

Hautarzt

<https://doi.org/10.1007/s00105-020-04663-7>

© Der/die Autor(en) 2020

Marc Pouly¹ · Thomas Koller¹ · Philippe Gottfrois² · Simone Lionetti¹¹Informatik, Hochschule Luzern, Rotkreuz, Schweiz²Department of Biomedical Engineering, University of Basel, Allschwil, Schweiz

Künstliche Intelligenz in der Bildanalyse – Grundlagen und neue Entwicklungen

Bildanalytische Verfahren der künstlichen Intelligenz erreichen heute in vielen Fällen die für den industriellen und klinischen Einsatz geforderte Genauigkeit und übertreffen sogar teilweise die Leistung menschlicher Experten. Die Autoren beschreiben ihre rasante technische Entwicklung, wie sich Deep Learning von klassischen Ansätzen des maschinellen Lernens unterscheidet, erörtern die Bedeutung von Transfer Learning für die Industrialisierung und folgen der aktuellen Forschung hin zu Explainable AI (Artificial Intelligence) und dem Potenzial von Artificial Creativity.


Im Januar 2017 sorgte eine Veröffentlichung der Stanford University in *Nature* für große mediale Aufmerksamkeit. Ein künstliche Intelligenz erreichte eine zu 21 menschlichen Dermatologen vergleichbare Sensitivität und Spezifität in der Unterscheidung zwischen „keratinozyte carcinomas“ und „benign seborrhoeic keratoses“ sowie zwischen „malignant melanomas“ und „benign nevi“ auf histologisch verifizierten Testbildern [1]. Mehr noch als ihre Leistungsfähigkeit beeindruckte zum damaligen Zeitpunkt die verwendete Methodik: Ein sog. neuronales Netz wurde auf 1,41 Mio. Bildern von Alltagsgegenständen, Tieren, Pflanzen und Personen trainiert und erlernte so eigenständig die Erkennung von charakteristischen Bildstrukturen. Anschließend wurde das gleiche System nochmals

auf 130.000 klinischen Bildaufnahmen mit über 2000 unterschiedlichen dermatologischen Erscheinungen spezialisiert. Drei Jahre später, im September 2019, veröffentlichte Google eine zur Differenzialdiagnose über die 26 in der Erstversorgung häufigsten Hauterscheinungen befähigte künstliche Intelligenz, die erstmals Bildaufnahmen in Kombination mit Anamnesedaten analysierte [17]. Eine gleichzeitig erschienene Meta-studie identifizierte 14 weitere Systeme, die in einer externen Validierung den Leistungen menschlicher Fachexperten zumindest gleichkamen [16]. Im Januar 2020 übertraf eine künstliche Intelligenz von DeepMind 6 Radiologen in der Früherkennung von Brustkrebs auf 500 histologisch geprüften Mammographie-aufnahmen, wobei die menschlichen Experten, wenn verfügbar, zusätzlich auf Anamnesedaten und frühere Aufnahmen zurückgreifen durften [19]. Im März 2020 bestätigte eine klinische Studie eine zu menschlichen Experten 3fach erhöhte Erfolgsrate einer künstlichen Intelligenz in der Früherkennung von Osteoporose bei Knochenbrüchen [10]. Mit diesen Forschungsergebnissen einhergehend, entwickelte sich auch deren Industrialisierung. Wegweisend dazu war die erstmalige Zulassung einer auf Bildanalyse mit künstlicher Intelligenz basierten Software zur Erkennung von diabetischer Retinopathie durch die amerikanische Arzneimittelbehörde im April 2018 [9]. Es handelte es sich dabei um einen Cloud Service, der ohne menschliches Zutun eine Bildaufnahme der Retina analysiert, den Schweregrad bestimmt und dem Erstversorger ei-

ne Handlungsempfehlung abgibt. Damit war „AI Big Business in Health“ geboren!

In diesem Beitrag möchten wir Sie zu einem kurzen Spaziergang durch die technische Entwicklung der modernen künstlichen Intelligenz einladen, die diese beeindruckenden Resultate ermöglicht hat und fortschreitend den klinischen Alltag immer mehr durchdringen wird. Alles begann zum Jahreswechsel 2016 – dem Annus Mirabilis der künstlichen Intelligenz.

2016 – Als Computer das Sehen lernten

Jedes Jahr treten die weltweit führenden Forschungsgruppen in der Bildanalyse an der berühmten ImageNet Challenge gegeneinander an. ImageNet ist eine öffentliche Bilddatenbank mit über 14 Mio. Bildern, die von freiwilligen Helfern in knapp 22.000 Kategorien eingeteilt wurden, wobei für den Wettbewerb nur 1000 dieser Kategorien verwendet werden [20]. Auf diesen Daten werden Modelle trainiert und später auf einem geheim gehaltenen Testset evaluiert. Da viele Bilder Objekte mehrerer Kategorien zeigen, was eine eindeutige Zuordnung unmöglich macht, oder z. B. Hunderassen aufgrund des Aufnahmewinkels nicht immer präzise unterschieden werden können, werden sowohl die Top-1- wie auch die Top-5-Fehlerraten der Modelle bestimmt. Bei der Top-5-Evaluation dürfen die Modelle also 5 Kategorien für das Bild vorschlagen, unter denen sich die richtige Zuordnung befinden muss. Die  **Abb. 1** zeigt die Resultate der ImageNet-Gewinner seit der erstmaligen Durchführung im

Die Abb. 4 und 5 entstammen unseren eigenen Arbeiten.

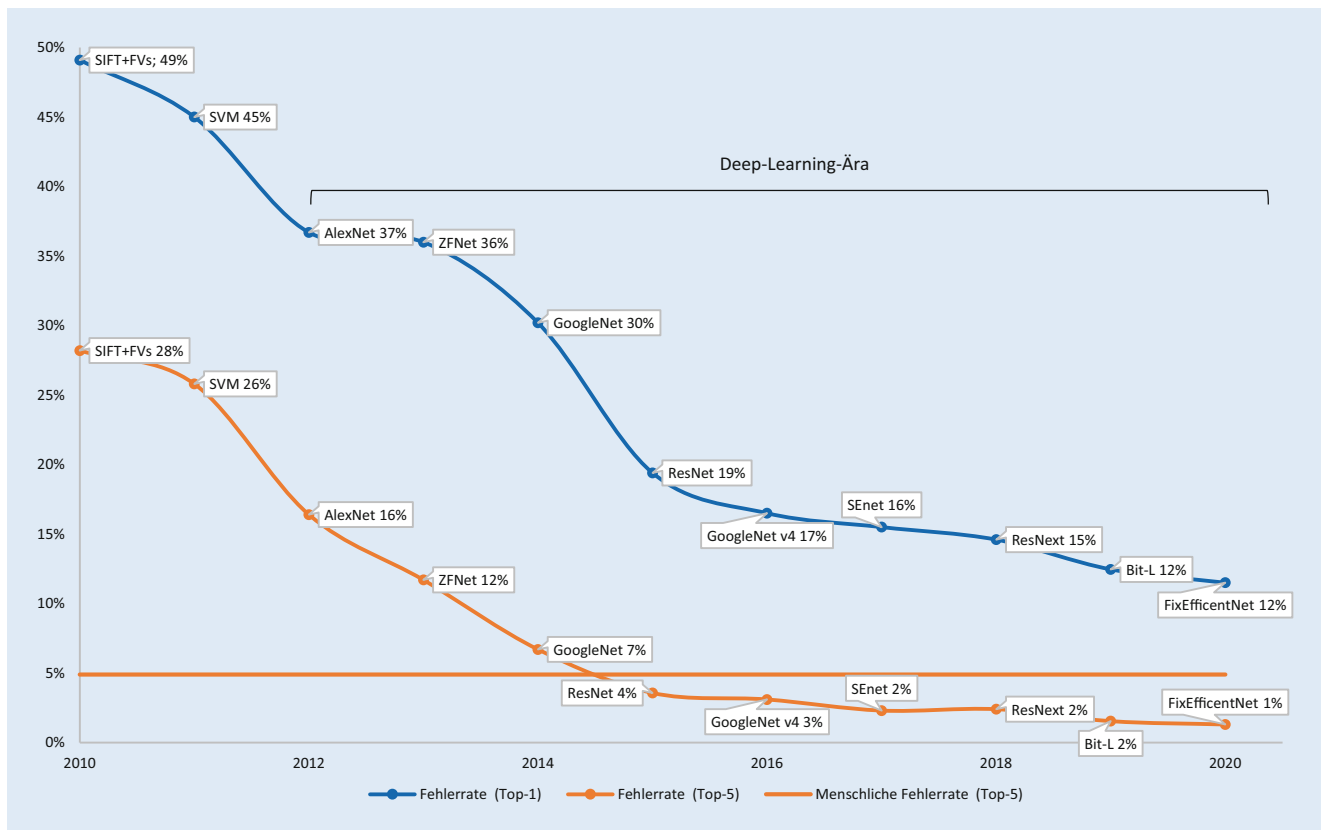


Abb. 1 ▲ Top-1- und Top-5-Fehlerrate der ImageNet-Gewinnerteams seit Beginn des Wettbewerbs im Jahr 2010. Das Top-1-Resultat von 2011 wurde nicht erhoben

Jahr 2010. Die menschliche Leistung in der Bestimmung der korrekten Kategorie wird auf eine Top-5-Fehlerrate von 5,1 % geschätzt [20]. Am 10.12.2015 verkündete die *New York Times* den großen Durchbruch [12]. Eine künstliche Intelligenz namens ResNet von Microsoft Research [4] gewann die ImageNet Challenge und übertraf dabei erstmals die Schallmauer der menschlichen Leistung.

» Eine moderne künstliche Intelligenz ist kein von Menschen erdachtes Regelwerk.

Ein digitales Bild ist ein zweidimensionales Raster von Farbwerten – sog. Pixel. Meist wird der Farbraum RGB (Rot-Grün-Blau) verwendet, sodass in jedem Pixel der Rot-, Grün- und Blauanteil der Farbe gespeichert wird. Die 3 Kanäle entsprechen dabei der Farbwahrnehmung des menschlichen Auges durch 3 verschiedene Zelltypen. Für einen Computer besteht ein Bild also nur aus einer Zahlenmatrix pro Farbkanal. Aus die-

sen Daten gilt es nun, eine Abbildung auf 1000 unterschiedliche Kategorien zu erstellen, die sich invariant bezüglich Skalierung oder Rotation der Bilder verhält. Bei einer durchschnittlichen Auflösung von 482×418 Pixel pro Bild in ImageNet scheiden durch Menschen erstellte und einprogrammierte Zuordnungsregeln als möglicher Ansatz sofort aus – oder können Sie aufgrund von 604.428 Zahlen einen Dalmatiner von einer Perserkatze unterscheiden?

Maschinelles Lernen aus Merkmalen und Beispielen

Inspiziert durch das menschliche Gehirn, analysiert eine künstliche Intelligenz Bilder in Zwischenstufen: Aus den 3 Farbwerten pro Pixel werden zuerst einfache Strukturen, sogenannte Features, extrahiert und zu immer komplexeren Merkmalen kombiniert, die schlussendlich die Erkennung einer Objektkategorie ermöglichen sollen. Die Grundlage der Feature-Berechnung bilden mathematische Filteroperatio-

nen. Mit diesen Verfahren lassen sich Linien, Kanten oder andere Grundstrukturen hervorheben. Die **Abb. 2** zeigt die Anwendung von Linien- und Kantentfilter auf eine Handaufnahme. Das geübte Auge erkennt dabei charakteristische Strukturen an den Außenseiten der Zeigefinger, die auf ein Handkzem hinweisen. Andererseits würde sich die Behaarung an den Handgelenken durch äußerst ähnliche Strukturen manifestieren, was uns einen guten Eindruck über die Schwierigkeit der Kombination von Filterwerten hin zu komplexeren Merkmalen vermittelt. Auch hier wäre die manuelle Erstellung von Zuordnungsregeln ein hoffnungsloses Unterfangen.

» Die Grundlage der Mustererkennung bilden mathematische Filteroperationen

Ein früher verbreitetes Verfahren zur Ableitung komplexerer Merkmale bestand darin, an markanten Stellen im Bild sog. Deskriptoren zu berechnen, welche die

Hautarzt <https://doi.org/10.1007/s00105-020-04663-7>
© Der/die Autor(en) 2020

M. Pouly · T. Koller · P. Gottfrois · S. Lionetti

Künstliche Intelligenz in der Bildanalyse – Grundlagen und neue Entwicklungen

Zusammenfassung

Hintergrund. Seit 2017 berichten Wissenschafts- und Populärmedien immer wieder von bildanalytischen Verfahren der künstlichen Intelligenz, die in der medizinischen Diagnostik zu menschlichen Experten vergleichbar gute Resultate erzielen. Mit der erstmaligen Zulassung eines solchen Systems durch die amerikanische Arzneimittelbehörde 2018 begann ihr Einzug in den klinischen Alltag.

Fragestellung. Dieser Beitrag gibt einen Überblick der wichtigsten Entwicklungen der künstlichen Intelligenz für bildanalytische Verfahren in klinischen Anwendungen mit Fokus auf die Dermatologie.

Material und Methode. Am Beispiel der ImageNet Challenge wird gezeigt, dass klassische Ansätze des maschinellen Lernens stark auf menschlicher Expertise beruhen und dass ihre Performance und Skalierbarkeit

praktischen Anforderungen nicht genügen konnten. Mittels Deep Learning, einem auf neuronalen Netzen beruhenden Verfahren, konnten diese Limitierungen und insbesondere die Abhängigkeit von menschlicher Expertise überwunden werden. Wir beschreiben wichtige Eigenschaften von Deep Learning, den methodischen Durchbruch von Transfer Learning und berichten über vielversprechende Entwicklungen hin zu generativen Modellen.

Ergebnisse. Mittels Deep Learning erreichen bildanalytische Verfahren in vielen Fällen die für den industriellen und klinischen Einsatz geforderte Genauigkeit. Zudem gestaltet sich ihre Industrialisierung weitestgehend barrierefrei. Derzeitige Entwicklungen fokussieren sich daher weniger auf die nochmalige Verbesserung der Genauigkeit, sondern nehmen sich den Herausforderungen

der Interpretierbarkeit und Anwendbarkeit unter Alltagsbedingungen an. Durch die Weiterentwicklung zu generativen Modellen werden gänzlich neuartige Anwendungen möglich.

Schlussfolgerungen. Deep Learning hat in vielerlei Hinsicht beeindruckende Erfolge vorzuweisen und gilt heute nicht nur in der Bildanalyse als das Standardverfahren schlechthin. Dieser Durchbruch der künstlichen Intelligenz ermöglicht eine rasch anwachsende Zahl von klinischen Anwendungen und entwickelt sich fortwährend zu einem unverzichtbaren Werkzeug in der modernen Medizin.

Schlüsselwörter

Computerunterstützte Bildanalyse · Deep Learning · Visuelle Merkmale · Diagnostische Bildanalyse · Bildanalytische Verfahren

Artificial intelligence in image analysis—fundamentals and new developments

Abstract

Background. Since 2017, there have been several reports of artificial intelligence (AI) achieving comparable performance to human experts on medical image analysis tasks. With the first ratification of a computer vision algorithm as a medical device in 2018, the way was paved for these methods to eventually become an integral part of modern clinical practice.

Objectives. The purpose of this article is to review the main developments that have occurred over the last few years in AI for image analysis, in relation to clinical applications and dermatology.

Materials and methods. Following the annual ImageNet challenge, we review classical methods of machine learning for image analysis and demonstrate how these

methods incorporated human expertise but failed to meet industrial requirements regarding performance and scalability. With the rise of deep learning based on artificial neural networks, these limitations could be overcome. We discuss important aspects of this technology including transfer learning and report on recent developments such as explainable AI and generative models.

Results. Deep learning models achieved performance on a par with human experts in a broad variety of diagnostic tasks and were shown to be suitable for industrialization. Therefore, current developments focus less on further improving accuracy but rather address open issues such as interpretability and applicability under clinical conditions.

Upcoming generative models allow for entirely new applications.

Conclusions. Deep learning has a history of remarkable success and has become the new technical standard for image analysis. The dramatic improvement these models brought over classical approaches enables applications in a rapidly increasing number of clinical fields. In dermatology, as in many other domains, artificial intelligence still faces considerable challenges but is undoubtedly developing into an essential tool of modern medicine.

Keywords

Computer-assisted image analysis · Deep learning · Visual features · Diagnostic imaging · Image analysis applications

Eigenschaften des Bildes an diesen Stellen abstrahieren. Diese Deskriptoren lassen sich z. B. aus Histogrammen der Orientierung und Stärke von verschiedenen Filterantworten ableiten. Alle Deskriptoren eines Bildes wurden dann in einem Merkmalsvektor zusammengefasst, der damit gewissermaßen als Fingerabdruck des Bildes diente und in einer Art Lexikon abgelegt wurde. In gleicher Art und Wei-

se konnte danach ein Testbild mit unbekannter Kategorisierung in einen Merkmalsvektor überführt und mit dem Lexikon abgeglichen werden. Die Kategorie des ähnlichsten Merkmalsvektors wurde übernommen. Computer „sahen“ also anhand des Abgleichs mit Beispielen in einem Lexikon. Das offensichtlichste Problem dieser Methodik war die explodierende Größe des Lexikons und

damit einhergehend der Rechenaufwand für den Abgleich. Es mussten also andere Möglichkeiten als das explizite Abspeichern und Absuchen von Merkmalsvektoren gefunden werden.



Abb. 2 ▲ Linien- und Kantenfilter machen charakteristische Bildstrukturen sichtbar

Zeitalter der Support Vector Machine

Das automatisierte Lernen von Zuordnungen anhand von in der Regel durch menschliche Experten erstellten Trainingsdaten wird als überwachtes Lernen bezeichnet (engl. „supervised learning“). Das oben skizzierte Verfahren mit dem Lexikonabgleich entspricht dabei einem Nearest-Neighbors-Modell. Andere Modelle der damaligen Zeit, wie z. B. Entscheidungsbäume, waren zwar effizienter, konnten jedoch nur lineare oder eben stückweise lineare Abbildungen von Merkmalsvektoren auf Objektkategorien lernen, was sich ebenfalls als nicht hinreichend herausstellte. Die vielversprechendste Methode war die sog. Support Vector Machine (SVM). Dieser Algorithmus unterteilt den Raum der Merkmale mittels Hyperebenen so, dass der Abstand zwischen einem Merkmalsvektor und der Grenze zur nächsten Bildkategorie maximiert wird. Dabei mussten nur noch diese minimalen Abstandsvektoren gespeichert werden und eben keine umfangreichen Lexika von Referenzbeispielen mehr. Das Problem der Linearität wurde umgangen, indem die Merkmalsvektoren mit dem sog. Kerneltrick in einen höher dimensionalen Raum projiziert wurden. Bis 2012 war die Support Vector Machine in Kombination mit manuell ausgewählten Bildfiltern und Deskriptoren die vorherrschende

Technik in der Bilderkennung und erreichte gemäß [Abb. 1](#) doch eine beachtlich tiefe Top-5-Fehlerrate von unter 26% bei 1000 möglichen Kategorien. Auch die Autoren dieses Beitrags haben sich noch 2015 derselben Techniken bedient, um Handekzem auf Fotografien erkennen und quantifizieren zu können [\[22\]](#). Aus über 30 Bildfiltern wurden charakteristische Hautstrukturen sichtbar gemacht, zu komplexeren, handverlesenen Merkmalen kondensiert und mittels der Support Vector Machine auf die 3 Kategorien Handekzem, gesunde Haut und Hintergrund abgebildet.

Deep-Learning-Revolution

In diesen frühen Jahren der ImageNet Challenge konnten selbstlernende Algorithmen keine rohen Pixelwerte verarbeiten. Die Support-Vector-Machine war also darauf angewiesen, dass menschliche Experten eine geeignete Auswahl von Merkmalen zur Verfügung stellten. Darüber hinaus stießen diese Systeme hinsichtlich Rechenzeit und Speicherverbrauch schnell an die Grenzen der damaligen Computertechnologie. Es konnten folglich nur wenige Bildfilter zur Berechnung der Merkmale verwendet werden, die verarbeitbare Menge an Trainingsbildern war stark eingeschränkt, und insbesondere der Umgang mit hochauflösenden Bildern mit Millionen von Pixeln war eine immense

Herausforderung. Zudem kamen diese Systeme nie an die für industrielle und medizinische Anwendungen geforderte Leistungsfähigkeit heran. Auch unsere damalige Anwendung zur Erkennung und Quantifizierung von Handekzem auf Bildern mit einer durchschnittlichen Auflösung von 1500×1000 Pixel konnte trotz performanter Hardware auf nur wenigen dutzend Bildern trainiert werden – viel zu wenige, um die vielfältigen Ausprägungen von Handekzem mit unterschiedlichen Hauttypen und Anomalien wie Ringe, Tattoos oder fehlenden Gliedmaßen repräsentativ zu erfassen. Zudem dauerte die Analyse eines einzigen Testbilds über 3 min. Zusammengefasst skalierte die Support-Vector-Machine nicht für große Datenmengen, und mit der Wahl der Merkmale basierte dieses Verfahren zu stark auf menschlicher Expertise. Den Wendepunkt markierte das Jahr 2012, als mit AlexNet [\[15\]](#) erstmals ein Deep-Learning-Verfahren die ImageNet Challenge gewann, und zwar gleich mit einer Verbesserung um 10% in der Top-5-Fehlerrate gegenüber dem Vorjahr.

Mit neuronalen Netzen zum Paradigmenwechsel

Deep Learning basiert auf sog. neuronalen Netzen, einem durch das biologische Neuron inspirierten Verfahren des maschinellen Lernen, dessen Grundzüge bereits 1943 von einem Neurophysiologen und einem Mathematiker zusammen skizziert wurden [\[18\]](#). Ein einzelnes künstliches Neuron, wie in [Abb. 3](#) dargestellt, wendet eine nichtlineare Transferfunktion auf eine gewichtete Summe seiner Inputwerte an. Nun werden Tausende solcher künstlichen Neuronen in Schichten hierarchisch angeordnet, sodass der Output einer früheren Schicht den Input einer späteren Schicht liefert. Inspiriert wurde diese Architektur auch durch den Konnektionismus, der versucht, mentale Phänomene durch Verbindungen, Signale und Interaktionen zwischen einer großen Anzahl einfacher Einheiten zu beschreiben. Die Gesamtheit der Gewichte aller einzelnen Neuronen bilden dabei die Parameter des Systems, deren Werte in einem

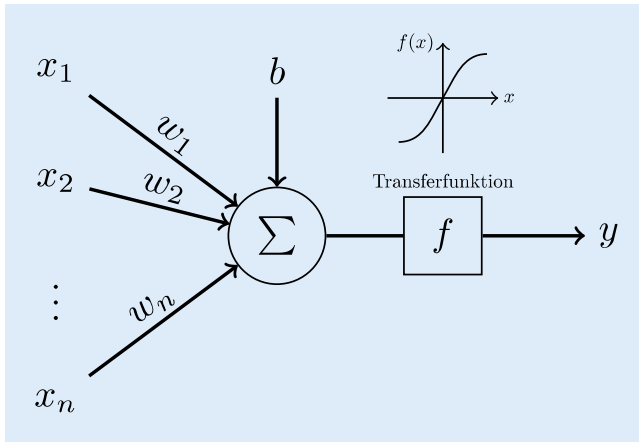


Abb. 3 ◀ Ein künstliches Neuron berechnet eine nicht-lineare Funktion einer gewichteten Summe seiner Eingabewerte, $y = f(w_1x_1 + \dots + w_nx_n + b)$, wobei x_i die Eingabewerte und w_i die aus den Trainingsdaten gelernten Gewichte darstellen

sehr rechenintensiven Trainingsprozess namens Backpropagation unter Zuhilfenahme bereits kategorisierter Trainingsdaten ermittelt werden. Es kann gezeigt werden, dass solche Netze prinzipiell jede Funktionen, also Abbildung auf Kategorien, approximieren können.

Die hierarchische Anordnung künstlicher Neuronen in einem Deep-Learning-Netz erlaubte fortan die direkte Verarbeitung der rohen Pixelwerte. Damit entfiel die menschliche Auswahl von Bildfilter als mögliche Fehlerquelle in der Bilderkennung. Werden die gelernten Gewichte in einem neuronalen Netz über die verschiedenen Schichten hinweg visualisiert, so fällt auf, dass frühere Schichten eigenständig primitive Bildstrukturen zu erkennen lernen und damit Filteroperationen imitieren. In späteren Schichten werden diese Informationen automatisch zu komplexeren Merkmalen kondensiert und schlussendlich auf eine Wahrscheinlichkeitsverteilung über die möglichen Bildkategorien abgebildet. Computer begannen also, die menschliche Leistungsfähigkeit zu übertreffen, nachdem sie von der expliziten Eingabe menschlichen Wissens befreit werden konnten. Yoshua Bengio, Geoffrey Hinton und Yann LeCun, die Pioniere dieses Fachgebiets, wurden denn auch 2018 mit dem ACM Turing Award, dem Nobelpreis der Informatik, ausgezeichnet [5]. Seit 2012 wurde jede ImageNet Challenge durch den Einsatz von Deep Learning gewonnen.

Obwohl sich der Trainingsprozess von neuronalen Netzen enorm rechenintensiv gestaltet, so ist die spätere Anwendung von trainierten Netzen effizienter

als die klassischen Verfahren des maschinellen Lernens. Dies ermöglichte die Bildsegmentierung, also die Kategorisierung einzelner Pixel anstelle ganzer Bilder bis hin zur Echtzeitanalyse von Bewegtbildern in Videos – und dies sogar mit der eingeschränkten Leistungsfähigkeit eines Smartphone-Prozessors. So ist es nicht erstaunlich, dass andere algorithmische Verfahren der Bildanalyse weitestgehend in der Bedeutungslosigkeit verschwanden. Aus technologischer Sicht beruht der durchschlagende Erfolg von Deep Learning v. a. auf der Tatsache, dass neuronale Netze sehr viel größere Datenmengen verarbeiten und damit statistisch besser verallgemeinern können [2]. Die mathematischen Operationen in einem neuronalen Netz lassen sich als Additionen und Multiplikationen von Matrizen auffassen und dadurch sehr viel effizienter auf spezialisierten Grafikprozessoren ausführen. Dementsprechend verdankt die moderne künstliche Intelligenz ihren Durchbruch zumindest teilweise der Computerspielindustrie. Ein zweiter Katalysator war natürlich die rasante Durchdringung unseres Alltags durch das Social Web und damit einhergehend die explosionsartige Ausweitung und Verfügbarkeit von Bild- und Textdaten im Internet.

Big Business durch künstliche Intelligenz

Nicht nur für uns Menschen gehört das Sehen zu den wichtigsten Voraussetzungen, um erfolgreich mit unserer Umgebung interagieren zu können. Als eine künstliche Intelligenz namens Al-

phaGo 2017 den damaligen Weltmeister Ke Jie im Brettspiel GO schlug, war dies zweifellos eine Meisterleistung und damit ein historisches Ereignis in der Entwicklung der künstlichen Intelligenz [8]. Das zuverlässige Kategorisieren und Segmentieren von Objekten auf Bilder hat aber ungleich weitreichendere Bedeutung nicht nur für medizinische Anwendungen, sondern gleichsam für die Entwicklung autonomer Fahrzeuge, der automatisierten Schädlingsbekämpfung in der Landwirtschaft oder der optischen Qualitätskontrolle in der produzierenden Industrie. So gesehen, markiert der Durchbruch von 2016 in der Bildanalyse auch gleichzeitig den Beginn von „Big Business with Artificial Intelligence“. Im Übrigen schlug AlphaGo bereits 2016 den topklassierten Spieler Lee Sedol und nutzte dabei dieselben Techniken („convolutional neural networks“) zur Analyse unzähliger auf öffentlichen Spielservern gesammelter Brettsituationen.

Mit dem Einzug der künstlichen Intelligenz in die medizinische Bildanalyse veränderte sich auch deren angestammte Forschungsmethodik. Über Nacht erschienen Google, Microsoft, Facebook und OpenAI auf den Autorenlisten medizinischer Publikationen; es werden medizinische Datenbanken kuratiert und für Forschung und Wettbewerbe öffentlich zur Verfügung gestellt. Beispielsweise beinhaltet ISIC-2019 über 25.000 dermatoskopische Bildaufnahmen annotiert mit 9 diagnostischen Kategorien, wie z. B. „melanoma“, „melanocytic nevus“, „dermatofibroma“ oder „actinic keratosis“ [6]. Die ISIC Challenge mit den Disziplinen „lesion segmentation“, „lesion attribute detection“ und „disease classification“ wird seit 2016 alljährlich durchgeführt.

Aktuelle Herausforderungen

Neuronale Netze manifestieren sich dem Menschen gegenüber als Blackbox. Sie berechnen ihre Resultate aufgrund von Millionen von Parameterwerten, die zuvor in einem oftmals tagelangen Rechenprozess unter Miteinbezug von zehntausenden durch Menschen annotierten Bildern, Texten oder anderen Daten festgelegt wurden. Zu Recht stellt sich dabei

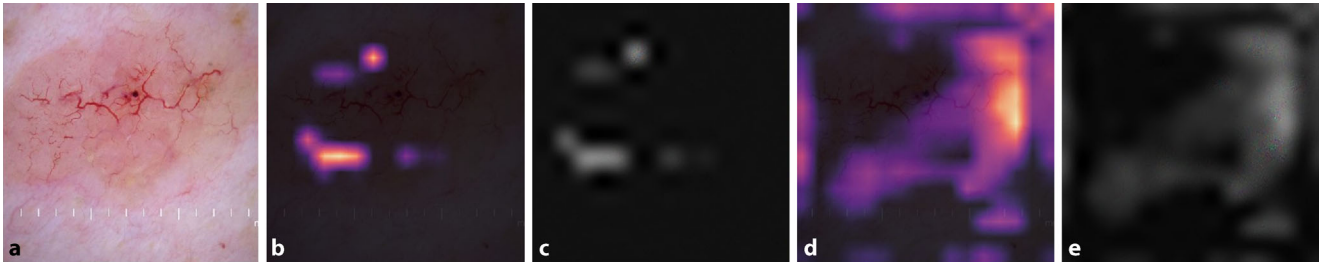


Abb. 4 **a** Originalbild aus [7], mit freundl. Genehmigung, ©David.moreno72, CC BY-SA 4.0, <https://creativecommons.org/licenses/by-sa/4.0/>. **b, c** Wärmebilder (GradCam und Guided Backpropagation) für die Diagnose („basal cell carcinoma“) (0,816). **d, e** Wärmebilder (GradCam und Guided Backpropagation) für die Diagnose („melanocytic nevus“) (0,045)

die Frage, *was* ein neuronales Netz nun tatsächlich gelernt hat, *wie* es zu einem bestimmten Resultat kommt und *inwiefern* sich dieses gegenüber menschlichen Experten rechtfertigen oder gegenüber Laien erklären lässt. Handelt es sich bei der Anwendung um eine optische Entriegelungsfunktion einer Katzentür, so sind die meisten Kunden durchaus gewillt, diese Intransparenz zu akzeptieren. Interessanterweise stimmt diese Aussage sogar für höchst sicherheitsrelevante Systeme. Vielleicht nutzen auch Sie die Smartphone-Gesichtserkennung für den Zugriff auf Ihre höchst privaten Bilder, Nachrichten oder Passwörter und vertrauen damit einem neuronalen Netz, dessen erlerntes Wissen sich nicht einmal seinen Entwicklern erschließt.

Weitaus differenzierter gestaltet sich diese Beziehung in medizinischen Anwendungen. Wir können zwar die Fehlerrate von neuronalen Netzen auf klinisch validierten und repräsentativen Testdaten erheben, in Wettbewerben mit menschlichen Fachexperten vergleichen, darauf hinweisen, dass neuronale Netze Dutzende von Diagnosen pro Sekunde reproduzierbar liefern und sich als medizinischer Service in Smartphone-Apps über die Welt verteilen lassen. Es bleibt zu Recht die Frage nach der Erklärbarkeit. Explainable Artificial Intelligence (XAI) ist damit auch eines der derzeit aktivsten Forschungsfelder in der künstlichen Intelligenz. Neben der Modellkomplexität ist diese Intransparenz der Tatsache geschuldet, dass neuronale Netze nur Korrelationen berechnen, jedoch keine Kausalitätsbeziehung ableiten können. Erste Ansätze basieren auf Visualisierungstechniken, die auf dem Originalbild die Bereiche mit dem stärksten Einfluss

auf die Klassifizierung hervorheben. Die **Abb. 4** zeigt ein dermatoskopisches Bild, für das ein Deep-Learning-Modell eine Wahrscheinlichkeit von 81,6% für die Diagnose „basal cell carcinoma“ und 4,5% für „melanocytic nevus“ prognostiziert. Die beiden Wärmebilder zeigen, welche Bildbereiche die entsprechenden Diagnosen stützen.

» XAI ist eines der derzeit aktivsten Forschungsfelder in der künstlichen Intelligenz

Weiterführende Ansätze trainieren hierarchische Modelle, die z.B. auf einer dermatoskopischen Aufnahme zuerst differenzialdiagnostische Merkmale wie „pigment network“, „streaks“, „globules“ oder „milia-like cyst“ identifizieren und erst danach eine Diagnose ableiten. Damit benötigt dieser Ansatz aber auch eine hinreichende Menge von auf dieser Granularität annotierten Trainingsdaten, und damit steigt unweigerlich der Aufwand in der Datenbeschaffung. Inspiriert durch den klinischen Alltag und befähigt durch die erstaunlichen Fortschritte in der Deep-Learning-basierten Textgenerierung soll sich die modernste Form der künstlichen Intelligenz nicht mehr nur mit der Voraussage der wahrscheinlichsten Diagnose begnügen, sondern dazu gleich einen medizinischen Bericht zu Händen des menschlichen Arztes verfassen. In der Medizin fehlen dazu aktuell noch die nötigen Trainingsdaten, jedoch hat Facebook Research bereits gezeigt, dass neuronale Netze von einer Fotoaufnahme eines Gerichts das dazu passende Kochrezept verfassen können [21].

Die zentralste Herausforderung überhaupt ergibt sich aus dem beinahe unstillbaren Datenhunger moderner Deep-Learning-Modelle. Je mehr Parameter ein Modell hat, desto grösser ist seine Lernkapazität. Folglich werden mit immer mehr Rechenressourcen immer größere Modelle trainiert zum Preis eines immer weiter ansteigenden Datenbedarfs. Gerade in der Medizin ist das verfügbare Bildmaterial jedoch beschränkt, sei es aufgrund von seltenen Erscheinungen, Datenschutzaspekten, besonders aber aufgrund der mangelnden Verfügbarkeit menschlicher Experten zur Annotation von Trainingsdaten. Somit war es denn auch keine algorithmisch-mathematische Erkenntnis, die den wichtigsten und eingangs bereits erwähnten Durchbruch hinsichtlich Industrialisierbarkeit von Deep-Learning-Modellen ermöglichte, sondern ein brillanter und auf viel Intuition beruhender technischer Kniff. Frühe Schichten in neuronalen Netzen lernen einfache Bildstrukturen, spätere Schichten kombinieren diese zu komplexen Merkmalen und leiten schlussendlich eine Wahrscheinlichkeitsverteilung über mögliche Kategorien daraus ab. Der Kniff besteht nun darin, dass zum Erlernen einfacher Bildstrukturen wie Linien und Kanten gar keine medizinischen Daten benötigt werden. Vielmehr lassen sich diese Modelle auf riesigen, allgemeinen Bilddatenbanken einmalig vortrainieren. Dann werden die Parameter der frühen Schichten eingefroren und nur die späteren Schichten auf einem domänenspezifischen Datensatz nochmals trainiert, wobei durch die damit erreichte Reduktion der Parameter sehr viel weniger Bilder und Rechenleistung benötigt werden. Wir bezeichnen dieses

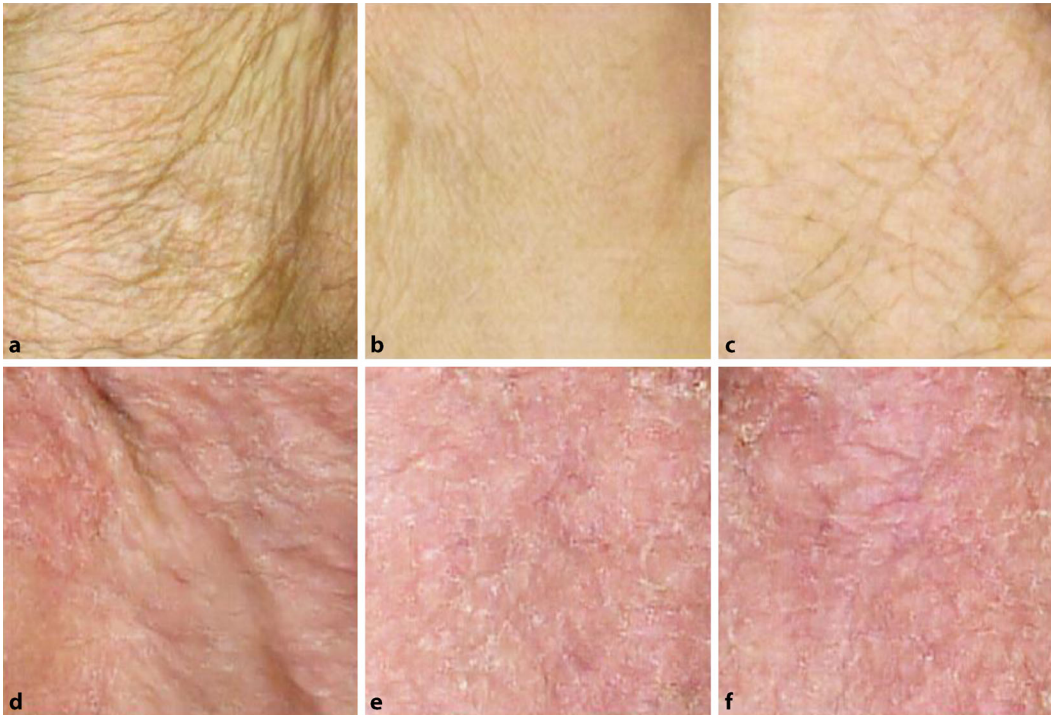


Abb. 5 ◀ a–f Durch eine künstliche Intelligenz generierte Bilder menschlicher Haut. (Aus [23])

Vorgehen als Transfer Learning. Google, Facebook und Co. stellen mittlerweile ganze Bibliotheken von vortrainierten Modellen zur Verfügung, für deren Training die wenigsten Forschungsinstitutionen überhaupt über die nötige Rechenleistung verfügen würden und die wir nun mit wenigen Tagen Aufwand und überschaubaren Datenmengen anwendungsspezifisch anpassen können.

Generative Modelle

Der Name von Facebooks Chief AI Scientist Yann LeCun ist im Zusammenhang mit dem Turing Award für seine epochalen Beiträge zur Deep-Learning-Forschung bereits gefallen. Seine viel zitierte Aussage bezieht sich auf ein von Ian Goodfellow 2014 vorgeschlagenes Konzept namens Generative Adversarial Networks (GAN) zur künstlichen Generierung von Fotografien [3], in den Populärmedien auch Deepfakes genannt:

The coolest idea in Deep Learning in the last 20 years [11].

Goodfellow ließ 2 neuronale Netze, einen „Fälscher“ und einen „Polizisten“, in einem kompetitiven Spiel gegeneinander antreten. Der Fälscher generiert künstliche Bilder, der Polizist muss echte Fo-

tografien von gefälschten Bildern unterscheiden können. Nun wurden die Optimierungsfunktionen beider Netze so gekoppelt, dass, wenn sich der Fälscher im Laufe des Trainingsprozesses verbessert und damit realistischere Bilder generiert, sich der Polizist bei der Unterscheidung zwischen echt und unecht ebenfalls verbessern muss. Verbessert sich umgekehrt der Polizist, muss der Fälscher seinerseits nachziehen und noch bessere Fälschungen generieren usw. Gelangt dieser Prozess in einen Gleichgewichtszustand, wurde eine künstliche Intelligenz erschaffen, die Bilder aufgrund erlernter statistischer Strukturen generieren kann, die eine andere künstliche Intelligenz ihrerseits nicht mehr von echten Bildern unterscheiden kann.

» Aus Artificial Intelligence wurde Artificial Creativity

GANs unterstützen heute nicht nur den menschlichen Designprozess von Möbelstücken oder Kleider, sondern finden zunehmend Einzug in die medizinische Bildanalyse: GANs generieren fotorealistiche Bilder von Gewebe und Organen, wie in **Abb. 5** gezeigt, übertragen pathologische Erscheinungen oder entfernen diese auf Patientenbilder, verändern eth-

nische Eigenschaften oder lassen Gliedmaßen nachwachsen.

So groß wie die Erfolge von Deep Learning und anderen Methoden der künstlichen Intelligenz in der medizinischen Bildanalyse der letzten Jahre, so groß sind unsere Erwartungen und gleichsam die damit verbundenen Herausforderungen. Immer wieder übertreffen technische Systeme die menschliche Leistungsfähigkeit unter Laborbedingungen und scheitern dann doch im klinischen Alltag an so unterschiedlichen Aspekten wie Akzeptanz, Prozessintegration, Bedienbarkeit in Stresssituationen oder den technischen Gegebenheiten [14]. Nur zu gerne schließen wir uns also dem Zitat von Michael Abramoff, dem Entwickler des eingangs erwähnten ophthalmologischen Deep-Learning-Systems und damit Empfänger der ersten Zulassung der amerikanischen Arzneimittelbehörde, an:

There is much more to Healthcare than Algorithms [13].

Fazit für die Praxis

- Bildanalytische Verfahren der modernen künstlichen Intelligenz erreichen heute in vielen Fällen die für den

industriellen und klinischen Einsatz geforderte Genauigkeit und über-treffen sogar teilweise die Leistung menschlicher Experten.

- Trotz zahlreicher Herausforderungen in Bezug auf ihren Datenhunger, mangelnde Interpretierbarkeit und Anwendbarkeit unter klinischen Alltagsbedingungen entwickelt sich diese Technologie fortwährend zu einem unverzichtbaren Werkzeug in der modernen Medizin.

Korrespondenzadresse



Prof. Dr. Marc Pouly
Informatik, Hochschule
Luzern
Suurstoffi 1, 6343 Rotkreuz,
Schweiz
marc.pouly@hslu.ch

Funding. Open access funding provided by Lucerne University of Applied Sciences and Arts

Einhaltung ethischer Richtlinien

Interessenkonflikt. M. Pouly, T. Koller, P. Gottfroid und S. Lionetti geben an, dass kein Interessenkonflikt besteht.

Für diesen Beitrag wurden von den Autoren keine Studien an Menschen oder Tieren durchgeführt. Für die aufgeführten Studien gelten die jeweils dort angegebenen ethischen Richtlinien.

Open Access. Dieser Artikel wird unter der Creative Commons Namensnennung 4.0 International Lizenz veröffentlicht, welche die Nutzung, Vervielfältigung, Bearbeitung, Verbreitung und Wiedergabe in jeglichem Medium und Format erlaubt, sofern Sie den/die ursprünglichen Autor(en) und die Quelle ordnungsgemäß nennen, einen Link zur Creative Commons Lizenz beifügen und angeben, ob Änderungen vorgenommen wurden.

Die in diesem Artikel enthaltenen Bilder und sonstiges Drittmaterial unterliegen ebenfalls der genannten Creative Commons Lizenz, sofern sich aus der Abbildungslegende nichts anderes ergibt. Sofern das betreffende Material nicht unter der genannten Creative Commons Lizenz steht und die betreffende Handlung nicht nach gesetzlichen Vorschriften erlaubt ist, ist für die oben aufgeführten Weiterverwendungen des Materials die Einwilligung des jeweiligen Rechteinhabers einzuholen.

Weitere Details zur Lizenz entnehmen Sie bitte der Lizenzinformation auf <http://creativecommons.org/licenses/by/4.0/deed.de>.

Literatur

1. Esteva A, Kuprel B, Novoa RA et al (2017) Dermatologist-level classification of skin cancer with deep neural networks. *Nature*. <https://doi.org/10.1038/nature21056>
2. Goodfellow I, Bengio Y, Courville A (2016) *Deep learning*. MIT Press, Cambridge, Massachusetts
3. Goodfellow IJ, Pouget-Abadie J, Mirza M et al (2014) Generative adversarial nets. In: Ghahramani Z, Welling M, Cortes C, Lawrence ND, Weinberger KQ (Hrsgs) *Advances in Neural Information Processing Systems 27*, Curran Associates, Inc., S 2672–2680. <https://papers.nips.cc/paper/5423-generative-adversarial-nets>
4. He K, Zhang X, Ren S, Sun J (2016) Identity mappings in deep residual networks. *Lect Notes Comput Sci (including Subser Lect Notes Artif Intell Lect Notes Bioinformatics)*. https://doi.org/10.1007/978-3-319-46493-0_38
5. <https://awards.acm.org/about/2018-turing>. Zugegriffen: 23. Mai 2020
6. <https://challenge2019.isic-archive.com>. Zugegriffen: 24. Mai 2020
7. https://en.wikipedia.org/wiki/Basal-cell_carcinoma#/media/File:Dermoscopy_nodular_basal_cell_carcinoma.jpg. Zugegriffen: 3. August 2020
8. <https://events.google.com/alphago2017/>. Zugegriffen: 23. Mai 2020
9. <https://www.fda.gov/news-events/press-announcements/fda-permits-marketing-artificial-intelligence-based-device-detect-certain-diabetes-related-eye>. Zugegriffen: 23. Mai 2020
10. <https://www.mdmag.com/medical-news/ai-model-flags-patients-at-risk-osteoporosis>. Zugegriffen: 23. Mai 2020
11. <https://www.nature.com/news/astromers-explore-uses-for-ai-generated-images-1.21398>
12. <https://www.nytimes.com/2015/12/11/science/an-advance-in-artificial-intelligence-rivals-human-vision-abilities.html>. Zugegriffen: 23. Mai 2020
13. <https://www.technologyreview.com/2020/04/27/1000658/google-medical-ai-accurate-lab-real-life-clinic-covid-diabetes-retina-disease/>. Zugegriffen: 23. Mai 2020
14. <https://www.technologyreview.com/2020/04/27/1000658/google-medical-ai-accurate-lab-real-life-clinic-covid-diabetes-retina-disease/>. Zugegriffen: 23. Mai 2020
15. Krizhevsky A, Sutskever I, Hinton GE (2012) ImageNet classification with deep convolutional neural networks. In: Pereira F, Burges CJC, Bottou L, Weinberger KQ (Hrsgs) *Advances in Neural Information Processing Systems 27*, Curran Associates, Inc., S 1097–1105. <http://papers.nips.cc/paper/4824-imagenet-classification-with-deep-convolutional-neural-networks.pdf>
16. Liu X, Faes L, Kale AU et al (2019) A comparison of deep learning performance against health-care professionals in detecting diseases from medical imaging: a systematic review and meta-analysis. *Lancet Digit Heal*. [https://doi.org/10.1016/S2589-7500\(19\)30123-2](https://doi.org/10.1016/S2589-7500(19)30123-2)
17. Liu Y, Jain A, Eng C et al (2020) A deep learning system for differential diagnosis of skin diseases. *Nat Med*. <https://doi.org/10.1038/s41591-020-0842-3>
18. McCulloch WS, Pitts W (1943) A logical calculus of the ideas immanent in nervous activity. *Bull Math Biophys*. <https://doi.org/10.1007/BF02478259>
19. McKinney SM, Sieniek M, Godbole V et al (2020) International evaluation of an AI system for breast cancer screening. *Nature*. <https://doi.org/10.1038/s41586-019-1799-6>
20. Russakovsky O, Deng J, Su H et al (2015) Imagenet large scale visual recognition challenge. *Int J Comput Vis*. <https://doi.org/10.1007/s11263-015-0816-y>
21. Salvador A, Drozdal M, Giro-I-Nieto X, Romero A (2019) Inverse cooking: recipe generation from food images. *Proc IEEE Comput Soc Conf Comput Vis Pattern Recognit*. <https://doi.org/10.1109/CVPR.2019.01070>
22. Schnürle S, Pouly M, Vor Der Brück T et al (2017) On using support vector machines for the detection and quantification of hand eczema. *ICAART 2017 – Proc 9th Int Conf Agents Artif Intell*. <https://doi.org/10.5220/0006125000750084>
23. Zenodo: <https://zenodo.org/record/3873159#.XuDalbx70>. Zugegriffen: 3. August 2020