

Krestel, Ralf

Book Chapter — Published Version

Neuronale Netze als Künstler - Oder KI als Werkzeug und Partner?

Suggested Citation: Krestel, Ralf (2022) : Neuronale Netze als Künstler - Oder KI als Werkzeug und Partner?, In: Bodrožić-Brnić, Kristina Fitzek, Herbert (Ed.): Kreativität und Künstliche Intelligenz: Der Mensch als treibende Kraft der KI, BSP Business and Law School, Berlin, pp. 41-55

This Version is available at:

<http://hdl.handle.net/11108/568>

Kontakt/Contact

ZBW – Leibniz-Informationszentrum Wirtschaft/Leibniz Information Centre for Economics
Düsternbrooker Weg 120
24105 Kiel (Germany)
E-Mail: info@zbw.eu
<https://www.zbw.eu/de/ueber-uns/profil-der-zbw/veroeffentlichungen-zbw>

Standard-Nutzungsbedingungen:

Dieses Dokument darf zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden. Sie dürfen dieses Dokument nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, aufführen, vertreiben oder anderweitig nutzen. Sofern für das Dokument eine Open-Content-Lizenz verwendet wurde, so gelten abweichend von diesen Nutzungsbedingungen die in der Lizenz gewährten Nutzungsrechte.

Terms of use:

This document may be saved and copied for your personal and scholarly purposes. You are not to copy it for public or commercial purposes, to exhibit the document in public, to perform, distribute or otherwise use the document in public. If the document is made available under a Creative Commons Licence you may exercise further usage rights as specified in the licence.

Neuronale Netze als Künstler - Oder KI als Werkzeug und Partner?

RALF KRESTEL, ZBW - LEIBNIZ-INFORMATIONSZENTRUM WIRTSCHAFT & CAU KIEL

Deep Learning Methoden haben rasante Erfolge in der Entwicklung von Technologie-Lösungen hervorgebracht. Wenn es darum geht, Künstliche Intelligenz in der Generierung von Kunst einzusetzen, stehen den Kunstschaffenden vielfältige Formen der Anwendungs- und Entwicklungsarbeit zur Verfügung. Dahinter steht das bestmögliche Ausschöpfen technologischer Potenziale. Der vorliegende Artikel eröffnet vielfältige Bezüge zwischen Digitalisierung und Kreativität und wirft einen Blick hinter die Fassade von KI-basierter Kunstproduktion.

Seit Beginn der Industrialisierung werden immer mehr Tätigkeiten Maschinen übertragen, die bis dato menschlicher Arbeitskraft bedurften. Mit Erfindung des Computers waren diese Tätigkeiten nicht mehr nur auf körperliche Arbeit beschränkt, sondern auch geistige Arbeit konnte nun in erhöhtem Maße durch Maschinen verrichtet werden. Man denke hier beispielsweise an Taschenrechner oder Schachcomputer. Ein weiterer wichtiger Schritt hin zu einer Künstlichen Intelligenz war die Entwicklung von Methoden

des maschinellen Lernens. Diese erlauben mit Hilfe von großen Datenmengen und geeigneten Lernalgorithmen ein selbstständiges Lernen und bedürfen nicht mehr dem expliziten Definieren von Regeln durch Programmierer:innen. Selbstfahrende Autos und maschinelle Übersetzung sind Beispiele von erfolgreichen Anwendungen von maschinellem Lernen. Diese Tätigkeiten sind geistig anspruchsvoll und können nicht einfach nur durch einfache Berechnungen sinnvoll verrichtet werden. Sie erfordern auch von Menschen eine Ausbildung, das heißt, auch Menschen müssen diese Tätigkeiten erlernen.

Wenn ein Computer nun komplexe, geistige Aufgaben lösen kann, kann er dann nicht auch Kunst erschaffen? Zumindest die Erzeugung von Artefakten, die auf den ersten Blick wie Kunst aussehen, ist in allen Kunstgattungen möglich. Beispiele finden sich unter anderem in Literatur (Roemmele, 2016; Ghazvininejad, 2016), Musik (Zulić, 2019; Lopez-Rincon, 2018) und den bildenden Künsten (Colton, 2012; Elgammal, 2017).

Computer können kunstähnliche Objekte erschaffen, aber es fehlen scheinbar zwei ganz entscheidende Fähigkeiten: wahres Verständnis und wahre Kreativität, wobei Kreativität ein Verständnis der Welt im Allgemeinen und der Kunstwelt im Speziellen wohl voraussetzt. Davon ist die KI-Forschung noch weit entfernt und es ist zurzeit noch unklar, inwieweit Computer tatsächlich Kreativität ähnlich der des Menschen besitzen können (Hertzmann, 2018; Still 2019). Zumindest was das Verständnis anbelangt, gibt es einige Ansätze, die sich unter dem Begriff des *Semantic Computing* zusammenfassen lassen und in klar umrissenen Szenarien eine Art semantisches Wissen zeigen.

Frühe KI-Forschung hat mit dem Turing-Test ein Verfahren entwickelt, das es ermöglichen soll, eine menschenähnliche, künstliche Intelligenz zu bewerten: Eine Person soll im Dialog mit einer Maschine oder einem Menschen nicht in der Lage sein die Beiden auseinanderhalten zu können. Für den Fall einer kunsterzeugenden Intelligenz wurde dieses Verfahren angepasst, indem man einer Kunstexpertin Bilder von menschlichen Künstler:innen, sowie Werke einer KI vorlegt, und der/die Expert:in dürfte es nicht gelingen, diese entsprechend zuzuordnen (Gangadharbatla, 2021). Diese Art des Turing-Tests vernachlässigt aber gerade die Prüfung der wesentlichen Fähigkeiten: Verständnis und Kreativität. Anhand der Erzeugung von Literatur kann man dies gut veranschaulichen: Texte können syntaktisch korrekt sein, ohne jedoch semantisch sinnvoll zu sein. Ein bekanntes Beispiel ist der Satz von Noam Chomsky „Colorless green ideas sleep furiously“, auf Deutsch „Farblose grüne Ideen schlafen wütend“ (Chomsky, 1957). Die Syntax ist korrekt, aber der Satz ergibt keinen Sinn. Ähnlich verhält es sich mit der Kreativität. Gute Dichter:innen können sinnvolle, neue Metaphern erschaffen, was als kreativ wahrgenommen wird. Eine KI kann dies noch nicht eigenständig. In Anlehnung an das Theorem der endlos tippenden Affen (die irgendwann durch Zufall die Werke Shakespeares erzeugen würden) würde eine KI heutzutage viel weniger Ausschuss produzieren und es wäre deutlich einfacher, interessante Schöpfungen unter den zufälligen Erzeugnissen ausfindig zu machen (Meyer, 2017). Was wir daher bei der Erzeugung von Literatur, Musik, und bildender Kunst durch KI sehen, ist die Imitation von Kunst. Dies entspricht nicht unserem heutigen Verständnis von Kunst, findet aber Anknüpfungspunkte an das Kunstverständnis in der Antike, z. B. bei Platon, wo Kreativität nicht erstrebenswert war, und Kunst eine Nachahmung der Natur sein soll (Tatarkiewicz, 1979).

Wenn Computer nun in der Lage sind, etwas zu erschaffen was keine Kunst ist, wie kann man daraus eventuell Kunst machen? Zum einen kann man KI als ein Werkzeug sehen, das Künstler:innen in einem künstlerischen Prozess nutzen können. Diese Sicht erinnert stark an die Diskussion um Malerei und Fotografie. Auch dort erzeugt nicht das

Werkzeug Kunst, sondern die Künstler:innen benutzen das Werkzeug, um Kunst zu erschaffen. Zum anderen kann man die existierende KI-Methoden erweitern, damit diese eine Art von Kreativität und Verständnis entwickeln. Insbesondere wollen wir uns der Frage widmen, inwieweit durch den Einsatz von KI-Grenzen der künstlerischen Arbeit verschoben werden oder Begriffe wie Kreativität und Kunst durch die Anwendung auf KI-Methoden eine Bedeutungsänderung erfahren.

Der Begriff der künstlichen Intelligenz ist und war einem ständigen Wandel unterworfen und umfasst eine ganze Reihe von mehr oder weniger „intelligenten“ Methoden. In der aktuellen Diskussion meint man meist Methoden des maschinellen Lernens und insbesondere tiefe, neuronale Netze, das sogenannte Deep Learning, wenn man von KI spricht. Dies ist aber tatsächlich nur ein kleiner Teil der KI-Methoden, welche jedoch für Aufgaben der Wahrnehmung (Bild, Text, Audio) besonders gute Ergebnisse liefern. Im Folgenden wollen wir uns daher auf eine bestimmte Art von neuronalen Netzen beschränken, den *Generative Adversarial Networks*, nachdem wir nochmal einen kurzen Blick auf maschinelles Lernen allgemein werfen.

Maschinelles Lernen für die Kunst

Der Einsatz von Computern in der Kunst ist nicht neu. Digitale Kunst, oder Computerkunst im allgemeinen und generative Kunst im speziellen, nutzen oft Computer, um Kunst zu erschaffen. Das maschinelle Lernen führt nun jedoch in der Kunst wie auch in der Informatik allgemein einen kompletten neuen Ansatz ein. Während die generative Kunst ebenso wie klassische Algorithmen typischerweise ein Programm abarbeiten, ist die Idee des maschinellen Lernens eine fundamental andere. Beim Programmieren werden Regeln festgelegt, die der Computer abarbeitet. Dabei bekommt das Programm eine Eingabe, zum Beispiel die Zahlen „4“ und „2“ und gibt die Zahl „6“ aus. Die festgelegte Regel in diesem Fall wäre die Addition (Eingabe A + Eingabe B). Beim Maschinellen Lernen fehlt die Angabe der Regel und der Computer muss selbst die Regel finden, welche aus der Eingabe die Ausgabe erzeugt. Dafür ist neben der Eingabe auch die gewünschte oder beobachtete Ausgabe notwendig. Da anhand eines Beispiels diese Regel nicht eindeutig bestimmt werden kann (die Regel könnte in diesem Beispiel auch $(2 \times \text{Eingabe A}) - \text{Eingabe B}$ lauten), sind mehrere Trainingsbeispiele notwendig. Je komplizierter die Regeln sind, das heißt je mehr Möglichkeiten und Kombinationen existieren, desto mehr Trainingsdaten sind notwendig, um möglichst nahe an die gesuchte Regel heranzukommen. Das Additionsbeispiel aus dem Bereich der Mathematik

dient hier nur zur Veranschaulichung. Einsatz findet das maschinelle Lernen eher in anderen Bereichen, wo die möglichen Eingaben deutlich komplexer sind, zum Beispiel in der Bildverarbeitung von Fotos bestehen aus tausenden Pixel.

Was bedeutet das nun für die Generierung von Kunst mit Hilfe von maschinellem Lernen? Während bei generativer Computerkunst der oder die Künstler:in die Regeln festlegt und dadurch Einfluss auf den generativen Prozess hat, entfällt dies beim maschinellen Lernen. Hier generiert der Computer die Kunst selbstständig mit Hilfe der zuvor selbst gelernten Regeln. Als Eingabe braucht der Computer nur Trainingsdaten, also Beispiele von Gemälden, Zeichnungen, etc. Dieser konzeptionelle Unterschied zwischen klassischem Programmieren und maschinellem Lernen ist gewaltig und hatte zu außerordentlichen Erfolgen in vielen Bereichen geführt. Eine weitere Steigerung hat das Gebiet der künstlichen Intelligenz durch die Entwicklung von tiefen, neuronalen Netzen, sogenanntes Deep Learning, erfahren. Verschiedene Architekturen dieser Netze können für die Bild- und Sprachanalyse genutzt werden. Zur Erzeugung von Daten jeglicher Art, also auch von Kunst, finden spezielle Deep-Learning-Architekturen Anwendung, auf die wir im folgenden Teil detaillierter eingehen werden.

„Kreative“ Neuronale Netze

Durch neue Entwicklungen und gesteigerte Rechenleistung konnte die alte Idee von künstlichen, neuronalen Netzen weiterentwickelt werden und es entstanden tiefe Netze mit vielen Schichten von künstlichen Neuronen. Der Durchbruch des Deep Learning wurde Anfang der 2010er Jahre erreicht mit Ergebnissen, die über denen von Menschen lagen, insbesondere für perzeptorische Aufgaben der Bild- und Sprachverarbeitung. Diese großen Deep-Learning-Architekturen wurden genutzt, um zu lernen, wie Eingabedaten auf Ausgabedaten abgebildet werden: Ein Modell der Daten wird erzeugt. Das Ziel war dabei meistens die Klassifikation, zum Beispiel das Erkennen von Hunden oder Katzen auf Fotos (Ciresan, 2012).

Diesen Prozess kann man aber auch umdrehen. Ein Modell, was auf das Erkennen von Hunden trainiert wurde, kann nun rückwärts angewendet werden: Gegeben ein Foto, welche Pixel müssen wie verändert werden, damit das Modell einen Hund erkennt? Die so veränderten Bilder haben eine stark psychedelische Komponente und erinnern an Traumsequenzen (McCormick, 2015). Daher wurde diese Methode Deep Dream (Mordvintsev, 2015) genannt. Wählt man weißes Rauschen als Eingabe, erzeugt diese Methode Bilder von zuvor erlernten Objekten, siehe Abbildung 1.



Abbildung 1: SEQ Abbildung | *ARABIC1: Ein von DeepDream erzeugtes Bild. (aus Berov, 2016)

Eine weitere wichtige Methode der Anwendung von Deep Learning im Bereich der Kunst ist der Neural Style Transfer (NST) (Gatys, 2016). Hierbei werden der Inhalt eines Bildes und dessen Stil getrennt, sodass der Stil ausgetauscht werden kann. Beispielsweise kann man den Inhalt eines beliebigen Fotos mit dem Stil von

Van Goghs Sternennacht kombinieren, siehe Abbildung 2. Dabei erzeugt das Modell zwar keine neue Kunst, aber die Fähigkeit Inhalt und Stil zu trennen war ein wesentlicher Schritt hin zu einem feingranulareren Umgang der Netze und zu besserem konzeptionellen Verständnis.

Während die beiden beschriebenen Methoden auf Convolutional Neural Networks (CNN) (Gu, 2018) aufbauen, die primär zur Klassifikation entwickelt wurden, hat sich mit den Generative Adversarial Networks (GAN) (Goodfellow, 2014) eine weitere, neuronale Netzwerk-Architektur etabliert, die speziell zur Erzeugung von Daten, z. B. von Bildern, entwickelt wurde.

Ein GAN besteht dabei aus zwei neuronalen Netzen, dem Generator-Netzwerk und dem Diskriminator-Netzwerk (Auswahl-Netzwerk). Die Aufgabe des

Diskriminators ist zu unterscheiden, ob ein Eingabebild ein echtes Bild ist, oder ein vom Generator erzeugtes Bild. Mit jedem Bild, was der Generator erzeugt und der Diskriminator als Fälschung entlarvt, lernt der Generator ein Bild zu erzeugen, was ein wenig „echter“ aussieht als das vorherige. Dieser iterative Prozess beginnt mit einem Bild von



Abbildung 2: Beispielanwendung des Neural-Style-Transfer-Algorithmus: oben links das originale Foto und in klein jeweils das Bild dessen Stil zur Manipulation genutzt wurde. (aus Gatys, 2016)

zufälligen Pixeln (Rauschen) und führt dazu, dass nach einer Weile der Diskriminator nicht mehr unterscheiden kann, ob das ihm vorgelegte Bild nun ein Echtes oder Erzeugtes ist. Ein Beispiel eines von einem GAN erzeugten Bild ist das Werk Edmond de Belamy, welches durch seine Versteigerung bei Christie's nicht nur hohe Aufmerksamkeit erhielt, sondern auch Fragen nach Urheberschaft und geistigem Eigentum von computergenerierter Kunst neu aufwarf (McCormack, 2019).

Während bei diesen Standard-GANs der menschliche Einfluss auf das erzeugte Bild auf die Auswahl der Trainingsbilder beschränkt ist, haben neue Entwicklungen nicht nur neue Anwendungen ermöglicht (Creswell, 2018), sondern bieten auch die Möglichkeit, stärker auf den Erzeugungsprozess Einfluss zu nehmen, zum Beispiel durch zusätzliche Bedingungen, die für das erzeugte Bild gelten sollen (Mirza, 2014). So lässt sich beispielsweise der Stil des zu erzeugenden Bildes vorher festlegen, siehe Abbildung 3.

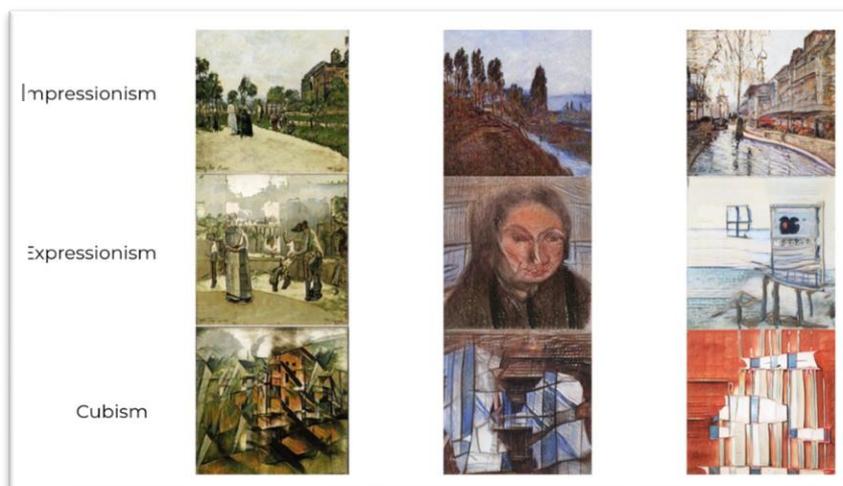


Abbildung 3: Generierte Bilder aus drei verschiedenen, zufälligen Eingaben und jeweils einem Kunststil als Bedingung. (aus REF: MP, 2021)

Aktuelle Forschung (Dobler, 2022) basierend auf einer Netzwerkarchitektur von NVIDIA (Karras, 2020) ermöglicht sogar die Eingabe von mehreren zusätzlichen Bedingungen, die das zu erzeugende Bild erfüllen soll. Zum Beispiel kann das Modell

Metadaten mit den Trainingsdaten assoziieren, die dann als weitere Eingabe eine gezieltere Bildgenerierung ermöglichen, siehe Abbildung 4.

Diese Modelle sind sehr groß und beschreiben einen latenten Raum möglicher Bilder. Dieser Raum wird durch das Training strukturiert, sodass sich in diesem sehr hochdimensionalen Raum bestimmte Eigenschaften in Dimensionen, d. h. bestimmten Richtungen in diesem Raum widerspiegeln. Dies wurde zuerst bei Fotos von Gesichtern entdeckt, als sogenannter Smile-Vektor (Radford, 2015). Übertragen auf ein Modell zur Generierung von Kunst mit zusätzlichen Metadaten, kann man sich in dem latenten Raum entlang einzelner Dimensionen bewegen, und damit direkt auf die zu erzeugenden Bilder Einfluss nehmen, siehe Abbildung 5 (aus Dobler, 2022).



Abbildung 4: Generierte Bilder mit mehreren Bedingungen.
(aus REF: MP, 2021)



Abbildung 5: Eine Annäherung an Leonardo da Vincis Mona Lisa im latenten Raum (Mitte) und davon Richtung Ehrfurcht (nach links) und Furcht (nach rechts) im latenten Raum bewegend generierte Bilder. (aus REF: MP, 2021)

Dies erlaubt das Explorieren des gelernten, latenten Raums und damit nicht nur mehr Kontrolle über den Erzeugungsprozess, sondern auch interessante Einsichten in die Trainingsdaten. Durch Analyse des Raums kann zum Beispiel das Charakteristische einer Kunstrichtung oder, wie im Beispiel der Abbildung, Emotionen beim Kunstbetrachter rein datengetrieben erforscht werden. Dafür sind entsprechend annotierte Trainingsdaten notwendig. Für das Lernen von Emotionen von Kunstwerken zur Strukturierung des latenten Raumes wurden 80.000 Kunstwerke von Menschen mit ihren assoziierten Emotionen annotiert (Achlioptas, 2021).

Der künstlerische Prozess

Die Möglichkeiten, mit Hilfe künstlicher Intelligenz fotorealistische Bilder zu erzeugen, sind beeindruckend. Was wie zuvor erläutert noch fehlt, ist die Kreativität und das Verständnis. Letzteres ist nicht nur zur Erzeugung von Kunst sehr wichtig, sondern auch in sehr praktischen Anwendungen von Deep Learning, wo eine Fehleinschätzung ernste Konsequenzen hat, zum Beispiel bei der Erkennung von Tumorzellen. Ein wesentlicher Nachteil sehr großer, neuronaler Netzwerke ist ihre Unübersichtlichkeit und Komplexität und daraus resultierende Unmöglichkeit des Nachvollziehens einzelner Entscheidungen des Modells. Man spricht daher von Black-Box-Modellen, d. h., man sieht nur die Daten, die man hineingibt und das Ergebnis, das am Ende herauskommt; eine sinnvolle Introspektion ist nicht möglich. Ob und gegebenenfalls was das Modell „verstanden“ hat, ist nicht nachvollziehbar.

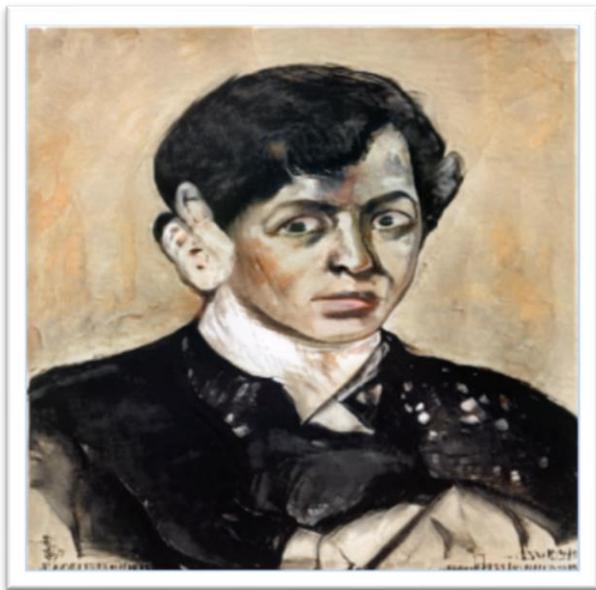


Abbildung 6: Generiertes Portrait mit 4 Ohren; das Konzept „Ohr“ wird von diesem neuronalen Netz noch nicht verstanden. (aus Dobler, 2022)

Die aktuelle Forschung beschäftigt sich daher intensiv mit den Möglichkeiten, Ausgaben von Deep-Learning-Modellen zu erklären. In der Praxis könnte dies folgendermaßen aussehen: Bei der Bilderkennung können Modelle beispielsweise diejenigen Pixel hervorheben, aufgrund deren sie glauben, eine Giraffe statt eines Elefanten zu erkennen. Und dies sollte auch bei einem rosa Elefanten noch funktionieren oder bei einem Elefanten, den man nur von hinten sieht. Dafür ist ein komplexes Verständnis von Elefanten notwendig. Je-

doch ist selbst das offenkundig noch weit von einem Verständnis entfernt, welches nötig wäre, um bedeutende Kunst zu erschaffen, wo es nicht nur um sachliches Verständnis geht, sondern auch um das Erkennen komplexer Zusammenhänge, Emotionen und tradiertem Wissen auf unterschiedlichen Ebenen. Hin und wieder wird das fehlende Weltverständnis in einzelnen erzeugten Bildern von kunstgenerierenden Modellen evident und führt zu skurrilen Ergebnissen, siehe Abbildung 6 (aus Dobler, 2022).

Erschwerend kommt bei der Erschaffung von Kunst hinzu, dass es nicht nur einem Verständnis der Welt bedarf, sondern auch einem Verständnis des Menschen und dessen Wahrnehmung, mithin einer echten, holistischen (künstlichen) Intelligenz. Die bespro-

chenen Methoden entstammen alle dem Bereich des maschinellen Lernens. Insbesondere braucht man Trainingsdaten mit Hilfe derer die Modelle lernen. Für die oben vorgestellten GANs bedeutet dies, dass die Bilder, die man dem Netzwerk zum Trainieren zeigt, den groben konzeptionellen Raum festlegen, indem das Modell neue Bilder erzeugen kann. Dies hat zur Folge, dass immer nur ähnliche Bilder oder Kombinationen von Bildern aus den Trainingsdaten erzeugt werden können, da das Ziel beim Trainieren die Täuschung des Diskriminators war, so dass dieser das erzeugte Bild für ein echtes Bild aus den Trainingsdaten hält. Es wird hierbei - wie bei fast allen Methoden des maschinellen Lernens - eine Zielfunktion optimiert, d. h. das Modell, welches aus Millionen von Parametern besteht, ändert, solange diese Parameter bis die resultierenden Bilder aussehen, als könnten sie Teil der Trainingsbeispiele sein. Dies ist ein rein technisches Unterfangen und setzt nicht nur kein Verständnis der Welt oder auch nur der Daten voraus, sondern schließt Kreativität (die Erschaffung von etwas konzeptionell Neuem) von vornherein aus (Stanley, 2015; Amabile, 1983; Hertzmann, 2018; Still, 2019). Um dieser konzeptionellen Kreativlosigkeit zu entrinnen, versuchen Wissenschaftler Kreativität explizit in den Generierungsprozess mit einzubeziehen, indem das neuronale Netz Eigenschaften von Kunststilen erlernt, um dann bewusst bei der Generierung davon abzuweichen (Elgammal, 2017).

Das (noch) fehlende Verständnis und die damit einhergehende fehlende Intention führt zwar zu (teilweise) sehr guter Imitation, aber inwieweit man die Imitation oder den Anschein von Kreativität tatsächlich Kreativität im engeren Sinn nennen darf, bleibt offen. Legt man eine weniger strenge Definition von Kreativität zugrunde, können Computer durchaus kreativ sein (Jordanous, 2014). Insbesondere kann Computerkreativität genutzt werden, um menschliche Kreativität zu erforschen oder vielleicht sogar zu erklären. Auch ist die Wechselwirkung von Kunst und Kreativität je nach Kunstbegriff (und Kreativitätsbegriff) nicht eindeutig spezifiziert. Möglich, dass man über den Umweg der Computerkreativität - was immer man darunter letztlich versteht - etwas über menschliche Kreativität lernen kann. Selbst wenn man Computern die Kreativität grundsätzlich abspricht, so können ihre Ergebnisse doch mindestens als Inspiration für menschliche Künstler:innen dienen. Oder man betrachtet Computer im Allgemeinen und Deep-Learning-Methoden im Speziellen als Werkzeuge des/der Künstlers/Künstlerin. Dies negiert jedoch die erschaffende Qualität gerade bei GANs, die etwas erzeugen, ob nun kreativ oder nicht. Eine treffendere Betrachtung stellt das Zusammenwirken von generativen Algorithmen und menschlichen Künstler:innen in den Mittelpunkt. Man spricht dann von *Co-Creation* des Künstlers und einer KI (Cizek, 2019).

Interaktion von Kunstschaffenden und KI

Künstler:innen können auf unterschiedliche Weise Einfluss auf die Ergebnisse eines GANs nehmen. Abhängig vom Zeitpunkt des konkreten Eingreifens, kann man drei unterschiedliche Arten der Einflussnahme unterscheiden:

1. Beim Design des Systems
2. Während des Lernprozesses
3. Während des Erzeugens

Die Ausgestaltung der GAN-Architektur bildet die Ausgangslage. Hier wird festgelegt, zu was das GAN theoretisch in der Lage ist, sprich: die Mächtigkeit des Modells. Dies geschieht zum einen über das Festlegen von Hyperparametern, wie zum Beispiel die Anzahl der Dimensionen des latenten Raums, die Anzahl der Neuronen pro Schicht, sowie die Anzahl von Schichten des neuronalen Netzes. Auch die Art und Anzahl der Verbindungen zwischen den Schichten prägt das neuronale Netz. Darüber hinaus muss festgelegt werden, wie die Eingabe und Ausgabe aussehen soll. Das heißt, die Anzahl der Pixel für die Trainingsbilder und die Repräsentation dieser Pixel, zum Beispiel als RGB-Werte. Diese Faktoren legen fest, was das Netz lernen und erzeugen kann. Nutzt der Künstler eben RGB-Werte, so können die erzeugten Pixel eines Bildes auch nur Farben aus diesem Bereich darstellen. Vereinfacht könnte man sagen, dass das Design, beziehungsweise die Designentscheidungen den latenten Raum aufspannen, indem neue Werke erzeugt werden können. Erwartet die GAN-Architektur nur Pixel, die entweder schwarz oder weiß sind, beinhaltet der latente Raum maximal alle denkbaren Kombinationen von schwarzen und weißen Pixel, aber keine farbigen. Dabei gilt zu bedenken, dass, je mächtiger das Modell ist, desto mehr Parameter gelernt werden müssen. Dies wiederum bedeutet, dass eine größere Anzahl an Trainingsbeispielen notwendig ist, um komplexere Zusammenhänge zu erlernen. Zur Veranschaulichung dazu folgende Rechnung: Selbst bei sehr kleinen Bildern, sagen wir 5 mal 5 Pixel und nur der Wahl zwischen Schwarz und Weiß, ergeben sich mehr als 33 Millionen mögliche, unterschiedliche Bilder. Der Eingaberaum beinhaltet also 33 Millionen Punkte und jeder Punkt entspricht einem unterschiedlichen Bild. Die Menge der sinnvollen Bilder, das heißt der Bilder, die nicht nach zufälligem Rauschen aussehen, ist viel, viel kleiner. Und je größer der Raum der möglichen Bilder gewählt wird, umso schwieriger ist es sinnvolle Bilder zu finden und eine Struktur in diesem Raum zu finden. Genau das ist die Aufgabe des GANs: Einen latenten Raum zu erzeugen und so zu strukturieren, dass wesentliche Merkmale der Trainingsbeispiele gut repräsentiert werden und die Trainingsbeispiele sinnvolle Nachbarschaftsbeziehungen in diesem Raum ausbilden, das

heißt, semantisch ähnliche Bilder in diesem Raum nahe beieinander liegen. Dieser latente Raum ist eine komprimierte Version des Eingaberaums und wird definiert durch die Trainingsbeispiele. Dies bringt uns zur zweiten Art der Einflussnahmemöglichkeiten: der Auswahl der Trainingsbeispiele während des Lernprozesses. Diese hat wahrscheinlich den größten Einfluss auf die Ergebnisse. Sie gilt nicht nur für GANs oder das Erzeugen von Kunst, sondern für alle Methoden des maschinellen Lernens. So möchte man im Allgemeinen möglichst objektive Ergebnisse erzielen. Zum Beispiel beim Einsatz von KI für die Vorhersage von Rückfallwahrscheinlichkeiten ehemaliger Straftäter. Durch einen unausgewogenen Trainingsdatensatz, in dem zum Beispiel die Hautfarbe erfasst und verwendet wurde, lernt die KI ein mögliches Bias in Bezug auf die Hautfarbe mit. Damit sind keine fairen Vorhersagen zu erzielen. Was im Kontext der Strafverfolgung vermieden werden muss, nämlich Diskriminierung aufgrund der Hautfarbe durch ungeeignete Auswahl und Repräsentation von Trainingsbeispielen, kann bei GANs genutzt werden, um das Ergebnis in bestimmte Richtungen zu beeinflussen. Der Künstler kann durch die Auswahl der Trainingskunstwerke beliebige latente Räume vom Netzwerk lernen lassen. Ein Beispiel ist die Erzeugung eines Portraits, indem das GAN nur auf Portraitbildern trainiert wird. Interessanter ist sicherlich die Kombination mehrerer Arten von Bildern, zum Beispiel barocke Portraits und kubistische Stadtansichten, um dann den latenten Raum „dazwischen“ zu explorieren. Moderne GAN-Architekturen erlauben zusätzlich das Definieren von Bedingungen, zum Beispiel in Form von Stichwörtern, die den Inhalt definieren, das Genre, usw. (siehe Abbildung 4). Dadurch lernt das GAN nicht nur einen strukturierten, latenten Raum, sondern auch das Assoziieren von Bereichen dieses Raums mit bestimmten Bedingungen.

Schließlich bleibt noch die Phase des tatsächlichen Generierens eines Bildes mit dem trainierten GAN. Hier kann der Künstler unmittelbar Einfluss nehmen – entweder durch die Selektion bestimmter erzeugter Bilder und dem Verwerfen von Bildern, die dem Künstler als nicht-geeignet erscheinen, oder durch das Arrangieren und Komponieren von größeren Werken. Denkbar ist auch die Einbindung der Selektion und Komposition in den Erzeugungsprozess des GANs. Eine Art des *Human-in-the-Loop-Konzepts*, bei dem der Mensch als weitere Eingabequelle für Algorithmen dient. Dies könnte auf vielfältige Art und Weise geschehen, zum Beispiel indem der Künstler die Position bestimmter Objekte auf der virtuellen Leinwand und auch die Farbgebung verschiedener Bereiche manuell festlegt. Hier ist der Kreativität des Künstlers keine Grenze gesetzt. Verschiedenste Projekte und Workshops aus dem Bereich Co-Creation von KI und Mensch zeigen die Möglichkeiten in diesem Kontext schon jetzt (z. B. auf <https://medium.com/artists-and-machine-intelligence>, <https://computationalcreativity.net/workshops/cocreative-iccc20/>, <https://neuripscreativityworkshop.github.io/2020/>).

Fazit

Unter dem Sammelbegriff der künstlichen Intelligenz werden mathematische bzw. statistische Methoden zusammengefasst, die aus Trainingsbeispielen Muster erkennen. Diese Algorithmen des maschinellen Lernens haben durch die Entwicklung von Deep-Learning-Methoden in den letzten Jahren enorm an Bedeutung gewonnen. Durch das Trainieren sehr großer neuronaler Netze mit sehr vielen Trainingsbeispielen ist es möglich geworden, semantische Informationen in komplexen Eingabedaten, wie beispielsweise Text oder Bilder, zu finden. Generative Adversarial Networks (GANs) können diese Informationen nutzen, um etwa Bilder zu erzeugen, die denen der Trainingsmenge ähneln. Man könnte also von Amateurlünstler:innen sprechen, der/die nicht künstlerisch geschult ist, und von (kreativen) Eingaben abhängig ist (Shokry, 2021): zum Einen von den Programmierern, die die Mächtigkeit und den Raum der möglichen zu generierenden Bilder festlegen, zum Anderen von den Künstlern, die durch die Auswahl der Trainingsbilder und das konzeptionelle Steuern des Erzeugungsprozesses die „Amateurkunst“ veredeln. Um wirklich Kunst zu erschaffen, müssen jedoch die Schlüsselherausforderungen der allgemeinen KI gelöst werden, d. h. ein Verständnis über die Welt (und den Menschen) und darauf aufbauend die Fähigkeit, kreativ und intentional Neues zu erschaffen.

Als Werkzeug und zur Analyse können die heutigen Methoden des maschinellen Lernens gute Dienste leisten. Datengetriebene kunsthistorische Forschung kann durch die Auswertung und Erforschung der gelernten latenten Räume bisher unentdeckte Muster finden. Zu Schulungszwecken können wesentliche Eigenschaften von Kunststilen oder Künstler:innen herausgearbeitet werden. Durch das Interagieren von Laien sowie Expert:innen mit neuronalen Netzen können interessante Fragestellungen entstehen und von Menschen erzeugte Kunst neu betrachtet und bewertet werden. Für Kunstschaffende ergibt sich die Möglichkeit, auf das (oberflächliche) Wissen über zehntausende Gemälde in Form eines gelernten Modells zugreifen zu können und in Co-Creation-Setups im Zusammenwirken mit KI-Methoden Neues zu erschaffen.



Prof. Dr. Ralf Krestel ist Professor für Information Profiling and Retrieval an der ZBW – Leibniz-Informationszentrum Wirtschaft und der Christian-Albrechts-Universität zu Kiel. Er forscht zu Themen der Künstlichen Intelligenz mit Fokus auf Sprachverarbeitung und Textanalyse. Weitere Themen sind das Erzeugen von Kunst mit neuronalen Netzen und der Einsatz von Deep-Learning-Methoden in der Bildverarbeitung von Kunstwerken.

Literaturverzeichnis

Achlioptas, P., Ovsjanikov, M., Haydarov, K., Elhoseiny, M., & Guibas, L. J. (2021). *Artemis: Affective language for visual art*. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 11569-11579.

Amabile, Teresa (1983), *The social psychology of creativity*. Springer.

Anna Jordanous (2014). *What is computational creativity? The Creativity Post* (10.04.2014) https://www.creativitypost.com/science/what_is_computational_creativity

Berov, L.; & Kuhnberger, K.-U. (2016). Visual hallucination for computational creation. In *Proceedings of the Seventh International Conference on Computational Creativity*. 107-114.

Chomsky, N. (1957). *Syntactic Structures*. The Hague: Mouton.

Meyer, R. (2017). When a Robot Names a New Color of Paint. In *The Atlantic* (22.05.2017). <https://www.theatlantic.com/technology/archive/2017/05/when-a-robot-names-a-new-color-of-paint/527421/>

Ciresan, D.; Meier, U.; Schmidhuber, J. (2012). *Multi-column deep neural networks for image classification*. In *Proceedings of the Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE, 3642–3649. doi:10.1109/cvpr.2012.6248110

Cizek, K.; Uricchio, W & Wolozin S. (2019). *Collective wisdom - part 6: media co-creation with non-human systems*. *MITP Works in Progress* (03.06.2019). <https://wip.mit-press.mit.edu/pub/collective-wisdom-part-6/release/1>

Colton, S. (2012). *The painting fool: Stories from building an automated painter*. In *Computers and creativity*. Springer, Berlin, Heidelberg. 3-38.

- Creswell, A., White, T., Dumoulin, V., Arulkumaran, K., Sengupta, B., & Bharath, A. A. (2018). *Generative adversarial networks: An overview*. In *Signal Processing Magazine*, 35(1), 53-65. IEEE.
- Dobler, K., Hübscher, F., Westphal, J., Sierra, A., de Melo, G., & Krestel, R. (2022). *Art Creation with Multi-Conditional Style GANs*. arXiv preprint. arXiv:2202.11777.
- Elgammal, A., Liu, B., Elhoseiny, M., & Mazzone, M. (2017). *CAN: Creative adversarial networks generating „Art“ by learning about styles and deviating from style norms*. In *Proceedings of the 8th International Conference on Computational Creativity (ICCC)*. Georgia Institute of Technology.
- Gangadharbatla, H. (2021). *The Role of AI Attribution Knowledge in the Evaluation of Artwork*. In *Empirical Studies of the Arts*. February 2021 online. SAGE Publications. 1-19. doi:10.1177/0276237421994697
- Gatys, L. A., Ecker, A. S., & Bethge, M. (2016). *Image style transfer using convolutional neural networks*. In *Proceedings of the IEEE conference on computer vision and pattern recognition (CVPR)*. 2414-2423.
- Ghazvininejad M, Shi X, Choi Y, Knight K (2016) *Generating topical poetry*. In *Proceedings of the conference on empirical methods in natural language processing (EMNLP)*, ACL, 1183-1191.
- Goodfellow, I.; Pouget-Abadie, J.; Mirza, M.; Xu, B.; Warde-Farley, D.; Ozair, S.; Courville, A.; Bengio, Y. (2014). *Generative Adversarial Nets*. In *Proceedings of the International Conference on Neural Information Processing Systems (NIPS)*. 2672-2680.
- Gu, J., Wang, Z., Kuen, J., Ma, L., Shahroudy, A., Shuai, B., ... & Chen, T. (2018). *Recent advances in convolutional neural networks*. *Pattern Recognition*, 77, 354-377. Elsevier.
- Hertzmann, A. (2018). *Can computers create art?* In *Arts*, 7 (2), 18. Multidisciplinary Digital Publishing Institute. 1-25.
- Karras, T., Aittala, M., Hellsten, J., Laine, S., Lehtinen, J., & Aila, T. (2020). *Training generative adversarial networks with limited data*. *Advances in Neural Information Processing Systems (NeurIPS)* 33. 1-11.
- Lopez-Rincon, O., Starostenko, O., & Ayala-San Martín, G. (2018). *Algorithmic music composition based on artificial intelligence: A survey*. In *Proceedings of the International Conference on Electronics, Communications and Computers (CONIELECOMP)*, IEEE, 187-193. doi: 10.1109/CONIELECOMP.2018.8327197.

McCormack, J., Gifford, T., & Hutchings, P. (2019). *Autonomy, authenticity, authorship and intention in computer generated art*. In *International Conference on Evolutionary and Biologically Inspired Music, Sound, Art and Design*. 35-50. Springer.

Mirza, M., & Osindero, S. (2014). *Conditional generative adversarial nets*. arXiv preprint arXiv:1411.1784.

Mordvintsev, A.; Olah, C.; Tyka, M. (2015). *Inceptionism: Going Deeper into Neural Networks*. Via Google Research (17.06.2015). <https://web.archive.org/web/20150703064823/http://googleresearch.blogspot.co.uk/2015/06/inceptionism-going-deeper-into-neural.html>

Radford, A., Metz, L., & Chintala, S. (2015). *Unsupervised representation learning with deep convolutional generative adversarial networks*. arXiv preprint arXiv:1511.06434.

Rich McCormick (2015) (7 July 2015). *Fear and Loathing in Las Vegas is terrifying through the eyes of a computer*. In *The Verge* (07.05.2015). <https://www.theverge.com/2015/7/7/8904641/fear-and-loathing-clip-google-deep-dream-visualization>

Roemmele, M. (2016). *Writing stories with help from recurrent neural networks*. In *Proceedings of the AAAI Conference on Artificial Intelligence*. 4311-4312.

Shokry, Y. (2021). *AI art and the nature of creation*. In *Arts Management and Technology Laboratory* (22.07.2021) <https://amt-lab.org/blog/2021/7/ai-art-and-the-nature-of-creation>

Stanley, K. O., & Lehman, J. (2015). *Why greatness cannot be planned: The myth of the objective*. Springer.

Still, A., & d'Inverno, M. (2019). *Can machines Be artists? A Deweyan response in theory and practice*. In *Arts 8* (1), 36. Multidisciplinary Digital Publishing Institute. 1-13.

Tatarkiewicz, W (1979). *Geschichte der Ästhetik. Band 1: Die Ästhetik der Antike*. S. 143 f., 150–158, 164–166 (mit Zusammenstellung der Quellenzeugnisse zu Platons Kunstverständnis in Übersetzung S. 160–167).

Zulić, H. (2019). *How AI can change/improve/influence music composition, performance and education: three case studies*. *INSAM Journal of Contemporary Music, Art and Technology*, 1(2). 100-114.