

DIE <GESTALT> DER KI

Jenseits von Holismus und Atomismus

Für den Anfang ein Gesicht: Im Herbst 2017 veröffentlichte der Medienkünstler Mario Klingemann auf Twitter ein Bild,¹ das sich auf den ersten Blick recht vertraut, ja klischeehaft ausnahm (Abb. 1, links).² Das Lächeln war da, ebenso die leichte Drehung des Kopfes nach rechts und der bekannte Blick, der auf oder etwas hinter den die Betrachter_in gerichtet ist. Und doch, ein näheres Hinsehen weckte Zweifel, ob es sich hier wirklich um das wieder und wieder reproduzierte Antlitz der Gioconda handelte. Tatsächlich war es eine Schöpfung Klingemanns selbst, der das kurz zuvor veröffentlichte Deep-Learning-Modell Pix2Pix auf eine bearbeitete Bilddatei von Leonardos Gemälde angewandt hatte.

Klingemanns Version von Pix2Pix leistet, was bis dahin nur in Science-Fiction-Filmen wie *Blade Runner* (USA, HK 1982) möglich schien, in denen eine Figur auf das grobkörnige Bild einer Überwachungskamera zeigt und dem Computer aufträgt, es zu <verbessern> (*to enhance*).³ Dieses vormals fiktive *enhancement* erlaubt es nun auch in Wirklichkeit, den Datenverlust auszugleichen, der bei einem Bild mit per *downsampling* verschlechterter Auflösung auftritt, und hebt Details hervor, die im Original nicht zu erkennen waren. So bestand auch Klingemanns Input nicht aus der echten *Gioconda*, sondern einer weichgezeichneten Schwarz-Weiß-Version des Gemäldes (Abb. 1, Mitte). Aus diesem Bild, in dem die Züge des Gesichts praktisch unsichtbar sind, erstellte Pix2Pix den Output. Der direkte Vergleich mit dem Original zeigt deutlich die Unterschiede zwischen Klingemanns und Leonardos *Mona Lisa* (Abb. 1, rechts); die übertriebenen Details fallen auf, die *glitches* um die Augenlider, das wallende Haar, das eher an eine Shampoo-Werbung erinnert als an das Bild einer florentinischen Frau aus dem 16. Jahrhundert. Der Output ist keine Verbesserung des Originals, sondern eine Neuschöpfung, die auf einigen wenigen Merkmalen eines Gesamterscheinungsbildes basiert. Pix2Pix stellt damit nicht zerstörte Details wieder her – gemäß dem Prinzip der Entropie bleiben verlorene Informationen verloren –, sondern interpoliert plausibel ein Gesicht aus dem unscharfen

¹ Mario Klingemann in: Twitter, 26.11.2017, twitter.com/quasimondo/status/934709314375372801 (4.7.2020).

² Dieser Text geht aus einem Vortrag hervor, den ich am 3.2.2020 auf der Tagung «Things Beside Themselves: Mimetic Existences» am IKKM Weimar gehalten habe. Ich danke den Teilnehmer_innen für ihre Kommentare, Mario Klingemann für die Abdruckerlaubnis seiner Bilder, sowie Florian Sprenger und Jana Mangold für Verbesserungsvorschläge.

³ Die Architekturen neuronaler Netze werden meist als PDF auf dem Open-Access-Archiv [arXiv.org](https://arxiv.org) veröffentlicht und sind nicht *peer reviewed*, was einerseits die schnellstmögliche Publikation erlaubt und andererseits die mitgelieferten exemplarischen Ergebnisse sowie den über GitHub verfügbaren Code zur Bewertungsgrundlage macht. Zu Pix2Pix siehe: Phillip Isola u. a.: Image-to-Image Translation with Conditional Adversarial Networks, in: *ArXiv*, 21.11.2016, arxiv.org/abs/1611.07004 (6.5.2020).



Eingabebild, indem es auf ein Wissen darüber zurückgreift, *wie Gesichter normalerweise aussehen*.⁴ Und dies geschieht nicht, indem explizite Regeln darüber angewendet werden, aus welchen Elementen sich ein Gesicht zusammensetzt, sondern weil es ohne jegliche Anleitung gelernt hat, was mit großer Wahrscheinlichkeit <Gesichthaftigkeit> ausmacht. Somit ist Klingemanns Bild kein Kompositbild, also weder Collage noch ein Mittelwertbild im Stil von Francis Galton.⁵ Das leuchtet bei einem Blick auf ein anderes Gesichter erzeugendes künstliches neuronales Netz (KNN) namens StyleGAN₂ ein, das auf der Website thispersondoesnotexist.com abrufbar ist (Abb. 2). Wie die URL bereits andeutet, zeigen diese Bilder keine tatsächlich existierenden Personen, vielmehr werden die «Fotos» bei jeder Aktualisierung des Browser-Fensters dynamisch generiert. Diese Gesichter weisen genügend individuelle Merkmale auf, um zu vermuten, dass sie weder eine Collage noch eine bloße Ansammlung der häufigsten Merkmale aus einer Reihe sind. Was auch immer die Gesichter in diesem Prozess erzeugt – offensichtlich räumt es dem Ganzen Vorrang vor seinen Teilen ein. Es scheint, als hätten Pix2Pix und StyleGAN₂ die *Gestalt* eines Gesichts gelernt.

Im Folgenden möchte ich den Begriff der Gestalt aufgreifen. Dabei soll es weniger um menschliche Wahrnehmung gehen, die für gewöhnlich im Fokus

Abb. 1 v.l.n.r. Output und Input von Mario Klingemanns Anwendung des Pix2Pix-Deep-Learning-Modells (2017) im Vergleich zu Leonardo da Vincis *La Gioconda* (*Mona Lisa*, ca. 1502/03)

Abb. 2 Generierte Porträts auf thispersondoesnotexist.com, 2019

⁴ Dies im vorliegenden Fall freilich unter Maßgabe des verwendeten Trainingsdatensatzes, der Klingemann zufolge nur weibliche Gesichter umfasst. Siehe Mario Klingemann, in: Twitter, 26.11.2017, twitter.com/quasimondo/status/934709314375372801 (4.7.2020).

⁵ Siehe Suzanne Bailey: Francis Galton's Face Project. Morphing the Victorian Human, in: *Photography and Culture*, Bd. 5, Nr. 2, 2012, 189–214.

der Gestaltpsychologie steht, sondern um die konzeptuellen Voraussetzungen, die bei der Repräsentation und Emergenz nicht-ableitbarer Ganzheiten in diskreten Systemen eine Rolle spielen. Der Begriff <Gestalt> soll helfen, einige der Annahmen zu erörtern, die dem Verständnis eines Teilgebiets Künstlicher Intelligenz zugrunde liegen, das unter dem Begriff des *deep learning* zusammengefasst und mithilfe mehrschichtiger Perzeptrone implementiert wird.⁶ Ich werde argumentieren, dass solch maschinelles Lernen jenseits der beiden philosophischen Linien des Atomismus und Holismus konzeptualisiert werden sollte, deren Argumente normalerweise mobilisiert werden, um die Möglichkeit oder Unmöglichkeit künstlicher Intelligenz zu diskutieren. Stattdessen wäre es angemessener, entweder von einem <gemischten Typ> oder von etwas ganz anderem zu sprechen. Und ich plädiere dafür, diesen gemischten Typ nicht nur als technische Beschreibung, sondern auch als konzeptuelles Werkzeug für nicht-technische Bereiche zu verwenden – als *intuition pump*, wie Daniel Dennett es nennt.⁷

Dass im Folgenden gerade das Gesicht als beispielhaft für Gestalteffekte herangezogen wird, ist kein Zufall. <[D]as menschliche Gesicht mit seiner unvergleichlichen Situationsbedeutung>⁸ ist – von Georg Simmels ästhetischer Einheitsfunktion über Emmanuel Lévinas' konstitutive Verbindung von <Antlitz und Ethik> bis zu Hans Beltings Bildanthropologie – ein Untersuchungsgegenstand mit eigener philosophischer, kunst- und kulturwissenschaftlicher Geschichte.⁹ Als Musterbeispiel maximal irreduziblen Sinns, gar als <anthropogenetischer Urtypus> von Bedeutsamkeit überhaupt,¹⁰ eignet es sich besonders gut, um daran die möglichen Korrespondenzen und Inkongruenzen von technischen Strukturen und lebensweltlicher Sinnerwartung zu untersuchen.

I. Holismus, Atomismus, Gestalt

Auf Klingemanns <Gioconda> den Begriff Gestalt anzuwenden, bedeutet zunächst, die Definition der Gestaltpsychologie aufzurufen. Zugleich aber wird die Beziehung zwischen Teilen und Ganzem auch jenseits der Gestalttheorie metonymisch verwendet, um eine Trennung in der Geschichte der Philosophie zwischen zwei Traditionen oder Denkrichtungen zu beschreiben: dem Holismus und dem Atomismus.¹¹

<Atomismus> meint die Überzeugung, dass jedes Ding einschließlich seiner spezifischen Eigenschaften durch die Analyse seiner konstitutiven Elemente erklärt werden kann und dass eine solche Erklärung abschließend ist. Die modernen Anhänger_innen des Atomismus stehen in der Tradition Gottlob Freges; der logische Atomismus Bertrand Russells, die Bildtheorie des frühen Wittgenstein, der logische Positivismus des Wiener Kreises und die Sinnesdatentheorie G. E. Moores und A. J. Ayers sind die wichtigsten Ansätze dieser Tradition. Der Atomismus ist objektivistisch, reduktionistisch und empiristisch. Er neigt dazu, eher auf die semantische als die pragmatische Dimension von Wissen zu sehen, eher auf *knowing-that* als auf *knowing-how*, wie Gilbert Ryle es formuliert hat.¹²

⁶ Christoph Engemann, Andreas Sudmann (Hg.): *Machine Learning. Medien, Infrastrukturen und Technologien der Künstlichen Intelligenz*, Bielefeld 2018 sowie die Schwerpunkttheft Zeitschrift für Medienwissenschaft, Bd. 11, Nr. 2, 2019 und *spheres*, Bd. 5, Nr. 5, 2019.
⁷ Daniel C. Dennett: *Consciousness Explained*, New York 1991, 440.

⁸ Hans Blumenberg: *Ausblick auf eine Theorie der Unbegrifflichkeit*, in: ders.: *Ästhetische und metaphorologische Schriften*, hg. v. Anselm Haverkamp, Frankfurt/M. 2001, 193–209, hier 195.

⁹ Georg Simmel: *Die ästhetische Bedeutung des Gesichts*, in: *Das Individuum und die Freiheit. Essays*, Berlin 1984, 140–145; Emmanuel Lévinas: *Totalität und Unendlichkeit. Versuch über die Exteriorität*, München 2002, 277–317; Hans Belting: *Faces. Eine Geschichte des Gesichts*, München 2013. Für einen kulturwissenschaftlichen Zugriff siehe auch Thomas Macho: *Vorbilder*, München 2011 sowie Sigrid Weigel (Hg.): *Gesichter. Kulturgeschichtliche Szenen aus der Arbeit am Bildnis des Menschen*, Paderborn 2013.

¹⁰ Blumenberg: *Ausblick*, 195.

¹¹ Im Folgenden beschränke ich mich auf eine (überblickshafte) Diskussion dieser Begriffe in der Philosophie, die aber auch in vielen anderen Feldern (Physik, Biologie) geführt worden ist, siehe Michael Esfeld: *Holismus und Atomismus in den Geistes- und Naturwissenschaften. Eine Skizze*, in: Alexander Bergs, Soelwe I. Curdts (Hg.): *Holismus und Individualismus in den Wissenschaften*, Frankfurt/M. 2003, 7–21.

¹² Gilbert Ryle: *Knowing How and Knowing That*, in: *Proceedings of the Aristotelian Society*, Bd. 46, Nr. 1, 1946, 1–16.

¹³ Immanuel Kant: *Prolegomena zu einer jeden künftigen Metaphysik*, AA IV, § 39, 322.

<Holismus> dagegen benennt die entgegengesetzte Überzeugung, ein Ding sei *nicht* erschöpfend durch die Eigenschaften seiner konstitutiven Elemente zu erklären. In dieser Traditionslinie kommt das Ganze begrifflich oder kausal vor seinen Teilen. Verwandt sind Terme wie <Struktur> oder, bei Kant, <System> im Gegensatz zum atomistischen <Aggregat>.¹³ Vertreten wird der Holismus im 20. Jahrhundert vor allem von der hermeneutischen Phänomenologie Heideggers und Merleau-Pontys sowie vom späten Wittgenstein und seinen Anhänger_innen. Für einen solchen Holismus geht es weniger um explizites propositionales als um implizites pragmatisches und weltkonstitutives Wissen.¹⁴

Der Begriff der Gestalt ist in diesem Sinne holistisch.¹⁵ Christian von Ehrenfels, der 1890 den Terminus der <Gestaltqualitäten> prägte, stellte fest, dass die Wahrnehmung dessen, was eine Melodie ausmacht, eine Ganzheit ist, die sich nicht auf die Abfolge einzelner Töne reduzieren lässt. Er wandte sich damit von einer bloßen Assoziationspsychologie ab, die rein atomistisch-kausal argumentierte.¹⁶ Rasch avancierte der Gesichtssinn zum zentralen Untersuchungsfeld der Gestaltpsychologie des 20. Jahrhunderts, wie sie vor allem in der Berliner Schule um Wolfgang Köhler, Kurt Koffka und Max Wertheimer sowie von deren Schüler_innen der zweiten Generation, z. B. Wolfgang Metzger, immer wieder erläutert wurde.¹⁷ Köhler bemerkte, ähnlich wie kurz darauf Heidegger,¹⁸ dass wir «kein indifferentes Mosaik [wahrnehmen]; es ist vielmehr charakteristisch für unser Sehen, Hören usw., daß es fortwährend Einheiten und Gruppen zeigt, die, in sich fest, aus ihrer Umgebung relativ ausgesondert erscheinen».¹⁹ Solche Einheiten, die «Gestalten», weisen eine innere Kohärenz und Stabilität auf, die Wertheimer «Prägnanz» nannte.²⁰ Sie basieren nicht auf «unabhängige[n] Elementarempfindungen» – einem Aggregat atomistischer Sinnesdaten –, sondern setzen sich aus «lokalen Zustände[n]» zusammen, die «von ihrer Zugehörigkeit, Lage und Rolle in den Gestalten abhängig» sind.²¹ «Einsicht», die plötzliche Wahrnehmung von Gestaltkonfigurationen, wurde in Köhlers Affenexperimenten zudem zu einem Maß für Intelligenz.²² Solche Gestalten, als visuelle Figuren wie als Konstellation der Einsicht, sind nicht-ableitbare Bedeutungseinheiten, die *holistisch* verstanden werden müssen.

Eine dieser nicht-ableitbaren Einheiten ist das Gesicht. So stellt Metzger fest, dass man, um ein Gesicht in seinem Ausdruck wahrnehmen zu können, es als Ganzes betrachten muss. Wo es noch hilfreich sein mag, die Aufmerksamkeit auf die Bewegung der Brauen oder des Mundes zu lenken, ist es dem expressiven Verständnis jedoch abträglich, sozusagen noch weiter in das Gestaltganze <hineinzuzoomen> und seine Teile zu isolieren. Keines dieser Teile

sagt etwas über das aus, worauf es bei einem Gesicht im Leben eigentlich ankommt; ob es zum Beispiel ein überhebliches, herrisches, hartes, verschlossenes, feindseliges oder ein weiches, warmes, aufgeschlossenes und teilnehmendes Gesicht ist. Diese entscheidenden Züge werden am deutlichsten oder überhaupt nur sichtbar, wenn man es aus genügendem Abstand *als ganzes* betrachtet.²³

¹⁴ Charles Taylor formuliert eine Atomismus/Holismus-Trennung für die Theorie der Bedeutung. Er stellt der *enframing theory* als Bedeutungslehre der atomistischen Tradition die *expressive-constitutive theory* als jene der holistischen Richtung gegenüber. Für Letztere heißt etwas zu artikulieren, überhaupt erst die Wahrnehmung dieses Merkmals zu ermöglichen. Das eine ist konstitutiv für das andere, aber nur, weil der Kontext dieser Operation nicht neutral ist, sondern bereits von einem komplexen Hintergrundwissen geprägt ist, das nicht propositionaler Natur ist, sondern in einer expressiven Praxis, in Lebensformen wurzelt. Charles Taylor: *Theories of Meaning*, in: ders.: *Human Agency and Language. Philosophical Papers I*, Cambridge 1985, 247–292.

¹⁵ Die Gestaltpsychologie ist Teil der holistischen Linie, aber nicht alle Holistiker_innen sind Anhänger_innen der Gestalttheorie, siehe etwa, trotz unleugbarer Einflüsse, Merleau-Pontys Kritik an Husserls Rezeption der Gestaltpsychologie und an ihr selbst, Maurice Merleau-Ponty: *Phänomenologie der Wahrnehmung*, Berlin 1966, 74.

¹⁶ Christian von Ehrenfels: Über Gestaltqualitäten, in: *Vierteljahrsschrift für wissenschaftliche Philosophie*, Bd. 13, 1890, 249–292.

¹⁷ Vgl. zur Geschichte der Gestaltpsychologie (auch zur ihrer Emigrations- wie Kollaborationsgeschichte im Nationalsozialismus) Mitchell G. Ash: *Gestalt Psychology in German Culture, 1890–1967. Holism and the Quest for Objectivity*, Cambridge 1995.

¹⁸ Martin Heidegger: Der Ursprung des Kunstwerks, in: ders.: *Holzwege*, Frankfurt / M. 2003, 1–74.

¹⁹ Wolfgang Köhler: Bemerkungen zur Gestalttheorie, in: *Psychologische Forschung*, Bd. 11, Nr. 1, 1928, 188–234, hier 188 f.

²⁰ Max Wertheimer: Untersuchungen zur Lehre von der Gestalt II, in: *Psychologische Forschung*, Bd. 4, Nr. 1, 1923, 301–350.

²¹ Köhler: Bemerkungen zur Gestalttheorie, 189.

²² Wolfgang Köhler: *Intelligenzprüfungen an Menschenaffen*, Berlin 1921.

²³ Wolfgang Metzger: Was ist Gestalttheorie?, in: Kurt Guss (Hg.): *Gestalttheorie und Erziehung*, Darmstadt 1975, 1–17, hier 2 f.

24 Vortrags- oder Protokollbände liegen für Erstere vor: Claus Pias (Hg.): *Cybernetics. The Macy-Conferences 1946–1953*, Zürich 2003; Lloyd A. Jeffress (Hg.): *Cerebral Mechanisms in Behavior. The Hixon Symposium*, New York 1951. Dem Dartmouth-Workshop folgte keine eigene Veröffentlichung, siehe aber Ronald R. Kline: *Cybernetics, Automata Studies, and the Dartmouth Conference on Artificial Intelligence*, in: *IEEE Annals of the History of Computing*, Bd. 33, Nr. 4, 2011, 5–16. Für einen historischen Überblick siehe Nils J. Nilsson: *The Quest for Artificial Intelligence. A History of Ideas and Achievements*, Cambridge 2010; Steve Joshua Heims: *The Cybernetics Group 1946–1953. Constructing a Social Science for Postwar America*, Cambridge 1991.

25 Für eine kundige Einführung siehe Melanie Mitchell: *Artificial Intelligence. A Guide for Thinking Humans*, New York 2019, 17–34.

26 Das einflussreichste System dieser Art war der General Problem Solver (GPS), siehe Allen Newell, J. C. Shaw, H. A. Simon: *Report on a General Problem-Solving Program*, in: *Proceedings of the International Conference on Information Processing*, Paris 1959, 256–264.

27 Siehe Pamela McCorduck: *Machines Who Think. A Personal Inquiry into the History and Prospects of Artificial Intelligence*, Natick 2004, 417–521.

28 Frank Rosenblatt: *The Perceptron. A Probabilistic Model for Information Storage and Organization in the Brain*, in: *Psychological Review*, Bd. 65, Nr. 6, 1958, 386–408; Nilsson: *The Quest for Artificial Intelligence*, 64–74; siehe auch Matteo Pasquinelli: *Machines that Morph Logic. Neural Networks and the Distorted Automation of Intelligence as Statistical Inference*, in: *Glass Bead*, Bd. 1, Nr. 1, 2017, www.glass-bead.org/article/machines-that-morph-logic (6.5.2020). Zu den benachbarten Arbeiten um Heinz von Foerster siehe Jan Müggenburg: *Lebhafte Artefakte. Heinz von Foerster und die Maschinen des Biological Computer Laboratory*, Konstanz 2018.

29 Siehe David Bates: *Creating Insight. Gestalt Theory and the Early Computer*, in: Jessica Riskin (Hg.): *Genesis Redux. Essays in the History and Philosophy of Artificial Life*, Chicago 2007, 237–260 sowie Heims: *The Cybernetics Group 1946–1953*, 201–247.

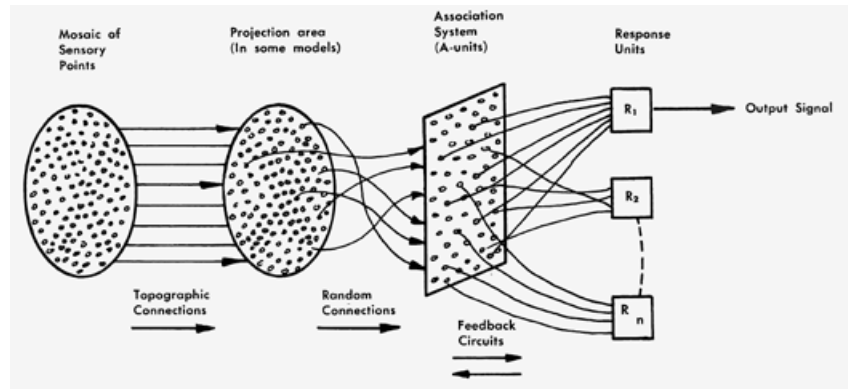
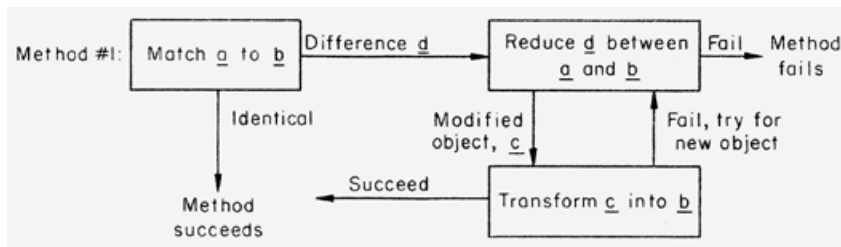
Wenn Pix2Pix und StyleGAN2 in der Lage sind, Gesichter als Ganzes abzuleiten, als Gestalten aus der Distanz gesehen – heißt das, dass ihnen eine holistische Logik zugrunde liegt? Wie aber sollte ein digitales System, das auf der symbolischen Operation diskreter Zeichen basiert und sich nicht zuletzt bei der Bilddarstellung einer diskreten Pixelmatrix bedient (die ja nichts anderes ist als Köhlers «Mosaik»), nicht-ableitbare Gestaltganzheiten zu schaffen in der Lage sein? Um diese Frage zu beantworten, lohnt ein Blick auf die Geschichte von KI-Systemen, von denen KNNs nur ein Paradigma sind, und auf die ihnen zugrunde liegenden Konzeptualisierungen.

II. Gestalt vs. KI

Die Geschichtsschreibung der KI-Forschung hebt ihren Ursprung in den USA der 1940er und 50er Jahre in einem Klima neobehavioristischer, reduktionistischer und empirischer Psychologie hervor, die mit den Methoden und Anliegen der Kybernetik eine produktive Konfluenz einging. Das Hixon-Symposium 1948, die Macy-Konferenzen zwischen 1946 und 1952 und der Dartmouth-Workshop über Künstliche Intelligenz 1956 sind wichtige Stationen dieser Geschichte. Vor allem Letzterer konstituierte die KI-Forschung als eigenständiges Gebiet und bestimmte die Parameter, unter denen sie in der Folgezeit betrieben wurde.²⁴

Die bis heute fortwirkende Unterscheidung zweier Typen von KI – symbolisch und subsymbolisch – hat ihren Ursprung in diesem Workshop.²⁵ Symbolische Ansätze, in Dartmouth favorisiert, nähern sich Künstlicher Intelligenz unter Annahme des atomistischen Paradigmas. Intelligenz wird hier in erster Linie als die Manipulation von Symbolen verstanden, die atomistische Fakten repräsentieren. Es wurde in sogenannten Expertensystemen implementiert, die eine Wissensdatenbank solcher Fakten mit einer Schlussregeln enthaltenen Inferenzmaschine verschränkten, was erlaubte, aus der Kombination der Fakten Folgerungen abzuleiten.²⁶ Expertensysteme erschienen anfangs äußerst vielversprechend, doch kam ihre Entwicklung spätestens in den 70er Jahren im ersten «KI-Winter» zum Stillstand.²⁷

Die heutigen KNNs wie Pix2Pix und StyleGAN2 gehören nicht zur symbolischen, sondern zur subsymbolischen Familie der KI, die lose auf dem Modell des Gehirns als einem Netzwerk von Neuronen und Synapsen basiert. Aufbauend auf den Vorarbeiten von Warren McCulloch und Walter Pitts entwarf Frank Rosenblatt 1958 das Perzeptron, das erste künstliche neuronale Netz, das in der Lage war, einfache visuelle Muster zu erkennen.²⁸ Meint «lernen» bei Expertensystemen die Erweiterung der Wissensdatenbank, sind Perzeptrone auf Wiederholungen innerhalb der zu lernenden Domäne angewiesen; folgt das Expertensystem linearen Wenn-dann-Strukturen (Abb. 3), hat die Schaltung des Perzeptrons einen parallelen Aufbau und kommt ohne die Trennung von Fakten und Regeln aus (Abb. 4). Bereits ihrer Struktur nach steht das



Paradigma der symbolischen KI einer atomistischen, das der subsymbolischen KI einer holistischen bzw. einer Gestaltlogik nahe.

Dieser Differenz war man sich von Beginn an bewusst. Wie David Bates und Steve Joshua Heims gezeigt haben, stand die US-amerikanische KI-Forschung den Gestaltideen der europäischen Emigrant_innen nach anfänglichem Interesse bald feindselig gegenüber.²⁹ So sprach sich Wolfgang Köhler, der noch 1948 neben dem KI-Pionier Warren McCulloch auf dem Hixon-Symposium vorgetragen hatte,³⁰ 1951 in einer Rezension von Norbert Wiens Buch *Cybernetics* gegen die Vorstellung aus, elektronische Rechenmaschinen seien hilfreiche Analogien für menschliche Intelligenz, weil Ersteren als diskret operierenden Systemen die gestalthafte <Einsicht> der Letzteren abgehe.³¹ Die atomistische Psychologie der KI-Forscher_innen schien ihm als Wiederkunft der empiristischen Psychologie des 19. Jahrhunderts, die zu ersetzen die Gestaltpsychologie ursprünglich angetreten war.³² Dennoch verloren die nicht-symbolischen Modelle, zumal Rosenblatts Perzeptron, gegenüber dem symbolischen Ansatz bald an Akzeptanz. Marvin L. Minsky und Seymour Papert, zwei der wichtigsten Symbolisten, veröffentlichten 1969 eine (sachlich inkorrekte) Kritik des Perzeptrons, die es bis in die 80er Jahre hinein ins Abseits drängte.³³ Nicht zuletzt formulierten sie ihre Einwände auch als Verteidigung einer atomistischen gegen eine <unwissenschaftliche> holistische Erkenntnistheorie, gewissermaßen in Umkehrung der Köhler'schen Vorbehalte, in der Absicht, «to dispel what we feared to be the

Abb. 3 Schema einer Methode zur «means-ends analysis» in: Allen Newell, John Shaw, Herbert A. Simon: «Report on a General Problem-Solving Program», 1959

Abb. 4 Frank Rosenblatt: «The Design of an Intelligent Automaton», 1958

³⁰ Wolfgang Köhler: Relational Determination in Perception, in: Jeffress (Hg.): *Cerebral Mechanisms in Behavior*, 200–243.

³¹ Wolfgang Köhler, Rezension von *Cybernetics, or Control and Communication in the Animal and the Machine* von Norbert Wiener, in: *Social Research*, Bd. 18, Nr. 1, 1951, 125–130.

³² Siehe Bates: *Creating Insight*, 239–249.

³³ Marvin L. Minsky, Seymour Papert: *Perceptrons. An Introduction to Computational Geometry*, Cambridge 1969.

first shadows of a <holistic> or <Gestalt> misconception that would threaten to haunt the fields of engineering and artificial intelligence as it had earlier haunted biology and psychology».³⁴

Einer der ersten Philosophen, der eine holistisch fundierte Kritik des symbolischen Ansatzes formulierte, war Hubert L. Dreyfus. In einer Reihe von Essays und in seinem Buch *What Computers Can't Do* (1972) argumentierte er, der symbolische Ansatz sei zur Begründung menschenähnlicher Intelligenz prinzipiell unfähig.³⁵ Dreyfus mobilisierte eine Reihe holistischer Argumente. Zentral war, dass Menschen nicht nur über eine verkörperte Intelligenz verfügten, sondern auch auf ein handlungskonstitutives implizites Hintergrundwissen zurückgriffen; ihr Wahrnehmen und Erkennen beruhe darauf, dass sie immer schon in eine Situation eingebettet seien. Hier appropriierte Dreyfus (nicht von ungefähr einer der wichtigsten US-amerikanischen Heidegger-Interpreten) das In-der-Welt-Sein als «being-in-a-situation».³⁶ Menschlicher intelligenter Umgang mit der Welt basiere häufiger, wie Charles Taylor (ein wichtiger Weggefährte Dreyfus') Heidegger auslegt, auf *task-rightness*, situativer Angemessenheit, als auf *semantic rightness*, logisch formalisierbarer Kongruenz von Daten und abstraktem Weltmodell.³⁷ Damit einem Computer Intelligenz zugesprochen werden könnte, bräuchte er dieses implizite Hintergrundwissen, das nur durch die tatsächliche Begegnung mit der Welt im «being-in-a-situation» gewonnen werden könne. Für Dreyfus folgt der Schluss: «[B]eing-in-a-situation turns out to be unprogrammable in principle using presently conceivable techniques.»³⁸

Dieser Einwand war überzeugend, solange die atomistische Annahme die Grundlage der «presently conceivable techniques» der KI-Forschung bildete. Weniger sicher war sich Dreyfus jedoch in seiner Kritik, als es um die Architektur des Perzeptrons ging. Als in den 1980er Jahren KNNs, die im Grunde nur mehrschichtige Perzeptrone sind, erneut an Popularität gewannen,³⁹ räumte auch Dreyfus ein, dass sie seiner holistischen Vorstellung von Wahrnehmung näherkamen. Doch blieb er nach wie vor skeptisch, ob sie tatsächlich als Baustein künstlicher maschineller *Intelligenz* taugten.⁴⁰

Zieht man die Fähigkeiten heutiger KNNs heran, hätte aber auch Dreyfus zugeben müssen, dass ein prinzipieller, nicht nur ein gradueller Fortschritt gemacht ist. In *What Computers Can't Do* hatte er eine Reihe von Aufgaben formuliert, die ein System meistern müsste, um als intelligent gelten zu können. Eine der wichtigsten unter ihnen war eine Version des Gestaltproblems, die er «perspicuous grouping» nannte. Unter dieser <deutlichen Gruppierung> verstand Dreyfus die Fähigkeit, Objektreihen anhand von zwischen ihnen herrschenden Ähnlichkeiten zu bilden, sozusagen ihre *kollektive Gestalt* zu erfassen. Neben Wittgensteins Idee der Familienähnlichkeit zog er zur Beschreibung solcher Gruppenschemata Köhlers Begriff der Einsicht (*insight*) heran. Familienähnlichkeiten wie kollektive Gestalt ließen sich nicht durch das Aufrechnen atomistischer Eigenschaften und den Vergleich von Merkmalslisten erfassen:

³⁴ Ebd., 19 f.

³⁵ Hubert L. Dreyfus: *What Computers Can't Do. A Critique of Artificial Reason*, New York 1972. Der Symbolismus, so Dreyfus, beruht auf einer Reihe atomistischer Annahmen: der biologischen Annahme, dass das Gehirn mit einem digitalen Computer identifiziert werden kann, der ontologischen, dass die Welt aus isolierbaren Fakten besteht, und der erkenntnistheoretischen, dass der Verstand solche Fakten verarbeitet. Für einen solchen Atomismus kann Denken durch explizite Regeln formalisiert werden – *knowing-how* lässt sich als *knowing-that* ausdrücken, ebd., 67–142.

³⁶ Dreyfus: *What Computers Can't Do*, 200.

³⁷ Charles Taylor: Heidegger on Language, in: Hubert L. Dreyfus, Mark A. Wrathall (Hg.): *A Companion to Heidegger*, London 2007, 433–455.

³⁸ Dreyfus: *What Computers Can't Do*, 215.

³⁹ Nämlich infolge der Publikation des Bandes *Parallel Distributed Processing*, der Minskys und Paperts Falschdarstellungen korrigierte – vor allem die Fähigkeit von Perzeptronen, das ausschließende Oder (XOR) zu modellieren, David E. Rumelhart, James L. McClelland, PDP Research Group: *Parallel Distributed Processing. Explorations in the Microstructure of Cognition*, Cambridge 1986.

⁴⁰ Hubert L. Dreyfus, Stuart E. Dreyfus: *Making a Mind versus Modeling the Brain: Artificial Intelligence Back at a Branchpoint*, in: *Daedalus*, Bd. 117, Nr. 1, 1988, 15–43. «Neural network modeling may simply be getting a deserved chance to fail, as did the symbolic approach.» (Ebd., 37) Ein gegenwärtiger Ansatz, der genau dies behauptet, findet sich bei Brian Cantwell Smith: *The Promise of Artificial Intelligence. Reckoning and Judgment*, Cambridge 2019.

[P]atterns as complex as artistic styles and the human face reveal a loose sort of resemblance which seems to require a special combination of insight, fringe consciousness, and ambiguity tolerance beyond the reach of digital machines.⁴¹

Damit wären wir wieder beim Gesicht und abermals bei der Frage: Wenn die Gestalt eines Gesichts, seine Familienähnlichkeit mit anderen Gesichtern, nicht durch eine digitale Maschine konzeptualisiert oder als Eigenschaftsliste zusammengefasst werden kann, wie ist dann ein KNN dazu in der Lage, das auf einer digitalen Maschine ausgeführt wird?

III. Gestalt als <latent space>

An dieser Stelle ist es nötig, wieder einen Schritt zurückzutreten. Dreyfus' übergeordneter Fokus – menschenähnliche Künstliche Intelligenz – ist für die Beantwortung der gestellten Frage zunächst unerheblich. Wie Matteo Pasquinelli betont, zielen KNNs nicht so sehr auf die Simulation von Kognition, sondern von Wahrnehmung.⁴² Heute spielen vor allem <tiefe> KNNs ihre Fähigkeiten bei Mustererkennungsaufgaben aus. Und Dreyfus' <perspicuous grouping> ist zunächst nur eine besonders anspruchsvolle Art von *pattern recognition*, die Familienähnlichkeiten ausmacht, ohne über explizites Wissen über sie zu verfügen. Auch Rosenblatts Perzeptron war entsprechend seiner Aufgabe, Muster zu erkennen, nach dem Sehnerv des Auges, nicht dem Großhirn modelliert und aus drei Hauptelementen aufgebaut:⁴³ der Eingabeschicht, einer verborgenen Schicht (*hidden layer*) und der Ausgabeschicht. Moderne KNNs, sogenannte *deep neural networks*, folgen weiterhin diesem Aufbau, verfügen aber über eine Vielzahl von verborgenen Schichten, die aus künstlichen <Neuronen> bestehen, die als logische Gatter fungieren und über <Synapsen> miteinander verbunden sind. Diese Synapsen wiederum haben durch eine im Trainingsprozess kalibrierte Gewichtung Einfluss auf die Aktivierungsstärke des nächsten Neurons.⁴⁴ Das Ziel eines KNN ist es, eine Funktion zu entwerfen, die die Input-Daten einem gewünschten Output anpasst, und diese Funktion auf zukünftige Inputs anzuwenden, um deren Output vorherzusagen. Was StyleGAN2 betrifft, so ist das KNN hier darauf trainiert, Variationen seiner Eingaben auszugeben – seine Eingaben sind eine große Menge von Gesichtern und die Ausgaben neue Gesichter. Tatsächlich verwenden StyleGAN2 und, bis zu einem gewissen Grad, auch Pix2Pix eine Unterform von KNNs, ein *generative adversarial network* (GAN), in dem diese Prozesse getrennt sind, die Grundstruktur aber dieselbe ist.

Die <Daten>, aufgrund derer diese Funktion generiert wird, müssen allerdings erst zur Verarbeitung vorbereitet und aus allen möglichen muss eine begrenzte Menge von Merkmalen (*features*) ausgewählt werden. Bei Techniken des *supervised learning* geschieht das manuell, wenn etwa in einem Datensatz aus Porträtfotos Bilder unterschiedlich markiert werden, je nachdem, ob eine Person lächelt oder nicht. Beim *unsupervised learning* hingegen erfolgt

⁴¹ Dreyfus: *What Computers Can't Do*, 32. Es ist erstaunlich, wie selten der Nähe von Gestaltpsychologie zu neuronalen Netzen nachgegangen wird. Eine Ausnahme ist Uwe Seifert, Randolph Eichert, Lüder Schmidt: *Logic, Gestalt Theory, and Neural Computation in Research on Auditory Perceptual Organization*, in: Marc Leman (Hg.): *Music, Gestalt, and Computing. Studies in Cognitive and Systematic Musicology*, Berlin u. a. 1997, 70–88.

⁴² Pasquinelli: *Machines that Morph Logic*. KNNs sind also, mit John Searles Unterscheidung gesprochen, bislang noch eine schwache, keine starke KI, siehe John R. Searle: *Minds, Brains, and Programs*, in: *Behavioral and Brain Sciences*, Bd. 3, Nr. 3, 1980, 417–457.

⁴³ Zur Differenz von Rosenblatts stochastisch operierendem Perzeptron und dem nur Bool'scher Logik folgenden Neuronensimulationen von McCulloch und Pitts siehe Müggenburg: *Lebhaftes Artefakte*, 115.

⁴⁴ Siehe hierzu und im Folgenden Pedro Domingos: *The Master Algorithm. How the Quest for the Ultimate Learning Machine Will Remake Our World*, New York 2015, 93–120; Ethem Alpaydin: *Machine Learning. The New AI*, Cambridge 2016, 85–110.

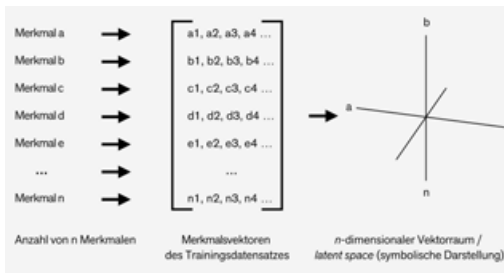


Abb. 5 Merkmalsextraktion und Latenzraumkartierung, Grafik des Autors

Abb. 6 Beispiel für die Interpolation des *latent space*. Oben links das Eingabebild, oben rechts das erste, unten links das zweite Zielbild; die restlichen Bilder sind die Interpolationen des Modells. Visualisierung als J-Diagramm in einer Studie von Tom White, 2016

das Auffinden solcher Merkmale automatisch. Jede der Schichten des KNN hat dann die Aufgabe, aus dem Input der jeweils vorangegangenen Schicht markante Muster zu extrahieren, gewissermaßen <Prägnanzen> zu finden. Da dies progressiv zwischen den Schichten geschieht, ist hier ein Abstraktionsprozess am Werk: Die erste Schicht mag eine Kombination von einigen wenigen Pixeln betrachten und das Ergebnis dann an die nächste Schicht weitergeben, die nun eine Kombination von einigen Kombinationen von Pixeln betrachtet, und so weiter; während die erste Schicht nur Kanten (hohe Kontraste zwischen den Pixeln) erkennen kann, gruppiert die zweite Schicht bereits Kanten in einfache Formen (gerade Linien oder Kurven), die dritte in Teile von Objekten, die vierte in Objekte und so weiter.⁴⁵ Die n aus der Eingabe abgeleiteten Merkmale werden auf einen n -dimensionalen Vektorraum abgebildet (Abb. 5). In diesem Vektorraum ist es möglich, die Interdependenz aller Merkmale zu berechnen und die Anzahl der Merkmale auf einen niedrigerdimensionalen Raum, den *latent space*, zu reduzieren. Dieser Prozess der Abstraktion wird als *dimensionality reduction* bezeichnet, «because it reduces a large number of visible [or explicit] dimensions (the pixels) to a few implicit ones (expression, facial features)».⁴⁶

Im Falle von zweidimensionalen Gesichtern könnte man sich den *latent space* wie in der Darstellung von Abb. 6 vorstellen. Das Modell kann nun die expliziten Dimensionen der Eingabe aus den impliziten, die im Modell verfügbar sind, wiederherstellen; genau das macht Pix2Pix, wenn es die Gesichtszüge des unscharfen Eingabebildes interpoliert und mit den wahrscheinlichen Zuständen seines Modells auffüllt. Es kann aber auch, einfach durch eine zufällige Auswahl von Initialwerten, neue Ausgaben aus den Dimensionen des *latent space* erzeugen. Auf diese Weise gibt StyleGAN² jedes Mal, wenn die Seite thispersondoesnotexist.com aktualisiert wird, ein neues Gesicht aus. Was beide Operationen gemeinsam haben, ist, dass ihre Grundlage die *Gesamtstruktur* dessen ist, was modelliert wird – ein Gesicht.⁴⁷

⁴⁵ Siehe Yann LeCun, Yoshua Bengio, Geoffrey Hinton: Deep learning, in: *Nature*, Bd. 521, Mai 2015, 436–444; Alpaydin: *Machine Learning*, 75 und 99f.

⁴⁶ Domingos: *The Master Algorithm*, 211. Siehe auch Vladan Joler, Matteo Pasquinelli: *The Nooscape Manifested. AI as Instrument of Knowledge Extractivism*, 2020, nooscape.ai (2.5.2020).

IV. «Quasi-analog» und «quasi-holistisch»

In Bezug auf Dreyfus' Kriterium des «perspicuous grouping» scheinen dies in der Tat beeindruckende Ergebnisse zu sein. Zwar stellt kein einziges der extrahierten Merkmale ein Gesicht als solches dar, doch kann man argumentieren, dass «Gesichthaftigkeit» – die Abstraktion, die die Gestalt eines Gesichts darstellt – in der Gesamtheit des *latent space* selbst angesiedelt ist.

Einige Punkte unterstützen diese Beobachtung: Erstens enthält, wie angedeutet, ein KNN kein explizites Wissen. Anders als bei der Trennung zwischen Wissensdatenbank und Inferenzmaschine ist das <Wissen> in einem neuronalen Netz nicht an einer bestimmten Stelle lokalisiert, sondern als statistisches Modell im gesamten System verteilt.⁴⁸ KNNs generalisieren also, ohne Begriffe zu bilden.⁴⁹ Zweitens folgt ein neuronales Netz auch nicht dem Paradigma logischer Deduktion oder explizit festgelegten Regeln, die sequenziell ausgeführt werden; vielmehr operiert es durch statistische Induktion, und es ist das System als Ganzes, das die Berechnungen vornimmt.⁵⁰ Drittens könnte man daher argumentieren, dass neuronale Netze nicht nur Outputs produzieren, die Menschen als Gestalten wahrnehmen, sondern dass sie intern bereits, als statistische Modelle, mit einer Gestaltlogik operieren – und zwar auf eine Art und Weise, die weder eine bloße Collage noch nur ein Mittelwert ist.

All das legt nahe, dass KNNs nicht im schlichten Sinne als atomistisch bezeichnet werden können. Allerdings deutet auch eine Reihe von Punkten darauf hin, dass es voreilig wäre, ein KNN tatsächlich holistisch zu nennen: Erstens ist das Netzwerk, zumindest auf der operationellen Ebene, immer noch hierarchisch strukturiert: Auch wenn der entstehende *latent space* mehr oder weniger abstrakte Merkmale enthält, geht er doch im Prozess der Abstraktion von Teilen zu Ganzem über und nicht umgekehrt. Zweitens bedeutet die Tatsache, dass neuronale Netze keine symbolische Repräsentation verwenden, nicht, dass sie überhaupt keine Repräsentation verwenden; auch das statistische Modell repräsentiert. Drittens kann der *latent space* zwar jede mögliche Interpolation eines Gesichts zeigen. Doch was möglich ist, hängt sowohl von der Zahl der Merkmale ab, die im Trainingsset vorhanden sind, als auch von der Auswahl und Gewichtung der extrahierten Merkmale; *dimensionality reduction* ist schließlich reduktiv. Über die im *latent space* gespeicherte «Gesichthaftigkeit» hinaus gäbe es also immer noch Gesichter, die Menschen als solche erkennen würden, die aber durch das Modell nicht interpoliert werden könnten. Und schließlich kann man darauf hinweisen, dass ein zentrales Merkmal des Gestaltkonzepts darin besteht, ein Verstehen und nicht nur ein Erkennen der Sache zu implizieren. Wenn Metzger schreibt, dass eine Gestalt etwas darüber aussagt, «worauf es bei einem Gesicht im Leben eigentlich ankommt», ob es arrogant oder freundlich ist, dann ist freilich *dieses* Wissen im *latent space* des Modells nicht repräsentiert. An dieser Stelle beginnt die Unterscheidung zwischen Wahrnehmung und Kognition wieder zu verschwimmen.⁵¹ Das bringt uns entweder zu Dreyfus' Frage

⁴⁷ Man kann sich den *latent space* vereinfacht als Reihe von Schiebereglern zur Änderung jedes beliebigen Merkmals des Gesichts vorstellen, die – vom Gesamtfarbtönen über die Richtung der Lichtquelle bis zum Gesichtsausdruck – alle Abstraktionsgrade der Merkmalsextraktion umfassen. Zudem ist es möglich, durch diesen hochdimensionalen Latenzraum zu «reisen» (*manifold traversal*) und dabei jede mögliche Konfiguration zu interpolieren, siehe: Robert Luxemburg: StyleGAN2 Interpolation Loop, 12.12.2019, youtube.com/watch?v=6E1_dgYlIfc (6.5.2020). Für das Beispiel von Tom White (Abb. 6) siehe ders.: Sampling Generative Networks, in: ArXiv, 6.12.2016, arxiv.org/abs/1609.04468 (4.7.2020).

⁴⁸ Pasquinelli: *Machines that Morph Logic*; Andreas Sudmann: Szenarien des Postdigitalen: Deep Learning als MedienRevolution, in: Engemann, Sudmann (Hg.): *Machine Learning*, 55–73, hier 66–68.

⁴⁹ Leistet jede Schicht eine Abstraktion der Merkmale der vorherigen Schicht, ist das noch keine Begriffsarbeit und resultiert in keiner Theorie, siehe Dreyfus u. a.: *Making a Mind versus Modeling the Brain*, 36.

⁵⁰ Pasquinelli weist darauf hin, dass bereits Rosenblatt selbst in *Principles of Neurodynamics* dieses Ganze als eine emergente Qualität im Sinne von Gestalt betrachtete. Siehe Pasquinelli: *Machines that Morph Logic*.

⁵¹ Darauf weist auch Smith hin, Smith: *Promise*, 7, 24–27 und 56f.

zurück, ob eine starke KI überhaupt möglich ist, oder lenkt unsere Aufmerksamkeit darauf, dass selbst bei einer schwachen KI, und selbst wenn sie *unsupervised* lernt, Kognition im Spiel ist, die in den vom Menschen gesetzten Parametern codiert ist – dass also kein Training wirklich unüberwacht ist.⁵²

KNNs können also *weder* als vollständig atomistisch *noch* als vollständig holistisch konzeptualisiert werden. Sie scheinen etwas dazwischen zu sein. John von Neumann nannte dieses Dazwischen ein gemischtes System, ein «mixed system». Von Neumann prägte den Ausdruck, als er 1958 in seinem posthum erschienenen Buch *The Computer and the Brain* Unterschiede und Ähnlichkeiten zwischen Zentralnervensystem und digitalem Automaten diskutierte. Das Gehirn sei strukturell digital, Neuronen würden über diskrete Zustände aktiviert. Doch während der Computer, als Instanz von Turings Universeller Maschine, seriell und deterministisch arbeite, habe das Gehirn eine parallele Struktur und seine Operationen basierten auf statistischen Zuständen. Das Gehirn sei also ein «mixed system».⁵³ Während Neurowissenschaftler_innen heute vorsichtig sind, die Operationen des Gehirns als «digital» zu bezeichnen, beschreibt von Neumanns «mixed system» KNNs ausgesprochen gut.

Andreas Sudmann hat diese Schlussfolgerung kürzlich noch genauer herausgearbeitet. Er hebt hervor, dass neuronale Netze bislang noch auf der nach von Neumann benannten digitalen Architektur realisierte analoge Strukturen seien. Auch sie funktionierten eher parallel als seriell. Während ihre «Neuronen» als logische Gatter in der Tat diskret operieren, seien die in ihren «Synapsen» verteilten Gewichtungswerte keine binären Zustände, sondern als Gleitkommazahlen gespeichert. Aus diesem Grund schlägt Sudmann vor, neuronale Netze weder vollständig digital noch vollständig analog, sondern «postdigital» oder – klarer, wie ich finde – «quasi-analog» zu nennen.⁵⁴

Wenn «quasi-analog» die technische Struktur moderner neuronaler Netze bezeichnet, so wäre «quasi-holistisch» der Begriff zur Beschreibung ihrer Konzeptualisierung. Weder folgen die neuronalen Netze vollständig dem atomistischen Paradigma noch sind sie wirklich holistisch. Sie sind, konzeptuell gesprochen, ein quasi-holistisches gemischtes System und kombinieren Eigenschaften beider Paradigmen. KNNs zeigen deutlich, dass die Unterscheidung zwischen Atomismus und Holismus zu starr ist, um dieses Phänomen wirklich angemessen zu erfassen. Dazu wäre eine dritte Klasse nötig.

Ein Konzept, das für eine solche dritte Option in Frage kommt, ist das der «Assemblage».⁵⁵ In Anlehnung an Gilles Deleuze' und Félix Guattaris Idee des *agencement* (Gefüge) – der mit *assemblage* ins Englische übersetzt wurde – hat Manuel DeLanda diesen Begriff vorgeschlagen, um etwas *zwischen* dem atomistischen und dem holistischen Ansatz zu konzeptualisieren. Beschreibt «Atomismus» Ganzheiten als mechanische Aggregate von isolierten Elementen, so bezeichnet «Holismus» «relations of interiority», in denen jedes Element in einem organischen Konstitutionsverhältnis zum Ganzen steht. Im Gegensatz zu beiden interpretiert DeLanda Assemblagen als

⁵² Siehe Matteo Pasquinelli: How a Machine Learns and Fails – a Grammar of Error for Artificial Intelligence, in: *spheres*, Bd. 5, Nr. 5, 2019, spheres-journal.org/how-a-machine-learns-and-fails-a-grammar-of-error-for-artificial-intelligence (6.5.2020).

⁵³ John von Neumann: *The Computer and the Brain*, New Haven 1958, 58–60.

⁵⁴ Sudmann: Szenarien des Postdigitalen, 66. Der Begriff des Postdigitalen für quasi-analoge Strukturen scheint mir schlecht gewählt, wird darunter doch bereits so viel anderes verstanden, dass eine weitere Ausdehnung nicht wünschenswert ist, siehe Hannes Bajohr: *Experimental Writing in its Moment of Digital Technization: Post-Digital Literature and Print-on-Demand Publishing*, in: Annette Gilbert (Hg.): *Publishing as Artistic Practice*, Berlin 2016, 100–115.

⁵⁵ Siehe für die neuere Diskussion Marcus Twellmann: *Assemblage* (Collage, Montage). Für einen neuen Formalismus, in: *Deutsche Vierteljahrsschrift für Literaturwissenschaft und Geistesgeschichte*, Bd. 93, Nr. 2, 2019, 239–261.

durch «relations of exteriority» charakterisiert.⁵⁶ Eine Assemblage bildet immer noch ein Ganzes mit Eigenschaften, die in ihren Teilen nicht zwingend vorhanden sind – aber die Elemente behalten auch einen Grad an Autarkie, den die holistische Sichtweise nicht zulässt. Aus diesem Grund, so DeLanda, weisen Assemblagen nicht-lineare Kausalitäten auf, die eher statistischer als deterministischer Natur sind. Während sein Beispiel ein chemischer Prozess wie die Katalyse ist, wäre das Gewichtungsmo- dell eines neuronalen Netzes ein weiterer Fall von nicht-linearem und statistischem Verhalten.⁵⁷ Auch wenn man mit DeLanda nicht in allem übereinstimmen mag, muss man doch konzedieren, dass der Begriff Assemblage zumindest *ein* Kandidat ist, mit dem eine dritte Option zwischen Holismus und Atomismus formuliert werden könnte. Ein künstliches neuronales Netz ist damit, wie Deleuze und Guattari es in *Tausend Plateaus* ausdrücken, ganz wörtlich eine «Maschine zur Erschaffung des Gesichts».⁵⁸

V. Nachtrag. Umkehr der Blickrichtung – KNNs als «intuition pump»

In meiner Diskussion der Gestalteigenschaften künstlicher neuronaler Netze habe ich versucht, die konzeptuellen Annahmen eines technischen Systems zu extrahieren, die ich quasi-holistische Assemblagen nenne. Damit könnte die Diskussion der KNNs an dieser Stelle abgeschlossen sein. Dennoch möchte ich, in aller Kürze, eine weitere Möglichkeit nachtragen, wie die Verschränkung von Begriff und technischem System produktiv gemacht werden kann, indem man die Blickrichtung umkehrt: Die assemblageartige, quasi-analoge, quasi-holistische Sichtweise auf KNNs kann, um einen Begriff von Daniel Dennett zu übernehmen,⁵⁹ als «intuition pump» dienen, die dabei helfen soll, einige traditionellere Probleme nicht-atomistischer Begriffe neu zu denken.

Neben dem Gesicht, das diesem Essay als Leitmotiv diente, nannte Hubert L. Dreyfus den Begriff des *Stils* als Domäne gestalthafter Erkenntnis.⁶⁰ In der Tat wird Stil, wo er nicht rein formalistisch verstanden ist, oft als holistischer Begriff gedacht, der zwar hermeneutisch zu durchdringen sein mag, sich aber gegen die atomistische Aufrechnung von Merkmalen sperrt.⁶¹ Doch bereits StyleGAN2 bringt KNNs und zumindest visuellen Stil zusammen. StyleGAN2 enkodiert nicht nur ein quasi-holistisches statistisches Modell von Gesichtern, sondern ermöglicht auch, eine bestimmte Konfiguration des *latent space* – einen <Stil> – auf ein anderes Bild zu übertragen, indem es erst die spezifische Merkmalsverteilung des Inputs extrahiert und dann auf die Merkmalsvektoren des Referenzbildes abbildet. Diese Methode verwendet auch eine andere GAN-Implementierung namens CycleGAN, die Stile bestimmter Maler auf Fotos transponiert und umgekehrt (Abb. 7).⁶²

Man mag diese Ergebnisse nicht sonderlich beeindruckend finden, doch dieser Prozess legt nahe, dass <Stil> als gestalthafte Familienähnlichkeit nicht notwendig eine irreduzible, holistische Angelegenheit sein muss. Damit soll nicht gesagt

⁵⁶ Manuel DeLanda: *A New Philosophy of Society. Assemblage Theory and Social Complexity*, London 2006, 9f.

⁵⁷ Ebd., 13–15.

⁵⁸ Gilles Deleuze, Félix Guattari: *Tausend Plateaus. Kapitalismus und Schizophrenie II*, Berlin 1992, 246.

⁵⁹ Dennett: *Consciousness Explained*, 440.

⁶⁰ Dreyfus: *What Computers Can't Do*, 32.

⁶¹ Siehe für einen historischen Überblick Hans Ulrich Gumbrecht: *Schwindende Stabilität der Wirklichkeit. Eine Geschichte des Stilbegriffs*, in: ders., K. Ludwig Pfeiffer (Hg.): *Stil. Geschichten und Funktionen eines kulturwissenschaftlichen Diskurselements*, Frankfurt/M. 1986, 726–788; die semiotische Stiltheorie betont gerade die konstitutive Expressivität des Stils, die in Lebensformen wurzelt, siehe etwa Dick Hebdige: *Subculture. The Meaning of Style*, London 1979; die Schwierigkeit, Stil zu definieren und zu katalogisieren, ist am besten illustriert in Susan Sontag: *Notes On Camp*, in: *Against Interpretation and Other Essays*, New York 1978, 275–292.

⁶² Siehe Jun-Yan Zhu u. a.: *Unpaired Image-to-Image Translation using Cycle-Consistent Adversarial Networks*, in: *ArXiv*, 30.3.2017, arxiv.org/abs/1703.10593, Abb. 7 (6.5.2020). Entsprechende Versuche für textlichen Stiltransfer sind dagegen noch weniger mächtig, siehe Xiangyang Li u. a.: *Review of Text Style Transfer Based on Deep Learning*, in: *ArXiv*, 6.5.2020, arxiv.org/abs/2005.02914 (19.5.2020).

sein, Stil lasse sich ganz in atomistischer, empirischer Weise quantifizieren, wie es die Stilometrie in den 70er Jahren versuchte und wie es die Digital Humanities heute wieder annehmen, noch bedeutet es eine einfache Trennung von Form und Inhalt im Sinne eines bloßen *ornatus*. Aber Stil nicht als echt holistisch, sondern nur als quasi-holistisch zu rekonzeptualisieren – als eine Assemblage mit *einigen* beweglichen Teilen, die sich *bis zu einem gewisse Grad* bestimmen lassen –, könnte eine Neubewertung dieses so suspekt gewordenen Begriffs ermöglichen.⁶³

Ein weiterer jener scheinbar irreduziblen holistischen Begriffe ist <Stimmung>.⁶⁴ Dass er auch in Heideggers Existenzialanalytik eine Rolle spielt und damit fest in der holistischen Tradition verankert ist, macht diesen Fall besonders interessant. Der Anglist und Mathematiker Peli Grietzer hat sich in seinem Aufsatz «A Theory of Vibe» KNNs als Konzeptualisierungsmodellen genähert. Grietzer verwendet ein bestimmtes neuronales Netz, den Autoencoder, als *intuition pump* für den holistischen Begriff der Stimmung – oder, wie er es lieber ausdrückt, dem *vibe*.⁶⁵ Ein Autoencoder ist ein simples KNN, das darauf trainiert ist, seinen Input als Output zu reproduzieren. Da seine *hidden layers* aber <kleiner> als die äußeren Schichten sind, die Daten einen Flaschenhals maximaler Dimensionsreduktion durchlaufen, komprimiert der Autoencoder die Merkmale des Inputs extrem, bevor er sie wiederherstellt. (Abb. 8)

Stellt man sich nun vor, so Grietzers Gedankenexperiment, eine Reihe ästhetischer Objekte werde auf diese Weise derart komprimiert, dass sie fehlerfrei reproduziert werden kann, bestünde das Kompressionsmodell schlicht in einer Auflistung der möglichen Variationen seiner allgemeinen Merkmale. Dieses Merkmalskollektiv beschreibt Grietzer als *vibe*. Der *vibe* ist in allen ästhetischen Objekten dieser Reihe vorhanden, allerdings nie in Reinform. Grietzer sieht im *vibe* eine besonders modernistische, materialistische Qualität ästhetischer Werke, die auf Symbolik verzichteten und stattdessen eine abstrakte Stimmung durch die Wiederholung struktureller Ähnlichkeiten, die in allen Werken des Kanons vorhanden seien, repräsentierten. Umgekehrt sei das Erfassen eines *vibe* der realen Welt durch seine Idealisierung als *vibe* eines literarischen Werks – ein offensichtliches Beispiel ist das Kafkaeske – selbst eine Variante jener mathematischen Abbildung, die sich in einem Autoencoder abspielt; auch der *die Leser_in* codiert.⁶⁶

Wie das doppelte Gesicht der Gioconda, mit dem dieser Text begann, sind auch Stil und Stimmung Beispiele dafür, wie für gewöhnlich holistisch begriffene Phänomene auch als quasi-holistische Assemblagen zu denken sind. Auch wenn dies hier nur andeutungsweise geschehen konnte, skizzieren diese Formen doch eine ganze Gruppe von Bereichen, in denen die seltsame dritte Position von KNNs – jenseits der Dichotomie von Holismus und Atomismus – nicht nur eine Technologie bezeichnet, sondern ein eigenständiges konzeptuelles Werkzeug sein kann, das es fruchtbar zu machen gilt.

⁶³ So wäre es etwa denkbar, Ernst Gombrichs Antwort auf «the riddle of style», das «Schema», als quasi-holistische Assemblage zu begreifen. Auch Gombrich argumentiert nicht-atomistisch, da die Erkenntnis von Schemata immer noch Menschen, also bereits gestaltverarbeitenden Systemen vorbehalten ist, siehe Ernst H. Gombrich: *Art and Illusion*, London 1961, 3, 60. Ich danke Jana Mangold für den Hinweis.

⁶⁴ Siehe Hans Ulrich Gumbrecht: *Stimmungen lesen. Über eine verdeckte Wirklichkeit der Literatur*, München 2011.

⁶⁵ Peli Grietzer: *A Theory of Vibe*, in: *Glass Bead*, Bd. 1, Nr. 1, 2017, www.glass-bead.org/article/a-theory-of-vibe, Abb. 8 (6.5.2020).

⁶⁶ Genauer ausgeführt in Peli Grietzer: *Ambient Meaning. Mood, Vibe, System*, [Diss.] Cambridge 2017.

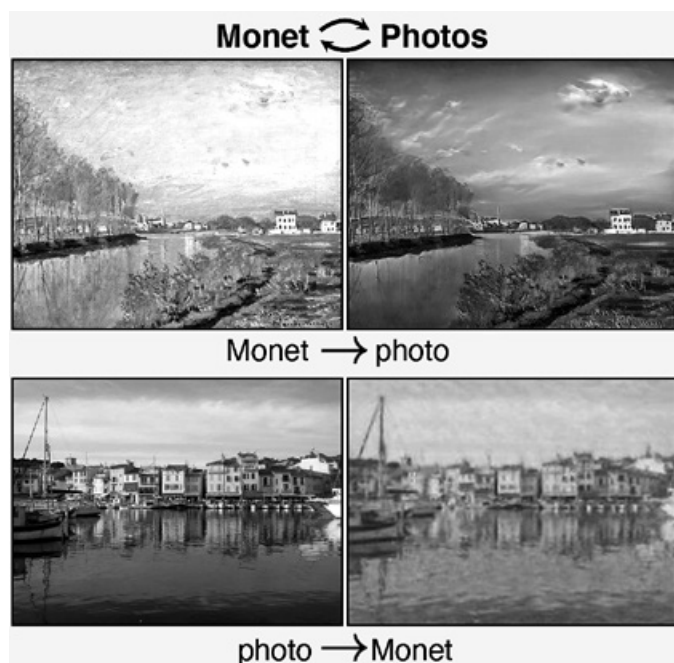


Abb. 7 Style transfer mit CycleGAN in einer Studie des Berkeley AI Research (BAIR) Laboratory, 2018

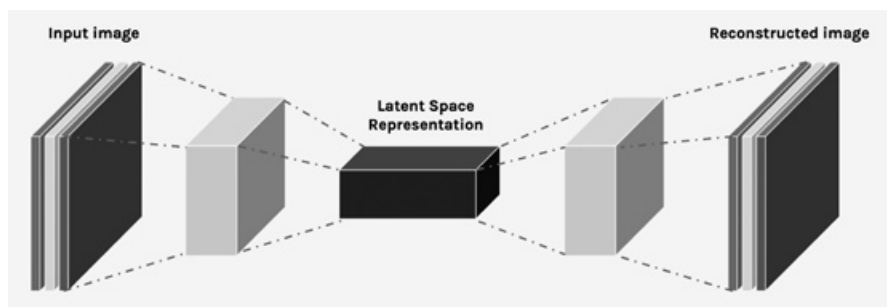


Abb. 8 Grafische Darstellung eines Autoencoders