

Wovon sprechen wir, wenn wir von Digitalisierung sprechen?

Gehalte und Revisionen zentraler Begriffe des Digitalen

Martin Huber, Sybille Krämer, Claus Pias
Symposienreihe „Digitalität in den Geisteswissenschaften“

Gefördert durch

DFG Deutsche
Forschungsgemeinschaft

IMPRESSUM

HERAUSGEBER

Martin Huber, Sybille Krämer, Claus Pias

KONTAKT

Julia Menzel

Digitalität in den Geisteswissenschaften

DFG-geförderte Symposienreihe

Universität Bayreuth

Universitätsstr. 30

95447 Bayreuth

www.digitalitaet.dfg@uni-bayreuth.de

1. Auflage April 2020

Wir danken der Deutschen Forschungsgemeinschaft (DFG Projekt Projektnummer 287972711) für die Förderung.

Computersehen. Elemente einer Medienarchäologie der Computer Vision und ihrer Sehstörungen

Birgit Schneider (Potsdam)

1. Die Schöne und der Algorithmus

Wenn ich Mitglied der globalen ‚Foto-Community‘ *EyeEm* werde, mache ich meine Fotografien zum Objekt einer umfassenden, ausschließlich algorithmisch betriebenen Sehschule. Mithilfe von *Computer Vision* und *Deep Learning*, so die Fachtermini der Branche, haben die Entwickler*innen dieser in Berlin ansässigen Fotoplattform ihren Algorithmen „beigebracht“, Objekte auf Fotos zu klassifizieren, daraus mitunter schlüssige Bildunterschriften zu generieren und sogar eine ästhetische Bewertung auf einer Skala von eins bis hundert vorzunehmen, den *Ästhetik-score*.

Der *Algorithm of Beauty*, wie die Firma dieses automatisierte ästhetische Urteil in einem Beitrag nannte,¹ trennt gelungene, ausdrucksstarke Fotos vom blässeren Durchschnitt. Während die einen „unique“ und „original“ sind, sind die anderen gewöhnlich, schlecht belichtet oder haben einen flachen Farbraum. Der algorithmische Foto-Kurator hebt nur jene Bilder aus der Masse der Millionen von Aufnahmen hervor, die diesen Standards genügen. Mitglieder bei *EyeEm* sind inzwischen mehr als 20 Millionen Profi- aber vor allem Amateurfotograf*innen weltweit. Nur Fotos mit hohem *Punktstand* erscheinen auf den ersten Seiten der Bilddatenbank.



Aesthetics moderate 62%

Suggested tags

environment wind turbine turbine sky fuel and power generation
renewable energy wind power environmental conservation cloud - sky
alternative energy field landscape nature land plant day rural scene
no people grass technology outdoors sustainable resources

Suggested caption

windmill on field against sky



Aesthetics really good 75%

Suggested tags

child childhood human body part portrait baby white background
body part looking at camera young offspring human face cut out
one person studio shot emotion cute human eye human head
innocence small

Suggested caption

portrait of cute baby

¹ Appu Shaji: Beauty & The Algorithm, in: Machina. A Curation of Real Photography by a Machine, EyeEm Magazine (November 2016).

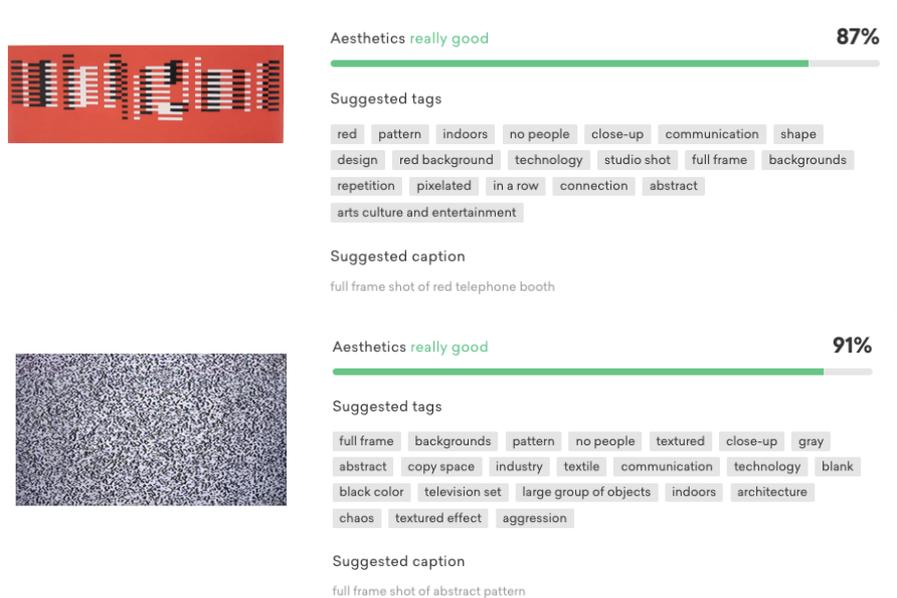


Abb. 1 a-d: Drei Tests des Eye-Em Algorithmus mit eigenen Bildern, dem Foto einer Windkraftanlage bzw. einer Puppe, einem Gemälde von Josef Albers und TV-Rauschen. Die automatisch vorgeschlagenen Bildunterschriften zeigen, wie gut der Algorithmus auf übliche Fotografien trainiert ist. Abstrakte Muster und Puppen werden nicht erkannt. Sie lauten: Windrad auf einer Wiese gegen den Himmel; Porträt eines süßen Babys; Vollbild einer roten Telefonzelle; Vollbild eines abstrakten Musters. (Screenshots: Birgit Schneider, 2019)

Zunächst als Sortierungs-App für unübersichtliche Fotoarchive auf Smartphones entwickelt, ist die Plattform seit 2014 in Kooperation mit *Getty Images* ein *market place*. Denn alle Fotograf*innen können ihre Bilder über den Dienst verkaufen – also für Werbekampagnen, Zeitungen und Bildbedarfe jeder Art. Die Fotografen und Fotografinnen werden zu 50% am Gewinn beteiligt ohne ihr Copyright am Bild zu verlieren. So wurde die Plattform zu einem ernst zu nehmenden Konkurrenten für andere Stock-Image-Anbieter. Das mächtige Potential der neuen Bildagentur liegt an der algorithmischen Analyse der Bilder mit Hilfe der künstlichen Intelligenz neuronaler Netze. Der Objekt-Klassifizierungs-Algorithmus des Startups heißt *EyeEm-Vision*.

Die Analyse besteht einerseits in einer automatischen, englischsprachigen Zuordnung von Schlagworten zu Bildern und Bildmotiven. Die neuronalen Netze haben gelernt, selbst kleine Details auf den Fotos zu erkennen und annotieren („taggen“) diese mit Schlagwörtern. Auf Abb. 1a z.B. erkennt der Algorithmus Windkraftanlagen, Felder, einen Himmel mit Wolken, eine Landschaft ohne Menschen und vieles mehr, woraus die Bildunterschrift „Windrad auf dem Feld vor Himmel“ generiert wird. Die Begriffe bestehen vor allem aus Nomen und Adjektiven. Der Algorithmus von *EyeEm-Vision* kann aber auch Gefühle und abstrakte Konzepte wie „Unschuld“, „Ekel“ oder „Liebe“ in den Bildern erkennen, wobei es auffällig wenige Bilder sind, die mit negativen Begriffen verknüpft sind; positive Bildaussagen scheinen in einer werbe- und wohlfühlorientierten Bilddatenbank zu überwiegen.

Während es bei der Objektklassifizierung darum geht, die Bildinhalte möglichst umfassend zu erkennen, geht es bei einer zweiten algorithmischen Analyse um die Ästhetik der Bilder; im Fall der Windanlage ist diese „moderate“ und liegt bei 62 Prozent. Der Algorithmus zur ästhetischen Analyse ist – wie im Übrigen jedes neuronale Netz, das Bilder erkennt – zunächst durch Menschen trainiert. In diesem Fall wurde dem Algorithmus Ästhetik durch ein

Team von Fotokurator*innen beigebracht. Im Ergebnis bewertet der Algorithmus das gesamte Spektrum der Stock Photography mit einem streng numerischen „score“ in automatisierter Weise.

Es sind mehrere Punkte, die zum Thema Computer und ihrer Sehkraft aus diesem Beispiel abgeleitet werden können. Erstens, die *EyeEm-Vision* wurde allein auf das Sehen fotografischer Bilder trainiert. Für *EyeEm-Vision* erschließt sich nur erfolgreich, was im dreidimensionalen Raum durch Linsen und durch die Gesetze der Optik zweidimensional abbildbar wurde, was also im fotografischen Raum Objektcharakter besitzt. Auf dem Auge abstrakter Malerei, der Kartografie, der Diagrammatik oder von Zeichnungen und Cartoons (um nur wenige andere Bildtypen zu nennen, die einen großen Teil des Universums der Bilder ausmachen) ist der Algorithmus blind oder zumindest sehschwach. Ich habe versucht, die Grenzen des Algorithmus mit gezielt gewählten Bildern auszuloten, um zu sehen, wo die Bilderkennung auf Abwege gerät. Es erzeugt eine unwillkürliche Komik, wenn die Computer Vision einen Puppenkopf als süßes Baby, die „programmierte“ Ästhetik des Künstlers Josef Albers als Telefonzelle und weißes Rauschen als Aggression mit 100 Prozent Ästhetik bewertet (Abb. 1b, c, d).

Dann wiederum wird durch das Beispiel von *EyeEm* deutlich, dass das Kerngeschäft der Computer Vision, wie im Übrigen auch aller Kunsthistoriker*innen, darin besteht, Bilder in Worte zu „zerfasern“, wie Vilém Flusser dies genannt hätte. Erkennen bedeutet in Worte fassen.² Oder besser: Bereits in Codezeilen zerfaserte Bilder werden in einem zweiten Schritt in für Menschen verständliche Begriffe übersetzt. Bei dieser Objektklassifizierung mittels neuronalen Netzen wird eine statistische Taxonomie zum Ordnungssystem eines Bildbestandes. Und zuletzt: Indem ich hier Begriffe wie „lernen“, „trainieren“ und „sehen“ benutze, übernehme ich damit die begrifflich behauptete Gleichsetzung dieser Vermögen von Menschen und Maschinen.

2. Elemente einer Medienarchäologie digitaler Methoden der Kunstanalyse

Bevor ich mich weiter mit dem Sehen von Computer Vision befasse, möchte ich einige Elemente einer Konzept- und Mediengeschichte bildanalytischer Verfahren anführen, die ich für die gegenwärtigen Debatten um Digital Humanities als wichtig erachte. Denn diese gründen den Blick noch einmal tiefer für das, was gegenwärtig unter dem Dach der Geisteswissenschaften mit digitalen Methoden jenseits vereinfachter Vorstellungen von digital und analog erzeugt werden soll. Die Perspektive baut auf der medienarchäologischen Erkenntnis auf, dass eine Historiografie des Digitalen in Techniken wie der Schrift oder in Stiftwalzen, in Diagrammen, Notationen, Lochkarten sowie anderen Codierungen zu finden ist, da diese Verfahren strukturell auf denselben digitalen Prinzipien aufbauen, wie sie später in den Computer eingewandert sind.

Eine Geschichte digitaler Methoden in den Geisteswissenschaften liegt bislang nur in Ausschnitten vor. Aus einer mediengeschichtlichen und wissenschaftshistorischen Perspektive muss diese Geschichte weiter zurückreichen als bis zum „Gründungsvater der digitalen Geis-

² „Es geht um ein Übersetzen von Vorstellungen in Begriffe, um ein ‚Erklären‘ der Bilder, ein Zerfasern der Bildflächen in Zeilen.“ Vilém Flusser: *Ins Universum der technischen Bilder*, Göttingen: European Photography 1996 [1985], S. 13.

teswissenschaften“, dem Jesuitenpater und Linguisten Roberto Busa, der 1949 ein erstes computergestütztes Editionsprojekt auf einem IBM begann, um einen Index aller Wörter im Werk von Thomas von Aquin zu erstellen. Oder bis zu den quantitativen Analysen des Physikers Wilhelm Fucks, Präsident der Universität Aachen und Atomenergieforscher, der die Methoden der Physik und Mathematik seit den 1950er Jahren auch auf Gegenstände der Musik, Literatur- und Kunstgeschichte anwendete, um Regeln und Gesetze der Kunst zu ergründen. In seinem Buch „Nach allen Regeln der Kunst“ (1968) stellte Fucks unter anderem Bildanalysen vor, für die er die Mosaiksteine der berühmten Fresken in der Kathedrale von Ravenna auszählte und dann, den fallenden Häufigkeiten ihrer Farbigkeit folgend, diese als Balkendiagramme darstellte. Im Diagramm, so Fucks, zeige sich der Verwandtschaftsgrad der Bilder. Indem Fucks bereits eine digitalisierte – also abzählbare und in Elemente zerlegte – Bildform nutzte, nämlich die des Mosaiks, war es ihm möglich, Kunst auch ohne die Hilfe eines Computers zum Gegenstand von Arithmetik zu machen (Abb. 3). Seine „objektiven Einsichten aus quantitativer Analyse“ beschreibt er im Vorwort als zwingende Konsequenz jeder Zählbarkeit. „Heute steht die Methode ‚Beobachtung, Experiment und zusammenfassende mathematische Beschreibung‘ prinzipiell jedem zur Verfügung. [...] Wer immer irgend etwas richtig zählt oder mißt, Dinge oder Vorgänge gleich welcher Art, gewinnt allemal objektive Erkenntnisse. ‚Objektiv‘ soll hier heißen: mitteilbar mit Zustimmungszwang.“³ Da Satzlängen, Worthäufigkeiten, Übergangshäufigkeiten von Tonhöhen und Mosaiksteine zählbar sind, eröffnen auch Kunstwerke einen Zugang zur Objektivität.

Was bereits der Ansatz Wilhelm Fucks verdeutlicht: Wer mit medienarchäologisch geschultem Blick nach historischen Vorläufern einer quantitativen Bestimmung kultureller Fragen sucht, müsste die Geschichte digitaler Methoden noch viel weiter fassen und auch auf den Einsatz quantitativer Methoden in den Geisteswissenschaften vor dem Computer hin befragen. Das epistemische Programm digitaler Erkenntnismethoden war nämlich bereits unabhängig vom und vor jedem Gebrauch von Großrechnern vorhanden. Dies zeigt insbesondere die Geschichte diagrammatischer Verfahren. Beim Blick in die Geschichte der Datenvisualisierungen wird deutlich, wie bereits im 18. und 19. Jahrhundert mit mathematischen und visualisierenden Methoden versucht wurde, kulturgeschichtlichen Gegenständen ein quantitatives und objektivierbares Wissen abzutrotzen, das mit traditionellen historischen oder philologischen Methoden nicht zu erlangen war. Im Rückblick entsteht der Eindruck, dass die Geschichte der grafischen, diagrammatischen Methoden und der digitalen Methoden zum Teil dieselben sind. Auf Grundlage dieser Vermutung stelle ich im Folgenden einige Funde vor, die ich als Ansätze einer Geisteswissenschaft mit digitalen Methoden vor 1949 anführen würde. Dies sind eher Neben- als Hauptfunde. Jedoch legen diese gerade aufgrund ihrer Nebensächlichkeit die anhaltende Bedeutung von quantitativen und messtechnischen Verfahren zur Analyse von Kunst lange vor dem Computer nahe.

3. A New Chart of History

Für die Geschichte der Diagrammatik stellt die Chronographie *A New Chart of History* des US-amerikanischen Theologen, Physikers und Chemikers Joseph Priestley einen wichtigen Ankerpunkt dar. Priestley publizierte seine „neue Darstellung der Geschichte“ im Jahr 1769

³ Wilhelm Fucks: *Nach allen Regeln der Kunst*, Stuttgart: Deutsche Verlagsanstalt 1968, S. 8.

als ausfaltbare Tafel (Abb. 2).⁴ Die Grafik begründete das Feld der Synopsen (wörtl. Gleichschau, Zusammenschau), welches das damals herrschende Geschichtsbild bildlich neu darzustellen und didaktisch zu vermitteln wusste.

Priestley hatte für die Chronographie das historische Wissen seiner Zeit diagrammatisch ausgewertet, wobei ihn vor allem Jahreszahlen und das Anwachsen und Verschwinden von Weltreichen sowie die Lebensdaten von Herrschern interessierten. Der anschwellenden Masse an historischen Informationen, die sich in den Archiven ansammelten, stellte Priestley mithin die Chronographie entgegen. Die Synopsen stehen so für den Versuch, in den disparaten Daten der geschichtlichen Archive Muster zu erkennen und die geschichtlichen Verläufe in grafischer Form vorstellbar und vergleichbar zu machen. Indem im Zentrum des Verfahrens quantifiziertes Wissen steht, muss das Diagramm als frühes Beispiel einer Anwendung grafisch-quantitativer Verfahren auf geisteswissenschaftliche Gegenstände gelten.

Auch Priestleys Ergebnis lässt sich durchaus in den Begriffen der digitalen Geisteswissenschaften beschreiben. Der Zeitstrahl auf der Abszisse leistet eine Quantifizierung anhand von Jahreszahlen. Auch die Ordinate ist geordnet, jedoch nicht so streng quantitativ wie die Abszisse. So gibt die Y-Achse die veränderliche Größe von Reichen an, ohne auszuweisen, wie das Maß der Reiche genau entschieden wurde. Im Bezug beider Größen und im historischen Verlauf – Ausdehnung von Reichen in der Zeit – entstehen die kantigen, farbigen Flächen, die Reiche ein- und auszuschachteln scheinen. Die von politischen Landkarten bekannten Völker und Reiche erscheinen in dieser Methode in gänzlich neuer Gestalt und im Prozess der Geschichte.

Durch die grafische Darstellung historischer Verläufe entlang einer streng linearen Zeitachse wird ein bestimmtes Geschichtsbild vermittelt: Dieses erzählt eine koloniale Weltgeschichte von Großreichen und ihren Herrschern aus der Sicht Europas. Ins Zentrum der Aufmerksamkeit gerückt wird der Aufstieg und der Verfall von Staaten. Insbesondere der Erfolg und die Dauer bestimmter Reiche werden plastisch, wenn diese sich in expandierenden Formen über das Blatt ausbreiten und so auch eine zukünftige Geopolitik andeuten.



Abb. 2: Joseph Priestley: Historische Weltkarte. Visual Analytics-Projekt aus dem 18. Jahrhundert.

Abb. 3: Wilhelm Fucks: Nach allen Regeln der Kunst, dva Stuttgart, 1968. Farben, geordnet nach fallenden Häufigkeiten am Beispiel eines bereits digitalisiert vorliegenden Bildes, nämlich des Mosaiks im Baptisterium der Kathedrale Hl. Petrus, Ravenna.

⁴ Joseph Priestley: A description of a new chart of history, containing a view of the principal revolutions of Empire, that have taken place in the world, London: J. Johnson 1770.

4. Die Koeffizienten der Stilgeschichte

Auch die Kunstgeschichte erkannte die Potentiale, die grafische und quantitative Verfahren bieten. Diese stehen im Kontext einzelner Versuche, Baudenkmäler und Kunststile in tabellarischer und kartografischer Form zu erfassen wie z.B. in der Mitte des 19. Jahrhunderts durch den Kunsthistoriker Franz Mertens.⁵ Hier jedoch soll ein Gedanke hervorgehoben werden, der sich in Gottfried Sempers Stillehre findet, einem Buch, das wegweisend für die Kunstgeschichte war – und das sich wie eine Vision für spätere Visual Analytics in der Kunstgeschichte liest. Semper ist interessiert daran, die bei der Entstehung von Kunstwerken „hervortretende Gesetzmäßigkeit und Ordnung“ zu ergründen. Denn diese seien als „allgemeine Prinzipien“ für eine „empirische Kunstlehre“ wirksam.⁶ Wie bereits Priestley geht es mit hin auch Semper um die Erforschung von Stilgeschichte und Stilentwicklung in historisch-geografischem Verlauf. Um das „innere Gesetz“ hervortreten zu lassen, das „durch die Welt der Kunstformen gleich wie in der Natur walte“ steht Semper eine bestimmte Sicht auf die historische Entwicklung der Kunstlehre vor Augen. Und hier kommt nun ein mathematischer Vergleich ins Spiel, wenn Semper das mathematische Bild einer Funktionskurve einschaltet: Die Kunstlehre durchwandere „[...] das Feld der Geschichte, die Kunstwerke der verschiedenen Länder und Zeiten nicht als Thatsachen auffassend und erklärend, sondern sie gleichsam entwickelnd, in ihnen die nothwendig verschiedenen Werthe einer Funktion, die aus vielen variablen Coëffizienten besteht, nachweisend“.⁷ Er skizziert also die Idee, dass die Kunstgeschichte einer Funktion⁸ folge, die „nachweist“, wie die Kunst im historischen Verlauf von verschiedenen Einflüssen bestimmt wurde.

Im Vergleich von Semper drängt es sich auf, den x-Wert als zeitliche Ordnung der Stilgeschichte zu betrachten, während der y-Wert den Stil zwischen Niedergang und Werden und alle relevanten Einflüsse verzeichnen könnte. Von den wichtigsten „thätigen Momente[n]“,⁹ die eine Kunstform bestimmen, handelt Sempers gesamtes Buch. Dies sind primär die von ihm breit dargestellten Gebrauchs- und Schutzbedürfnisse, die Konsequenzen des verwendeten Materials sowie der jeweiligen technischen Verfahren; aber auch geografische Besonderheiten bestimmter Regionen sowie Katastrophen politischer und natürlicher Art benennt Semper als Gründe für das Entstehen und den Niedergang von Kulturen und ihren Artefakten. Würde man die Funktion als Kurve anzeichnen (Abb. 4), ließe sich im Verlauf der Kurve das seit Giorgio Vasari (1511-1574) geschilderte Auf und Ab der Stilgeschichte zwischen den Polen von Krise, Höhepunkt und Neugestaltung sichtbar machen. Das innere Naturgesetz der stilgeschichtlichen Formentwicklung würde sich in der charakteristischen Form der Höhen und Talfahrten einer Linie als Bild offenbaren.

⁵ Wolfgang Cortjaens/Karsten Heck (Hg.): Stil-Linien diagrammatischer Kunstgeschichte (Transformationen des Visuellen, Band 2), Berlin/München: Deutscher Kunstverlag 2014.

⁶ Gottfried Semper: Der Stil in den technischen und tektonischen Künsten oder praktische Ästhetik. Ein Handbuch für Techniker, Künstler und Kunstfreunde, Bd. 1: „Die Textile Kunst für sich betrachtet und in Beziehung zur Baukunst“, München: Bruckmann's 1878 [1860], S. VIII.

⁷ Ebd.

⁸ Für die Mathematik gilt eine Funktion (auch ‚Abbildung‘ genannt) als Relation zwischen zwei Mengen, die jedem Element der einen Menge – genannt Variable oder x-Wert – genau ein Element der anderen Menge – genannt abhängige Variable oder y-Wert – zuordnet. Koeffizienten (von lat. *coefficiente* ‚mitwirken‘) wiederum sind Zahlenausdrücke, die die Variablen weiter bestimmen.

⁹ Semper 1878, S. VIII.

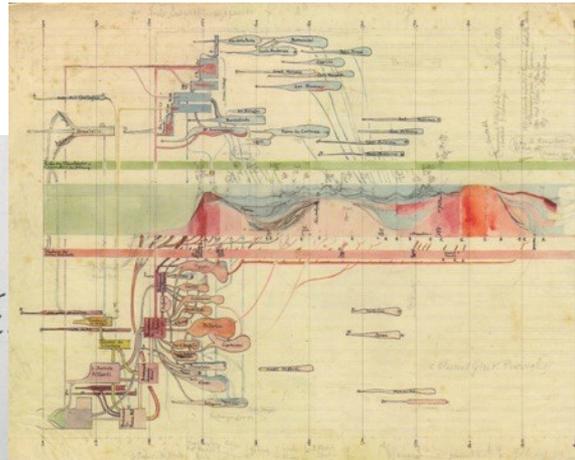
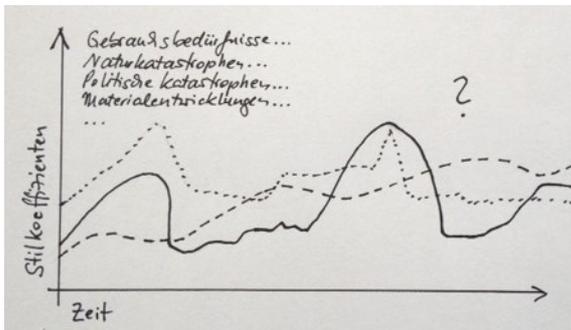


Abb. 4: Stilgeschichte als Funktion von Zeit und Stilkoeffizienten nach Semper? Die Kunstlehre durchwandere „das Feld der Geschichte, die Kunstwerke der verschiedenen Länder und Zeiten nicht als Thatsachen auffassend und erklärend, sondern sie gleichsam entwickelnd, in ihnen die nothwendig verschiedenen Werthe einer Funktion, die aus vielen variablen Coëffizienten besteht, nachweisend“. Skizze: Birgit Schneider.

Abb. 5: Heinrich von Geymüller: Perioden und Phasen des Renaissance-Stils in Frankreich, 1890er Jahre. Quelle: Heinrich von Geymüller (1839-1909). Architekturforscher und Architekturzeichner, hrsg. Josef Ploder und Georg Germann, Basel: Friedrich Reinhardt 2009.

Wissend, dass der kurze Satz aus Sempers Stillehre hier bereits einer ‚Überinterpretation‘ zum Opfer gefallen ist, kann die Idee Sempers dennoch für die generelle Frage nach den Wahrheitsformen digitaler, quantitativer Geisteswissenschaften stehen. Semper deutete an, auch die Stilgeschichte der Kunst ließe sich den Ansprüchen einer mathematischen Rationalität gerecht machen und ihr Gesetz ließe sich mit diesen Mitteln im Bild einer einzigen Funktion fassen. Während der Architekt und Kulturhistoriker Semper sein Wissen einerseits auf Erfahrung, Beobachtung und Fachwissen gründete und intuitiver Spekulation folgte, nimmt er andererseits einen – wenngleich tastenden – Bezug auf den herrschenden Denkstil des 19. Jahrhunderts, als quantitative Verfahren zu einem neuen Königsweg der Evidenzerzeugung wurden. Messdaten und Statistik begannen für das Ideal einer objektiven Wissenschaft und eines bis heute geltenden „Trust in Numbers“ (Theodore Porter) einzustehen. Die Ausdruckskraft der auf Zahlen basierenden Linien ist nüchtern und kühl, sie repräsentieren das Ideal der im 19. Jahrhundert entwickelten, disziplinierten Moral der Objektivität.¹⁰

Die Übertragung mathematischer Methoden auf die Stilgeschichte, wie sie später realisiert wurde und auch aktuell praktiziert wird, erzeugt jedoch fundamentale Schwierigkeiten, die Semper noch nicht weiter entwickeln musste. So könnte das Bild einer kunsthistorischen Funktionskurve nur dann mit Erfolg gezeichnet werden, wenn auch entschieden werden könnte, wie die „variablen Coëffizienten“, also die zahlreichen Einflüsse auf die Kunstformen, exakt quantifiziert werden können. Denn der Stil als Gesamtbild unzähliger Faktoren und Einzelwerke widersetzt sich der „eisernen Sprache“ der mathematischen Verfahren, wie sie die abendländische Rationalität hervorbrachte. Auch wenn sich die Zeitachse, geordnet nach den Jahreszahlen der Geschichtszeit, noch recht eindeutig beziffern lässt, gilt dies nicht für die Ordinate, die den Ausschlag der Kurve bestimmt. Der unklar konturierte Forschungsgegenstand der Stilentwicklung in Abhängigkeit zur Geschichtszeit lässt sich nicht auf das Bild der harten Mathematik der Funktionskurvenmethodik bringen. Eine zweidimensionale Bewertung im Auf und Ab einer Kurve erzeugt zudem Vereinfachungen, die gleichzeitig ge-

¹⁰ Lorraine Daston/Peter Galison: Objektivität, Frankfurt a. M.: Suhrkamp 2008.

troffene Bewertungen und Vorurteile verbergen. Eine Einschätzung des Stils ist nicht das Resultat von Berechnungen und Daten, sondern gründet bis heute auf den Idealen kunsthistorischer Kennerschaft, die immer wieder neu ausgehandelt werden müssen.¹¹

Derjenige Kunsthistoriker, der eine Koeffizienzkurve für die Stilgeschichte der Architektur dreißig Jahre später dann tatsächlich entwarf, war Heinrich von Geymüller, ein Schweizer Kunst- und Architekturhistoriker, der zahlreiche Handbücher der Architektur sowie Monographien zur Architektur der Renaissance publiziert hatte (Abb. 5).¹² Die bunten Formen, die im Kreuzungsfeld des Koordinatennetzes seines Diagramms entstanden sind, sind nicht eckig wie bei Priestleys Chronographie, sie erinnern eher an pumpende Organe oder eine sonderbare hydraulische Maschine.¹³ Hier deutet sich die generelle Begrenzung der Visual Analytics an, also z.B. die Unmöglichkeit, die Stilgeschichte der Renaissance im zweidimensionalen Rahmen eines Koordinatensystems erschöpfend darzustellen.

5. Fernsehen als Mittel der Kunstanalyse

Die Beispiele bislang waren alle Ergebnis einer Arbeit mit Stiften, Papier und Lineal. Ein automatisiertes, maschinelles Verfahren zur Analyse von Kunst, das den menschlichen Analysen vermeintlich zurückdrängt, tauchte spätestens mit der ersten Idee des elektrischen Fernsehens auf. Wobei vielfältige Anwendungen eines möglichen Fernsehens bereits Jahrzehnte bevor dieses zu einem Alltagserlebnis wurde in Form von technischen Spekulationen erprobt wurde.

Die Schrift *Die Zukunft des elektrischen Fernsehens* von 1892 erschien im Rahmen einer zweiseitigen Publikation mit dem Titel *Ein Blick auf die großen Erfindungen des zwanzigsten Jahrhunderts*.¹⁴ Fernsehhistorisch ist die Publikation am ehesten mit den zum Teil prognostischen Abhandlungen von Adriano de Paiva (1880) und Constantin Senlecq (1881) vergleichbar.

In seiner Schrift entwarf Maximilian Pleßner, der hauptberuflich Hauptmann der preußischen Armee war und im Übrigen für seine frühe Version des Ohropax bekannt ist, zahlreiche Anwendungsgebiete von elektrischen Verwandlungsphänomenen. Diese reichen von freikünstlerischen über ästhetisch-analytische bis hin zu angewandten Gebieten. Pleßner knüpfte große Hoffnungen an die Optophonie für die wissenschaftliche Erkenntnis. So widmete er einen breiten Teil seiner Schrift den damals noch hypothetischen Möglichkeiten mittels einer Selenzelle – dem vielleicht ersten elektrischen Sensor, der Licht in elektrische Schwingung verwandelte – Klänge in Bilder umzuwandeln oder sich Bilder als Klänge anzuhö-

¹¹ Zu Bildvergleich und Kennerschaft vgl. Max Friedländer: *Von Kunst und Kennerschaft*, Leipzig: Reclam 1992 [1946]; Lena Bader/ Marin Gaier/ Falk Wolf (Hg.): *Vergleichendes Sehen*, München: Fink 2010.

¹² Karsten Heck hat die Diagrammatik dieser Idee herausgearbeitet. Karsten Heck: *Formen des Stils*. Heinrich von Geymüllers grafische Methoden der Stilgeschichtsschreibung, in: Cortjaens/Heck 2014, S. 168-185.

¹³ Vgl. hierzu umfassend Karsten Heck: *Der Schreibtisch als Denkraum*. Heinrich von Geymüllers *Tableau Graphique*, in: Heinrich von Geymüller (1839-1909). *Architekturforscher und Architekturzeichner*, hrsg. v. Josef Ploder und Georg Germann, Basel: Friedrich Reinhardt 2009, S. 54-63.

¹⁴ Vgl. Birgit Schneider: *Maximilian Plessner: Die Zukunft des elektrischen Fernsehens 1892*, in Dieter Daniels/Sandra Naumann (Hg.): *See this Sound. Audiovisiology II, Essays. Histories and Theories of Audiovisual Media and Art*, Linz/Leipzig: Verlag der Buchhandlung König 2011, S. 193-195.

ren. Die Hoffnung, die er an diese Umwandlungsphänomene knüpfte, bestand darin, dass letztlich eine „Einheitlichkeit des Schönen“ zwischen der Klangkunst und der räumlich-bildlichen Kunst herrsche.¹⁵ Diese Einheit lasse sich alsbald mit den Mitteln der Optophonetik wissenschaftlich untersuchen und beweisen.

Seine knappen Überlegungen zu einer „akustischen Schönheitskonkurrenz“ müssen deshalb als Idee einer experimentellen Klang- und Bildanalyse gelten.¹⁶ Eine ästhetische Analyse mittels Optophonie würde zeigen, dass das Klangerlebnis eines Apollo als ebenso schön empfunden würde wie der Anblick der steinernen Statue. Die Fassaden altgriechischer Bauwerke wollte Pleßner akustisch mit den Fassaden späterer Epochen vergleichen; mit derselben Methode trachtete er „Tongemälde“ von Werken der bildenden Künste, wie beispielsweise von Tizian, den naturalistischen „Hässlichkeiten“ seiner Zeit akustisch gegenüberzustellen.¹⁷ Ohne hier auf die interessante Schrift weiter eingehen zu können, zeigt sich bei Pleßner abermals wie ein ingenieurstechnischer Blick auf Kunst für die Hoffnung leitend ist, mit Kurven ließe sich Kunstwerken eine kontrollierbare Wahrheit abtrotzen, die einen Zugang zum Reich der Objektivität eröffne.

Knapp vierzig Jahre nach Pleßners rein spekulativen Fernsehvisionen begann Fritz Wilhelm Winckel, der heute als ein Pionier der elektronischen Musik gehandelt wird, als Student der Fernmeldetechnik und Akustik in Berlin im Labor von Dénes von Mihály mit Fernsehtechnik zu experimentieren. 1930 verfasste er eine Publikation zum Fernsehen, die auf der Basis der ersten konkreten Übertragung von Fernsehbildern die Technik und Aufgaben des Fernsehens für ein breites Publikum skizzierte. Darin berichtet er unter anderem von seinen im selben Jahr im Labor von Mihály unternommenen Versuchen, Musik in Bilder umzuwandeln. Hierzu verwendete er einen mechanischen Fernseher, der mithilfe einer Nipkow-Scheibe Bilder zeilenweise in 1200 Lichtimpulse zerlegte (Abb. 6). An den Fernseher schloss er ein Radio an. Da beide Medien elektrische Schwingungen in einem ähnlichen Spektrum verarbeiteten, musste es möglich sein, akustische Impulse optisch darzustellen.¹⁸

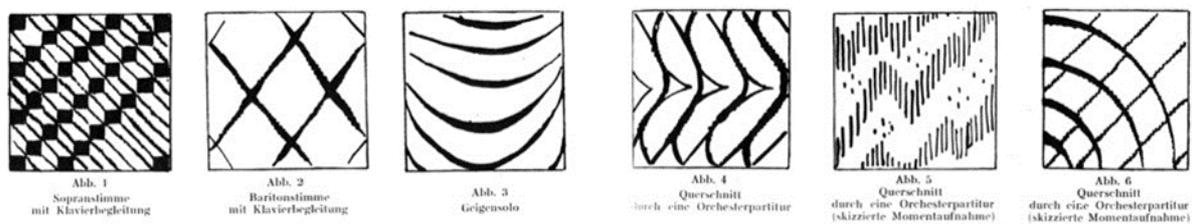


Abb. 6: Fritz Wilhelm Winckel: Visualisierung von Musik mit Hilfe einer Nipkowscheibe. Aus einem Artikel in der Zeitschrift: *Fernsehen*, Nr 4, 1930.

Ähnlich wie Pleßner deutet Winckel in den abschließenden Sätzen seines Artikels eine Verschiebung des Experiments aus dem Kontext der Kunst in das Feld der angewandten Technik

¹⁵ Maximilian Plessner: Ein Blick auf die grossen Erfindungen des 20. Jahrhunderts, Heft 1: Die Zukunft des elektrischen Fernsehens, Berlin: Dümmler 1892, S. 51.

¹⁶ Ebd., S. 52.

¹⁷ Ebd.

¹⁸ Vgl. ausführlich zu diesen Versuchen Birgit Schneider: Von hörenden Augen und sehenden Ohren: Medienästhetik unterschiedlicher Verhältnisse von Ton und Bild, in: Daniels/ Naumann 2011, S. 174-190.

an. „Die Synthese der Kunst auf elektrischem Wege führt uns ebenso zur Analyse, zur eindeutigen, objektiven Beurteilung eines Kunstwerkes als Ergänzung und Kontrolle des persönlichen, hin- und herpendelnden Geschmacks.“¹⁹

In allen genannten Beispielen wird Kunst auf mathematischem oder ingenieurstechnischem Weg zum Gegenstand objektivierender Verfahren gemacht. Die technischen und grafischen Methoden sind an die Hoffnung geknüpft, die Ästhetik zu standardisieren. Entgegen Immanuel Kants Auffassung, dass ästhetische Urteile nicht dem Bereich der Vernunft zugeordnet werden könnten, sollen diese rationalisiert werden. Zusammenfassend lassen sie diese Bestrebungen so deuten: Das mit ihnen verknüpfte Ziel besteht darin, in den Kurven „die Sprache der Phänomene selbst“ zu erkennen, vergleichbar mit den Verfahren, die Étienne-Jules Marey in seiner Schrift zu den grafischen Methoden in den experimentellen Wissenschaften skizzierte.²⁰ Die Rationalisierung erfolgte im Modus einer durch Maschinen oder quantifizierte Verfahren beglaubigten und hergestellten Objektivität. Diese ist übermoralisch, diszipliniert und standardisiert, sie verfährt präzise, kühl und messend. Auch die hier vorgestellten Versuche entsprechen folglich dem Ideal, das die Wissenschaftshistoriker*innen Galison und Daston in der historischen Entstehung im 19. Jahrhundert als mechanische Objektivität für unterschiedliche Medien und Wissensgebiete beschrieben haben.²¹

Die aus den Elementen der Medienarchäologie abgeleitete These lautet demnach, dass auch die gegenwärtigen Bestrebungen, Kunst zu quantifizieren und zum Gegenstand automatisierter Verfahren zu machen, immer mit der Geschichte der mechanischen Objektivität und dem Positivismus dieser Verfahren verbunden sind. Das epistemologische Paradigma dieser Praxis ist empirisch und statistisch. Es verschiebt die Theoriebildung in technische Verfahren, die dann selbst wieder erklärt werden müssen. Die Datenbilder, die aus derartigen Verfahren folgen, also die Kurven, Balken und Karten, sind ebenfalls Teil jenes epistemologischen Paradigmas, das dem der Geisteswissenschaften fremd ist, die ihr Wissen bislang nicht auf derartige Verfahren gründen.

Der Knackpunkt dieses Paradigmas für die Geisteswissenschaften liegt jedoch in der *Kritischen Theorie*. Nach dieser von der Frankfurter Schule skizzierten Theorie sind Fakten keine gegebenen Tatsachen, sondern gesellschaftlich fabriziert. Aus ihnen ließe sich sogar das Unrecht gesellschaftlicher Herrschaft entziffern, wie es auch in der Kritik der digitalen Geisteswissenschaften als „neoliberal tools“ formuliert ist.²² Dieses Unrecht könne nur entschlüsselt werden, wenn man um die Gemachtheit der Fakten weiß und diese analysiert. Die Kritische Theorie ist zudem an einem Vernunftbegriff orientiert, der nicht in der Zweck-Mittel-Rationalität aufgeht. Vielmehr steht er der instrumentellen Vernunft entgegen, also zum Beispiel der Bürokratisierung und der technischen Naturbeherrschung. Im Rahmen der naturwissenschaftlichen Methoden hat diese Sichtweise oftmals keinen zentralen Platz oder sie verweigert sich sogar dieser Perspektive. Mit diesem Problem müssen sich „digitale Geisteswissenschaften“ auseinandersetzen.

¹⁹ Fritz Wilhelm Winkel: Technik und Aufgaben des Fernsehens, Berlin: Rothgessler & Diesing, 1930, S. 61.

²⁰ Étienne-Jules Marey: La méthode graphique dans les sciences expérimentales – La circulation du sang à l'état physiologique et dans les maladies, Paris: G. Masson 1878.

²¹ Daston/Galison 2008.

²² Vgl. Daniel Allington/Sarah Brouillette/David Golumbia: Neoliberal Tools (and Archives): A Political History of Digital Humanities, Los Angeles Review of Books (Mai 2016), open access.

Abschließend für diesen Abschnitt soll ein letztes Beispiel für die Medienarchäologie digitaler Verfahren stehen, das die epistemischen Fallstricke von Kunstbetrachtungen mittels Linien ironisierte. Die herausstechende Kurven-Abbildung aus Laurence Sternes Roman *The life and opinions of Tristram Shandy* (1762) zeigt eine literarische Adaption diagrammatisch aufgeladener Linienornamentik (Abb. 7). Das Buch erzählt in neun aufeinander folgenden Bänden (1759–1767) die Lebensgeschichte des Tristram Shandy als autobiographischen Roman, wobei Sterne die Vorstellung einer Chronologie oder eines roten Fadens, den eine Lebensgeschichte – oder ein Buch – erwarten lässt, systematisch untergräbt. Kapitel brechen plötzlich ab oder werden an unerwarteten Stellen fortgesetzt. Um die Struktur des Buches seiner Leserschaft zu veranschaulichen, zeichnete Sterne im sechsten Band seines Romans verschiedene arabeskenförmige Kurven. Diese sollen die Handlungsverläufe von Band eins bis fünf darstellen, sind also eine diagrammatische Veranschaulichung des Romans und seiner verschränkten und springenden Zeit-, Handlungs- und Raumebenen. Der Erzähler Shandy verlässt an dieser Stelle des Buches die Handlung und begibt sich auf die Metaebene der Romanstruktur, die er mit der Semantik von Beschriftungen, Schnörkeln, zickzackförmigen und ausgestülpten Linien karikiert – wobei die Diagramme eine Klarheit erzeugen, die sie beim zweiten Blick sogleich wieder verdunkeln. Es bleibt die Frage, ob ein Lebenslauf nicht eher aus nachträglichen Meinungen und Zufällen zusammengesetzt ist, als aus Chronologie, Planungen und logischen Abfolgen.

An dieses Beispiel anschließen möchte ich die Frage, was sich verändert, wenn Geisteswissenschaftler*innen ihre Erkenntnisse ebenfalls in Kurven und Balkendiagrammen publizieren? Welche Logik gerät so in die Wissensproduktion der Geisteswissenschaften? Besteht hier ein unüberwindbarer Methodenkonflikt, der die langsamen, qualitativ-fragenden Geisteswissenschaften mit digitalen Methoden in quantitative, ergebnisorientierte, schnelle Ergebnisse zu überführen hofft? Und wenn dieses Wissen aus dem hier historisch skizzierten objektiven Paradigma kommt, wie lässt sich dieses für die Geisteswissenschaften produktiv machen oder umwerten, ohne das kritische Potential – die Kritik genau dieser Methoden – zu verlieren? Aus der Linienornamentik Laurence Sternes leite ich ab: Nur wer um den Objektivierungsanspruch diagrammatischer, quantitativer Verfahren und ihren zu kritisierenden Positivismus sowie der Grenzen dieser Verfahren weiß, kann mit diesen arbeiten. Die Satire Laurence Sternes kann dabei helfen, die mit digitalen Verfahren erzeugten Ergebnisse und ihre Gestalt der Deutungshoheit aus kritischer Distanz auf ihre Wahrheitsform zu hinterfragen und insbesondere den Anteil der Medialität an den Ergebnissen mit zu reflektieren. Nur so lässt sich dem Vorwurf der Technikfetischisierung digitaler Methoden, wie er zum Kern des Neoliberalismus gehört, begegnen. Und nur so lässt sich vielleicht die Generalisierung digitaler Methoden in die Breite mit dem kritischen Denken in die Tiefe verbinden.

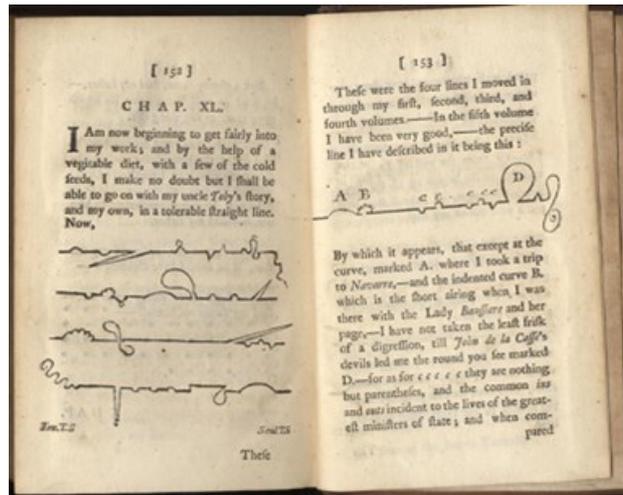


Abb. 7: Doppelseite aus Laurence Sternes Roman *The life and opinions of Tristram Shandy* (1762).

6. Der Blick der Computer Vision

Wieviel menschliches Sehen steckt in Computer Vision und wieviel davon ist Analogie? Was für ein Konzept vom menschlichen Sehen braucht man, um Computer Vision zu denken? Was ‚sieht‘ Computer Vision? Diese Fragen stehen im Zentrum des zweiten Teils dieses Artikels, der weniger durch die Lektüre von Büchern zu künstlichen neuronalen Netzen gespeist ist,²³ sondern versucht, sich mit der heuristischen Methode der Störung dem ‚Sehen‘ der Computer Vision anzunähern, um den Preis, das andere virulente Fragen an Computer Vision und Künstliche Intelligenz im Hintergrund bleiben.

Nachdem bereits Ende der 1950er Jahre mit dem Perzeptron-Modell die Funktionsweise eines künstlichen neuronalen Netzes samt der induktiven Idee einer Lernregel als sehende Maschine vorgestellt worden war, kam dieser Forschungszweig ins Stocken. Stattdessen forschten vor allem US-Wissenschaftler*innen in die Richtung einer symbolischen künstlichen Intelligenz, die nicht auf der Idee des Lernens, sondern einer automatisierten symbolischen Deduktion eines von Beginn an vollkommenen Algorithmus basierte.²⁴

Das Buch, das das Interesse an neuronalen Netze für den aufkeimenden Bereich der Computer Vision erneut befeuerte, war ein kognitionswissenschaftliches Buch aus dem Jahr 1982 über das menschliche Sehen. Es trug den Titel *Vision – A Computational Investigation into the Human Representation and Processing of Visual Information* und stammte von dem Psychologen und Informatiker David Marr.²⁵ Die Publikation behandelte die neuronale Vernetzung beim Sehen im Gehirn, entfaltete diese jedoch als generellen Betrachtungsrahmen in Begriffen der Informatik und in Analogie zum maschinellen Sehen. Es resultierte aus For-

²³ Stellvertretend für viele andere seien genannt: Christoph Engemann, Andreas Sudmann (Hg.): *Machine Learning. Medien, Infrastrukturen und Technologien der künstlichen Intelligenz*, Bielefeld: transcript 2018.

²⁴ Vgl. Matteo Pasquinelli: *Pasquinelli, Matteo: Machines that Morph Logic: Neural Networks and the Distorted Automation of Intelligence as Statistical Inference* (zitiert aus der deutschen Übersetzung: *Neuronale Netzwerke und das Unberechenbare. Über die Automatisierung von Intelligenz als Statische Interferenz*), Glass-Bead, Site 1: Logic Gate, the Politics of the Artificial Mind (2017), open access.

²⁵ David Marr: *Vision. A Computational Investigation into the Human Representation and Processing of Visual Information*. Cambridge, Mass./London: MIT Press 1982.

schungen, die David Marr mit seinem Team bereits seit den 1960er Jahren auf der Basis der Ähnlichkeitsannahme von organischem Sehen und maschinellm Sehen vorangetrieben hatte. Die Analogie zwischen künstlichen Neuronen und Nervenzellen gründet wiederum auf dem Aktionspotential, das beide besitzen, wenn eine bestimmte Reizintensität bzw. eine Eingabegröße auf sie trifft.

Der Mensch erkennt sich in der Technik. Die Projektion verläuft dabei in beide Richtungen, wie der Philologe und Techniktheoretiker Ernst Kapp bereits 1877 ausführte. Die Erforschung der Wahrnehmung war immer verbunden mit Metaphern und Modellen, welche die Geschichte der Wahrnehmung zu einer parallel verlaufenden Mediengeschichte der Metaphern macht. Die wachsüberzogene Schreibtafel Platons oder der fühlende Seidenwebstuhl bei Diderot oder das wie ein Nervennetz verdrahtete Telegrafennetz bei Samuel Morse – die jeweils aktuellsten technischen Medien dienten als Modelle für die Kognition und andersherum.²⁶ Sie machten Bewusstsein, Wahrnehmung, sinnliche Erkenntnis, Kognition und die Nervenleitungen vorstellbar. Friedrich Kittler hat hierzu den Dichter und Philosophen Jean Marie Guyau zitiert, der 1880 schrieb:

So ist denn das menschliche Gehirn mit vielen verschiedenen Gegenständen verglichen worden. Nach Spencer hat es eine gewisse Ähnlichkeit mit mechanischen Klavieren, die eine unbegrenzte Anzahl von Melodien reproduzieren können. Taine macht aus dem Gehirn eine Art Druckerei, die ohne Unterlaß unzählige Klischees herstellt und auf Vorrat legt. [...] Alles wohl überlegt, ist das feinste Instrument (zugleich Empfänger und Motor), mit dem sich das menschliche Gehirn vergleichen ließe, vielleicht der von Edison neuerdings erfundene Phonograph.²⁷

Der „Analogieschluss“ besäße in der Wissenschaft „eine beträchtliche Bedeutung; ja vielleicht bildet die Analogie, sofern sie das Prinzip der Induktion ist, die Grundlage aller physischen und psychophysischen Wissenschaften“.²⁸

Auch für die visuelle Wahrnehmung gibt es eine Mediengeschichte der Analogieschlüsse. Als Modell für das Sehen war es spätestens seit *La Dioptrique* (1637) von René Descartes bis ins 19. Jahrhundert die Camera Obscura, die zum Verständnis diente. Heute wirkt im Kompositbegriff ‚Computer Vision‘ sowie im dazugehörigen Vokabular vormals menschlicher Eigenschaften wie Sehen/Vision, Erkennen, Lernen, Trainieren oder Intelligenz der Analogieschluss in beide Richtungen – gleichermaßen erhellend wie vernebelnd. Die Semantik der Maschine greift auf zentrale Bereiche der Semantik menschlicher Kognition und Wahrnehmung über und andersherum. Computer beginnen zu verstehen, zu lernen und zu fühlen (‚sensing‘), so wie Menschen Informationen verarbeiten oder im Umkehrschluss in manchen Fähigkeiten als roboterhaft adressiert werden.

²⁶ Zu diesem Zusammenhang gibt es zahlreiche Publikationen. Z.B. Laura Otis: *Networking. Communicating with Bodies and Machines in the Nineteenth Century* (Studies in Literature and Science), Ann Arbor: Univ. of Michigan Press 2001.

²⁷ Zitiert nach Friedrich Kittler: *Grammophon, Film, Typewriter*, Berlin: Brinkmann & Bose 1986, S. 49/50.

²⁸ Zitiert nach Kittler, ebd. S. 50.

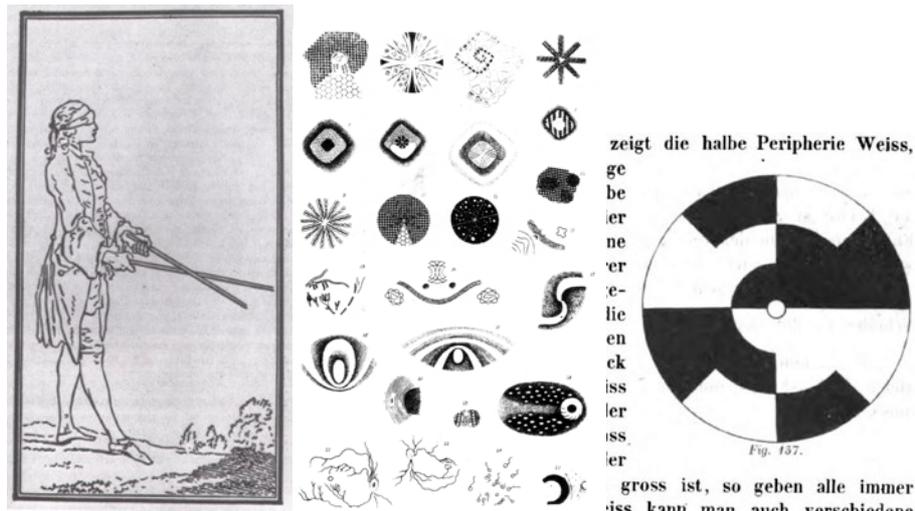


Abb. 8: Blinder mit zwei Taststöcken, aus René Descartes, *Dioptrique*, in: *Discours de la méthode*, Paris 1724.

Abb. 9: Nachbilder, aus: Jan Evangelista Purkinje: *Beobachtungen und Versuche zur Psychologie der Sinne*, 1823.

Abb. 10: Rotierende Scheiben zur Testung der Sehgeschwindigkeit von schnell wiederholten Lichteindrücken, aus Hermann von Helmholtz: *Handbuch der physiologischen Optik*, Leipzig 1867.

7. Sehstörungen als Zugang zum Sehen und zur Computer Vision

Eine Möglichkeit, etwas über das Sehen, Hören oder Tasten zu erfahren, bestand seit dem 18. Jahrhundert darin, dies über die Störungen der Sinne zu tun. Die vielleicht radikalste Version dieser Heuristik besteht darin, das Sehen aus dem Blindsein, also *ex negativo* zu ergründen. Was die Kunst über Jahrhunderte aus der Idee eines blinden Sehens zog, hat der Kunsthistoriker Peter Bexte ausführlich dargestellt.²⁹ Innerhalb der Vorstellung eines Sehens ohne Augen gewinnen alle Ebenen des Tastens – mit Stöcken oder Händen – an Bedeutung (Abb. 8). Sie bereiten die Vorstellung eines haptischen Sehens im Gegensatz zur Optik vor, das sich, so mein Vorschlag, in gewinnbringender Weise auf Computer Vision übertragen lässt. Denn auch hier findet eine Trennung in *visio* und *pictura* – Blick („Vision“) und Bild – statt und stellt gleichzeitig die Beziehung dieser beiden in Frage.

Diderot beeindruckte ein „Blindgeborener aus Puisaux“, weil dieser tastend fast alles „sehen“ kann. Sehen, so legt Diderot in seinem „Brief über die Blinden zum Gebrauch für die Sehende“ nahe, sei keine Eigenschaft, für die Menschen mit funktionierenden Augen ein Monopol besäßen.³⁰ Auch ohne Sehvermögen hat der Blinde aus Puisaux ein beeindruckendes Verständnis von Optik, von der Funktion von Spiegeln, vom Schreiben, von Mustern und Geometrie sowie von der perspektivischen Darstellung. Er besitzt zudem ein genaues Empfinden für Symmetrien und Proportionen und damit auch eine ästhetische Urteilskraft für Bildformen, denn fast alle Charakteristika des Sehens (außer der Farbe) lassen sich auch über den Raumsinn tastend wahrnehmen und vorstellen.

Mit Diderots Idee verwandt sind die späteren Experimente, bei denen seit Ende des 19. Jahrhunderts Täuschungen, Verzerrungen und blinde Flecken zu den Ankerpunkten der Wahrnehmungsforschung wurden, um von diesen aus das Sehen zu ergründen. Derartige

²⁹ Peter Bexte: *Blinde Seher. Wahrnehmung von Wahrnehmung in der Kunst des 17. Jahrhunderts*, Amsterdam/Dresden: Verlag der Kunst 1999.

³⁰ Denis Diderot: *Brief über den Blinden, zum Gebrauch für die Sehenden von 1749*. In: Denis Diderot: *Philosophische Schriften*, hrsg. v. Alexander Becker, Frankfurt/M.: Suhrkamp 2013, S. 49–99.

Experimente wurden beispielsweise durch Forscher wie Thomas Young, Jan Evangelista Purkinje, Ernst Mach, aber auch Hermann von Helmholtz (Abb. 10), Gustav Theodor Fechner oder Johannes Müller für die Erkenntnistheorie fruchtbar gemacht. Besonders beeindruckend ist die Geschichte von Purkinje, der sich in Experimenten immer wieder auf die Augen drückte und seine Pupillen grellen Blitzen aussetzte, um aus den so verursachten Bildeindrücken, den Nachbildern, die Funktionsweise des Auges zu entschlüsseln (Abb. 9).³¹

Was für eine Art von Sehen scheint in den Störungen von Computer Vision auf, die auf künstlichen neuronalen Netzen (KNN) basiert, also welches Sehen zeigt sich dort, wo dieses nicht funktioniert? Welche Nachbilder, welche Sehstörungen hat diese Form der Computer Vision? Welche blinden Flecken? Gibt es auch hier basale Reize, trübe Blicke, Unschärfen und störendes Flimmern oder Blitzen? Welches vorangelegte Raster bringen KNN bereits mit? Wobei im Folgenden speziell das gegenwärtig besonders erfolgreiche programmierte Sehen auf der Basis von sogenannten *Convolutional Neural Networks* im Zentrum stehen wird, das wiederum von neuronalen Netzen und der schichtweisen „Verarbeitung“ von Reizen im Gehirn inspiriert ist.

8. Computer Vision mittels *Convolutional Neural Networks*

Um der Frage der Störung weiter nachzugehen, muss zunächst die Basis der *Convolutional Neural Networks* (convolutional: ‚gefaltet‘, ‚faltend‘) erklärt werden. Mittels neuronalen Netzen können sehr komplexe Konzepte der visuellen Wahrnehmung „erlernt“ werden. Sie sind Teil eines Gebietes, das heute als „Maschinelles Lernen“ einen besonders erfolgreichen Teil der ‚Künstlichen Intelligenz‘ ausmacht.

Es waren diese Prinzipien der *Deep Convolutional Neural Networks* (CNN), welche die Fähigkeiten des maschinellen Lernens für die Bilderkennung in noch vor wenigen Jahren kaum vorstellbare Bereiche verschoben. Der in der Einleitung beschriebene Fotodienst *EyeEm* und seine Algorithmen zur automatischen Bilderkennung und -bewertung sind die Realität von Computer Vision auf dieser Basis knapp zwanzig Jahre nach der Tagung *Suchbilder* 2001 in Berlin, auf der bereits ausführlich über *Image Search* aus einer medienwissenschaftlichen Perspektive diskutiert wurde.³² 2001 war auch das Jahr, in dem der erste funktionierende Gesichtserkennungsalgorithmus von Paul Viola und Michael Jones für eine Webcam entwickelt wurde. In den letzten fünf bis zehn Jahren sind neue Computer-Vision-Methoden entwickelt worden, die die früheren Vorstellungen von Medien-, Kunst- und Bildwissenschaftler*innen bei Weitem übertreffen, so dass sich diese nur ihre vom auf die Bilder Gucken trockenen menschlichen Augen reiben können. Zu nennen ist hier die Bildersuche von *Google Search by Image*, bei der man nicht mehr allein mit Schlagworten, sondern mit Aby Warburgs Schlagbildern nach optisch ähnlichen Bildern suchen kann. Zu nennen ist weiterhin die Arbeit des *Google Cultural Institute*, das mit seinem immensen Korpus an digitalisierten glo-

³¹ Jonathan Crary: *Techniken des Betrachters. Sehen und Moderne im 19. Jahrhundert*, Dresden/Basel: Verlag der Kunst 1996.

³² Stefan Heidenreich/ Wolfgang Ernst/ Ute Holl (Hg.): *Suchbilder. Visuelle Kultur zwischen Algorithmen und Archiven*. Berlin: Kadmos 2003.

balen Museumsbeständen Technikforschung als Kulturwissenschaft betreibt.³³ Lev Manovich nennt den neuen Zweig computerbasierter Analysen von allen möglichen kulturellen, digitalen und online verfügbaren Datensets – egal ob im Auftrag von Unternehmen wie Spotify oder Google oder im Rahmen kulturwissenschaftlicher Forschungen – übergreifend „Media Analytics“.³⁴

Die Begriffe ‚maschinelles Lernen‘ und ‚künstliche neuronale Netze‘ werden oft synonym gebraucht. Engemann und Sudmann beschreiben den maschinellen Lernprozess als einen besonderen Prozess der Systemverbesserung. „Gegeben ist ein Input (x) sowie ein Output (y). Ziel ist es, für irgendeinen Input x (z.B. Bilder bzw. Pixels) den dazu passenden Output y (z.B. den Inhalt eines Bildes) möglichst genau vorherzusagen. Dafür wird das maschinelle Lernsystem im Rahmen einer sogenannten Lernphase auf Grundlage riesiger Mengen von Beispieldaten (anhand gelernter Übereinstimmungen von x, y) so lange trainiert, bis das System auf der Basis der Lernerfahrung aus dem Training in der Lage ist, auch neue, unbekannt Inputdaten zutreffend zu generalisieren.“³⁵ Luciana Parisi hebt hervor, dass die Verfahren des maschinellen Lernens „im großen Maße nicht-deduktive und heuristische Methoden [adaptierten], die es Maschinen erlaubten, von unsicheren und unvollständigen Informationen zu lernen.“³⁶ Das „maschinelle Denken“ baut mithin auf *induktiven* Verfahren auf, sie sind die Basis einer evolutionären Veränderung mit der Zeit.³⁷

Nach dem Training in einer „Seherschule“ wie dem *ImageNet* lassen sich mittels CNN und ihren Filtertechniken, die nach dem Prinzip der bekannten Photoshop-Filtern funktionieren, wichtige Merkmale wie Formen, Farben und Strukturen und deren Relationen zueinander erkennen.³⁸ Während Filter bereits seit langem zum Bestandteil der Computer Vision gehören – wie zum Beispiel in OCR-Programmen, um Schrift zu erkennen – sind die CNN darauf ‚trainiert‘, sich die optimale Auswahl und Konfiguration von Hunderten bis Tausenden von Filtern auf Basis von Millionen von Trainingsbildern „selbst beizubringen“. Ein Raster, das jede Computer Vision mittels CNN prägt – und mithin zeigt, wie das Begreifen der Computer Vision vorstrukturiert und eben keine unbeschriebene Wachstafel ist, sind die vortaxonomierten Bilddatenbanken, mit denen die Netze trainiert werden. Analog zum Aphorismus ‚wir sehen nur, was wir wissen‘ erkennt Computer Vision besonders erfolgreich, was zuvor in der Datenbank war.

³³ Margarete Pratschke zum Beispiel betrachtet seit einiger Zeit die Forschungsnarrative vom Google Cultural Institute unter dem Titel Google Cultural Institute. Berechnende Erkenntnis – Zur Ökonomisierung der Geisteswissenschaften kritisch.

³⁴ Lev Manovich: Media Analytics & Gegenwartkultur, in: Christoph Engemann/Andreas Sudmann (Hg.): Machine Learning. Medien, Infrastrukturen und Technologien der künstlichen Intelligenz, Bielefeld: transcript 2018, S. 269-288.

³⁵ Zum Begriff des Lernens im Kontext von KNN vgl. Christoph Engemann, Andreas Sudmann: Zur Einführung. Medien, Infrastrukturen und Technologien des maschinellen Lernens, in: Engemann/Sudmann 2018, S. 10, S. 9-36. Außerdem zum Thema Kognition jenseits von Kategorien wie Mensch und Maschine Kathryn Hayles; Matteo Pasquinelli: Augmented Intelligence Traumas, Meson Press, Lüneburg 2015.

³⁶ Luciana Parisi: Das Lernen lernen oder die algorithmische Entdeckung von Information, in: Engemann/Sudmann 2018, S.91-113, hier S. 98.

³⁷ Ebd. S. 99.

³⁸ Hier findet sich eine verständliche Beschreibung dieser Filter, der so genannten „image-kernels“: <http://setosa.io/ev/image-kernels/> (vom Software-Ingenieur Victor Powell).

9. Die Sehschule für Computer *ImageNet*

Es lohnt sich, die Bilder, mit denen die CNN trainiert werden, genauer zu betrachten. Eine der größten Forschungsbilddatenbanken zu diesem Zweck ist die für andere Forscher*innen frei nutzbare Bildersammlung *ImageNet*.³⁹ Diese beinhaltet derzeit über 14 Millionen Fotografien von einzelnen Objekten unterteilt in Sparten wie „animal“, „plant“, „person“, „fungus“, „sport,athletics“, „natural objects“, „material“, „instrumentation“, „scene“ und „food“. Diese sind als Taxonomie geordnet, viele der abgebildeten Gegenstände sind umrahmt. Die Datenbank wurde aus online zugänglichen Bildbeständen aufgebaut, die Fotografien sind von Hand sortiert. Mitarbeiter*innen des WordNet-Projektes der Universität Princeton versahen die Bilder mit ca. 22.000 Kategorien in englischer Sprache. Alle Fotos sind einem Substantiv zugeordnet. Jede Kategorie, z.B. „Ballon“, „Katze“ oder „Tisch“, ist wiederum mit mehreren hundert unterschiedlichen Fotos belegt, die ein Objekt in unterschiedlichen Ansichten, Typen und Winkeln zeigen. Die Klasse „Fungus“ zum Beispiel besteht aus über 300 Pilzsorten, von denen jede mit mehreren hundert Bildern belegt ist (Abb. 11).

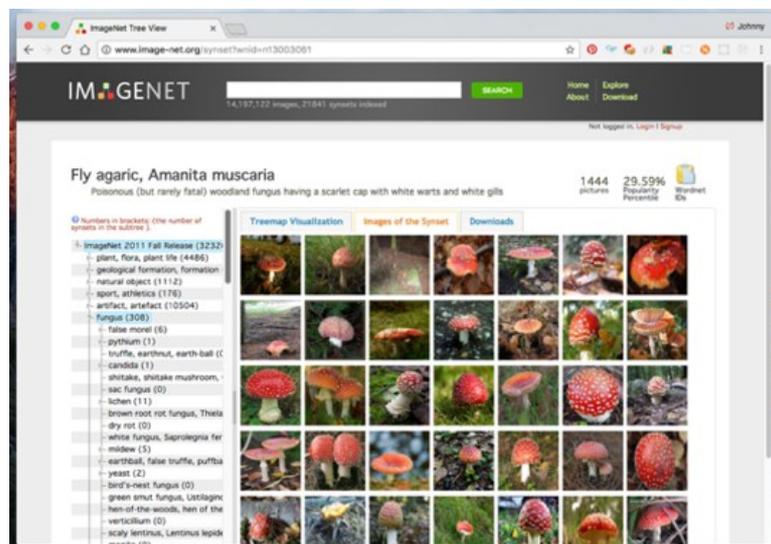


Abb. 11: Fotos von Fliegenpilzen in unterschiedlichen Formen und Perspektiven aus der Bilddatenbank ImageNet. Diese stellt Fotografien für das Training eines neuronalen Netzes zur Objekterkennung auf Bildern zur Verfügung. Quelle: Screenshot.

Das *ImageNet* ist wie *EyeEm* Vision vor allem auf sichtbare Dinge im Makro- und Mesobereich trainiert – also vor allem auf die fotografisch erfassbare Welt aus Licht und Farbe. Eine Stilgeschichte, wie sie die Kunstgeschichte beinhaltet, wird in fototechnischer Schärfe und Ausleuchtung verschluckt, sie kommt im *ImageNet* nicht vor. Stattdessen treten die Fotografien gewissermaßen stilfrei als die dargestellten Dinge selbst auf. Indem die Objekte allein mit Substantiven verknüpft sind, wird wiederum deutlich, dass abstrakte Objekte wie z.B. Stimmungen oder Gefühle nicht im Interesse der Bilddatenbank stehen. Die größte Gruppe der Bilder nimmt die Kategorie „people“ ein.

Deutlich wird beim Blick auf die Bilder, dass die Fotos vor allem aus der westlich-zivilisierten Welt und hier vor allem aus den USA stammen. Wenn z.B. eine Jurte abgebildet ist, dann mit

³⁹ <http://www.image-net.org>.

großer Wahrscheinlichkeit eine, die auf einem westlichen Zeltplatz errichtet wurde oder die aus den Reisefotografien eines westlichen Touristen stammt. Auf die Mehrheit von weißen „people“ in den Trainingsdatensätzen für Computer Vision und die Frage nach einprogrammierten Rassismen und anderen Vorurteilen wurde an anderer Stelle bereits hingewiesen.⁴⁰ Jede Sehschule ist nur so gut wie ihre menschlichen Sammlungsdirektor*innen. Die kulturelle Vielfalt einer Datenbank wie *ImageNet* ist nicht einfach nur begrenzt, sondern ein maximal vielfältiger Blickwinkel auf mannigfaltige Welten ist gar nicht das Ziel der Datenbank. Wäre ein solches Ziel auf der Grundlage von Fotografien überhaupt zu erreichen? In dieser Hinsicht muss aber auch das A Priori der englischen Sprache derartiger Bilddatenbanken bei der Klassifikation mit *WordNet* bedacht werden bzw. die standardisierte lateinisch-binäre Nomenklatur für Tier-, Pflanzen- und Pilzarten, die der Taxonomie als Grundlage dient und die kulturell und sprachlich unterschiedlichen Bezeichnungen für beispielsweise den „Fliegenpilz“ ausklammert.

10. Der statistisch-taxonomische Blick der KI

Wenn Convolutional Neural Networks Bilder erkennen, wird, stark vereinfacht gesprochen, jedes Bild in ein Raster von wenigen Pixeln Kantenlänge zerlegt. Jedes Viereck steht nun für ein Merkmal. Gemeinsam bilden die Quadrate die sogenannte „Feature Map“, also eine Karte mit unterschiedlichen Merkmalen. Für jedes der Quadrate wird berechnet, mit welcher Wahrscheinlichkeit das dort befindliche Muster z.B. einem Autoreifen *ähnlich* ist (Abb. 12). Das Ergebnis wird deshalb auch in Prozent angegeben. Das Grundprinzip des Erkennens ist die Suche nach Ähnlichkeiten, also der Bildvergleich. Erst am Schluss werden die Merkmale mit einem Substantiv, also z.B. „Autoreifen“, verknüpft.



Abb. 12: Ansicht des CNN-Fensters, in diesem Fall mit einer Convolutional Unit, die Autoräder erkennt. Quelle: <https://github.com/yosinski/deep-visualization-toolbox>.

Abb. 13: Prozess der stufenweisen Mustererkennung mittels CNN. Aus dem Artikel von Matthew D. Zeiler und Rob Fergus: *Visualizing and Understanding Convolutional Networks*, Springer International Publishing Switzerland 2014.

Mit CNN ist ein Detailblick am Werk, der Segment für Segment fließbandartig vorgeht und anhand der Bilddatenbank induktiv überprüft, ob das jeweilige Segment einem durch das Training bekannten Feature ähnelt. Zeile für Zeile wird detektiert, ob durch einen Bildausschnitt beispielsweise eine vertikale Kante verläuft oder eine Kurve. Der Vorgang der „Vision“ gelangt schichtweise von Pixeln zu Texturen und Mustern, von Texturen zu Formen und

⁴⁰ Für Spracherkennung mittels Machine Learning haben dies zum Beispiel Aylin Caliskan/Joanna J. Bryson/Arvind Narayanan untersucht: *Semantics derived automatically from language corpora contain human-like biases*, *Science* 356 (2017), S. 183–186.

Gestalten und erst dann zu klassifizierten Objekten, oder wie es der Verfasser des *Deep Learning Cookbooks* schreibt: „The lowest layers are looking for colors and simple patterns; if you go higher, the patterns become more complex.“⁴¹ Erst in den höherliegenden, späteren Schichten wird erkannt, ob sich die gefundenen Features zu bekannten Objekten zusammensetzen lassen, also z.B. Augen, Nasen oder Räder sein könnten (Abb. 13).⁴²

Beim ‚Sehen‘ zuschauen kann man den CNN bei Anwendungen der populären und seit November 2015 öffentlich nutzbaren Programmierschnittstelle *TensorFlow Object Detection* von *Google*, deren Weiterentwicklungen auf *GitHub* veröffentlicht werden, einem Online-dienst, der Software-Entwicklungsprojekte auf seinen Servern zur Verfügung stellt. Im Bild sieht man dann zum Beispiel, wie farbige Umrissformen die durch das CNN erfassten Objekte umgeben (Abb. 14). Unter den Rechtecken stehen die erkannten Begriffe, also z.B. „kyte“ und „person“.



Abb. 14: Objekt-Klassifizierung mit der TensorFlow Object Detection API, Quelle: Google, Open Source, 2016, https://github.com/tensorflow/models/tree/master/research/object_detection.[^]

Viele der Anwender*innen stellen die Erfolge ihrer Computer-Vision-Experimente mittels CNN auch als Filme online. In dieser Form lässt sich der Prozess der Bilderkennung in der Zeit beobachten, also gewissermaßen der Wahrnehmung bei der Wahrnehmung zusehen. In einem Testfilm kann man beobachten, wie das neuronale Netz im Millisekunden-Takt zwischen unterschiedlichen Begriffen hin- und herpendelt und Objekte in Abhängigkeit der Bewegung an immer neuen Stellen im Videobild findet. Die aufflackernden Rahmen wirken wie ein flackernder Geist, der sich nicht festlegen kann, was er sieht; besonders eindrucksvoll ist ein Film, bei dem hunderte schwarz-weiß gefleckte schottische Schäferhunde auf einer Hundeshow auf einer Wiese in einem Moment als Holstein-Kuh, im anderen als Hunde diagnostiziert werden. Hier sieht man, wie sich im zeitlichen Verlauf und abhängig von ihrer Bewe-

⁴¹ Douwe Osinga: What do computers see? Tricks to visualize and understand how neural networks see, O'Reilly (August 2018), <https://www.oreilly.com/ideas/what-do-computers-see> (Zugriffsdatum: 11.06.2019). Kurzfassung eines Kapitels aus dem Deep Learning Cookbook von Douwe Osinga, O'Reilly 2018 mit zusätzlichen Bildbeispielen.

⁴² Wichtig zu erwähnen ist auch, dass CNNs nicht auf Bilderkennung allein festgelegt sind und dass sie auch Muster detektieren können, die für menschliche Kategorisierungen nicht intelligibel sind, also eigenen Modellen folgen. Diesen Punkt erforscht insbesondere Paul Heinicker im Rahmen unseres Forschungsprojekts ANCI (siehe Anm. 1). Zu diesem Umstand vgl. Hito Steyerl: A Sea of Data: Apophenia and Pattern (Mis-) Recognition, in: e-flux, Nr. 72, April 2016, open access <https://www.e-flux.com>.

gung die statistische Ähnlichkeit von Objekten immer wieder verändert (Link: <https://www.youtube.com/watch?v=raZOwtJWrlo>).⁴³

Was in dieser Beschreibung deutlich wird, hat abermals mit dem induktiven Wissensmodus des maschinellen Lernens zu tun, der statistische Kalkulation ins Zentrum stellt. Parisi beschreibt diesen auf Wahrscheinlichkeiten basierenden Modus so: „Statistische Methoden konzentrieren sich auf die Funktion der Vorhersage, ermöglichen es jedoch gleichzeitig Algorithmen zu lernen und Muster wiederzuerkennen, indem sie an Datensätzen trainiert werden, die zwar zunehmend größer sind, die jedoch immer noch durch wissenschaftsgestützte Wahrscheinlichkeiten [...] begrenzt werden.“⁴⁴ Auch digitales Sehen auf der Grundlage von „neuronalen Netzen“ basiert auf Statistik. Das digitale Auge der CNN braucht die Optik nicht. Er ist in weiten Teilen ein messender und zählender Blick, ein statistisch-taxonomischer Nahblick, der ein Erkennen auf der Ebene von Wahrscheinlichkeiten erzeugt.⁴⁵ Das digitale Auge abstrahiert die Eigenschaften von Mustern, Farbverteilungen, Kontrasten, Sättigungen, Seitenverhältnissen oder Formen, indem es nach Differenzen sucht und Bildteile mit der erlernten Datenbank vergleicht – und ihre Deutung mit Begriffen vorhersagt.

11. Blinde Flecken und Sehstörungen der KI

So wie Helmholtz für das menschliche Sehen zahlreiche Prüfbilder in Form gemusterter Kreisscheiben entwarf (Abb. 11), gibt es Ansätze, die CNNs solchen Bildern auszusetzen, die ein Erkennen unmöglich machen und stattdessen das Sehen der Computer Vision selbst begreifbar machen. Ein Bildtyp sind Rauschbilder, also Bilder, die eine zufällige Verteilung von hellen und dunklen Blindpunkten aufweisen. Was aber geschieht, wenn eine CNN versucht, Kanten und Formen zu sehen, wo keine sind? In einem Artikel mit dem Titel *What do Computers See?* des Software-Entwicklers und Google-Mitarbeiters Douwe Osinga, der auch das *Deep Learning Kochbuch für Einsteiger* bei O'Reilly verfasst hat und Spezialist für *Style Transfer* ist, finden sich fruchtbare Antworten und Beispiele „to experiment with deep learning and image recognition“, die dem hier verfolgten Anliegen, Computer Vision aus der Störung und dem Blindsein zu verstehen, nahekommen.

Gleich zu Beginn vergleicht Osinga sein Experiment mit der Gehirnforschung, wenn er fragt: „So, how do we prod a neural net with an electrode?“ Das erste Experiment besteht darin, einem neuronalen Netz anstatt eines informativen Bildes Rauschen zu präsentieren und „es zu fragen, was es sieht“.⁴⁶ Hierzu wird das Netz so umprogrammiert, dass ein bestimmtes binär funktionierendes ‚Neuron‘ im Output potenziert wird, also maximale Aktivierung erhält. Im Ergebnis zeigt sich, wie ein CNN selbst dann Muster erkennt, wenn in der Vorlage keinerlei Muster angelegt sind. Im geloopten, nur wenige Momente dauernden Video erscheine kleine, Heringshaken-förmige Formen, die das Netz im Rauschen erkennt und Schicht für Schicht verstärkt (Abb. <https://www.oreilly.com/ideas/what-do-computers-see>). Aus dem Rauschen wird Information. In einem zweiten, darauf aufbauenden Versuch hat

⁴³ Die Tensorflow Object Detection API wurde durch den User „Rico C“ anhand eines auf Youtube gestellten Videos einer Hundeshow zur Aufstellung eines Weltrekords getestet.

⁴⁴ Parisi 2018, S. 99.

⁴⁵ Vgl. hierzu den Artikel von Manuela Koelke, Matteo Pasquinelli: Neuronale Netzwerke und das Unberechenbare. Über die Automatisierung von Intelligenz als statistische Interferenz, Manuskript 2018.

⁴⁶ Osinga 2018.

Osinga verschiedene Neuronen in unterschiedlichen Schichten eines CNN miteinander verbunden und verstärkt. Was hierbei sichtbar wird, ist der schichtweise Aufbau des Erkennens eines CNN, im Bild dargestellt als in Spalten angeordnete Bildkacheln von oben nach unten (Abb. 15). In der Bildabfolge zeigt sich, dass das CNN auf den ersten Schichten nach Farben und einfachen Mustern sucht (oberste Zeile), die Muster aber komplexer werden, je höher man geht (unterste Zeile).⁴⁷ In einem dritten Experiment, welches die Methode der Verstärkung mit einem Zoom verbindet, erscheinen schließlich plastisch wirkende, Fraktal-artige Muster. Alle Versuche lassen sich so deuten: Wenn Information und im Übrigen auch Kunst, wie es der Philosoph der Stuttgarter Schule Max Bense fasste, die unwahrscheinliche Ordnung ist, erkennt die Computer Vision durch die Verstärkung einzelner Neuronen unwahrscheinliche Ordnungen in der Entropie eines Rauschbildes.⁴⁸

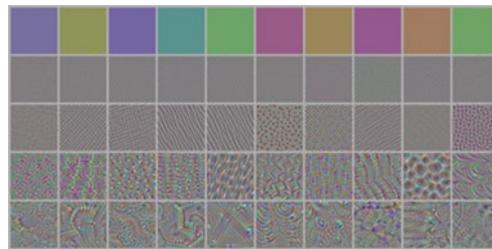


Abb. 15: „that way we get an idea of the sort of features that each layer is looking out for ... The lowest layers are looking for colors and simple patterns; if you go higher, the patterns become more complex.“ *What do computers see? Tricks to visualize and understand how neural networks see.* By Douwe Osinga, O'Reilly, August 15, 2018. (Quelle: <https://www.oreilly.com/ideas/what-do-computers-see>).

Dass die Mittel zur Analyse immer auch als Verfahren zur Synthese von neuen Kunstwerken verwendet werden können, war eine Erkenntnis, die bereits in den ersten Fernsehspekulationen von Pleßner und Winckel zu finden war. In der frühen generativen Informationsästhetik wurde diese Erkenntnis zu einem ästhetischen Leitprogramm.⁴⁹ Im Prozess der Umkehrung von Analyse in Synthese entstehen auch gegenwärtig mittels CNN neuartige Bilder. Abgeleitet aus den oben skizzierten, fehlgeleiteten ‚Sehtests‘ nach dem Prinzip „verstärken, was immer erkannt wird“ haben sich in den letzten Jahren Filme mit dem Titel *Deep Dreams* aus der Community der CNN-Entwickler*innen und speziell anhand der von Google-Entwickler*innen als Open-Access angebotenen Verfahren verbreitet. Die Resultate bestehen aus einer neuartigen, psychedelisch anmutenden, bunten Videos die sich von mäandernden, abstrakten Formen kontinuierlich in monsterartige Tierbilder verwandeln, die – so legt es der Titel *Deep Dreams* in Anlehnung an die Traumdeutung nahe – aus den Tiefen der CNN auftauchen – dem ‚Unterbewusstsein der Algorithmen‘ (Abb. 16).

⁴⁷ Osinga 2018.

⁴⁸ Max Bense: AESTHETICA. Einführung in die neue Ästhetik, Baden-Baden: Agis 1965, Kapitel: Ästhetische Prozesse als statistische S. 271.

⁴⁹ Bei Max Bense heißt es: „es ist klar, dass jeder generativen ästhetik, die natürlich eine ästhetische synthese ermöglicht, eine analytische ästhetik vorangeht, durch deren verfahren ästhetische strukturen aus vorgegebenen kunstwerken, die ihre träger sind, als ästhetische informationen präpariert werden [sic].“ Max Bense: Projekte generativer Ästhetik, in: Computer-Grafik, hrsg. Max Bense/Elisabeth Walther, Stuttgart: edition rot 1965, S. 11.



Abb. 16: *Journey on the Deep Dream*. Roger Luedecke, 2015. *Deep Dreams aus dem Computer oder doch eher die Tiere der Bilddatenbanken, mit denen die KI sehen lernt?* Quelle: <https://www.youtube.com/watch?v=SCE-QeDfXtA>.

Technisch basieren diese Experimente auf einer Kombination des sogenannten *Neural Style Transfer* mit den eben skizzierten Experimenten. Unter Stil-Übertragung sind solche neuronale Netze gefasst, die darauf trainiert wurden, den Stil eines Bildes für ein anderes Bild zu übernehmen. Auch hier werden Verfahren der CNN angewendet, bei denen ausgewählte Teile eines Netzes besonders stark aktiviert werden und Rauschbilder als Ausgangsbilder dienen. Indem die CNNs beispielsweise für die Erkennung von Katzen und Hunden trainiert sind, werden sie genau jene Attribute von Katzen oder Hunden im Rauschen erkennen und verstärken, die sie zuvor erlernt haben und im Modus der Stilübertragung auf die gefundenen Muster applizieren. Das Ergebnis sind dann jedoch keine normalen Katzen, sondern generative, unheimliche bewegte Katzenmuster, die Katzenattribute wie Augen, Schnauzen oder Fell zu sonderbaren Mustern verbinden. Osinga beschreibt diesen Vorgang so: „[W]e’ll end up with a picture that has the style of the original, but not the content. Just the eyes and hairs of the cats, but no real cats.“⁵⁰ Indem dieses Verfahren mit einem kontinuierlichen Zoom verbunden wird, stellt sich ein psychedelischer Effekt ein. Das Prinzip funktioniert auch mit anderen Ausgangsbildern, die entsprechend andere Bildwelten generieren, wenn sie z.B. Architekturen oder pflanzliche Strukturen als Attribute verwenden.

Auch bei den Deep-Dream-Filmen wurde der Vergleich zu menschlichen Wahrnehmungsformen gezogen, wovon bereits der Name „Tiefenträume“ zeugt. Darüber zogen Autoren*innen die Ähnlichkeit ihrer Ästhetik zu LSD- oder anderen Drogen-induzierten Halluzinationen als Argument für die funktionelle Ähnlichkeit zwischen künstlichen neuronalen Netzwerken und Schichten des visuellen Kortex heran. In einem Artikel wird eine Erzählung des Neurologen Oliver Sacks über eine blinde Frau zitiert, die bewegte Muster mit vergleichbarer Ästhetik wahrnahm. Obgleich sie blind war, beschrieb sie Wahrnehmungen aus Fragmenten

⁵⁰ Osinga: Gramzoom. <https://douwe.com/projects/gramzoom>, (Zugriffsdatum: 11.06.2019). Wenn CNN von neuronalen Netzen im Gehirn bei der Verarbeitung visueller Reize inspiriert wurde, sind dann beide auch in gleicher Weise limitiert? Zu dieser Annahme führt die Beobachtung, dass mit Fotos trainierte Netze ähnliche Schwierigkeiten bei der Bilderkennung haben wie Menschen. Für beide gilt, dass verdrehte Objekte die schwierigste Transformation für das Erkennen sind, gefolgt von Skalierung, Rotation in der Ebene und der richtigen Einschätzung der Position von Objekten, wie eine Forschergruppe aus Informatiker*innen und Neurowissenschaftler*innen herausfand. „Rotation in depth is by far the hardest transformation to handle, followed by scale, then rotation in plane, and finally position (much easier).“ Saeed Reza Kheradpisheh et al: Deep Networks Can Resemble Human Feed-forward Vision in Invariant Object Recognition, *Nature* (Sept. 2016). <https://www.nature.com/articles/srep32672>, (Zugriffsdatum: 11.06.2019).

von Gegenständen, die wie ein exklusiver Stummfilm vor ihrem inneren Auge ablaufen.⁵¹ Das Argument bei diesem Vergleich: Neuronale Netze erzeugen nicht nur in der Computer Vision, sondern auch – selbst ohne funktionsfähige Sehorgane – im Gehirn Bilder.

Anstelle weiter die Frage zu verfolgen, was eine solche Ähnlichkeitshypothese für die Vorstellung von „Künstlicher Intelligenz“ bedeutet, soll das letzte Beispiel der Deep Dreams abschließend auf den Begriff des „optisch Unbewussten“ bei Walther Benjamin bezogen werden. Mit dieser Wortverbindung hatte Benjamin jene Anteile der neuen Bildmedien gefasst, die sich aufgrund ihrer Dauer, Intensität und Frequenz nicht mehr mit menschlichen Augen sehen lassen, weil sie die „Merkzeuge“, wie Jakob von Üexküll es in seiner Abhandlung über die Um- und Innenwelt der Tiere genannt hatte, übersteigen. Indem Benjamin die limitierte „Merkwelt“ des Menschen auf das Kino bezog, konnte er die Besonderheit der Bewegtbilder für die menschliche Wahrnehmung herausstellen.⁵² Mit dem optisch Unbewussten fasste Benjamin folglich, wie die Kamera in der Lage ist, Dinge auf eine neuartige Weise zu zeigen und für menschliche Sinne erfahrbar zu machen, wenn z.B. Großaufnahme oder Zeitlupe Anteile der Wirklichkeit sichtbar machen, die ohne Film für Menschen nicht beobachtbar wären.

Der Begriff des Unbewussten meint in der Psychoanalyse und Tiefenpsychologie das Triebhaft-Unbewusste. Oftmals kommt es dem Verdrängten und Abgewehrten gleich, also jenen Bereichen der Psyche, die nicht direkt zugänglich sind, dieser aber zugrunde liegen. Verfahren der Psychoanalyse können diese Bereiche an die Oberfläche bringen – aber ist dies auch etwas, das für Bildmedien beansprucht werden kann, wie der Name „Deep Dream“ suggeriert? Indem die CNN als „Hineinseher“ gebraucht wird, könnte man sagen, dringt Computer Vision in neuer Weise in die Wirklichkeit ein. Die Idee des Unbewussten wiederum ist mit dem Unterbewussten verbunden, also der Idee, dass sich dieses der Wahrnehmung entzieht und sich im Hintergrund abspielt. Dementsprechend wären die Hunde und Katzen, wäre das Animalische der vielen Augen und Schuppen das Verdrängte und Abgewehrte. Es handelt sich dabei jedoch nicht um das Abgewehrte der CNN, sondern jener Menschen, die bestimmte Bilder für das Training des Algorithmus machten. Deep Dream zeigt die Phantasmen und Tiefenträume der CNN-Entwickler und ihrer Bilddatenbanken, gespeist aus dem unerschöpflichen Pool im Internet verbreiteter Bilder von Haustieren.

12. Ausblicke: Statistisch-taxonomische Blicke für die digitalen Geisteswissenschaften?

Im zweiten Teil dieses Beitrages habe ich versucht zu begründen, inwiefern der digitale Blick der CNN in weiten Teilen ein messender und zählender Blick ist, ein statistisch-taxonomischer Nahblick, der ein stufenweises Erkennen auf der Ebene von Wahrscheinlichkeiten erzeugt – und die CNN eben kein optischer Blick sind.⁵³ Mit der Figur des blinden Se-

⁵¹ Adrienne LaFrance: When Robots Hallucinate. What do Google's trippy neural network-generated images tell us about the human mind? in: The Atlantic, September 2015.

<https://www.theatlantic.com/technology/archive/2015/09/robots-hallucinate-dream/403498/>, (Zugriffsdatum: 11.06.2019).

⁵² Vgl. Jakob von Uexküll: Umwelt und Innenwelt der Tiere. Berlin: Springer 1909. Walter Benjamin gebraucht den Begriff Merkwelt im Aufsatz Das Kunstwerk im Zeitalter seiner Reproduzierbarkeit, Frankfurt a.M.: Suhrkamp 2007, S. 39.

⁵³ Auch Matteo Pasquinelli hat das maschinelle Sehen als „Thinking Eye“ bezeichnet. In seinem unveröffentlichten Vortrag „The Thinking Eye“ entwarf er eine Skizze vom biologischen Auge eines Frosches über das künstli-

hens vergleichbar kann Computer Vision auch ohne Augen wesentliche Merkmale von Bildern erkennen. Gleichzeitig verschleiert die Idee von KI, Machine Learning und Computer Vision, dass auch automatische Bilderkennung kein neutral-objektiver Vorgang sein kann, sondern immer nur so gut ist wie die Trainingsdatensätze und Aufgaben, die Menschen als Ausgangspunkte gewählt haben.

Abschließend soll noch ein Vergleich von Computer Vision mittels CNN und der kunsthistorischen Sehschule erlaubt sein. Kunsthistorisch geschulten Augen erinnert die Bilderkennung mittels Computer Vision auf der Basis von CNN und ihren Schichten nämlich nicht allein an die Schichten der visuellen Wahrnehmung, sondern auch an das Stufenmodell der ikonologischen Interpretation, wie es durch den Kunsthistoriker Erwin Panofsky entwickelt wurde.⁵⁴ Auch dieses geht von Form und Farbe zu Semantik.⁵⁵ Während die erste und teilweise auch Teile der zweiten Stufe dieses Modells bereits erfolgreich in Verfahren wie ‚Object Detection‘ (Objekterkennung) und ‚Image Classification‘ (Bildklassifizierung) digitalisiert wurden, ist die dritte Ebene der Ikonologie, also die Erforschung von Kontexten, Bedeutungen, Verborgenen, Impliziten, Topoi, Mythen, Symbolen und vor allem Historischem, bislang nicht in den Schichten der CNN erschienen. Sind Bildforscher*innen also in ihrem Alleinanspruch als Ikonolog*innen sicher, wenn Rechnen versus Hermeneutik immer noch eine zugunsten der menschlichen Expert*innen zu beantwortende Frage ist? Wenn also menschliche Expert*innen Bilder immer noch in anderer Weise „prozessieren“ als Maschinen? Die einfache Antwort wäre, dass der digitale Blick der formal-quantitative Blick sei, der Blick der Bildforscher*innen jedoch qualitativ-hermeneutisch verfare. Wer jedoch in der Praxis mit ‚mixed methods‘ operiert, stellt fest, dass eine solche Trennung sehr oft nicht aufrechterhalten werden kann. Qualitative und quantitative Methoden mischen sich, die Archive und Bilddatenbanken basieren bereits auf Taxonomien, auf beiden Seiten wird gezählt, sortiert und verglichen, werden Bilder kategorisiert und in Texte übertragen.

Es ist aber eine andere Schere, die in der Mischung der Methoden aufgeht und die Claus Pias in einem Artikel bereits 2003 beschrieb: „[...] plötzlich [entsteht] durch Geräte ein Wissen von Bildern, zu dem die Kunstgeschichte keine Frage formulieren kann, auf die dieses Wissen eine Antwort wäre, ein Wissen, das einfach vom kunsthistorischen Diskurs nicht verarbeitbar ist.“⁵⁶ Das mittels Computer Vision erzeugte Wissen arbeitet mit Ähnlichkeiten, die bislang nicht von Interesse waren und findet Muster und Ordnungen, die bislang keine Beachtung fanden. Antworten, die nicht die Frage waren, fordern die Bildwissenschaft auch methodisch

che Auge im ersten neuronalen Netz von *Perceptron* bis zur „Multi-Layered Perceptron“ einer CNN. Auch er nennt die letzte Stufe „statistical ‚eye“.

⁵⁴ Margarete Pratschke hat den Einfluss von Panofsky ganz konkret auf die Digitalen Geisteswissenschaften in einem Artikel herausgestellt, in dem sie beschreibt, wie über dessen ehemaligen Schüler Edmund A. Bowles wichtige Schritte in diese Richtung gemacht wurden, als dieser Leiter des Humanities-Programms bei IBM wurde. Vgl. Margarete Pratschke: Wie Erwin Panofsky die digital humanities erfand. Für eine Geschichte und Kritik digitaler Kunst- und Bildgeschichte, in: kritische berichte, Heft 3 (2016), S. 53-62.

⁵⁵ Die Frage, was aus Kunstgeschichte unter digitalen Bedingungen wird, wurde bereits an vielen Orten diskutiert. Stefan Heidenreich hat sehr früh darauf hingewiesen, dass insbesondere die grundlegende Unterscheidung von Bildern des Kunsthistorikers Wölfflin in malerisch/haptisch (oder linear) erfolgreich von Computern zu detektieren ist und für die Bildsortierung verwendet werden kann. Stefan Heidenreich: Form und Filter – Algorithmen der Bilderverarbeitung und Stilanalyse, in: zeitenblicke 2 (2003), Nr. 1.

⁵⁶ Claus Pias: Maschinen/lesbar. Darstellung und Deutung mit Computern, in: Matthias Bruhn (Hg.): Darstellung und Deutung in der Kunstgeschichte (visual intelligence, Bd.1), Weimar: Verlag und Datenbank für Geisteswissenschaften 2000, S. 125-144.

heraus. Wer Bildforschung unter digitalen Bedingungen betreiben möchte, sieht sich gezwungen, mit diesen Antworten umzugehen.

Als Aufgabe für die Medienwissenschaft im Feld digitaler Methoden sehe ich die Analyse von maschinellem Lernen, einerseits auf der Ebene einer kritischen Historiographie und Epistemologie, wie sie bereits in Teilen vorliegt.⁵⁷ Hier eröffnet die Medienarchäologie einen Blickwinkel auf digitale Methoden, der historisch weiter zurückreicht als bis zum technikhistorischen Computer. Ich sehe jedoch andererseits auch großes Potential darin, durch den Methodenwechsel auf Experimente mit CNN andere Zugänge zu einem Verständnis von ‚Künstlicher Intelligenz‘ zu erlangen. Beide Wege – eine historiografisch und analytisch verfahrenende Medienkritik sowie ein explorativer Umgang mit KI eröffnen Zugänge jenseits der Mythisierung einer gottgleichen ‚Künstlichen Intelligenz‘, wie sie zurzeit zu beobachten ist. In beiden Fällen muss es darum gehen, die für diese Technik so wichtigen Analogieschlüsse und die Metaphorik von Mensch und Maschine in ihrer Ambivalenz zwischen Produktivität und Verdunkelung zu hinterfragen, zumal maschinelles Lernen die Beziehung von Menschen und Maschinen selbst grundlegend in Frage stellt. Ein Forschungsthema, das ich hieraus anregen würde, besteht aber auch darin, die Datenmüllplätze und Festplatten gescheiterter DH-Projekte in den Blick zu nehmen; es muss eine große Müllhalde sein.

Literaturverzeichnis

Allington, Daniel/Brouillette, Sarah/Columbia, David: Neoliberal Tools (and Archives): A Political History of Digital Humanities, Los Angeles Review of Books (Mai 2016), open access.

Bader, Lena/Gaier, Marin/Wolf, Falk (Hg.): Vergleichendes Sehen, München: Fink 2010.

Benjamin, Walter: Das Kunstwerk im Zeitalter seiner Reproduzierbarkeit. Frankfurt a.M.: Suhrkamp 2007.

Bense, Max: AESTHETICA. Einführung in die neue Ästhetik, Baden-Baden: Agis 1965.

Bense, Max: Projekte generativer Ästhetik, in: Computer-Grafik, hrsg. Max Bense/Elisabeth Walther, Stuttgart: edition rot 1965, 11.

Bexte, Peter: Blinde Seher. Wahrnehmung von Wahrnehmung in der Kunst des 17. Jahrhunderts, Amsterdam/Dresden: Verlag der Kunst 1999.

Caliskan, Aylin/Joanna J. Bryson/Arvind Narayanan: Semantics derived automatically from language corpora contain human-like biases, Science 356 (2017), 183–186.

Cortjaens, Wolfgang/Karsten Heck (Hg.): Stil-Linien diagrammatischer Kunstgeschichte (Transformationen des Visuellen, Band 2), Berlin/München: Deutscher Kunstverlag 2014.

Crary, Jonathan: Techniken des Betrachters. Sehen und Moderne im 19. Jahrhundert, Dresden/Basel: Verlag der Kunst 1996.

⁵⁷ Siehe z.B. Bernhard Dotzler: Down-to-earth resolutions. Erinnerungen an die KI als eine »häretische Theorie«, in: Engemann/Sudmann 2018, S. 39-54 sowie andere Beiträge in diesem Band.

Daston, Lorraine/Peter Gallison: Objektivität, Frankfurt a. M.: Suhrkamp 2008.

Diderot, Denis: Brief über den Blinden, zum Gebrauch für die Sehenden von 1749, in: Denis Diderot: Philosophische Schriften, hrsg. v. Alexander Becker, Frankfurt/M.: Suhrkamp 2013, 49-99.

Dotzler, Bernhard: Down-to-earth resolutions. Erinnerungen an die KI als eine »häretische Theorie«, in: Engemann, Christoph/Sudmann, Andreas (Hg.): Machine Learning. Medien, Infrastrukturen und Technologien der künstlichen Intelligenz, Bielefeld: transcript 2018, 39-54.

Engemann, Christoph/Andreas Sudmann (Hg.): Machine Learning. Medien, Infrastrukturen und Technologien der künstlichen Intelligenz, Bielefeld: transcript 2018.

Engemann, Christoph/Sudmann, Andreas: Zur Einführung. Medien, Infratsrukturen und Technologien des maschinellen Lernens, in: Engemann, Christoph/Sudmann, Andreas (Hg.): Machine Learning. Medien, Infrastrukturen und Technologien der künstlichen Intelligenz, Bielefeld: transcript 2018, 9-36.

Flusser, Vilém: Ins Universum der technischen Bilder, Göttingen: European Photography 1996 [1985].

Friedländer, Max: Von Kunst und Kennerschaft, Leipzig 1992 [1946].

Fucks, Wilhelm: Nach allen Regeln der Kunst, Stuttgart: Deutsche Verlagsanstalt 1968.

Heck, Karsten: Formen des Stils. Heinrich von Geymüllers grafische Methoden der Stilgeschichts-schreibung, in: Cortjaens, Wolfgang/Heck, Karsten (Hg.): Stil-Linien diagrammatischer Kunstgeschichte (Transformationen des Visuellen, Band 2), Berlin/München: Deutscher Kunstverlag 2014, 168-185.

Heck, Karsten: Der Schreibtisch als Denkraum. Heinrich von Geymüllers Tableau Graphique, in: Heinrich von Geymüller (1839-1909). Architekturforscher und Architekturzeichner, hrsg. v. Josef Ploder und Georg Germann, Basel: Friedrich Reinhardt 2009, 54-63.

Heidenreich, Stefan: Form und Filter - Algorithmen der Bilderverarbeitung und Stilanalyse, in: zeitenblicke 2 (2003), Nr. 1.

Heidenreich, Stefan/Ernst, Wolfgang/ Holl, Ute (Hg.): Suchbilder. Visuelle Kultur zwischen Algorithmen und Archiven. Berlin: Kadmos 2003.

Kheradpisheh, Saeed Reza et al: Deep Networks Can Resemble Human Feed-forward Vision in Invariant Object Recognition, Nature (Sept. 2016).

Kittler, Friedrich: Grammophon, Film, Typewriter, Berlin: Brinkmann & Bose 1986.

Koelke, Manuela/Pasquinelli, Matteo: Neuronale Netzwerke und das Unberechenbare. Über die Automatisierung von Intelligenz als statistische Interferenz, Manuskript 2018.

LaFrance, Adrienne: When Robots Hallucinate. What do Google's trippy neural network-generated images tell us about the human mind? in: The Atlantic, September 2015. <https://www.theatlantic.com/technology/archive/2015/09/robots-hallucinate-dream/403498/>, (Zugriffsdatum: 11.06.2019).

Manovich, Lev: Media Analytics & Gegenwartkultur, in: Engemann, Christoph/Sudmann, Andreas (Hg.): Machine Learning. Medien, Infrastrukturen und Technologien der künstlichen Intelligenz, Bielefeld: transcript 2018, 269-288.

Marey, Étienne-Jules: La méthode graphique dans les sciences expérimentales – La circulation du sang à l'état physiologique et dans les maladies, Paris: G. Masson 1878.

Marr, David: Vision. A Computational Investigation into the Human Representation and Processing of Visual Information, Cambridge, Mass./London: MIT Press 1982.

Osinga, Douwe: What do computers see? Tricks to visualize and understand how neural networks see, O'Reilly (August 2018), <https://www.oreilly.com/ideas/what-do-computers-see> (Zugriffsdatum: 11.06.2019).

Otis, Laura: Networking. Communicating with Bodies and Machines in the Nineteenth Century (Studies in Literature and Science), Ann Arbor: Univ. of Michigan Press 2001.

Parisi, Luciana: Das Lernen lernen oder die algorithmische Entdeckung von Information, in: Engemann, Christoph/Sudmann, Andreas (Hg.): Machine Learning. Medien, Infrastrukturen und Technologien der künstlichen Intelligenz, Bielefeld: transcript 2018, 91-113.

Pasquinelli, Matteo: Augmented Intelligence Traumas, Lüneburg: Meson Press 2015.

Pasquinelli, Matteo: Machines that Morph Logic: Neural Networks and the Distorted Automation of Intelligence as Statistical Inference (zitiert aus der deutschen Übersetzung: Neuronale Netzwerke und das Unberechenbare. Über die Automatisierung von Intelligenz als Statische Interferenz), Glass-Bead, Site 1: Logic Gate, the Politics of the Artificial Mind (2017), open access.

Pias, Claus: Maschinen/lesbar. Darstellung und Deutung mit Computern, in: Matthias Bruhn (Hg.): Darstellung und Deutung in der Kunstgeschichte (visual intelligence, Bd.1), Weimar: Verlag und Datenbank für Geisteswissenschaften 2000, 125-144.

Plessner, Maximilian: Ein Blick auf die grossen Erfindungen des 20. Jahrhunderts, Heft 1: Die Zukunft des elektrischen Fernsehens, Berlin: Dümmler 1892.

Pratschke, Margarete: Wie Erwin Panofsky die digital humanities erfand. Für eine Geschichte und Kritik digitaler Kunst- und Bildgeschichte, in: Kritische Berichte, Heft 3 (2016), 53-62.

Priestley, Joseph: A description of a new chart of history, containing a view of the principal revolutions of Empire, that have taken place in the world, London: J. Johnson 1770.

Schneider, Birgit: Maximilian Plessner: Die Zukunft des elektrischen Fernsehens 1892, in: Daniels, Dieter / Naumann, Sandra (Hg.): See this Sound. Adiovisiology II, Essays. Histories and Theories of Audiovisual Media and Art, Linz/Leipzig: Verlag der Buchhandlung König 2011, 193-195.

Schneider, Birgit: Von hörenden Augen und sehenden Ohren: Medienästhetik unterschiedlicher Verhältnisse von Ton und Bild, in: Daniels / Naumann (Hg.): See this Sound. Adiovisiology II, Essays. Histories and Theories of Audiovisual Media and Art, Linz/Leipzig: Verlag der Buchhandlung König 2011, 174-190.

Semper, Gottfried: Der Stil in den technischen und tektonischen Künsten oder praktische Ästhetik. Ein Handbuch für Techniker, Künstler und Kunstfreunde, Bd. 1: „Die Textile Kund für sich betrachtet und in Beziehung zur Baukunst“, München: Bruckmann's 1878 [1860].

Shaji, Appu: Beauty & The Algorithm, in: Machina. A Curation of Real Photography by a Machine, EyeEm Magazine (November 2016).

Steyerl, Hito: A Sea of Data: Apopenia and Pattern (Mis-) Recognition, in: e-flux, Nr. 72, April 2016, open access.

Von Uexküll, Jakob: Umwelt und Innenwelt der Tiere, Berlin: Springer 1909.

Winckel, Fritz Wilhelm: Technik und Aufgaben des Fernsehens, Berlin: Rothgiesser & Diesing, 1930.

Abbildungsverzeichnis

Abb. 1a-c: Drei Tests des Eye-Em Algorithmus mit eigenen Bildern, dem Foto einer Windkraftanlage, einer Puppe, einem Gemälde von Josef Albers und TV-Rauschen. Die automatisch vorgeschlagenen Bildunterschriften zeigen, wie gut der Algorithmus auf übliche Fotografien trainiert ist. Abstrakte Muster und Puppen werden nicht erkannt. Sie lauten: Windrad auf einer Wiese gegen den Himmel; Porträt eines süßen Babys; Vollbildaufnahme einer roten Telefonzelle; Vollbildaufnahme eines abstrakten Musters. Screenshots © Eigene Grafik, 2019.

Abb. 2: Joseph Priestley: Historische Weltkarte. Visual Analytics-Projekt aus dem 18. Jahrhundert. Staatsbibliothek Berlin – Preußischer Kulturbesitz.

Abb. 3: Wilhelm Fucks: Nach allen Regeln der Kunst, dva Stuttgart, 1968. Farben, geordnet nach fallenden Häufigkeiten am Beispiel eines bereits digitalisiert vorliegenden Bildes, nämlich des Mosaiks im Baptisterium der Kathedrale Hl. Petrus, Ravenna.

Abb. 4: Stilgeschichte als Funktion von Zeit und Stilkoeffizienten nach Semper? Die Kunstlehre durchwandere „das Feld der Geschichte, die Kunstwerke der verschiedenen Länder und Zeiten nicht als Thatsachen auffassend und erklärend, sondern sie gleichsam entwickelnd, in ihnen die nothwendig verschiedenen Werthe einer Funktion, die aus vielen variablen Coëffizienten besteht, nachweisend“. Skizze © Eigene Grafik, 2019.

Abb. 5: Heinrich von Geymüller: Perioden und Phasen des Renaissance-Stils in Frankreich, 1890er Jahre. Quelle: Heinrich von Geymüller (1839-1909). Architekturforscher und Architekturzeichner, hrsg. Josef Plode und Georg Germann, Basel: Friedrich Reinbrardt, 2009.

Abb. 6: Fritz Wilhelm Winckel: Visualisierung von Musik mit Hilfe einer Nipkowscheibe. Aus einem Artikel der Zeitschrift: Fernsehen, Nr. 4, 1930.

Abb.7: Doppelseite aus Laurence Sternes Roman „The life and opinions of Tristram Shandy“ (1762).

Abb. 8 und 9: Wahrnehmung der Wahrnehmung. Camera Obscura als Modell des Sehens und Blinder mit zwei Taststöcken, aus René Descartes, Dioptrique, in: Discours de la méthode, Paris 1724.

Abb. 9: Nachbilder, aus: Jan Evangelista Purkinje: Beobachtungen und Versuche zur Psychologie der Sinne, 1823.

Abb. 10: Rotierende Scheiben zur Testung der Sehgeschwindigkeit von schnell wiederholten Lichteindrücken, aus Hermann von Helmholtz: Handbuch der physiologischen Optik, Leipzig 1867.

Abb. 11: Fotos von Fliegenpilzen in unterschiedlichen Formen und Perspektiven aus der Bild-datenbank ImageNet. Diese stellt Fotografien für das Training eines neuronalen Netzes zur Objekterkennung auf Bildern zur Verfügung. Screenshot © Eigene Grafik, 2019.

Abb. 12: Ansicht des CNN-Fensters, in diesem Fall mit einer Convolutional Unit, die Autoräder erkennt. <https://github.com/yosinski/deep-visualization-toolbox>: Objekt-Klassifizierung mit der TensorFlow Object Detection API, Google, Open Source, 2016, https://github.com/tensorflow/models/tree/master/research/object_detection (Zugriffsdatum: 11.06.2019).

Abb. 13: Aus dem Artikel von Matthew D. Zeiler und Rob Fergus: Visualizing and Understanding Convolutional Networks, Springer International Publishing Switzerland, 2014.

Abb. 14: Objekt-Klassifizierung mit der TensorFlow Object Detection API, Google, Open Source, 2016, https://github.com/tensorflow/models/tree/master/research/object_detection (Zugriffsdatum: 11.06.2019).

Abb. 15: „that way we get an idea of the sort of features that each layer is looking out for ... The lowest layers are looking for colors and simple patterns; if you go higher, the patterns become more complex.“ What do computers see? Tricks to visualize and understand how neural networks see. By Douwe Osinga, O'Reilly, August 15, 2018. <https://www.oreilly.com/ideas/what-do-computers-see> (Zugriffsdatum: 11.06.2019).

Abb. 16: Journey on the Deep Dream. Roger Luedecke, 2015. Deep Dreams aus dem Computer oder doch eher die Tiere der Bilddatenbanken, mit denen die KI sehen lernt? <https://www.youtube.com/watch?v=SCE-QeDfXtA> (Zugriffsdatum: 11.06.2019).