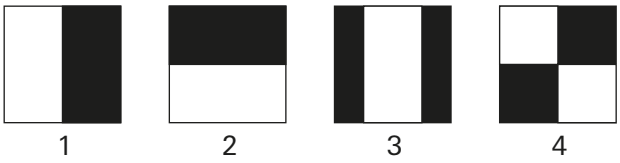
Welche Gebiete werden sie noch erobern? Und welche Rolle werden sie in der Wissenschaft spielen?

Ursprünglich war der Aufbau eines künstlichen neuronalen Netzes durch biologische Vorbilder motiviert, wie sie zum Beispiel im menschlichen Gehirn zu finden sind. Die Bezeichnung hat sich gehalten, obwohl heute die meisten künstlichen neuronalen Netze praktisch nichts mehr mit ihren natürlichen Gegenstücken zu tun haben. Nicht die Struktur, sondern die Funktion ist entscheidend. Deswegen nähert sich dieser Artikel dem Thema nicht, wie sonst üblich, über den Aufbau eines neuronalen Netzes, sondern über ein klassisches Problem der künstlichen Intelligenz: das Erkennen menschlicher Gesichter.

Standardaufgabe: Bilderkennung

Gegeben sei ein Bild in der heute üblichen Form: ein Farbwert zu jedem Element aus einer rechteckigen Anordnung von Bildpunkten (Pixeln). Gesucht sind die Positionen und Größen aller Gesichter, die in dem Bild enthalten sind – für einen Menschen trivial, für einen Computer jedoch extrem schwierig.

Im Jahr 2001 entwickelten Paul Viola und Michael Jones einen Algorithmus zur Gesichtslokalisation, der wegen seiner Schnelligkeit und Robustheit bis heute in Kameras – zum Scharfstellen auf Gesichter – Verwendung findet. Das Verfahren arbeitet mit einem Satz von Schablonen (»Basismustern«), die auf verschiedene Merkmale des menschlichen Gesichts ansprechen. Ein einfacher Basismustersatz könnte zum Beispiel wie folgt aussehen:

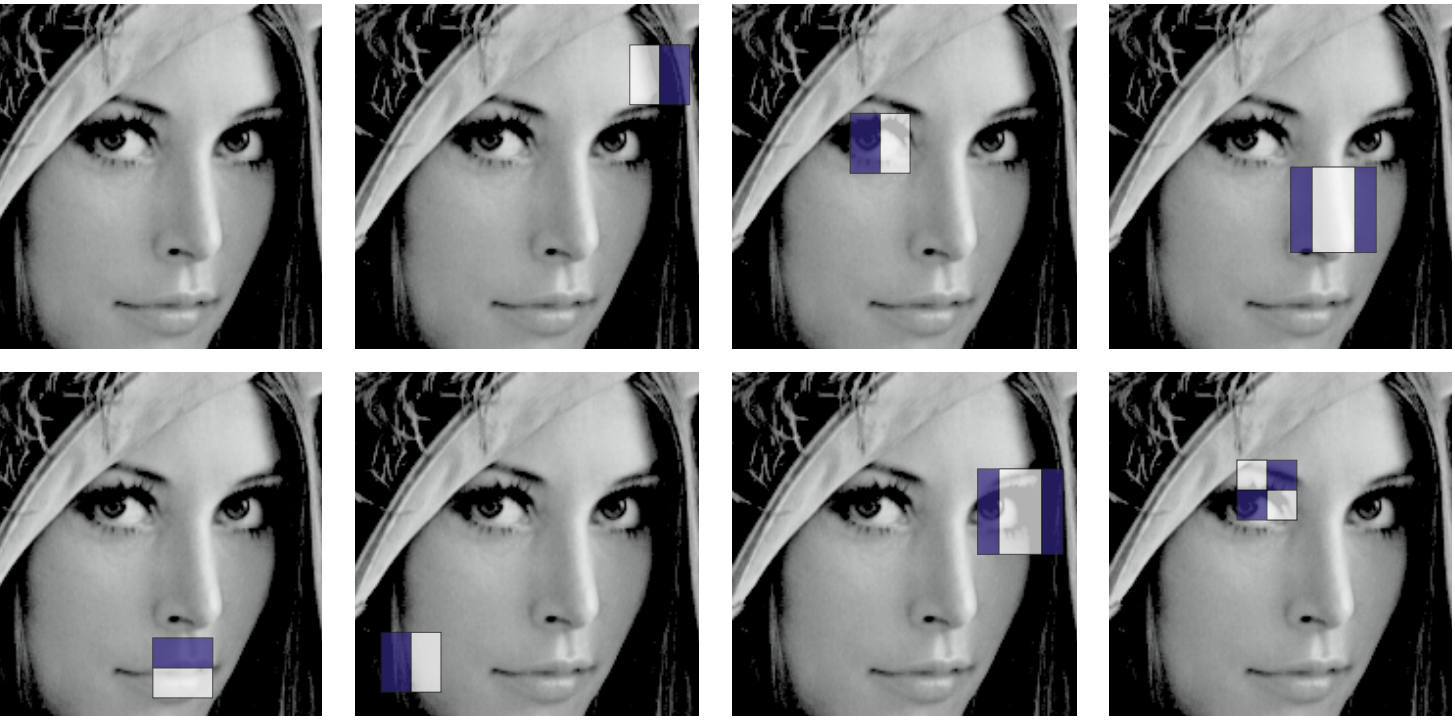


SPEKTRUM DER WISSENSCHAFT, NACH: CHRISTOPH ANGERER

Der Algorithmus setzt ein solches Basismuster der Reihe nach an jede Stelle des Bildes; dann addiert er die Helligkeiten der Pixel, die unter die schwarzen Teile des Musters geraten, und subtrahiert davon die Helligkeiten unter den weißen Teilen. (Das kann für jeden der drei Farbwerte rot, grün und blau separat geschehen, was aber am Prinzip nichts ändert.) Je extremer – positiv oder negativ – der so errechnete Wert ist, desto besser gibt das Muster den Helligkeitsverlauf an dieser Stelle wieder. Muster 3 beispielsweise wurde so gestaltet, dass es auf die Farbverläufe des menschlichen Nasenrückens anspricht, das heißt dort einen hohen Wert liefert. Im Gegensatz dazu würde Muster 2 eher auf eine Augenpartie ansprechen (Bild unten).

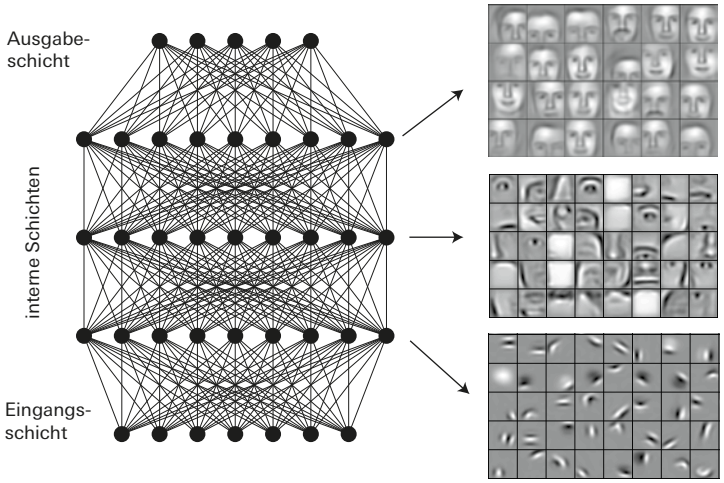
Der Algorithmus vermutet nun überall da ein Gesicht, wo alle Basismuster ähnlich stark reagieren und vor allem in der richtigen räumlichen Anordnung auftreten. Die letzte Forderung realisiert der Algorithmus im Prinzip genau so, wie er die Basismuster einsetzt: Die mit Hilfe der Basismuster errechneten Werte bilden ihrerseits wieder eine in einem rechteckigen Schema angeordnete Menge von Zahlen – formal dasselbe wie ein Bild. Über dieses Bild lässt der Algorithmus ein neues, der gesuchten räumli-

In dem Bild »Lenna«, das den Praktikern der Bildverarbeitung als Standardübungsobjekt dient, findet der Algorithmus von Paul Viola und Michael Jones zahlreiche Stellen, an denen eines seiner Basismuster (Bild oben, hier violett-weiß statt schwarz-weiß) »anspricht« (einen hohen Wert ausgibt). Aber nur dort, wo diese Stellen in der richtigen räumlichen Anordnung vorkommen, meldet der Algorithmus ein Gesicht.



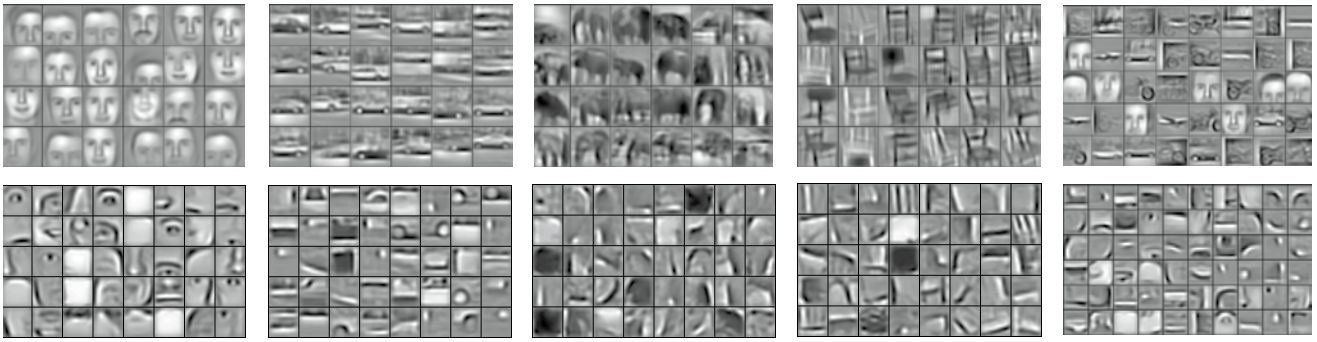
SPEKTRUM DER WISSENSCHAFT / CHRISTOPH PÖPPE, MIT DATEN VON HTTP://SIPI.USC.EDU/DATABASE/DATABASE.PHP?VOLUME=MISC&IMAGE=12

LINKS: SPEKTRUM DER WISSENSCHAFT / CHRISTOPH PÖPPE;
RECHTS: MIT FRDL. GEN. VON HONGLAK LEE; LEE, H. ET AL.: CONVOLUTIONAL DEEP BELIEF NETWORKS FOR SCALABLE UNSUPERVISED LEARNING OF HIERARCHICAL REPRESENTATIONS. IN: ICML 09 PROCEEDINGS OF THE 26TH ANNUAL INTERNATIONAL CONFERENCE ON MACHINE LEARNING, S. 609-616, 2009, TEILE AUS FIG. 3 UND 2



Ein neuronales Netz zur Gesichtserkennung hat Muster gelernt. In den untersten Schichten sprechen die Muster noch auf relativ einfache Merkmale an; je höher die Schicht, desto komplexer werden die Muster.

Ob ein neuronales Netz auf (von links nach rechts) Gesichter, Autos, Elefanten, Stühle oder auf eine Mischung verschiedenster Objekte (Gesichter, Autos, Flugzeuge, Motorräder) trainiert worden ist, macht in der untersten Schicht (untere Bildzeile) noch keinen großen Unterschied. Erst in den oberen Schichten (obere Bildzeile) machen sich die verschiedenen Lernstoffe bemerkbar.



MIT FRDL. GEN. VON HONGLAK LEE; LEE, H. ET AL.: CONVOLUTIONAL DEEP BELIEF NETWORKS FOR SCALABLE UNSUPERVISED LEARNING OF HIERARCHICAL REPRESENTATIONS. IN: ICML 09 PROCEEDINGS OF THE 26TH ANNUAL INTERNATIONAL CONFERENCE ON MACHINE LEARNING, S. 609-616, 2009, FIG. 3

chen Anordnung der Bildelemente entsprechendes Muster wandern. Auf diese Weise filtert er aus den zahlreichen Stellen, an denen einzelne Basismuster ansprechen, diejenigen heraus, an denen sie das alle zugleich und in der richtigen räumlichen Anordnung tun.

Vom Feature Engineering zum Network Engineering

Die Viola-Jones-Methode ist historisch bedeutend als eines der frühesten und schnellsten Verfahren zur Gesichtserkennung, leidet jedoch unter erheblichen Schwächen. So versagt sie, wenn ein Gesicht im Profil zu sehen ist, schräg im Bild liegt oder eine ungewöhnliche Größe hat. Dem wäre zwar von Fall zu Fall mit einem erweiterten Satz von Basisfunktionen und entsprechend höherem Rechenaufwand abzuhelfen, aber solche Nachbesserungen bleiben Flickwerk. Vor allem jedoch sind es menschliche Experten, welche die Basismuster und deren räumliche Verteilung entwerfen und nachjustieren müssen; und für eine neue Klasse von Objekten – zum Beispiel Autos – würde die ganze Arbeit von vorne anfangen. Der englische Fachbegriff für diesen Ansatz ist »feature engineering«, frei zu übersetzen als »Entwurfsarbeit mit Merkmalen«.

Dem gegenüber steht das »network engineering«: Der Experte denkt nicht mehr über Merkmale nach, sondern beschränkt sich darauf, die netzartige Struktur einer Mustererkennungsmaschine zu definieren. Diese Maschine ist anfänglich »leer«, also auch nicht fähig, irgendein Muster zu erkennen, sondern muss das erst lernen.

In seiner Struktur hat das Verfahren gewisse Gemeinsamkeiten mit der Viola-Jones-Methode. Man lässt ebenfalls einzelne Muster über das Eingabebild wandern und berechnet an jeder Position einen Wert, der angibt, wie das Bild an dieser Stelle zum Muster passt. Dieser Vorgang entspricht einer mathematischen Operation namens Faltung (englisch convolution); daher werden diese Maschinen als Convolutional Neural Network (CNN) bezeichnet, auch wenn außer der Faltung noch andere mathematische Operationen zum Einsatz kommen.

Im Gegensatz zur Viola-Jones-Methode sind die Muster allerdings nicht schwarz-weiß, sondern haben auch Grautöne. Das heißt: Die Helligkeitswerte der unter dem Muster liegenden Pixel werden nicht mit plus oder minus 1 multipliziert, bevor sie aufaddiert werden, sondern mit reellen Zahlen. Jedes Muster ist charakterisiert durch einen Satz dieser Zahlen, die als Gewichte, Parameter oder Koeffizienten bezeichnet werden. Und eben diese Parameter muss das Netzwerk erst lernen.

Ein Satz von Mustern, die gemeinsam über das Bild wandern, wird – in Anlehnung an das biologische Vorbild – als eine Schicht des neuronalen Netzes bezeichnet. Wie bei der Viola-Jones-Methode wird jedes Muster einer Schicht im Prinzip an allen Stellen des Bildes angelegt und liefert dort jeweils einen Wert. Diese Werte zusammengenommen ergeben wieder ein bildähnliches Objekt, das an die nächste Schicht weitergegeben wird. Die nachfolgenden Schichten eines solchen Netzwerks können somit auf

Dieser Vogel ist rot und braun gefärbt, mit gedrungenem Schnabel.

Der Vogel ist klein und gedrungen, mit Gelb auf dem Körper.

Vogel mit orangefarbenem Schnabel, weißem Rumpf, grauen Flügeln und Schwimmhäuten

MIT FRDL. GEN. VON DIMITRIS METAXAS; ZHANG, H. ET AL.: STACKGAN: TEXT TO PHOTO-REALISTIC IMAGE SYNTHESIS WITH STACKED GENERATIVE ADVERSARIAL NETWORKS. IN: ARXIV:1612.03242. 2017, TEILE AUS FIG. 3

Diese Vögel gibt es nicht in der Wirklichkeit. Zwei neuronale Netze, die gegeneinander arbeiten (»generative adversarial networks«), haben die Bilder jeweils nach einer Beschreibung in Worten erzeugt. Das Werk ist eine Gemeinschaftsarbeit von Han

übergeordnete Muster wie zum Beispiel räumliche Zusammenhänge ansprechen.

Im Gegensatz zu einem manuell entwickelten Basismustersatz findet ein neuronales Netz die relevanten Merkmale von selbst während eines Trainingsprozesses (siehe »Wie lernt ein neuronales Netz?«, unten). Dabei werden allerdings nur die einzelnen Parameter verändert, während die innere Struktur eines neuronalen Netzes in der Regel unangetastet bleibt. Im Beispiel würde man dem Algorithmus an Stelle eines relativ kleinen handentworfenen Basismustersets für Gesichter einen anfänglich leeren Speicherplatz für eine größere Anzahl anfangs zufälliger Muster zur Verfügung stellen. Der Lernalgorithmus füllt diese Muster mit Hilfe von Trainingsbildern, in denen die Positionen der Gesichter bekannt sind; und zwar passt er die Parameter in lauter kleinen Schritten an, um die Vorhersagegenauigkeit zu verbessern. Ein solches Netzwerk extrahiert selbstständig die für die aktuelle Aufgabe relevanten Erkennungsmuster (Bild S. 15 oben).

Die gelernten Muster auf der ersten Schicht haben in diesem Beispiel eine gewisse Ähnlichkeit mit einem Basismustersatz der Viola-Jones-Methode. Die höheren Schichten bringen diese Basismuster dann in eine räumliche Ordnung und modellieren so komplexere Formen, wie zum Beispiel Augen, Nasen oder auf der höchsten Ebene schließlich ganze Gesichter. Im Unterschied zum Feature Engineering kann jedoch ein lernfähiges Netzwerk (genug Trainingsdaten und Rechenleistung vorausgesetzt) deut-

Wie lernt ein neuronales Netz?

Man kann den Lernerfolg eines neuronalen Netzes durch eine Zahl ausdrücken. Im Beispiel der Gesichtserkennung sollen die Muster der obersten (Ausgabe-)Schicht auf Aussagen ansprechen wie »In diesem Punkt ist die Nasenspitze eines Gesichts« oder auch »Dieses Bild zeigt Angela Merkel«. Idealerweise hat ein fertig trainiertes Netz genau ein Element der obersten Schicht, das für die richtige Antwort einen hohen Wert ausgibt, während die Werte aller anderen Elemente bei null liegen.

Wie kann man erkennen, wie gut ein Netz trainiert ist? Man berechnet die so genannte Fehlerfunktion, indem man zum Beispiel für jedes Element der obersten Schicht die Differenz zwischen dem Sollwert – 1 für die richtige Antwort, 0 für alle anderen – und dem tatsächlich ausgegebenen Wert bildet und die Quadrate aller Differenzen aufaddiert. Je kleiner der Wert der Fehlerfunktion, desto besser hat das Netz gelernt. Neben der Summe über die einzelnen Fehlerquadrate sind auch andere Fehlerfunktionen in Gebrauch.

Der Wert der Fehlerfunktion hängt von den Eingabedaten und den Parametern aller Muster auf allen Schichten ab; das können mehrere tausend bis Millionen Variable sein. Lernen heißt nun, diese Parameter so zu verändern, dass der Wert der Fehlerfunktion kleiner wird.

Das gelingt, weil die Fehlerfunktion differenzierbar ist: Wenn man alle Parameter bis auf einen festhält, ist die Fehlerfunktion in Abhängigkeit von diesem einen Parameter eine glatte Kurve, die überall eine Tangente hat. Und obgleich die Ausgabe des Netzes und damit auch die Fehlerfunktion sehr kompliziert und über zahlreiche Zwischenergebnisse zu berechnen ist, lässt sich die Steigung der genannten Tangente, das heißt die Ableitung der Fehler-

Kleiner Vogel mit verschiedenen Weiß- und Brauntönen unter den Augen

Kleiner gelber Vogel mit schwarzem Kopffleck und kurzem, spitzem, schwarzem Schnabel

Kleiner Vogel mit weißer Brust und hellgrauem Kopf; Flügel und Schwanz schwarz

MIT FRDL. GEN. VON DIMITRIS METAXAS: ZHANG, H. ET AL.: STACKGAN: TEXT TO PHOTO-REALISTIC IMAGE SYNTHESIS WITH STACKED GENERATIVE ADVERSARIAL NETWORKS. IN: ARXIV:1612.03242, 2017, TEILE AUS FIG. 3

Zhang von der Rutgers University in New Brunswick (New Jersey) sowie Fachkollegen von der Lehigh University, der Chinese University of Hong Kong und der Forschungsabteilung des chinesischen Suchmaschinenbetreibers Baidu.

lich mehr Muster hinzuzufügen und über viele Schichten hinweg miteinander kombinieren, so dass es zum Beispiel nicht nur gedrehte und vergrößerte Gesichter, sondern gleichzeitig noch zahlreiche weitere Objektklassen erkennen kann. Moderne Netze (»deep networks«) haben oftmals dutzende, wenn nicht sogar hunderte Schichten und können dank ihrer Kapazität tausende unterschiedlicher Objektklassen voneinander unterscheiden (**Spektrum** September 2014, S. 62). Dabei verwenden sie, insbesondere auf den früheren Schichten, Muster über Objektklassen hinweg ganz automatisch wieder und kombinieren sie in späteren Schichten räumlich neu (Bild S. 15 Mitte).

Insgesamt hat der Schritt vom Feature Engineering zum Network Engineering Netzwerke hervorgebracht, die einfacher zu entwickeln sind und mehr Objektklassen mit höherer Genauigkeit auseinanderhalten können. In komplexen Bilderkennungsaufgaben mit vielen Unterkategorien sind diese Netze inzwischen sogar dem Menschen überlegen.

Feature Engineering dominierte nur deswegen bis vor wenigen Jahren in der Praxis, weil das Training eines komplexen neuronalen Netzes mehr Rechenleistung und größere Datenmengen benötigt, als damals zur Verfügung standen. Dies änderte sich 2012, als Alex Krizhevsky, Ilya Sutskever und Geoffrey Hinton von der University of Toronto eine neue Netzwerkarchitektur präsentierten, welche die Qualität der bisherigen Systeme weit übertraf. In dem etablierten Bilderkennungswettbewerb »Large Scale Visual

funktion nach dem Parameter, mit relativ geringem Zusatzaufwand berechnen.

Aus der so bestimmten Ableitung der Fehlerfunktion nach jedem Parameter gewinnt man eine Anweisung, in welche Richtung und wie stark man jeden Parameter verändern muss, um den Gesamtfehler etwas in Richtung Minimum zu treiben. Nach der Kettenregel aus der Differenzialrechnung berechnet man diese Ableitungen unter Verwendung der Zwischenergebnisse zuerst für die Parameter der letzten Schicht und dann Schicht für Schicht nach vorne, entgegen der Richtung, in der das Netz eigentlich gearbeitet hat. Die Information über den Fehler wird »nach rückwärts« verbreitet, weswegen das Verfahren Backpropagation genannt wird.

Man schickt nun eine genügende Anzahl von Trainingsdaten immer wieder vorwärts durch das Netz und passt jedes Mal die Parameter mittels Backpropagation an. Dann bilden mit der Zeit die einzelnen Teile des Netzwerks bestimmte Muster, die ihrerseits Muster in den Eingabedaten wiedergeben.

Eine natürliche Nervenzelle sendet einen Impuls aus (sie »feuert«), wenn die Summe der Eingangssignale einen gewissen Schwellenwert überschreitet, und tut sonst nichts. Wollte man dieses Verhalten in einem künstlichen neuronalen Netz nachbilden, müsste man in jedem Element eine Funktion einbauen, die für alle x-Werte unterhalb der Schwelle den Wert 0 und oberhalb den Wert 1 annimmt. Eine solche Stufenfunktion ist aber noch nicht einmal stetig, geschweige denn differenzierbar, und macht daher die Backpropagation undurchführbar. Stattdessen verwendet man für diese Aufgabe verschiedene nichtlineare Funktionen, die üblicherweise differenzierbar sind.

Recognition Challenge« (LSVRC) konnte ihr Netz die Fehlerrate von einem Jahr zum nächsten annähernd halbieren (von 26 auf 15 Prozent).

Die wichtigste Neuerung der drei Forscher bestand darin, im Trainingsprozess anstatt herkömmlicher Prozessoren (CPUs) Grafikkarten (GPUs) zu benutzen. Diese Chips, die ursprünglich zum Ansteuern von Bildschirmen entworfen wurden, besitzen zwar nur Miniprozessoren mit beschränkten Rechenfähigkeiten, davon jedoch extrem viele. So können GPUs zahlreiche Rechenoperationen gleichzeitig durchführen. Krizhevsky, Sutskever und Hinton gelang es, die unzähligen Berechnungen für das Training, namentlich den Backpropagation-Algorithmus (siehe »Wie lernt ein neuronales Netz?«, S. 16/17), so zu programmieren, dass GPUs sie bewältigen konnten. Die damit erreichte hohe Rechenleistung erlaubte es den Forschern, die Trainingszeit dramatisch zu reduzieren und zugleich die Komplexität des Netzwerks deutlich zu erhöhen.

Seit diesem überwältigenden Erfolg nehmen GPU-beschleunigte Netze stetig an Zahl und Komplexität zu. Inzwi-

Ein Zoo von Netzwerken

Der Aufbau künstlicher neuronaler Netze ist ursprünglich inspiriert von biologischen Nervennetzwerken, wie sie sich zum Beispiel im menschlichen Gehirn finden. Inzwischen haben sie sich von ihrem biologischen Vorbild weitgehend gelöst.

Es gibt eine fast unüberschaubare Vielzahl neuronaler Netze für die verschiedensten Einsatzgebiete, und fast wöchentlich kommen neue Varianten hinzu. Daher ist eine genaue und erschöpfende Einteilung schwierig. Ein mögliches Kriterium ist das Signal, das dem Netz während des Lernvorgangs den Fehler seiner Vorhersage mitteilt.

Die im praktischen Einsatz gängigsten Netze benutzen überwachtes Lernen (supervised learning). Dabei hat jeder Datensatz der Trainingsdaten einen Zielwert (Label), auf den das Netz trainiert wird. Der Nachteil dieser Vorgehensweise ist, dass es sehr aufwändig sein kann, Millionen von Datensätzen mit qualitativ hochwertigen Zielwerten bereitzustellen.

Beim unüberwachten Lernen (unsupervised learning) haben die Datensätze keine Zielwerte. Stattdessen versucht das Netz, strukturierte Muster in den Eingabedaten zu erkennen.

Bestärkendes Lernen (reinforcement learning) kommt hingegen ganz ohne Trainingsdaten aus. Stattdessen lernt das Netz »on the job«, indem es, während es die eigentliche Aufgabe ausführt, Rückmeldung über die Qualität der Vorhersagen bekommt und sich intern so anpasst, dass die Qualität mit der Zeit anwächst.

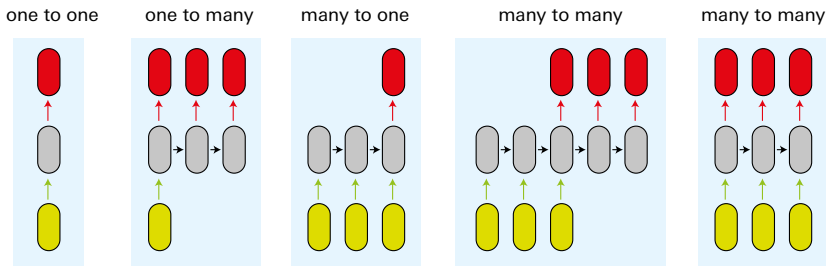
Eine weitere mögliche Kategorisierung orientiert sich an den Eingaben und Ausgaben der Netze:

Klassische Objekterkennung nimmt ein Bild als Eingabe und erzeugt eine Aussage über das abgebildete Objekt als Ausgabe (one to one). Die Erzeugung einer textuellen Bildbeschreibung bildet hingegen ein einzelnes Bild auf eine Sequenz von Wörtern ab (one to many). Umgekehrt kann auch eine solche Sequenz als Eingabe für ein Netz dienen, das dann eine Einschätzung der diesem Text zu Grunde liegenden Stimmung liefert (many to one).

Die Abbildung mehrerer (eventuell zeitbezogener) Eingaben auf mehrere Ausgaben (many to many) kommt beispielsweise in der klassischen Textanalyse vor.

Für zeitabhängige Eingaben, zum Beispiel Bildfolgen aus einem Film, werden häufig so genannte rückgekoppelte neuronale Netze (recurrent networks) verwendet: Deren Eingabe ist nicht nur der Datensatz des aktuellen Zeitschritts, sondern auch der interne Zustand des letzten Zeitschritts (und damit indirekt auch dessen Datensatz sowie der interne Zustand aller vorhergehenden Zeitschritte). Das führt zu einer Rückkopplung, die es dem Netz ermöglicht, Muster über mehrere Eingaben hinweg zu erkennen. Der Trainingsprozess unterscheidet sich jedoch nicht grundsätzlich von nicht rückgekoppelten Netzen.

Beide Klassifikationen sind unvollständig. Zahlreiche interessante Netzwerkarchitekturen passen in keines der Schemata, zum Beispiel generative Netzwerke, die Daten nicht nur verarbeiten, sondern erzeugen (Bilder S. 16/17 und 20; siehe auch **Spektrum** Dezember 2015, S. 86).



SPEKTRUM DER WISSENSCHAFT, NACH: CHRISTOPH ANGERER

MIT FREUNDLICHER GENEHMIGUNG VON JUN-YAN ZHU, J.-Y. ET AL.: UNPAIRED IMAGE-TO-IMAGE TRANSLATION USING CYCLE-CONSISTENT ADVERSARIAL NETWORKS. IN: ARXIV:1703.10593, 2017, TEILE AUS FIG. 11

Neuronale Netze können den Malstil eines Künstlers lernen und auf neue Bilder anwenden. Hier von links nach rechts die Tübinger Neckarfront im Foto sowie im Stil von Monet, van Gogh und Cézannne.

schen stellen sie die Mehrheit der Beiträge bei solchen Wettbewerben. Während das neuronale Netz von 2012, genannt AlexNet, noch aus acht Schichten mit 650 000 Neuronen und ungefähr 60 Millionen Parametern bestand, bringen es neuere Netze wie das ResNet-152 auf mehr als 100 Schichten, was auf den ungefähr zehnfachen Bedarf an Rechenleistung hinausläuft. Um solche »tiefen« Netzwerke trainieren zu können, mussten ihre Konstrukteure nicht nur zahlreiche Verbesserungen in den Algorithmen vornehmen, sondern auch die Hardware, in diesem Fall die GPUs, anpassen und die zugehörige Software neu entwickeln.

Wissenschaftliche Datenanalyse

Neuronale Netze bearbeiten heute riesige Datenmengen in kommerziellen Anwendungen und sind ein wichtiges Werkzeug der Datenanalyse. Doch die Anzahl der verfügbaren Daten ist in den letzten Jahren nicht nur im kommerziellen Sektor explodiert, auch die Wissenschaft produziert immer mehr Daten, die ausgewertet und aufbereitet werden wollen. Hier bieten neuronale Netze einzigartige neue Möglichkeiten.

Ein Beispiel ist die Suche nach Gravitationswellen im Advanced Laser Interferometric Gravitational Wave Observatory (Advanced LIGO). Die von Einstein bereits 1916 vorhergesagten Gravitationswellen wurden erst 2015 direkt nachgewiesen (**Spektrum** April 2016, S. 12; zum Nobelpreis für diese Leistung siehe **Spektrum** Dezember 2017, S. 24). Ihr messbarer Effekt ist winzig klein: Es kommt auf Längenveränderungen an, die einem Bruchteil des Protonendurchmessers entsprechen.

Diese winzigen Signale müssen jedoch aus einem reißenden Strom verrauschter Daten herausgefiltert werden. Für den ersten Nachweis erledigten diese Aufgabe noch klassische Datenverarbeitungsalgorithmen, die auf tausenden CPUs liefen. Auf Grund des großen Rechenaufwands kamen die Computer mit dem Filtern nicht nach.

Durch den Einsatz eines tiefen neuronalen Netzes ist es den Wissenschaftlern am advanced LIGO inzwischen gelungen, die Messgenauigkeit zu erhöhen und gleichzeitig die Anforderungen an die Rechenleistung um einen Faktor 1000 zu reduzieren, so dass die Datenanalyse nun in wenigen Mikrosekunden abläuft. Bei der Kollision zweier Neutronensterne, die sich am 17. August 2017 durch Gravitationswellen bemerkbar machte, gelang die Auswertung so rasch, dass die LIGO-Astronomen binnen Minuten zahlreiche andere Observatorien auf das Ereignis hinweisen konnten.

Große Mengen an Daten erzeugt auch das europäische Kernforschungszentrum CERN. In dessen größtem Teilchenbeschleuniger, dem Large Hadron Collider (LHC), finden um die 600 Millionen Teilchenkollisionen pro Sekunde statt, was einem Rohdatenstrom von mehr als einem halben Petabyte pro Sekunde entspricht; ausgedruckt wären das 250 Milliarden DIN-A4-Seiten. Um diese Datenmengen handhaben zu können, werden die Kollisionen in einem mehrstufigen Prozess gefiltert. Am Ende bleiben pro Sekunde 100 bis 200 potenziell interessante Ereignisse übrig. Aus diesen Rohdaten rekonstruieren die Forscher dann die Kollisionsprozesse.

Damit kommt der Trennschärfe der Filteralgorithmen eine entscheidende Bedeutung zu. Was diese verwerfen, bleibt unerforscht, da einmal ausgefilterte Ereignisse unwiderruflich verloren sind. Andererseits können die Algorithmen gewissermaßen nicht beliebig genau hinschauen, weil sie sonst den Datenstrom nicht in der verfügbaren Zeit bewältigen würden.

Mit dem geplanten High Luminosity Upgrade des LHCs soll die Anzahl der Kollisionen nochmals um einen Faktor 10 erhöht werden. Bereits heute ist abzusehen, dass dies neue Ansätze zur Datenanalyse erfordert. Neuronale Netze haben das Potenzial, diese Aufgabe in der notwendigen Größenordnung und Geschwindigkeit durchzuführen, ohne die relevanten Signale aus den Augen zu verlieren. Dies wäre die Basis für Entdeckungen, die unser Verständnis der fundamentalen Physik verändern könnten.

Bisher haben neuronale Netze ihre Stärken im Wesentlichen dort gezeigt, wo es Fragen zu beantworten gilt, für die keine gute Theorie zur Verfügung steht: Welches sind die charakteristischen Merkmale eines Gesichts (in Pixeln ausgedrückt)? Wodurch unterscheiden sich Signale von Gravitationswellen von den allgegenwärtigen Störungen? Woran erkennt man ein interessantes Kollisionsereignis im LHC im Gegensatz zu einem gewöhnlichen? Für die Gesichtserkennung haben die Experten mit dem Feature Engineering so etwas wie eine Theorie aufgestellt, jedoch nur mit beschränktem Erfolg. Dagegen hat ein neuronales Netz im Training eine Art Theorie der Gesichtserkennung gelernt – die in den Parametern seiner Muster besteht – und kann sie anwenden.

Neu und überraschend ist, dass neuronale Netze auch dort phänomenale Erfolge einfahren, wo es eine gut ausgearbeitete Theorie gibt.

Als Beispiel zur Einführung möge ein sehr einfacher physikalischer Prozess dienen: der Wurf im Schwerefeld

zahlreiche kleine violette Blütenblätter in kugelförmiger Anordnung

Diese Blüte ist pink, weiß und gelb mit gestreiften Blütenblättern.

dunkelpinke Blütenblätter mit weißen Rändern und pinken Staubfäden

weiß und gelb, mit gekrümmten, glattrandigen Blütenblättern

MIT FRDL. GEN. VON DIMITRIS METAXAS: ZHANG, H. ET AL.: STACKGAN: TEXT TO PHOTO-REALISTIC IMAGE SYNTHESIS WITH STACKED GENERATIVE ADVERSARIAL NETWORKS. IN: ARXIV:1612.03242, 2017. TEILE AUS FIG. 4

Die »adversarial networks« der KI-Forscher um Han Zhang können, entsprechend trainiert, auch Blüten von nicht existierenden Pflanzen realistisch darstellen.

der Erde. Die Theorie dazu lernt man in der Schule: Kraft ist Masse mal Beschleunigung; Letztere ist konstant und gleich 9,81 m/s^2. Aus diesen Zutaten erhält man mit ein wenig elementarer Differenzialrechnung eine Gleichung, die für jede Kombination aus Anfangsort und -geschwindigkeit die Bahn des geworfenen Gegenstandes angibt. Das ist der (nur selten realisierte) Idealfall, in dem man einen physikalischen Prozess aus elementaren Naturgesetzen (»ab initio«) herleitet.

Neuronale Netze sind parametrisierte Modelle

Allerdings konnten die Leute schon gezielt werfen, bevor Newton die theoretischen Grundlagen bereitstellte; und die Kanoniere der frühen Neuzeit hatten durch geduldiges Probieren Tabellen und gelegentlich sogar Formeln erstellt, mit deren Hilfe sie zu einem zu beschießenden Ziel den Anstellwinkel und die Feuerkraft ihrer Waffe bestimmen konnten. Viele physikalische Prozesse stellt man durch aus der Erfahrung gewonnene Formeln dar, in die man in der konkreten Situation noch spezifische Werte für gewisse Parameter – Reibungskoeffizient, elektrische Leitfähigkeit und so weiter – einsetzen muss.

Der offensichtliche Nachteil einer rein empirischen (Tabelle) oder semiempirischen (Formel mit Parameter) Methode ist, dass man sie schlecht verallgemeinern kann. Typischerweise funktioniert sie nur für Situationen, die jenen ähneln, für die man Experimente durchgeführt hat. Wer noch nicht mit sehr schweren Objekten gearbeitet hat, kann für die nächste Filmproduktion nicht vorhersagen, wie sich ein Auto verhalten wird, das über eine Klippe stürzt.

Semiempirische Methoden (auch parametrisierte Methoden genannt) versuchen gewissermaßen, einen Kompromiss zwischen ab-initio- und empirischen Methoden zu finden. Man modelliert einen physikalischen Prozess bewusst nicht vollständig explizit, sondern ersetzt das Modell durch einfacher zu berechnende Annäherungen.

Die Grenze zwischen semiempirischen und ab-initio-Modellen ist fließend: Die Erdbeschleunigung g ist keine Naturkonstante, sondern variiert ortsabhängig um einige Promille, ist also in Wahrheit ein Parameter eines semiempirischen Modells. Streng genommen müsste man deshalb für eine reine ab-initio-Berechnung eines gewöhnlichen Steinwurfs die ganze Erde samt abgeplatteter Kugelgestalt und Massenverteilung modellieren – eine absurde Vorstellung.

Um ein semiempirisches Modell aufzustellen, braucht es bislang noch einen Experten, der sich im jeweiligen Fachgebiet auskennt. Und selbst der liefert keine fertige Formel, sondern nur deren interne Struktur: eine Näherungsformel mit vorläufig unbekannten Parametern, die noch durch einen Feinabstimmungsprozess so einzustellen sind, dass die auf diese Weise modellierten Daten die Beobachtungen möglichst genau wiedergeben. Oftmals haben die Wissenschaftler eine gute Vorstellung über die Bedeutung einzelner Parameter; aber diese müssen keine direkte Entsprechung in dem darunterliegenden ab-initio-Modell haben.

Das erinnert bereits stark an den oben beschriebenen Trainingsprozess neuronaler Netze. Und das ist kein Zufall: Ein neuronales Netz ist nichts anderes als ein parametrisiertes Modell. Nur sind die Parameter weitaus zahlreicher und umfassen auch die interne Struktur der Formeln, die bislang der Experte vorgibt. Damit ist die Situation vergleichbar dem Übergang vom Feature Engineering zum Network Engineering: Das neuronale Netz übernimmt die Struktur seines Modells nicht vom Experten, sondern lernt sie anhand von zahlreichen Beispielen – und erzielt damit oft ein besseres Ergebnis, weil es weitaus mehr Möglichkeiten in Betracht zieht.

Es gibt jedoch einen gravierenden Unterschied zwischen den klassischen und den neuen Einsatzgebieten neuronaler Netze: In einem Gebiet wie der Bild- oder Spracherkennung ist der zu Grunde liegende »Prozess« (falls es denn überhaupt einen gibt) nicht bekannt. Neuronale Netze sind einfach das beste Werkzeug, das wir haben, um diese Probleme zu lösen. In weiten Teilen der Datenanalyse kann man ebenfalls den Strom der Daten nicht als Messwerte eines physikalischen Prozesses auffassen; allenfalls hat man ein statistisches Modell eines solchen. Wenn ein neuronales Netz jedoch eine ab-initio-Rechnung oder ein bekanntes semiempirisches Modell ersetzen soll, sind diese Prozesse nicht nur bekannt, sondern exakt mathematisch formuliert. Nunmehr ist das Netz nicht mehr konkurrenzlos, sondern muss sich an der Qualität der bereits vorliegenden Ergebnisse (der »ground truth«) messen lassen. Das ist nicht nur eine Herausforderung, sondern auch eine Chance: Das exakte Wissen über die Ground Truth könnte in der Zukunft die Entwicklung eines entsprechenden neuronalen Modells befördern.

Kann sich die Geschichte wiederholen? Werden tiefe neuronale Netze in Zukunft die parametrisierten Modelle aus der Physik ablösen und sogar komplexe ab-initio-Kalkulationen mit höherer Qualität und größerer Flexibilität ausführen können?

Das sieht, bei allen Vorbehalten gegenüber Vorhersagen, gegenwärtig ganz so aus. Als universelle Funktionsapproximatoren sind tiefe neuronale Netze gut ausgestattet, um gängige parametrisierte Modelle nachzubilden. Es scheint auch plausibel, dass sie bei einem Bruchteil des Rechenaufwands so manche ab-initio-Kalkulation mit gleicher Qualität ersetzen können, für die es bisher keine brauchbaren parametrisierten Beschreibungen gibt. Und dies ist in der Tat auf einzelnen Gebieten schon gelungen, insbesondere in der Quantenchemie und in der numerischen Strömungsmechanik.

Wie können wir wissen, ob ein neuronales Netz die Schrödinger-Gleichung gelernt hat?

Quantenchemische Simulationen modellieren chemische Reaktionen detailliert, indem sie die Bewegung jedes einzelnen Atoms in einem – möglicherweise sehr großen – Molekül beschreiben. Sie sind ein unverzichtbares Werkzeug für die Material- und Pharmaforschung (**Spektrum** November 2014, S. 86). Das grundlegende Naturgesetz für eine ab-initio-Rechnung ist bekannt: Es handelt sich um eine komplexe partielle Differenzialgleichung, die so genannte Schrödinger-Gleichung aus der Quantenmechanik. Um grundlegende Eigenschaften von Festkörpern und Molekülen, beispielsweise Bindungslängen und -energien, zu berechnen, müsste man eine komplexe Vielteilchen-Variante der Schrödinger-Gleichung lösen. Das übersteigt bei Weitem die Kapazität selbst der modernsten Computer; aber mit der so genannten Dichtefunktionaltheorie (DFT) lässt sich der Rechenaufwand drastisch reduzieren (**Spektrum** Dezember 1998, S. 24). Selbst mit DFT reicht die Rechenfähigkeit moderner Supercomputer allerdings nur für wenige tausend Atome aus.

Justin S. Smith, Olexandr Isayev und Adrian E. Roitberg von der University of Florida haben nun ein neuronales Netz entwickelt, das für ihre Beispielrechnungen die gleiche Genauigkeit wie DTF erreicht, jedoch den Rechenaufwand um erstaunliche sechs Größenordnungen reduziert. Eine Simulation, die zuvor Minuten dauerte, kann nun in Mikrosekunden durchgeführt werden. Es bleibt jedoch abzuwarten, ob diese Lösung auch außerhalb der getesteten Systeme funktioniert.

Forscher des Projekts »Google Brain« und, unabhängig davon, von Disney Research benutzen datengetriebene maschinelle Lernalgorithmen, um Strömungen zu berechnen. Auch für das Verhalten von Flüssigkeiten und Gasen gibt es ein Naturgesetz, und zwar die so genannten Navier-Stokes-Gleichungen (**Spektrum** April 2009, S. 78). Die Lösung dieser Gleichungen wird jedoch für alle bis auf die einfachsten Situationen schnell extrem rechenaufwändig. Den weit überwiegenden Teil der Rechenzeit verbraucht dabei das Lösen eines großen linearen Gleichungssystems, der Poisson-Gleichung. Die Google-Forscher trainierten Netze mit Hilfe simulierter Beispieldaten darauf, Näherungslösungen dieser Hilfsgleichungen zu finden – anhand von Beispielen gelöster Gleichungen.

Die Fachkollegen bei Disney modellieren das strömende Medium nicht als Kontinuum wie bei den Navier-Stokes-Gleichungen, sondern als Ensemble von sehr vielen Teilchen – Größenordnung Millionen. Ihre künstliche Intelligenz lernt, aus dem momentanen Zustand (Ort und Geschwindigkeit) aller dieser Partikel den Zustand eine gewisse Zeit später zu ermitteln. Die resultierenden Systeme rechnen um Größenordnungen schneller als die herkömmlichen, was völlig neue Einsatzgebiete eröffnet.

Nahezu in allen wissenschaftlichen Disziplinen lässt sich derzeit eine ähnliche Entwicklung beobachten: Ob Wetter- und Klimaforschung, Astronomie, Hochenergiephysik oder Molekularbiologie, überall finden sich ständig neue Einsatzgebiete neuronaler Netze.

Das wirft wichtige Fragen auf: Wie können wir sicher sein, dass das trainierte Netzwerk wirklich die Schrödinger-Gleichung löst und die Navier-Stokes-Gleichungen gelernt hat? Und können wir Fehlergrenzen oder andere Qualitätsmerkmale der Lösung angeben?

Eigentlich ist ein neuronales Netz so transparent wie nur möglich. Schließlich können wir jeden Parameter ablesen und jede der deterministischen Rechnungen genauestens nachvollziehen. Dennoch wird es oftmals als »Black Box« beschrieben, als schwarzer Kasten, der jeglichen Einblick in sein Inneres verwehrt. Das liegt daran, dass schwer nachzuvollziehen ist, wie das Netz die Parameter genau gelernt hat und welche Neurone im Netzwerk (falls überhaupt) welche Teile der Physik nachbilden. Man kann zwar Netze mit einzelnen Eingaben füttern und deren Ausgabe überprüfen, aber ein analytischer Ansatz, der solche Fragen erschöpfend beantworten könnte, fehlt derzeit noch gänzlich. Im Extremfall könnte es passieren, dass ein neuronales Netz Ereignisse stets korrekt voraussagt, für die es noch keine vollständige Theorie gibt – sagen wir Teilchenkollisionen am LHC. Also steckt in seinen Parametern eine physikalische Theorie, jedenfalls gemäß der klassischen Philosophie, nach der sich eine gute Theorie dadurch auszeichnet, dass sie genau solche Vorhersagen macht. Nur sind wir unfähig, aus diesen Parametern die Theorie herauszulesen und in Worte zu fassen.

Die steigende Popularität neuronaler Netze in der Forscherwelt lässt jedoch hoffen, dass derartige analytische Methoden in Zukunft entwickelt werden. Am Ende sind auch neuronale Netze lediglich ein Werkzeug von vielen im Arsenal wissenschaftlicher Methoden. Sie sollten stets als Ergänzung und nicht als Ersatz bestehender Ansätze angesehen werden. ◂

QUELLEN

Krizhevsky, A. et al.: ImageNet Classification with Deep Convolutional Neural Networks. https://papers.nips.cc/paper/4824-imagenet-classification-with-deep-convolutional-neural-networks.pdf

Sajjadi, M. S. M. et al.: EnhanceNet: Single Image Super-Resolution Through Automated Texture Synthesis. https://arxiv.org/abs/1612.07919

Zhang, H. et al.: StackGAN: Text to Photo-realistic Image Synthesis with Stacked Generative Adversarial Networks https://arxiv.org/pdf/1612.03242.pdf

Zhu, J.-Y. et al.: Unpaired Image-to-Image Translation using Cycle-Consistent Adversarial Networks. https://arxiv.org/abs/1703.10593