

Francis Hunger

Punktwolken. Scatterplots und Tabellen
als User-Interfaces Künstlicher ‚Intelligenz‘.

Training the Archive – Working Paper Series 5

Francis Hunger – Punktwolken. Scatterplots und Tabellen als User-Interfaces Künstlicher
,Intelligenz‘.

Training the Archive (Hrsg.), Aachen/Dortmund, January 2023

DOI: [10.5281/zenodo.7554463](https://doi.org/10.5281/zenodo.7554463)

Training the Archive



Ludwig Forum für Internationale Kunst Aachen

Jülicher Str. 97-109, 52070 Aachen

<http://ludwigforum.de/>

**Ludwig
Forum**

für Internationale Kunst
Aachen

Hartware MedienKunstVerein

Büro: Hoher Wall 15, 44137 Dortmund

www.hmkv.de

HMKV
Hartware MedienKunstVerein

Dieses Working Paper ist lizenziert unter der Creative Commons Attribution-NonCommercial 4.0 International License (CC-BY-NC 4.0): <https://creativecommons.org/licenses/by-nc/4.0/>.

Gefördert im Programm *Kultur Digital* der Kulturstiftung des Bundes

Ku/tur
Digita/

KULTURSTIFTUNG
DES
BUNDES

Gefördert von der Beauftragten der Bundesregierung für Kultur und Medien



Die Beauftragte der Bundesregierung
für Kultur und Medien

Working Paper 5: Punktwolken. Scatterplots und Tabellen als User-Interfaces Künstlicher ‚Intelligenz‘.

„Kategorien sind zentral für das In-der-Welt-Sein. Big Data schafft Kategorien keineswegs ab“ (Bowker 2014, 1797).

„Eine Kulturtechnik befördert die Leistungen der Intelligenz durch Versinnlichung und exteriorisierende Operationalisierung des Denkens“ (Krämer und Bredekamp 2003, 18).

Abstract

Scatterplots und tabellarische Strukturen sind die wichtigsten grafischen User-Interfaces für die diagrammatische Darstellung großer Bilddatensammlungen, die mit visuellen Erkennungswerkzeugen verarbeitet wurden. In dieser Studie werden verschiedene Medienvisualisierungen als Fallbeispiele diskutiert: ARTigo (LMU München), Imgs.ai (UC Santa Barbara/Bildarchiv Foto Marburg), iArt (Universitäten München und Hannover), Vikus Viewer (FH Potsdam), und X Degrees of Separation (Google). Darüber hinaus wird untersucht, wie diese Projekte die Visualisierungsalgorithmen PCA, t-SNE und UMAP im Verhältnis zu künstlichen gewichteten Netzen wie VGG19 oder CLIP einsetzen und es werden praktische Vorschläge für einen Umgang mit den identifizierten Problemen gemacht.

1 Herleitung

Dieses Working Paper dient der Evaluierung von Visualisierungskonzepten für The Curator’s Machine, ein Softwaremodul maschinellen Lernens, welche im Zuge des Forschungsprojektes Training the Archive entwickelt wird. The Curator’s Machine soll mit Hilfe von Verfahren der Mustererkennung und Künstlicher ‚Intelligenz‘ kuratorische Entscheidungen für große Bildmengen automatisieren (Bönisch 2021; Hunger 2021a; Hunger 2021b).

Die Ergebnisse von Machine-Learning-Verfahren werden in User-Interfaces angeordnet, welche auf diagrammatischen Prinzipien basieren. Ausgehend von der These, dass räumliche Verhältnisse die Wissensbildung in Diagrammen dominieren, ist die Aufgabe des vorliegenden Textes eine Untersuchung von Visualisierungen durchzuführen, die im Machine Learning eingesetzt werden. So entwickelt der Text eine Syntax der Visualisierungen, die mehrere für Training the Archive wesentliche diagrammatische Typen, die Tabelle, Skalendiagramm und Scatterplot und das K-Nearest-Neighbors-Verfahren identifiziert. Ein nächster Abschnitt beschreibt drei Schichten der Reduktion und Rekonstruktion von Wirklichkeit in

Visualisierungen, die auf gewichteten Netzen (Convolutional Neural Networks)¹ aufbauen. Zu diesen drei Schichten zählen 1.) die Datenerhebung, 2.) Statistik/Computation und 3.) die Visualisierung. Auf diese einleitenden Überlegungen folgt ein zentraler Abschnitt mit Fallbeispielen, die ich aufgrund ihrer inhaltlichen und technologischen Nähe zu unserem Vorhaben gewählt habe. In zahlreichen Abbildungen der User-Interfaces werden ARTigo, eine kunsthistorische Suchmaschine (Universität München), die experimentelle kunsthistorische Bildersuche *Imgs.ai* (UC Santa Barbara/Bildarchiv Foto Marburg), das interaktive Werkzeug zur Analyse von Bilddatenmengen *iArt* (Universitäten München und Hannover) sowie das Projekt *Vikus Viewer* mit kuratierten Datenmengen (FH Potsdam) und abschließend das pfadorientierte Bildexperiment *X Degrees of Separation* (Google) untersucht. Abschließend werden Schlüsse für den Prototypen von *Training the Archive* gezogen, die für ähnlich gelagerte User-Interfaces generalisierbar sind. Bevor der Text die Syntax von Medienvisualisierungen erkundet, gilt es zwei hier relevante Konzepte einführend zu verorten: Diagramme und Medienvisualisierungen.

1.) Diagramme werden als visuelle Argumente verwendet. Sie dienen der Visualisierung von Sachverhalten, indem grafische Elemente in räumliche Verhältnisse gesetzt werden, die ‚gelesen‘ werden können. Oft sind Diagramme Hybride aus Text und Grafikelementen. Die Stärke von Text-Grafik-Hybriden liegt darin, dass sie Zahlen und Kategorien, also quantitative und kategoriale Informationsobjekte unterschiedlicher Wissensdomänen in einem Schaubild vereint darstellen können. Diese Vereinigung in der Fläche ist eines der wichtigsten Wissenswerkzeuge der Menschheit (Krämer 2010). Ziele der Interfaces sind es, Hypothesen zu validieren oder falsifizieren und neue Forschungsfragen zu provozieren. Es lohnt sich daher, zu beobachten, wie diese Ziele durch den Versuchsaufbau selbst – die Interfaces und deren diagrammatische Operation – beeinflusst werden.

Menschliches Lesen, Sehen und Verstehen von Diagrammen basiert auf visuellen Hierarchien. In einer solchen visuellen Hierarchie lese ich ‚oben‘ als besser im Vergleich zu ‚unten‘. Erstens kommt vor zweitens, und zwar von oben nach unten. Links lese ich vor rechts, denn am linken Bund hat das Auge Halt. Die hier aufgezählten Hierarchien sind Beispiele aus dem Sprachraum des Autors. In anderen kulturellen und historischen Zusammenhängen sind die Hierarchien andere, wie beispielsweise im Arabischen oder Hebräischen, die von rechts nach links gelesen werden. Eines ist den Kulturtechniken des verstehenden Lesens und Schreibens gemein: Die räumliche Organisation des Aufgezeichneten und die oft unausgesprochenen, kulturtechnisch erlernten Hierarchien und Ordnungen ziehen sich durch das Lesen und das Gelesene.

Anhand von Beispielen aus dem Feld der Künstlichen ‚Intelligenz‘, der Computer-Vision und der Datenvisualisierung untersuche ich im Folgenden jene diagrammatischen Darstellungen, welche Bildformationen als visuelles Argument zugänglich machen. Die vorliegende Studie versteht sich daher als Teil von Interface-Kritik, deren Aufgabe es ist, Interfaces nicht

¹ Um den anthropomorphisierenden Begriff ‚Neural Network‘ zu vermeiden, verwende ich alternativ ‚gewichtetes Netzwerk‘. Hierzu und generell zu den technologischen Verfahren des Machine Learning in ‚Convolutional Neural Networks‘ siehe (Hunger 2021a, 2–7, 9 und FN 1).

allein als Übergang zwischen ‚dem Innen‘ und ‚dem Außen‘, sondern als „eigene Zonen von Aktivität“ aufzufassen (Galloway 2012, vii).

2.) Die hier diskutierten visuellen Verfahren sind „Medienvisualisierungen“ (Manovich 2020, 215), denn sie erforschen visuelle Muster in großen Datenmengen. Die räumlichen Anordnungen der Datenobjekte erzeugen visuelle Argumente. Doch am wichtigsten für das Konzept der Medienvisualisierungen: Die Datenobjekte werden nicht abstrahiert dargestellt, zum Beispiel als Punkte, Grapheme oder durch numerische Repräsentation, sondern als Icons, das heißt als verkleinerte, visuelle Referenzen auf die ursprünglichen Datenobjekte. Medienvisualisierungen zeichnen sich zudem durch Interaktivität und die Verfügbarkeit verschiedener Skalierungen aus: So kann in vielen User-Interfaces in die Bildsammlungen beziehungsweise in einzelne Bilder hereingezoomt werden. Dieses beständige Changieren zwischen Detail und Überblick erlaubt das Entstehen neuen Wissens in immer neuen visuellen Konfigurationen.

In den User-Interfaces, welche Bildsammlungen darstellen, zum Beispiel auf Websites oder KI-Software, werden die Bilddatensätze algorithmisch zu An-Ordnungen formiert. In diesen Mengen verschwindet das einzelne Bild zugunsten der Übersicht. Nicht mehr die Referenzialität des einzelnen Bildes ist von Interesse, sondern die Zusammenschau der Relationen der Bilder untereinander. Jedoch bringt die Darstellung massenhafter Bilder, die nunmehr als operationale Bilder auftreten, ganz eigene Problematiken mit sich, die in den Fallbeispielen aufgespürt werden sollen.

1.1 Syntax von Medienvisualisierungen

Im Folgenden werden die Visualisierungstypen ‚Tabellen‘, ‚Skalen-Diagramme und Scatterplot‘ und ‚K-Nearest-Neighbor‘ diskutiert, die für Medienvisualisierungen häufig Verwendung finden. An ihnen lassen sich bestimmte Stärken und Schwächen der Visualisierungen zur Wissensgenerierung identifizieren. Eine Reihe weiterer Visualisierungen, Netzwerkdiagramme, Karten, Infografiken und weitere werden an dieser Stelle nicht diskutiert, da sie auf die Visualisierungen im Rahmen von The Curator’s Machine nicht zutreffen.

Tabellen

Tabellen schaffen Überblick und erlauben das Eintragen, Nachschlagen und die Operationalisierung der Einträge. Durch den Eintrag in Zeilen und Spalten sowie die Operationen des Gruppierens, Sortierens und Summierens werden die Daten in immer neue Wissenskonstellationen gebracht, sodass sie Menschen als Denkwerkzeug zum Schaffen neuen Wissens dienen. Es kann zwischen statistischen, mathematischen, transaktionalen beziehungsweise prozessualen und wissensbildenden Tabellen unterschieden werden.

Statistische Tabellen verdanken ihren Aufstieg der Entwicklung der Nationalstaaten ab dem 17. Jahrhundert, in denen die Staatsstrukturen systematisch durch die Verzeichnung von Bevölkerung, Produktion und den Haushalt hinterlegt wurden. Hierfür wurden Daten intentional produziert, indem Beobachtung gesammelt und in tabellarische Raster übertragen wurden.

Mathematische Tabellen umfassen zum einen die aus empirischer Beobachtung gewonnenen Daten, zum Beispiel den Orbit von Himmelskörpern, und zum anderen berechnete Zahlenreihen, die für weitere Berechnungen verwendet werden konnten, zum Beispiel Quadratzahlen, Zins und Zinseszins, Logarithmen und trigonometrische Funktionen. Mathematische Tabellen wurden durch die Entwicklung von Rechenmaschinen und Computern abgelöst, welche diese Berechnungen ad hoc ermöglichen beziehungsweise in Tabellenkalkulationen und Datenbanken ablegen.

Wissensbildende Tabellen versammeln Wissen als Überblick. Sie erlauben es, Wissen zum Beispiel über kunstgeschichtliche, weltliche-historische oder politische Ereignisse chronologisch oder kategorial zu ordnen, und dienen einerseits dem Auswendig-Lernen und andererseits dem Generieren neuen Wissens. Sie knüpfen an die neuen Wissensordnung der entstehenden Museen, Akademien und Enzyklopädien ab dem 16. Jahrhundert an, welche unsere heutige Wissenschaft und Wissensordnungen begründeten.

86 SHOP MANAGEMENT

Machine shop
 Order for Tires.....
 Do work on Tire No
 As follows and per blue print

	Tem- plet	Size to be cut to	Depth of cut	Driving belt	Feed	Rate	Time this operation should take
Surface to be ma- chined
Set tire on machine ready to turn....
Rough face front edge
Finish face front edge
Rough bore front...
Finish bore front...
Rough face front I. S.C.....
Cut out filled
Rough bore front I. S.F.....
Rough face back edge
Finish face back edge
Finish bore back
Rough bore back
Rough face back I. S.F.....
Cut out filled
Cut recess
Rough turn thread..
Finish turn thread..
Rough turn flange..
Finish turn edge....
Clean fillet of flange.
Remove tire from machine and clean face plate

FIGURE 1. — TIRE-TURNING INSTRUCTION CARD

Abb. 1: Transaktionstabelle zur Produktion von Stahlreifen für Fahrzeuge in den Midvale Steel Works 1883. Die jeweilige Transaktion kann über die Order-Nr. adressiert werden. Die erste Spalte gibt den Ablauf der Transaktion von Beginn bis Ende vor (Taylor 1903, 86).

Transaktionale Tabellen sind grundlegende Management- und Prozesswerkzeuge geworden (beispielhaft siehe Abb. 1). Sie repräsentieren aktuelle Zustände als Transaktionen und wurden historisch beispielsweise für die Änderung von Kontenständen bei Banken, zur internen Arbeitsorganisation in Betrieben oder für die koordinierende Abwicklung von Dienstleistungen, wie zum Beispiel die Kombination von Flug- und Passagierdaten bei Fluggesellschaften eingesetzt. Die in ihnen festgehaltenen Transaktionsdaten beschreiben den jeweils aktuellen Zustand eines Gegenstandes oder Verfahrens und dienen der kontrollierten Abwicklung eines Vorganges von seinem definierten Anfang bis zu seinem definierten Ende. Tabellen zählen zu den Kulturtechniken der Wissensproduktion, ihr Gebrauch muss erlernt werden.

Folgende visuelle Argumente prägen die Tabelle: Erstens schafft das Tabellenraster ein Konstrukt, in welches Daten überhaupt eingetragen werden können. Zweitens springt das lesende Auge zwischen Zeilen und Spalten, um die einzelnen Informationsobjekte (die Zeile) nach verschiedenen Kategorien (je Spalte) zu überblicken. Drittens verweist Leerraum auf fehlende und noch zu erzeugende Daten und viertens verändern sich die räumlichen Verhältnisse innerhalb der Tabelle durch Neuordnung (Filtern, Ordnen, Gruppieren) der Daten im Raster (Hunger 2022, 76–128).

Tabellen etablieren einen vereinheitlichenden, gemeinsamen Denkraum, denn sie erlauben sowohl kategorial/qualitative als auch quantitative Daten in einem einheitlichen Raster zu kombinieren.

Skalen-Diagramme und Scatterplots

Skalen-Diagramme, zum Beispiel Balkendiagramme oder Scatterplots, bestehen aus grafischen Elementen (Linien, Rechtecken bzw. Balken, Punkten), die in einem zweidimensionalen Raum verteilt sind, welcher durch zwei Skalen mit quantitativen oder qualitativen Werten (X-Achse und Y-Achse) aufgespannt wird. Sie erlauben die Darstellung von Zeitreihen, Rankings, prozentuale Aufteilung, Abweichungen im Verhältnis zum Null-Wert und Verteilungen der Gesamtwerte (z. B. Glockenkurve als Normalverteilung). Üblicherweise verfügen die Skalen über gleichbleibende Abstände, sodass ein Raster entsteht (Wilke 2019, 13–25). Die visuellen Argumente von Skalen-Diagrammen orientieren sich an dem wechselnden Betrachten der Skalen und der als Grafiken eingetragenen Werte. Die grafischen Elemente erlauben als visuelle Relationen eine Reihe von Mustern zu identifizieren: Schwerpunkte, Verteilungen, Trends und Tendenzen, Vergleiche. Elemente können durch Farben gruppiert werden und somit jenseits der zweidimensionalen Verteilung zusätzlich dimensioniert werden (Tufté 1983, 28–39; Few 2005; Few 2012, 87–136).

Aus dem Erfahrungswissen des Grafikdesigns hat sich eine Reihe von Gebrauchsweisen von Diagrammen herausgeschält: So lassen sich Mengen am ehesten durch Balkendiagramme vergleichen. Sollen nicht allein numerische, sondern auch kategoriale Qualitäten verglichen werden, können gruppierte und gestapelte Balkendiagramme verwendet werden. Verteilungen werden bevorzugt als Histogramme, Density Plots oder violinenförmige Plots visualisiert. Für Proportionen, die sich an der Verwendung von Prozentangaben erkennen lassen, können Tortendiagramme und Balkendiagramme eingesetzt werden. Relationen sind als

Scatterplots, Bubble Chart und Slopegraph, als Matrix oder Korrelogramm darstellbar. Für geografische Daten eignen sich Karten und auch geografische Heatmaps (Wilke 2019, 37–44).

Zu den Skalen-Diagrammen zählen prinzipiell auch Scatterplots, welche Verteilungen innerhalb eines Datensatzes zeigen. Warum prinzipiell? Wie wir später sehen werden, entledigen sich die hier untersuchten User-Interfaces der Skalen. Damit einher geht auch ein Informationsverlust, den Bedeutungen, die bis dahin in den Skalen aufgehoben waren, werden unklarer. Bevor jedoch darauf eingegangen werden soll, wenden wir uns den ‚klassischen‘, also den mit Skalen versehenen Scatterplots zu.

Mittels Scatterplots wird das Augenmerk auf die Korrelationen, Häufungen, Muster, Trends und Ausreißer in einer Menge von Datenpunkten gelegt, indem zueinander Zugehöriges als Verdichtung, räumlich nah beieinander organisiert wird. „Zentral für den Großteil dieser Techniken ist es, die Bedeutung der Entfernung zwischen Objekten als Gradmesser für die Ähnlichkeit aufrecht zu erhalten“ (Sarıkaya und Gleicher 2018, 2). Dabei bedeutet Ähnlichkeit, dass sich mehrere Datenobjekte in mindestens einem Merkmal gleichen, wobei einerseits eine Gesamtähnlichkeit adressiert werden kann, andererseits eine partielle Ähnlichkeit, in der sich Datenobjekte in einzelnen Aspekten ähneln. Die visuellen Features von Scatterplots sind jedoch keine vollständigen ‚Beweise‘ im mathematischen Sinne, sondern dienen als Proxies für Korrelationen.

Scatterplots wurden definiert als „eine Darstellung von zwei Variablen, x und y , die unabhängig voneinander gemessen werden, um bivariate Paare (x, y) zu bilden, und die als einzelne Punkte auf einem Koordinatengitter dargestellt werden, das in der Regel durch eine horizontale und eine vertikale Achse definiert ist, wobei es keine notwendige funktionale Beziehung zwischen x und y gibt“ (Friendly und Denis 2005, 105). Anders gesagt: Der Zusammenhang von Daten ist in einer Tabelle oder einem Balkendiagramm dann gut erkennbar, wenn x und y miteinander linear verknüpft sind. Wenn aber – wie in den unten besprochenen Praxisbeispielen – die Funktion zwischen x und y mathematisch komplex ist, oder wenn x an mehrere andere Parameter geknüpft ist, dann kann dieser Zusammenhang in einer tabellarischen Darstellung nur schwer ‚gelesen‘ werden. Hier sind Scatterplots im Vorteil, denn deren Datenpunkte können als Proxies für Korrelationen durch Ansehen bewertet werden und zu neuen Erkenntnissen führen.

Genealogien führen die Entwicklung von Scatterplots zurück auf den Naturforscher und Eugeniker Francis Galton, der hervorgehend aus einer tabellaren Visualisierung ab den 1880er Jahren die Phänomene der Regression und der statistischen Korrelation aufgezeigt hatte (Friendly und Denis 2005, 109–113).² Es würde zu weit führen, die vielen weiteren Beiträge anderer Statistiker*innen und Grafiker*innen aufzuführen, doch ist zumindest der Astronom John Frederick W. Herschel zu erwähnen, der mit Hilfe eines Scatterplots, also einer grafischen Lösung, die elliptischen Bahnen von Doppelsternen trotz unvollständiger Daten vorhersagte. Die Begriffe ‚Scatter-Diagramm‘ oder ‚Scatterplot‘ kamen zwischen 1906 und 1920

² Auf die problematischen Genealogie von Korrelation und linearer Regression als eugenische Methoden Francis Galtons macht die Medientheoretikerin Wendy Chun aufmerksam (siehe Chun 2021, 59–66).

zunehmend in Gebrauch, als das Verfahren in Lehrbüchern der Statistik und Datenvisualisierung beschrieben wurde (Friendly und Denis 2005, 116–119).

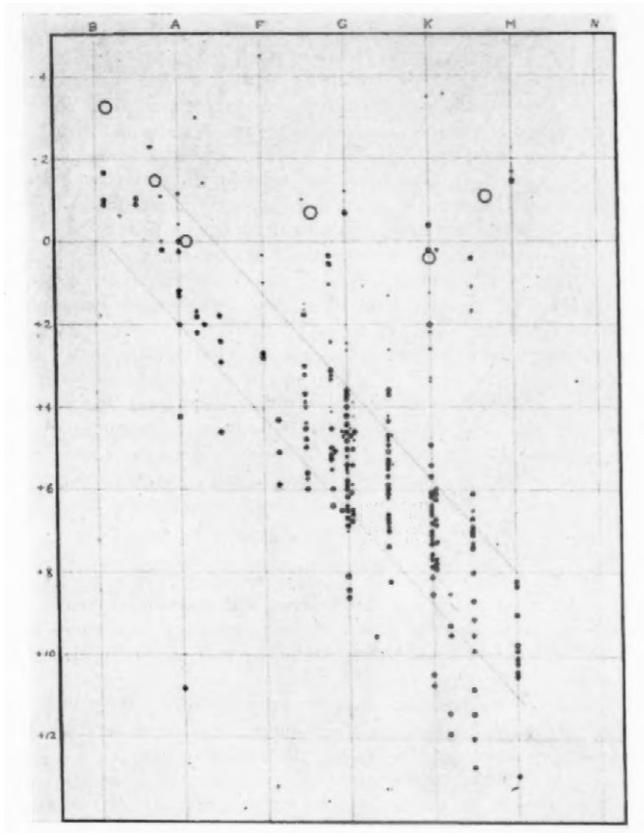


Abb. 2: Historischer Scatterplot: „Die Spektralklasse [B, A, F, G, K, M, N – bezogen auf das Farbspektrum des wahrnehmbaren Lichts, F.H.] erscheint als horizontale Koordinate, während die vertikale die [...] visuelle Helligkeit [das Maß für die Lichtenergie, F.H.] ist, die jeder Stern haben würde, wenn er auf eine Standardentfernung, die einer Parallaxe von 0".1 entspricht, gebracht würde“ (Russell 1914, 285). Sternriganten befinden sich oben und Zwergsterne eher unten rechts.

Die Dichte der jeweiligen Cluster zeigt den Grad der Korrelation innerhalb der jeweiligen Cluster an: Weiter verstreute Datenpunkte indizieren eine geringere Korrelation der Datenpunkte untereinander. Korrelationslinien (Regressionsanalyse) erlauben es, Trends sichtbar zu machen, ggf. Voraussagen treffen zu können und Abweichungen von den Trends zu visualisieren. „Im Gegensatz zu anderen grafischen Formen – Tortendiagrammen, Liniendiagrammen und Balkendiagrammen – bot das Streudiagramm einen einzigartigen Vorteil: die Möglichkeit, Regelmäßigkeiten in empirischen Daten (die als Punkte dargestellt werden) zu entdecken, indem man geglättete Linien oder Kurven hinzufügte, die so gestaltet waren, dass sie ‚nicht durch [die Punkte], sondern dazwischen hindurch gingen, um so von den Rohdaten zu einer theoriegestützten Beschreibung, Analyse und einem Verständnis zu gelangen“ (Friendly und Denis 2005, 128).

Das oben abgebildete Streudiagramm des Astronomen Henry Norris Russel zeigt beispielsweise von links nach rechts gelesen, dass Sterne der Klassen A und B (niedrige spektrale Farbe) in der Mehrheit sehr hell sind. Die meisten Sterne der Klassen K und M sind dunkler und im rötlichen Spektrum (höhere spektrale Farbe) angesiedelt. In der Mitte des

Diagramms finden sich rote Sterne hoher Helligkeit. Als Ausreißer finden sich blaue und rote Supergiganten und rote und weiße Zwergsterne.

Verdichtungen erzeugen hier also eine visuelle Ordnung, die dem Gruppieren ähnelt, wobei die Abgrenzung der Gruppen weniger scharf ist als beispielsweise im Raster der Tabelle, bei dem Spalten und Zeilen die Daten zueinander abgrenzen. Gruppieren erlaubt eine Zusammenfassung von Daten als ‚UND‘, beziehungsweise ‚ODER‘, sodass die visuelle Wissensoperation die Wissensobjekte anhand von Ähnlichkeiten, jedoch nicht notwendigerweise Übereinstimmungen eingruppiert. So können Schwerpunkte, Verteilungen, Tendenzen, Vergleiche und Ausreißer dargestellt werden.

Die Computer-Visualisierungsexperten Alper Sarikaya und Michael Gleicher haben eine Reihe von Wissensoperationen identifiziert, die mit Scatterplots möglich sind: die Identifikation und Lokation eines Objektes, den Vergleich von Objekten (Zugehörigkeit zu Objektklassen, zum Beispiel ‚Einrichtungsgegenstand: Tisch‘, oder ‚Kunsthistorische Epoche: Impressionismus‘), eine Exploration der Nachbarschaft und der Daten, die Suche nach einem Muster (Cluster, Korrelation) und deren Charakterisierung (dicht, weit streuend etc.) und die Identifikation von Anomalien und Außenseitern (Sarikaya und Gleicher 2018, 3). Sie zeigen zudem fünf typisierte Verteilungen in Scatterplots auf, welche häufig anzutreffen sind: 1. die zufällige Verteilung, 2. die lineare Korrelation, 3. das Bilden verschiedener Cluster, 4. eine Krümmung (z. B. in Form einer Sinuskurve), und 5. einander überlappende Daten-Punkte (Siehe Abb. 3).

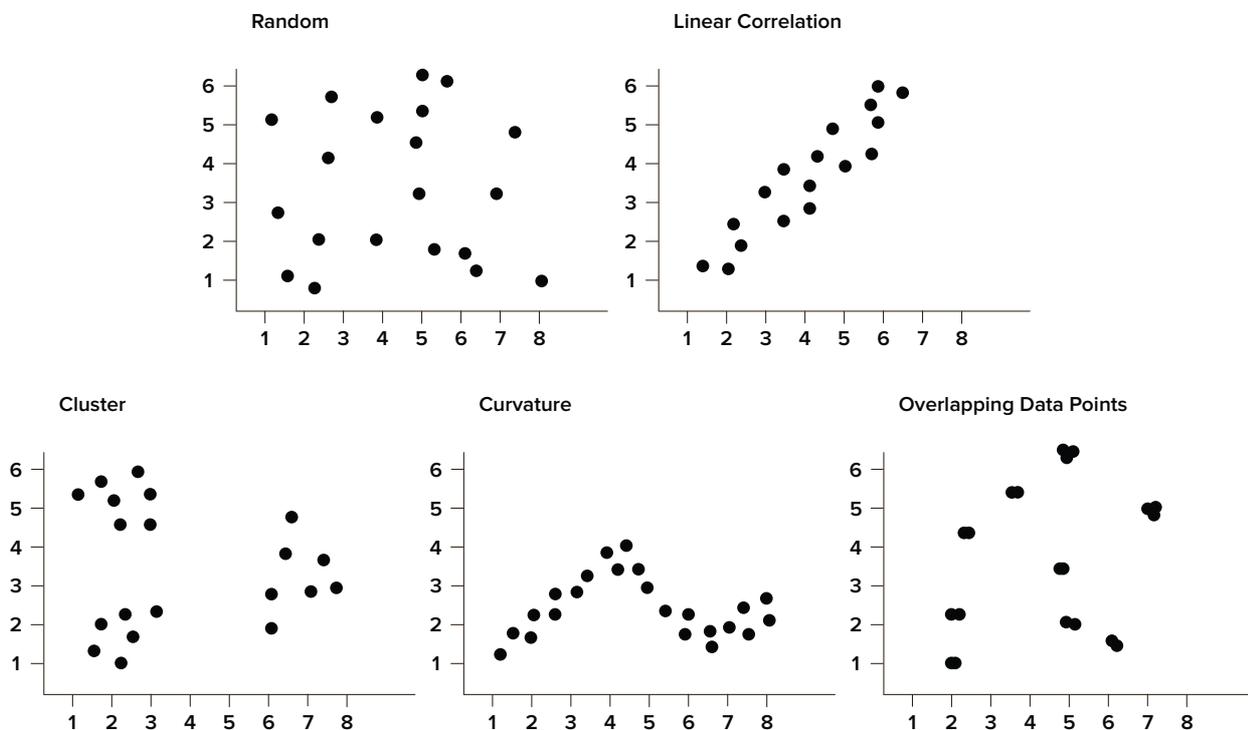


Abb. 3: Fünf typisierte Verteilungen in Scatterplots (zitiert nach Sarikaya und Gleicher 2018, 5).

K-Nearest-Neighbors

Ausgehend von den Scatterplots spielen in den visuellen Darstellungen von (Kunst-)Sammlungen mittels Machine Learning das Clustering und Pfad-Darstellungen eine zunehmende Rolle. Dabei wird die Zugehörigkeit von Datenobjekten im User-Interface sichtbar gemacht, zum Beispiel durch gemeinsame Farbgebung oder durch einzelne Datenpunkte verbindende Netzwerklinien.³ Die K-Nearest-Neighbors ergeben sich aus der mathematischen Ähnlichkeit, die als Nähe von Datenobjekten untereinander sichtbar gemacht wird. Dabei gibt K an, in wieviel klassifikatorische Teilmengen die Gesamtdaten unterteilt werden, bei $K=3$ entstehen also drei Cluster (Verdichtungen).

Die häufigsten Verwendungen des K-Nearest-Neighbor-Algorithmus sind derzeit das Clustering und die Pfad-Darstellung:

1.) Clustering: Das Clustering teilt sich wiederum in zwei unterschiedliche Methoden auf: Zum einen kann die Medienvisualisierung über diskretisierte, ‚strukturierte‘ Daten, also mit Schlüsselworten gelabelte Bilder, erfolgen. Dabei werden unbekannte Bilder anhand vorher gelabelter bekannter Bilder im mehrdimensionalen latenten Raum gewichteter („neuronaler“) Netze verortet. Je nachdem, welchem Merkmalsraum die unbekanntesten Bilder am nächsten liegen, erhalten sie auch das entsprechende Label. Beim Clustering, also einer Klassifikation in zum Beispiel drei Cluster, erfolgt demnach eine räumliche Zuordnung zu einem der Label ‚Katze‘, ‚Hund‘, oder ‚Pferd‘.

Jenseits diskreter Zuordnungen können K-Nearest-Neighbor-Algorithmen auch für ‚kontinuierliche Regression‘ eingesetzt werden, für Ähnlichkeitsberechnungen. Hierbei wird nicht ein diskretes Label vergeben, sondern ein kontinuierlicher Zahlenraum, ein Graph, anhand der Vergleichsdatenpunkte aufgespannt. Die Verortung eines Datenpunktes ergibt sich im Verhältnis zu den k -nächsten Nachbarn, also bei Betrachtung von $k=3$ aus dem Verhältnis, und zwar dem Mittelwert, zu den drei nächsten Nachbarn. Das Ergebnis ist ein Zahlenwert, welcher den Grad der Ähnlichkeit im Verhältnis zu den Nachbarn angibt, und der wiederum für die Verteilung in einem Scatterplot verwendet wird (Abb. 13).

2.) Pfad: Zwischen diesen Bilddaten wird nach dem Prinzip des nächsten Nachbarn im multidimensionalen Raum gewichteter Netze ein Pfad aufgespannt, der vom Start zum Ziel führt. Dieses Verfahren geht historisch zurück auf das Brückenproblem, welches 1736 durch Leonhard Euler erstmals beschrieben wurde. Der daraus sukzessive entwickelte Algorithmus liefert eine mathematische Lösung für den kürzesten Weg zwischen zwei Punkten in einer Stadt, während eine Reihe von Brücken überquert werden muss (Siehe Euler in: König 1936, 290–301). In Visual-Computing-Projekten wird der Pfad für das visuelle Durchschreiten des latenten Raumes verwendet, wie in dem Praxisbeispiele X Degrees of Separation dargestellt. Dabei legen die Betrachter*innen ein Anfangsbild und ein Zielbild fest und in mehreren Schritten, das heißt Zwischenbildern, stellt der Algorithmus anhand der nächsten Nachbarn

³ Eine aufschlussreiche Genealogie des K-Nearest-Neighbors-Algorithmus entwickelt James E. Dobson im Kapitel *The Cultural Significance of k-NN* (Dobson 2019). Dobson bezeichnet K-Nearest-Neighbor als Mittel, welches es erlaubte, Klassifikation direkt aus Daten abzuleiten, in scheinbarer Auslöschung subjektiver menschlicher Kategorisierungen.

eine Verbindung zwischen Start und Ziel her. Wie sich Clustering und Pfad konkret verhalten, diskutiere ich ausführlicher in einer der Fallstudien.

Tabellen, Scatterplots und K-Nearest-Neighbors – ihnen gemein ist die räumliche Verteilung als visuelles Argument, welches Wissen erzeugen soll. Nach dem Durchgang durch die Syntax von Diagrammen in Medienvisualisierungen mittels gewichteter Netze ist nun die Reduktion und Rekonstruktion von Realität als weiterer Aspekt zu erörtern.

1.2 Drei Schichten der Reduktion und Rekonstruktion von Wirklichkeit

Visualisierungen sind Reduktionen und Rekonstruktionen von Wirklichkeit. Doch sind sie nur die letzte und sichtbarste Schicht. Als zweite Schichtung ist die Computation und Statistik und als erste Schicht die Datenerhebung zu verzeichnen. Das ist deshalb relevant, weil jede Schicht an reduzierender und konstruktiver Informationsverarbeitung ihre jeweils eigenen politischen und ethischen Werte einführt, welche für die Betrachtenden unsichtbar bleiben. Dabei stehen – wie auch in den Praxisbeispielen zu sehen ist – Reduktion und Rekonstruktion in einem operativen Verhältnis zueinander. Die im Datenraum reduzierte Wirklichkeit erfährt in Diagrammen ihre Rekonstruktion und wird überhaupt erst dadurch operational.

Erste Schicht ‚Datenerhebung‘: Bereits bei der Erhebung von Daten wird anhand eines Informationsmodells entschieden, welche Anteile von Realität als Daten erhoben werden sollen und welche sprichwörtlich liegengelassen und ausgeschlossen werden. Diese Ausschlüsse sind nicht per se ‚schlecht‘, denn neues Wissen zu generieren erfordert Konzentration und Klarheit. Dabei kommt es zu einer räumlichen, zeitlichen oder räumlich-zeitlichen Segmentierung der Welt, wie Bowker/Star (1999, 10) vermerken. Bereits die Datenerhebung etabliert Standards und im Verlauf der Standardisierung wird Wirklichkeit reduziert und gleichzeitig auch rekonstruiert (vgl. Gitelman 2013; Bowker 2014, 1797). Doch Reduktion tritt nicht allein als Ausschluss auf, wie am Beispiel der Data Doubles (D. Haggerty und Ericson 2000) zu illustrieren ist. Data Doubles sind Repräsentationen einer Person oder eines Gegenstandes zu einem bestimmten Zeitpunkt. Es sind immer auch Rekonstruktionen: Da die Daten einer Person nie vollständig erfasst werden können, werden fehlende Informationen beispielsweise durch Korrelation mit Vergleichsgruppen kompensiert. Häufig werden Daten aus verschiedenen Quellen miteinander kombiniert, sodass auch hier von einer Rekonstruktion zu sprechen ist. Die Rekonstruktion von Wirklichkeit in Daten geschieht anhand eines Informationsmodells, also anhand eines Verständnisses, welcher Anteil von Wirklichkeit in die Berechnung aufzunehmen und welcher auszuschließen ist. Beispielsweise ist im Kopf einer Tabelle durch die Spaltenbeschriftung verankert, welche Daten eingetragen werden können. Da sich Personen und Gegenstände mit der Zeit verändern, sind Data Doubles per se nie identisch mit dem ‚Original‘. Sie sind historisch inkorrekt, unvollständig, fragmentiert und reduziert. Reduktion tritt also sowohl als Ausschluss als auch durch zeitliche Divergenz auf.

Zweite Schicht ‚Statistische Computation‘: Die gewichteten Netze künstlicher ‚Intelligenz‘ sind Reduktionsmaschinen. Sie reduzieren große Pixelmengen, zum Beispiel Bildgrößen von 5000×3000 Pixeln, auf computable Dimensionen wie im Falle des vortrainierten, gewichteten Netzwerkes ResNet 50 auf Dimensionen von 229×229 Pixel. Weil außerdem viele Module in Convolutional Neural Networks die Anzahl der Datenpunkte (Features) erweitern, müssen zwischen den Modulen immer wieder Reduktionen des mehrdimensionalen latenten Raumes erfolgen, um Rechenperformance zu gewährleisten. Dafür werden u.a. Algorithmen wie die ‚Principal Component Analysis (PCA)‘ eingesetzt. Wenn beispielsweise die Dimensionen von 1024 auf 256 reduziert werden sollen, dann filtert der Algorithmus jene Dimensionen mit der größten Varianz heraus und verwendet diese weiter. Dies bedeutet auch, dass inhaltliche Gründe für die Frage, welche Dimensionen reduziert werden sollen, irrelevant werden. Der Algorithmus nimmt die Reduktion allein anhand der maximalen Varianz vor, nicht anhand inhaltlicher Kriterien. Neben der Reduktion erfolgt durch Interpolation und Vereinigung von Datenmengen auch eine Rekonstruktion, besonders in jenen Modulen von Convolutional Neural Networks, welche die Anzahl der Features erweitern.

Dritte Schicht ‚Visualisierung‘: Die bereits in diesen unterliegenden Schichtungen reduzierten Daten werden in der Visualisierung durch grafische Repräsentation erneut reduziert. Daten, also sozial vermittelte Referenzen auf Fakten, werden zu Punkten im Skalen-Diagramm, zum Knotenpunkt und Relation im Netzwerkdiagramm oder zum Zelleneintrag in einer Tabelle oder einer Farbcodierung. Mit Hilfe von Algorithmen wie PCA, MDS, PCS, t-SNE oder UMAP werden die hochdimensionalen Räume gewichteter Netze für den visuellen Output dramatisch, und zwar auf nur zwei Dimensionen, die X- und Y-Achse in Scatterplot-Diagrammen, reduziert.

Die Referenz auf das ursprünglich erhobene Faktum wird schwächer, fast homöopathisch (Abb. 2).⁴

Schicht	Reduktion	Rekonstruktion
Datenerhebung	räumliche, zeitliche oder räumlich-zeitliche Segmentierung der Welt	Informationsmodell
Statistik/Computation in gewichteten Netzen	Algorithmische Reduktion von Datenmengen zugunsten von Rechengeschwindigkeit	Zusammenführung und Extrapolation
Visualisierung	Reduktion der Dimensionen gewichteter Netze für die zweidimensionale Darstellung Symbole, Graphen, Punkte	Bildrepräsentationen (Icons)

Abb. 4: Reduktion und Rekonstruktion von Wirklichkeit in auf gewichteten Netzen (Convolutional Neural Networks) basierenden Visualisierungen.

⁴ Diese Analyse gilt auch für dreidimensionale Darstellungen, die prinzipiell zweidimensionalen Reduktionen gleichen, und lediglich die dritte Dimension als weiteren Parameter in Reduktionsalgorithmen wie PCA, MDS, PCS, t-SNE oder UMAP einführen.

Doch ein neuer Typ der Visualisierung addiert nach diesen vielen Reduktionen neuen visuellen Reichtum: „Anstatt Text, Bilder, Videos oder andere Medien durch neue visuelle Zeichen wie Punkte, Rechtecke und Linien darzustellen, bauen Medienvisualisierungen neue Darstellungen aus den ursprünglichen Medien auf. Bilder bleiben Bilder, Text bleibt Text“ (Manovich 2020, 197f.). Auch wenn man Manovich in dieser Aussage nicht komplett folgen mag („Bilder bleiben Bilder, Text bleibt Text“), so weist er doch auf einen wichtigen Punkt hin: Die Operationalität ‚klassischer‘ Diagramme ist durch die symbolhafte Repräsentation abstrakter als es in Mediendiagrammen der Fall ist, die das Repräsentierte mittels Icons – ikonisch – abbilden. Durch die hohe Kapazität an Rechenleistung sind heute in Echtzeit Mediendarstellungen möglich, die vor den 2000er Jahren nur schwer zu leisten waren und die neue diagrammatische Lösungen für visuelles Wissen erlauben. So können, wie in den Praxisbeispielen auszuführen ist, Tabelle und Scatterplot nicht mit abstrakten Punkten, Linien und ähnlichem, sondern mit Bildrepräsentationen gefüllt werden und neben der räumlichen Verteilung eine zusätzliche Informationstiefe schaffen, die neue visuelle Interpretationen erlaubt (Abb. 12). Hier tritt eine ‚Rekonstruktion‘ von Wirklichkeit in diagrammatischen Darstellungen auf. Jede grafische Konstruktion erschafft einen neuen Raum, in den die Datenobjekte übertragen werden, und zwar anhand des Rasters, die dieser Raum vorgibt. Markant für diese Medienvisualisierung ist der Modus, nach dem die Datenobjekte dargestellt werden. Hier wird eine dimensionale Verkleinerung der ursprünglichen Bilder auf entweder die gleiche Höhe oder die gleiche Breite vorgenommen. Mit diesen Skalierungen geht, wie wir in den Praxisbeispielen sehen werden, eine ganze Reihe von Interpretationsproblemen einher.

Im ersten Teil dieser Studie wurde die Syntax verschiedener Medienvisualisierungen erörtert, mit dem Schwerpunkt auf Tabellen und Diagrammen, insbesondere Scatterplots und Clustering mittels K-Nearest-Neighbor-Verfahren. Während den genannten visuellen Erkenntniswerkzeugen räumliche Bezüge als wissensbildende Operation gemein sind, gibt es Unterschiede in der Art und Weise, wie sie gelesen werden. In einem weiteren Schritt identifizierten wir drei Schichten der Reduktion und Rekonstruktion von Realität: 1.) im Zuge der Datenerhebung, welche die Objekte aus der Realität in Datenstrukturen überführt, 2.) durch statistische Computation in gewichteten Netzen und die dortigen Dimensionalitätsreduktionen mittels PCA und 3.) im Zuge der Visualisierung während der die hochdimensionalen Räume auf zwei Dimensionen X und Y für Scatterplot-Diagramme mittels Algorithmen wie PCA, MDS, PCS, t-SNE oder UMAP reduziert werden. Schließlich stellte das Paper dar, welche Wissensoperationen für The Curator’s Machine relevant sind und wie räumliche Anordnung jeweils operationalisierbar wird.

2 Praxisbeispiele der KI-Visualisierung

Die Auswahl der folgenden Projekte erfolgte aufgrund ihrer Nähe zu Training the Archive in Bezug auf den Umgang mit großen Bildsammlungen. Die ausgewählten Beispiele wurden im Sinne einer konstruktiven Kritik geprüft. Sie wurden nicht ausgewählt, um ihre Leistungen oder Eigenschaften zu schmälern, sondern um die Notwendigkeit detaillierten Fachwissens herauszuarbeiten, wenn es um Informations- und Visualisierungsentscheidungen geht. Diskutiert werden im Folgenden die Projekte ARTigo, Imgs.ai, iART, Vikus Viewer und X Degrees of Separation, bevor in der Zusammenfassung ein weiteres Vorgehen formuliert wird.

2.1 ARTigo – Die Kunsthistorische Suchmaschine

ARTigo (<https://www.artigo.org/>) ist eine semantische Bildsuchmaschine, welche Datenbestände der Diathek des Instituts für Kunstgeschichte der LMU München durchsuchbar macht. ARTigo verwendet keine Algorithmen der künstlichen ‚Intelligenz‘, sondern lädt im Zuge eines Crowd-Intelligence-Verfahrens die Nutzer*innen des Webangebots ein, die Bilder mit strukturierten Meta-Daten zu versehen (Bry, Schefels und Schemainda 2018). Im Folgenden wird das User-Interface der Suchmaschine besprochen, ohne auf die zugrundeliegenden Datenerhebungen einzugehen. Bisher werden keine Machine-Learning-Verfahren eingesetzt, doch bearbeitet ARTigo ähnlich wie The Curator’s Machine Sammlungen von Kunstgegenständen und ist daher für unser Vorhaben relevant. Die Stärke von ARTigo im Unterschied zu vielen Projekten der künstlichen ‚Intelligenz‘ liegt in der Konzentration auf semantische Aspekte der Bilder – die Inhalte und Bedeutungen jedes einzelnen Bildes wurden annotiert und verifiziert. So bleibt die Referenzialität jedes einzelnen Bildes in der Datenbank belegbar erhalten. Von einer Übertragung semantischer Bedeutung auf andere Bilder mittels Machine-Learning wurde seitens der LMU bisher abgesehen, jedoch bietet sich der Datensatz dafür an.

Tabellarisches Raster

Die Suchergebnisse sind in der Standardansicht tabellenförmig untereinander angeordnet, wobei eine Zeile jeweils einen Findeintrag markiert (Abb. 5). Die Spalten bestehen aus einer zum Bild gehörigen Tag-Cloud, einer Abbildung und den klassischen kunsthistorischen Metadaten Künstler*in, Titel, Standort und Datierung. Damit folgt diese Auflistung jener Karteikartenlogik, wie sie häufig bei relationalen Datenbanken anzutreffen ist. Das treffendste Ergebnis auf die Anfrage wird im Zuge der visuellen Oben-Unten-Logik oben angeordnet und die einzelnen Einträge durch Weißräume voneinander getrennt. Diese Ansicht konzentriert sich auf das scrollende Sichten einzelner Einträge, um mittels der Filtermöglichkeiten eine Auswahl eines oder weniger Einträge zu treffen.

Gebirge

1945 Suchergebnisse 25 Ergebnisse pro Seite Zeige Suchformular Zeige Schlagworte Zeige Metadaten
 Suchanfrage verfeinern 1 2 3 4 5 6 7 8 9 10 > >>

<p style="font-size: small; margin: 0;">BÄUME GIPFEL EIS TANNE SCHATTEN HIMMEL WINTER BACH GEBIRGE HAUS LICHT KALT WOLKEN FLUSS LANDSCHAFT NATUR WALD SCHNEE HÜGEL BERGE</p>		<p>Künstler Adalbert Wex</p> <p>Titel Winterliches Tal im Gebirge</p> <p>Standort München / Galerie Bubenik</p> <p>Datierung 1884/1932</p>
<p style="font-size: small; margin: 0;">WEISS WOLKEN GIPFEL NEBEL LANDSCHAFT WINTER BERGSPITZE BAUM GEBIRGE LICHT STURM BÄUME FELSEN BERG SCHNEESTURM TANNE GRAU TANNEN HIMMEL SCHNEE</p>		<p>Künstler Caspar David Friedrich</p> <p>Titel Morgennebel im Gebirge</p> <p>Standort Rudolstadt / Thüringer Landesmuseum Heidecksburg</p> <p>Datierung 1806-1810</p>
<p style="font-size: small; margin: 0;">LICHT LANDSCHAFT BERG GEBIRGE FELS WIESE FRAU WOLKEN WEG STEIN SCHLUCHT PFLANZEN BERGE BÄUME FELSEN STRÄUCHER NATUR HIMMEL MENSCH GRÜN</p>		<p>Künstler Cari Spitzweg</p> <p>Titel Mäherinnen im Gebirge</p> <p>Standort Privatsammlung</p> <p>Datierung 1863/1867</p>
<p style="font-size: small; margin: 0;">SCHWARZ KIRCHTURM WEISS LANDSCHAFT BÄUME BERGE GRAS FELD GETREIDE SCHNEE KIRCHE BERG KORN WALD HIMMEL GRÄSER WEIZEN GEBIRGE ÄHREN HÜGEL</p>		<p>Künstler Frank Buchser</p> <p>Titel Roggenfeld im Gebirge</p> <p>Standort Basel / Kunstmuseum</p> <p>Datierung 1863</p>
<p style="font-size: small; margin: 0;">STADT MENSCHEN TAL HÄUSER WOLKEN BÄUME GEBIRGE BLAU</p>		<p>Künstler Jan Griffier</p> <p>Titel Wassermühle im Gebirge</p>

Abb. 5: Ergebnisanzeige von ARTigo in Tabellarischer Form mit drei Spalten und je einem Eintrag pro Zeile (Screenshot, 12.2.2022).

Durch Klick auf einen Begriff aus der Tag-Cloud lässt sich das Suchergebnis verfeinern. Die Begriffe fungieren als Filter. Im vorliegenden Beispiel wurde der Begriff ‚Gebirge‘ gesucht (Abb. 5). Durch Klick auf ‚Bäume‘ in der Tag-Cloud wird die Suchmenge auf jene Bilder eingeschränkt, für die in den Tags ‚Gebirge‘ und ‚Bäume‘ annotiert wurde. Ähnlich fungiert der Klick auf die Meta-Daten, die rechts vom Bild positioniert sind, sodass Suchen nach Gemälden mit ‚Gebirgen‘ möglich sind, die am Standort ‚München‘ vorhanden sind.

Die Suchen fungieren nach dem Prinzip der Closed-World-Assumption: Was die Suche finden kann, muss in der Datenbank enthalten sein. Etwas anderes kann nicht gefunden werden, das heißt, die Suchmaschine ist nicht nach außen verknüpft, beispielsweise durch Open-Data-Zugriffe oder Verknüpfungen mit bibliografischen Aggregatoren.⁵ Für diesen Umstand gibt es keine gesonderte Visualisierung, die Suchliste kommt entsprechend der Konvention schlicht an ihr Ende. Die vordefinierten Kategorien bestimmen das Suchgeschehen: Gesucht werden kann nach Titel, Künstler*in, Standort, Datum und Schlüsselworten.

Matrix-Ansicht

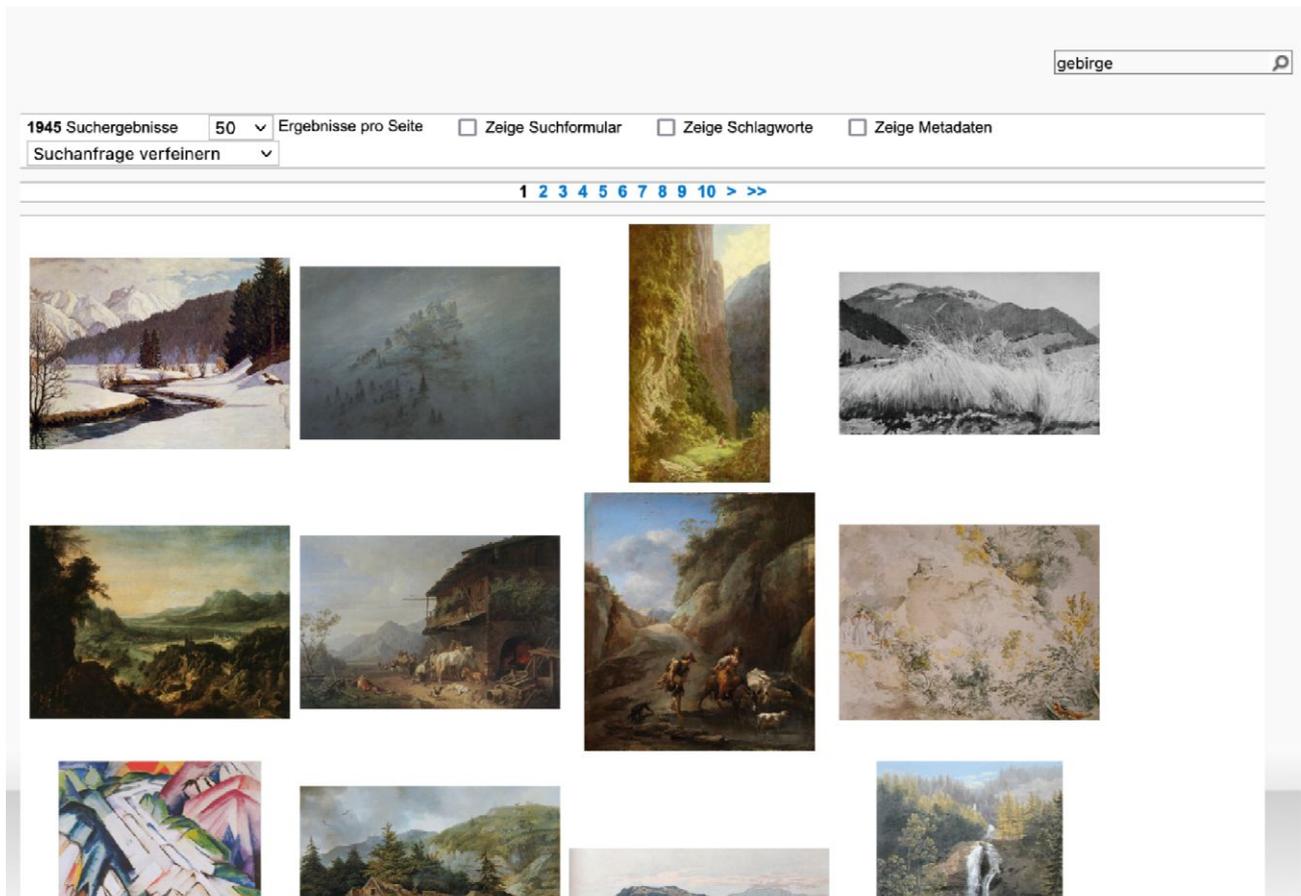


Abb. 6: ARTigo Interface in Matrixform (Screenshot, 12.2.2022).

Werden in dem User-Interface von ARTigo die Optionen „Zeige Metadaten“ und „Zeige Beschreibung“ abgewählt, ändert sich die Darstellungsform zu einer Matrix. Die Leserichtung dieser Matrix ist unklar und wird visuell nicht unterstützt (z. B. durch Abstände, die Zeilen erzeugen würden). Der Moduswechsel aus dem dreispaltigen Raster der Tabelle in eine neue Bildordnung wird durch das Interface nicht spezifisch kommuniziert. Grundsätzlich verändert sich hier die Anzeigereihenfolge der Ergebnisse, denn diese werden statt wie bisher von oben nach unten, nun von links nach rechts angezeigt, wobei das zutreffendste Ergebnis auf

⁵ Ein bibliografischer Aggregator ist ein „Dienst, der Metadaten von Kultur- und Wissenseinrichtungen sammelt, vereinheitlicht, verwaltet, vorhält und weitergibt“ (<https://pro.deutsche-digitale-bibliothek.de/glossar/aggregator>).

die Anfrage sich oben links befindet und die darauffolgenden Ergebnisse in der Zeile bis zu deren Umbruch von links nach rechts sortiert sind. Die verschiedenen Bildgrößen sind auf maximal 200px in Höhe und Breite begrenzt, wobei die Bilder verkleinert, jedoch nicht maßstäblich, unbeschnitten und im richtigen Seitenverhältnis dargestellt werden. Dadurch entsteht bei vielen Hochformaten der Eindruck, die Ergebnisse seien von oben nach unten zu lesen und bei vielen Querformaten der Eindruck, die Rasterung erfolge von links nach rechts. Insgesamt ist ARTigo ein wertvoller Schritt zur Erschließung kunsthistorischer Sammlungen mit Suchmaschinenfunktion. Doch zeigt sich deutlich, dass die Kooperation mit User-Interface-Expert*innen dem Projekt zusätzliche Expertise verschafft hätte.

2.2 Imgs.ai – Eine experimentelle kunsthistorische Bildsuche

Imgs.ai (<https://imgs.ai>) ist eine experimentelle Plattform zur Erkundung von Bildsammlungen mittels Machine-Learning, welche an der University of California, Santa Barbara und dem Deutschen Dokumentationszentrum für Kunstgeschichte – Bildarchiv Foto Marburg angesiedelt ist. Mit Hilfe der Softwaretechnik CLIP (Radford u. a. 2021) ist eine inhaltliche Suche anhand von Textanfragen, sogenannten Prompts, nach unklassifizierten Bildern und Bildinhalten möglich. Außerdem können Bilder als positive oder negative Schablonen für die weitere Suche gesetzt werden, und durch den Upload eigener Bilder lassen sich diese mit den visuellen Ordnungen von Museums- und Bildsammlungen vergleichen. Sortiert wird nach dem K-Nearest-Neighbor-Prinzip, sodass einander ähnliche Bilder aus der Sammlung gezeigt werden. Dies gibt Kurator*innen und Forschenden die Möglichkeit, anhand eigener Bilder oder Begriffe ähnliche Bilder in den zur Verfügung stehenden Museumssammlungen zu suchen, ohne dass die Sammlungen annotiert sein müssen.

Tabellarisches Raster

Das User Interface von Imgs.ai besteht aus einer Sektion mit zahlreichen Filtermöglichkeiten, wie zum Beispiel dem verwendeten Datensatz (Metropolitan Museum, Rijksmuseum), dem eingesetzten gewichteten Netz (VGG 19, raw, CLIP-vit, Poses),⁶ einer Upload-Möglichkeit für eigene Bilder als Suchterminus und einer Auswahl für die Spaltenbreite der Darstellung, zum Beispiel 124 Pixel Breite, und aus der Anzahl benachbarter Bilder zwischen 10 und 100. Darunter befindet sich die Darstellung der Bilddaten, die horizontal in Spalten angeordnet sind. Klickt man auf ein Bild, so kann man dessen URL im Internet erfahren, oder direkt zur hochauflösten Bilddatei auf den Websites der Herkunftsinstitutionen gelangen. Die verschiedenen Bildgrößen sind auf eine Spaltenbreite von 32px bis 224 px begrenzt, auswählbar mittels des Menüpunktes ‚size‘. Entsprechend werden alle Bilder in der Darstellung verkleinert, und unbeschnitten und im richtigen Seitenverhältnis dargestellt.

⁶ Pose kann menschliche Posen, vor allem das Verhältnis von Armen, Beinen zum Körper ermitteln (Toshev und Szegedy 2014); VGG 19 ist ein auf ImageNet vortrainiertes gewichtetes Netz aus 19 Layern (Simonyan und Zisserman 2015); CLIP ist auf kombinierte Bilddaten und Word-Embeddings (generiert aus Bildunterschriften) trainiert, u. a. durch das Training mit annotierten Flickr-Bildern (Radford u. a. 2021).

Positive:



Searching **358426** images in **Metropolitan**

Embedding **manhattan** **128** **60**



Abb. 7: Imgs.ai Interface mit der Radierung „Charles Carroll of Carrollton“ (ca. 1835) aus der Sammlung des Metropolitan Museum New York als Prompt. Als Embedding wurde hier ‚Poses‘ gewählt, sodass im Prinzip ähnliche Verhältnisse von Kopf, Schulter und Armen gefunden werden (Screenshot, 12.2.2022).

Positive:



Searching **358426** images in Metropolitan

Embedding vgg19 manhattan 128 60



Abb. 8: Gleiche Suchparameter wie in der vorhergehenden Abbildung, jedoch sind die K-Nearest-Neighbor nicht anhand des Poses-Netzwerkes, sondern anhand von VGG-19 berechnet (Screenshot, 12.2.2022).

In Abb. 7 ergibt sich die horizontale Orientierung allein aus den Bildformaten und erfährt keine weitere visuelle Unterstützung. In dieser vertikalen visuellen Logik ist das erste und der Suchanfrage nächste Bild links oben in Spalte 1 angeordnet. In der Spalte folgen weitere Bilder mit absteigender ‚Ähnlichkeit‘ von oben nach unten. Beim achten Bild erfolgt für das Auge ein Spaltensprung in Spalte 2, ganz nach oben.

Nach dem K-Nearest-Neighbor-Prinzip ist dieses Bild weiter entfernt vom ersten Bild als alle Bilder in Spalte 1, doch in dieser grafischen Lösung liegt es direkt neben dem ersten Bild.

Hier zeigt sich ein Problem der Matrixanordnung in vertikalen Spalten. Sie generieren eine spezifische Leseordnung. Die Darstellung in einem skalenorientierten Raster ist für die zugrundeliegende Wissensaufgabe grundsätzlich problematisch, denn Nähe und Entfernungsverhältnisse werden aufgrund des Rasters nur ungenügend wiedergegeben.

Wenn in dieser spaltenorientierten Matrix Bilder mit sehr ähnlichen Seitenverhältnissen auftreten, gerät die visuelle Ordnung zusätzlich ins Wanken, denn für das lesende Auge ist unklar, wie die Bildanordnungen orientiert sind: von oben nach unten oder von links nach rechts?

Anhand des Interface-Beispiels von img.ai wird deutlich, dass für K-Nearest-Neighbors eine matrixförmige Anordnung eher ungeeignet ist, da eine Divergenz aus berechneter Statistik und visueller Zuordnung entsteht.

2.3 iART – Ein interaktives Werkzeug zur Analyse von Bilddatenmengen

iART (<https://labs.tib.eu/iart/>) ist ein interaktives Analyse- und Retrieval-Tool für Bildsammlungen mit Schwerpunkt auf Kunstgeschichte. Das DFG-geförderte Forschungsprojekt der Ludwig-Maximilians-Universität München, der Universitätsbibliothek Hannover (TIB) und der Universität Paderborn ordnet die Bilddaten nach Gewichtungungen als K-Nearest-Neighbor.

Es soll drei verschiedene wissensgenerierende Modi zusammenführen: Mustererkennung und Klassifizierung mittels Deep Learning, Ähnlichkeitsanalyse durch Clustering und K-Nearest-Neighbor-Analyse, und die Bezugnahme auf Nutzer*innenpräferenzen durch Filterwerkzeuge. Nach Aussage der Projektentwickler*innen ist das Interface am Google-User-Interface-Design orientiert, da dieses gewohnt und eingeübt sei (Schneider und Kohle 2021, 6:50 min).

Als Daten wurden Sammlungen verwendet, die bereits durch Dritte klassifiziert wurden; das Projekt verzichtete also auf eigene, aufwändige Klassifikationsarbeiten. Die circa 1 Millionen Bilddaten stammen aus der Numismatik-Sammlung KENOM, der Wikimedia Commons-Kategorie „Art_by_Subject“, dem Rijksmuseum Amsterdam sowie weiteren kleineren Datensätzen mit circa 60.000 Bildern, darunter auch aus ARTigo (Springstein 2021). Die Kriterien für ein Auftreten bestimmter künstlerischer Positionen orientieren sich daran, wie oft ein*e Künstler*in in den Datensätzen vorhanden ist. Zum Beispiel dominiert bei der Suchanfrage nach „Mountain“ der deutsch-amerikanische Landschaftsmaler Albert Bierstadt (1830–1902), da dessen Bilder mit dem Schlagwort „mountain“ in der Sammlung Wikimedia Commons gut vertreten sind. Daraus folgt, dass romantische Bergdarstellungen von Bierstadt

zahlreich in den Suchergebnissen vertreten sind, und dies den Eindruck erwecken könnte, seine Arbeiten seien künstlerisch von herausragender Bedeutung gewesen (in Wirklichkeit waren sie nur gut verschlagwortet).

Die visuelle Suche basiert nicht allein auf den Meta-Daten selbst, sondern auch auf Embeddings. Embeddings sind mathematische Übersetzungen von Worten, die als Vektoren gefasst werden. Diese Vektoren werden zueinander mathematisch-räumlich ins Verhältnis gesetzt, um Bedeutungszusammenhänge zu modellieren, wobei Worte mit einander ähnlichen Vektoren einen einander ähnlichen Bedeutungszusammenhang ergeben (sollen). Dadurch kann beispielsweise angegeben werden, dass das Wort ‚painting‘ auf einer Skala bis 1, eine Ähnlichkeit zu ‚sculpture‘ von 0,845, zu ‚painter‘ von 0,842 und zu ‚watercolor‘ von 0,841 hat, während die Ähnlichkeit zu ‚create‘ nur noch 0,359 beträgt.⁷

Weil bestimmte Wort-Kombinationen der Meta-Daten in gewichteten Netzen nah beieinander liegen, werden die im latenten Raum verteilten Bilder entsprechend nah zueinander positioniert. Dabei nehmen die Autor*innen Bezug auf klassische Ordnungsprinzipien:

„Ordnungskriterien, die bereits in frühneuzeitlichen Wunderkammern üblich waren, wie Farbe, Material oder Funktion, können ebenso angewandt werden wie ikonografisch orientierte Ordnungsprinzipien, die zum Beispiel Objekte auf biblische Motive oder christliche Themen hin untersuchen“ (Springstein u. a. 2021, 1). Das Interface erlaubt es, verschiedene Embeddings vortrainierter gewichteter Netzwerke auszuwählen: CLIP, Wikimedia und Imagenet, wobei die genaueren Bezeichnungen der Embeddings (z. B. Versionen, Datensätze) im Interface nicht hinterlegt sind. Dadurch werden unterliegende Genealogien und Biases der Trainingsdatensätze für Forschende intransparent.

In der Folge sollen anhand von iART die zwei wesentlichen Darstellungsmodi Bildraster und Verteilung diskutiert werden.

Tabellarisches Raster

Im Raster werden die Ergebnisse nach Relevanz geordnet gezeigt und beim Überfahren eines einzelnen Bildes erscheinen zusätzliche Informationen wie Metadaten und Link zu einer hochauflösten Version (Abb. 9). Da die Bilder auf die gleiche Bildhöhe angeglichen werden, gehen die Größenverhältnisse der Bilder untereinander verloren. Die Bilder werden als eine fortlaufende Datenmenge gezeigt, wobei das erste Bild den idealtypischsten Berg darstellt und dann horizontal organisiert, zeilenweise der Grad des Berges in einem abnehmenden Ranking sinkt. Der Blick streift also immer von links nach rechts und springt dann in die nächste Zeile. Die Reihenfolge folgt aufgrund der Word-Embeddings im mehrdimensionalen Raum des gewichteten Netzes anhand der CLIP-Logik (Springstein u. a. 2021, 2).

Diese Relevanz-Ansicht kann alternativ auch nach Bildtitel oder chronologisch, nach Jahreszahl sortiert werden.

⁷ Siehe beispielsweise die Darstellung auf <https://embeddings.sketchengine.eu>.

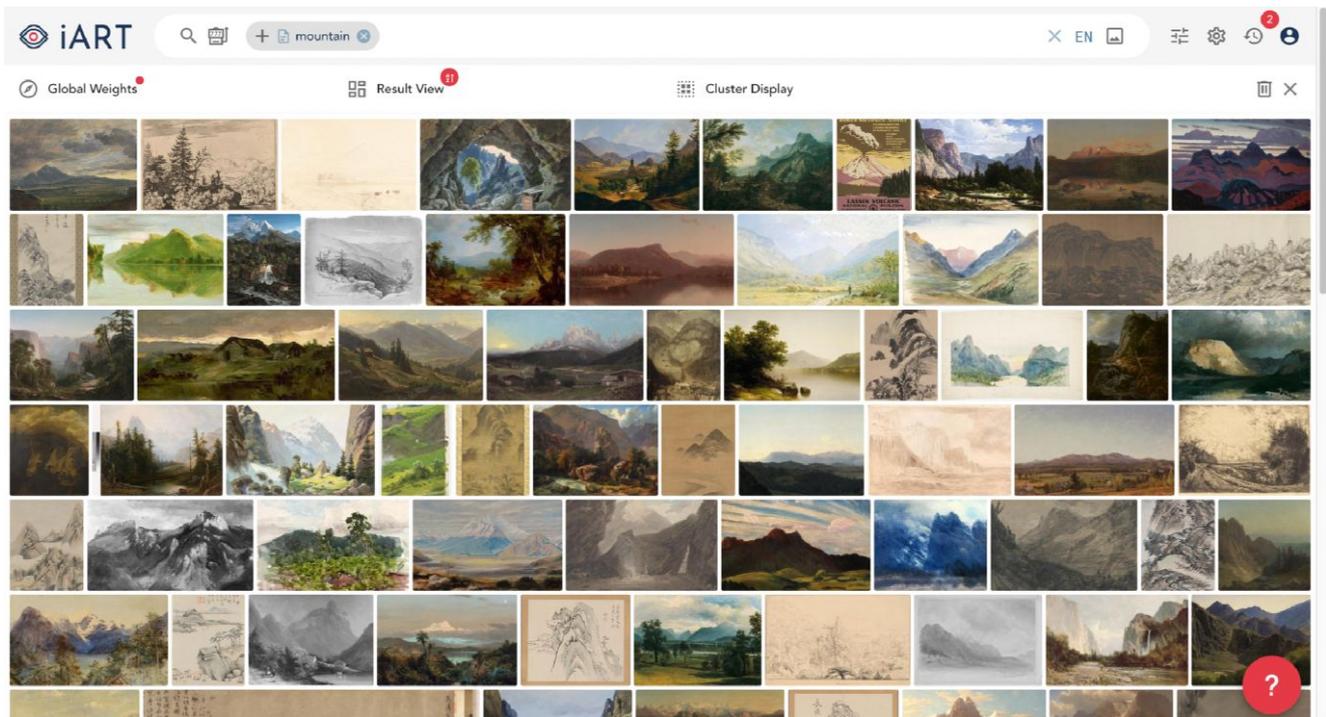


Abb. 9: iART, Suchbegriff ‚mountain‘ ohne Gruppierung/Cluster (Screenshot, 12.2.2022).

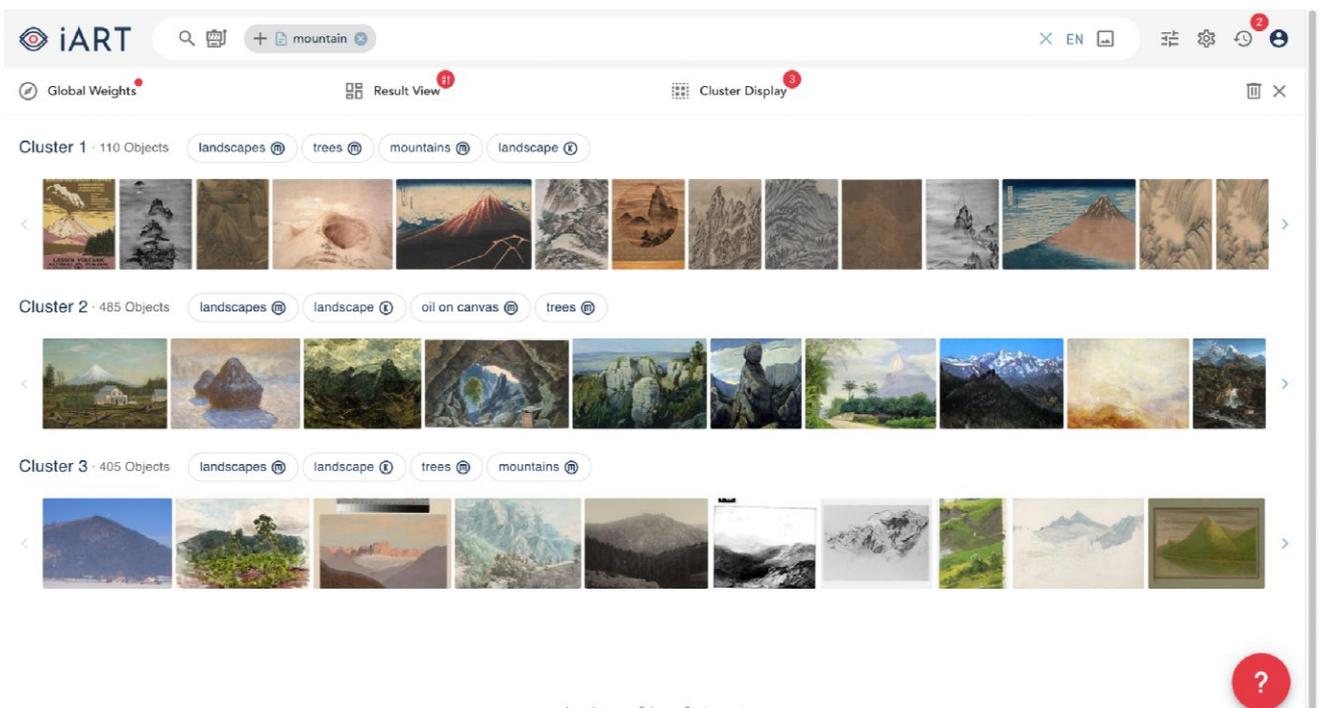


Abb. 10: Der Suchbegriff ‚mountain‘ unterteilt in drei Cluster (Screenshot, 12.2.2022).

Die Rasteransicht kann in einem weiteren Modus in Cluster gruppiert werden. Jedes gebildete Cluster erhält eine eigene Zeile, die endlos nach rechts scrollt. Diese Darstellung hat den Vorteil, dass das Auge tatsächlich einen kontinuierlichen Datensatz von links nach rechts durchstreift und nicht in die nächste Zeile umbrechen muss, wie am Beispiel des Interfaces von Imgs.ai diskutiert.

Die Cluster werden automatisch gebildet. In diesem Beispiel mit dem Suchbegriff ‚mountain‘. Die Cluster sind anhand sehr ähnlicher Stichworte identifiziert, die sich teils nur in der Reihenfolge unterscheiden:

Cluster 1: 110 Bilder – Landscapes, Trees, Mountains, Landscape

Cluster 2: 485 Bilder – Landscapes, Landscape, Oil on Canvas, Trees

Cluster 3: 405 Bilder – Landscapes, Landscape, Trees, Mountains

Bei der Sichtung fällt auf, dass in Cluster 1 viele Zeichnungen vorhanden sind, vornehmlich aus der japanischen Kultur. In diesen Bildern finden sich nahezu immer japanische Schriftzeichen. Cluster 2 zeigt hauptsächlich Ölmalerei und unterscheidet sich von den beiden anderen Clustern durch eine brillante, kontrastreiche Farbgebung. Das Cluster 3 enthält Wasserfarbenmalerei, Tuschezeichnungen, Kreidezeichnungen, einige Fotografien, aber auch Gravuren, deren Farbigkeit eher gedämpft und zurückgenommen ist. Die Farbsättigung könnte daher eines der Kriterien sein, nach denen die Cluster gebildet wurden. Für die Clustering können jedoch auch andere Faktoren eine Rolle spielen, beispielsweise die Sensitivität des zugrundeliegenden Sprachembeddings CLIP für spezifische Wortgruppen, oder die in Working Paper 2 diskutierte Tendenz gewichteter Netze, Bilddaten nicht anhand der Umrisse, wie Menschen es tun würden, sondern anhand der Texturen zu unterscheiden, dem sog. Texture-Bias.⁸ Wenn die japanischen Zeichnungen, Tuschezeichnungen und Drucke, die in Cluster 1 versammelt sind, keine japanischen Schriftzeichen enthalten würden, so würden sie womöglich durch den Algorithmus anders geclustert. Dies müsste experimentell ermittelt werden.

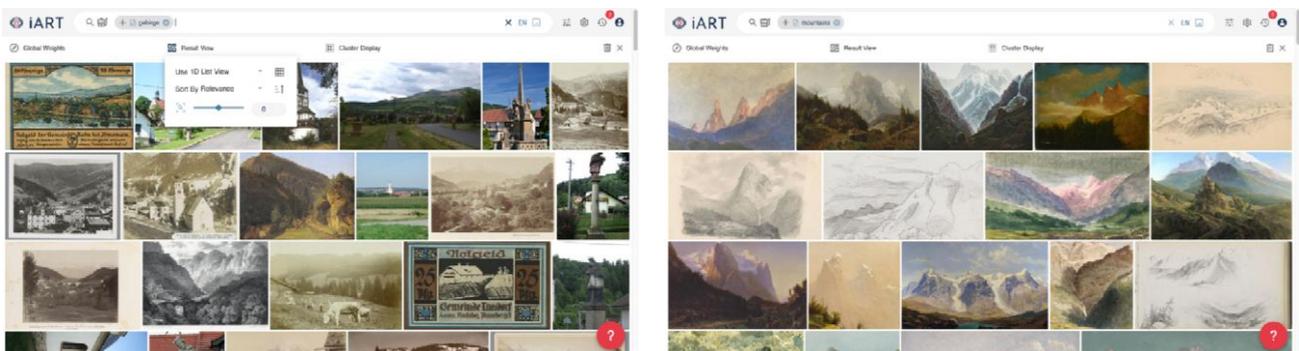


Abb. 11: Links: Ergebnisse der deutschsprachigen Suchanfrage ‚Gebirge‘, mit überwiegend Postkartenabbildungen; Rechts: Ergebnisse der englischen Suchanfrage ‚mountains‘ mit vorwiegend Malerei (Screenshot, 12.2.2022).

In der Eingabezeile gibt es durch das Kürzel EN einen Hinweis, dass eine Englischsprachige Eingabe erwartet wird. Anderssprachige Eingaben sind jedoch möglich und erzeugen überraschende Ergebnisse (Abb. 11).

⁸ Vgl. (Geirhos u. a. 2019).

Scatterplot

Das Interface von iART ermöglicht mit Hilfe des Scatterplots eine weitere Ansicht auf den gleichen Datensatz. Erreicht wird dies durch eine Verringerung der Dimensionen des latenten Datenraums mittels des Uniform Manifold Approximation and Projection Algorithmus (UMAP) (McInnes, Healy und Melville 2020), eine Weiterentwicklung des t-SNE-Algorithmus (Springstein u. a. 2021, 2).⁹ Damit wird der hochdimensionale Output gewichteter Netzwerke entlang der Varianzen einzelner Dimensionen reduziert, wobei die relativen Abstände der Bilder in der jeweiligen Dimension möglichst beibehalten werden.¹⁰ „UMAP leitet sich von dem Axiom ab, dass lokale Distanzen von größerer Bedeutung sind als größere Distanzen. [...] Die multidimensionale Skalierung zielt speziell darauf ab, die gesamte Distanzmatrix der Daten zu erhalten, und ist daher ein guter Kandidat, wenn alle Strukturskalen von gleicher Bedeutung sind.“ (McInnes, Healy und Melville 2020, 45). Wichtig hier ist die Feststellung, dass in UMAP alle dimensional Skalen ähnlich wichtig seien. Dies ist ein Unterschied zu PCA und t-SNE, welche die Neudimensionierung anhand der zwei Skalen mit der jeweils größten Varianz vornehmen, dann die X- und Y-Achsen ergeben.

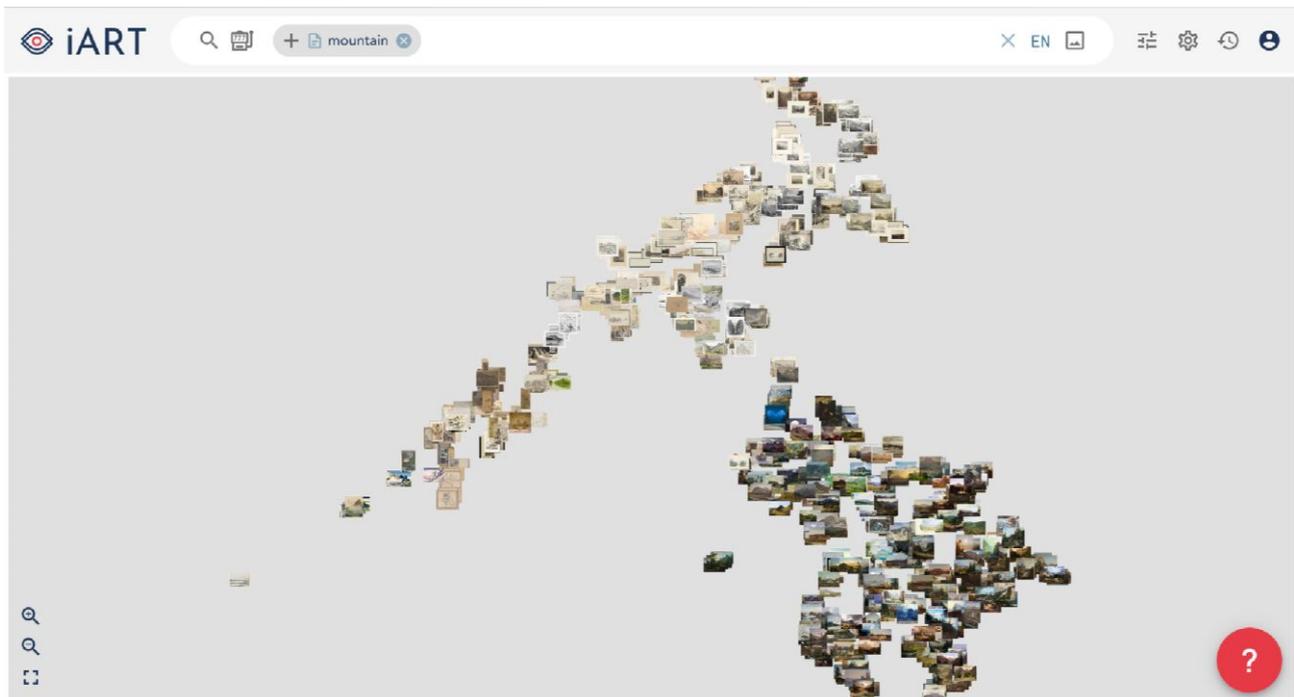


Abb. 12: Scatterplot für den Begriff ‚mountain‘ mit drei Clustern (Screenshot, 12.2.2022).

⁹ Die Autoren von UMAP geben zu bedenken, dass es UMAP im Vergleich an der starken Interpretierbarkeit des Reduktionsalgorithmus PCA (Principal Component Analysis) mangle, dass man also weniger starke Aussagen treffen könne, wie das Innere des Algorithmus Entscheidungen treffe. „In particular the dimensions of the UMAP embedding space have no specific meaning, unlike PCA where the dimensions are the directions of greatest variance in the source data“ (McInnes/Healy, 2018, 45).

¹⁰ Zum Verständnis für Statistik-Novizen seien die YouTube-Tutorials *Principal Component Analysis (PCA), Step-by-Step*, <https://www.youtube.com/watch?v=FgakZw6K1QQ> und *t-SNE, Clearly Explained*, <https://www.youtube.com/watch?v=NEaUSP4YerM> empfohlen.

Ein visueller Vergleich verschiedener Algorithmen zur Dimensionsreduktion, die für Scatterplots verwendet werden können, zum Beispiel von PCA, t-SNE, UMAP, durch das Forscherteam McInnes/Healy zeigt markant unterschiedliche visuelle Ergebnisse für den gleichen Datensatz. So unterscheiden sich die Dimensionsreduktionen von t-SNE und UMAP insofern, dass UMAPs reduzierte Dimensionen deutlichere Abgrenzungen der Gruppierungen untereinander vornehmen als bei t-SNE (McInnes, Healy und Melville 2020, Abb. 4). In der Praxis würde es zu weit führen, die Nutzer*innen im User-Interface über die Unterschiede aufzuklären, doch wären einige Hinweise, zum Beispiel durch Erklärvideos hilfreich, da das richtige Lesen von Scatterplots nicht im gleichen Maße eingeübt ist wie beispielsweise das Lesen von Tabellen (vergleiche Abb. 3).

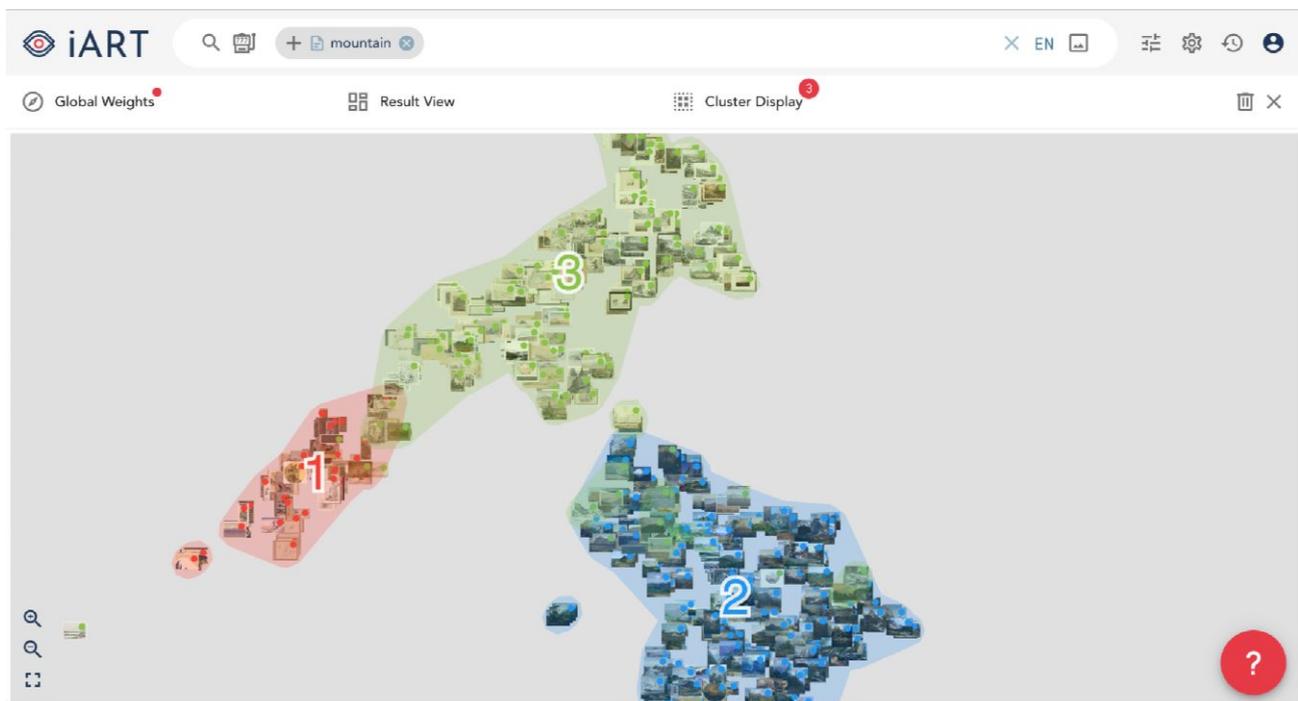


Abb. 13: Die Ansicht der Cluster kann zusätzlich durch farbige Punkte und Farbüberlagerungen verdeutlicht werden. In Cluster 2 befinden sich auch Bilder mit grünen Punkten, die also zu Cluster 3 gehören (Screenshot, 12.2.2022).

Um die Verteilungen und Zugehörigkeiten zu den jeweiligen Clustern sichtbar zu machen, werden in iART die Bilder des Scatterplots mit einem kleinen farbigen Punkt zusätzlich farblich codiert. Dadurch werden Cluster visuell erkennbar, aber vor allem wird deutlich, welche Ausreißer zu welchem Cluster gehören.

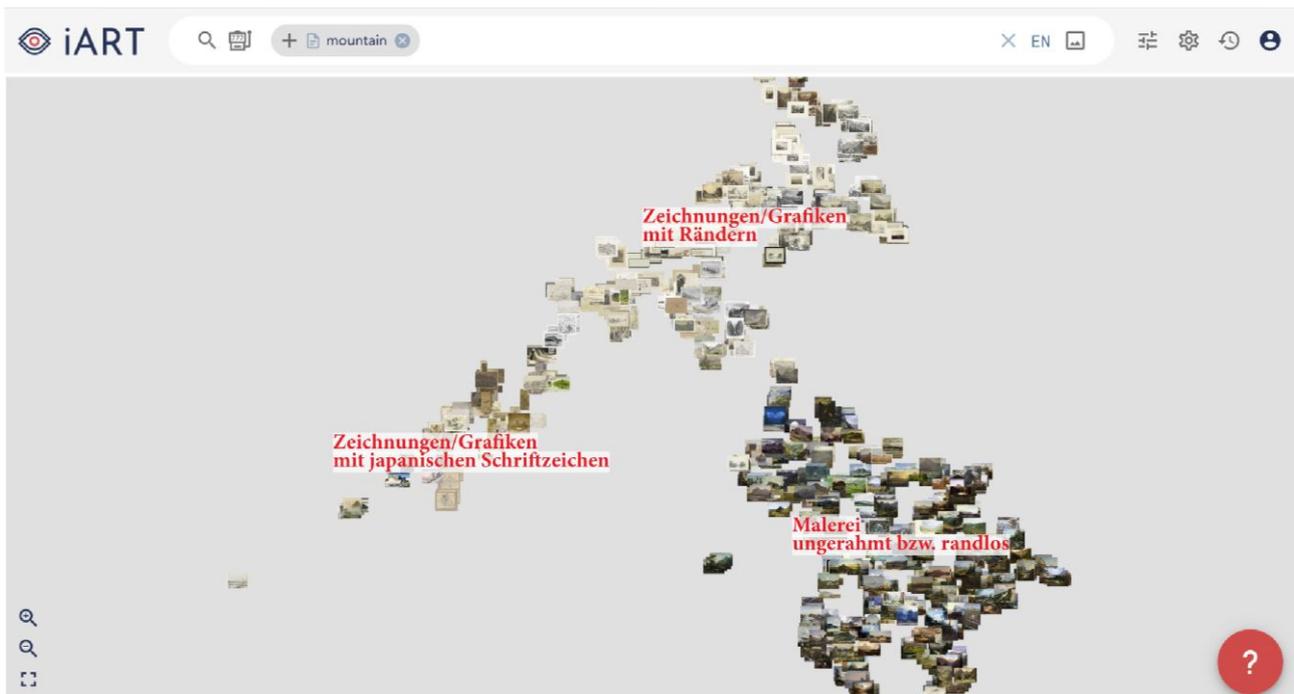


Abb. 14: Mittels CLIP-embedding bilden sich drei Cluster (Screenshot, 12.2.2022).

Der Vorteil des Scatterplots ist, dass prinzipiell alle in der Findmenge vorhandenen Bilder repräsentiert werden können. In der Rasteransicht wird hingegen die jeweilige Clustermenge nicht komplett dargestellt, da sie aufgrund der Bildschirmbreite begrenzt ist. Es ist nur zu sehen, was der Bildschirm fasst, ohne Zoom-Möglichkeit.

iART verfügt über zahlreiche weitere Möglichkeiten. So können prinzipiell eigene Bildinventare importiert werden, die Cluster-Darstellung kann mit Stichwort-Filtern und Suchbegriffen kombiniert werden, vor allem auch einer Bildsuche („Search by Image“), ähnlich dem Projekt Imgs.ai. Deren detaillierte Erörterung würde hier jedoch zu weit führen.

iART ist ein komplexes Forschungsmittel, welches einer besseren Dokumentation bedürfte, zum Beispiel durch Videotutorials. Während der Arbeit am vorliegenden Text wurde das Interface weiter verbessert und ergänzt. iART ist ein offenes Analysewerkzeug zur Exploration großer Bildmengen mit zahlreichen Nutzungsoptionen. Diese konnten hier nur in den Grundzügen dargestellt werden. Seitens des User-Interfaces von iART bedürften die vielfältigen Ordnungswerkzeuge weiterer Erklärung und Anleitung.

2.4 Vikus Viewer

Vikus Viewer (<https://vikusviewer.fh-potsdam.de/>) ist Teil eines Forschungsprojektes zur Darstellung großer Bildmengen auf Webseiten, angesiedelt am Urban Complexity Lab der Fachhochschule Potsdam 2014–2017.¹¹ Gestaltet und programmiert wurde Vikus Viwer

¹¹ Es handelt sich beim hier vorgestellten Vikus Viewer – Van Gogh um eines von zahlreichen visuellen

durch Christopher Pietsch, ursprünglich für das Forschungsprojekt Past Visions (Katrin Glinka, Christopher Pietsch und Marian Dörk). Diese Implementierung des Designers und Programmierers Christopher Pietsch baut auf Nikhil Thorats *Teachable Machine* auf, welche auf dem Machine-Learning Framework Tensorflow basiert. Das in 2017 veröffentlichte Projekt, welches seit 2015 entwickelt wurde, dürfte damit zu den frühzeitigen derartigen Ansätzen in Deutschland und weltweit zählen.

Im Zuge von Workshops, Vorträgen und Konferenzen wurden Interfaces und deren Generalisierbarkeit untersucht. Dabei griff die Forschungsgruppe auf zwei ‚klassische‘ Wissensordnungen zurück: 1.) auf eine zeitliche Ordnung und 2.) auf Schlagworte in den Metadaten der Objekte.

Im Folgenden soll eine weitere Iteration des Projektes Past Visions beschrieben werden, welche unter dem Titel Vikus Viewer – Vincent van Gogh auf 986 Zeichnungen und Gemälde von Vincent van Gogh (1853–90) aus der Sammlung des Van Gogh Museum, Amsterdam, basiert. Ebenso wie die anderen durch Vikus verwendeten Sammlungen wurde diese insofern kuratiert, als dass die Daten des Van Gogh Museums zusätzlich durch die Projektmitarbeiterin Viktoria Brüggemann kuratiert und korrigiert wurden. Im Zuge von User-Interface-Experimenten im Sommer 2014 führte der Designer und Programmierer Christopher Pietsch eine weitere Wissensordnung ein, und zwar in einem Scatterplot. Pietsch bezog sich auf seine Auseinandersetzung mit generativem Design und insbesondere Kyle McDonalds Module in Open-Framework, welches in der Sprache C++ als Teil von Creative-Coding-Praktiken auch Ähnlichkeitsanalysen ermöglichte.

Positiv hervorzuheben ist der Ansatz, die zugrundeliegende webbasierte Software als Open-Source auf der Plattform Gitlab zu veröffentlichen. Die auf Nachhaltigkeit abzielenden Überlegungen zeichnen sich auch in einem Konferenzbeitrag der Autor*innen ab, indem sie einige Minimalvoraussetzungen aufzählen: „Dazu zählen primär die zeitliche Einordnung und die Verschlagwortung entlang eines kontrollierten Vokabulars. Zusätzlich sollte der Bestand in ausreichend guter Qualität digitalisiert sein (und als jpg vorliegen), um die Zoomfunktion (von der Übersicht ins Detail) voll ausnutzen zu können. Die benötigten Sammlungsdaten müssen in einem standardisierten CSV Format vorliegen [...]“ (Glinka, Pietsch und Dörk 2017, 205).

Tabellarisches Raster

In der tabellarischen Darstellung wird der gesamte Datensatz nach Jahren geordnet dargestellt. Die Jahre wurden durch eine weitere Zeile kontextualisiert, in denen Schaffensperioden notiert sind. Darüber befinden sich in der Spalte die Icons der entsprechenden Bilder (Abb. 15). Innerhalb eines Jahres sind die Bilder zufällig angeordnet. Hier wäre auch eine Ordnung nach Ähnlichkeit, Hell-Dunkel-Werten oder alphabetischen Titeln denkbar gewesen. De facto sind in der Initialeinstellung die Bilder-Icons so klein, dass am ehesten noch

Forschungsprojekten dieser Arbeitsgruppe. Siehe auch <https://uclab.fh-potsdam.de/projects> und in: *Von der Wolke zum Pfad – Visuelle und assoziative Exploration zweier kultureller Sammlungen* (Brüggemann u. a. 2022).

Farbigkeit und Hell-Dunkel-Kontrast sichtbar sind. Jedoch erlaubt eine stufenlose Zoom-Funktion, andere Bildgrößen zu wählen, die mehr Details erkennen lassen.

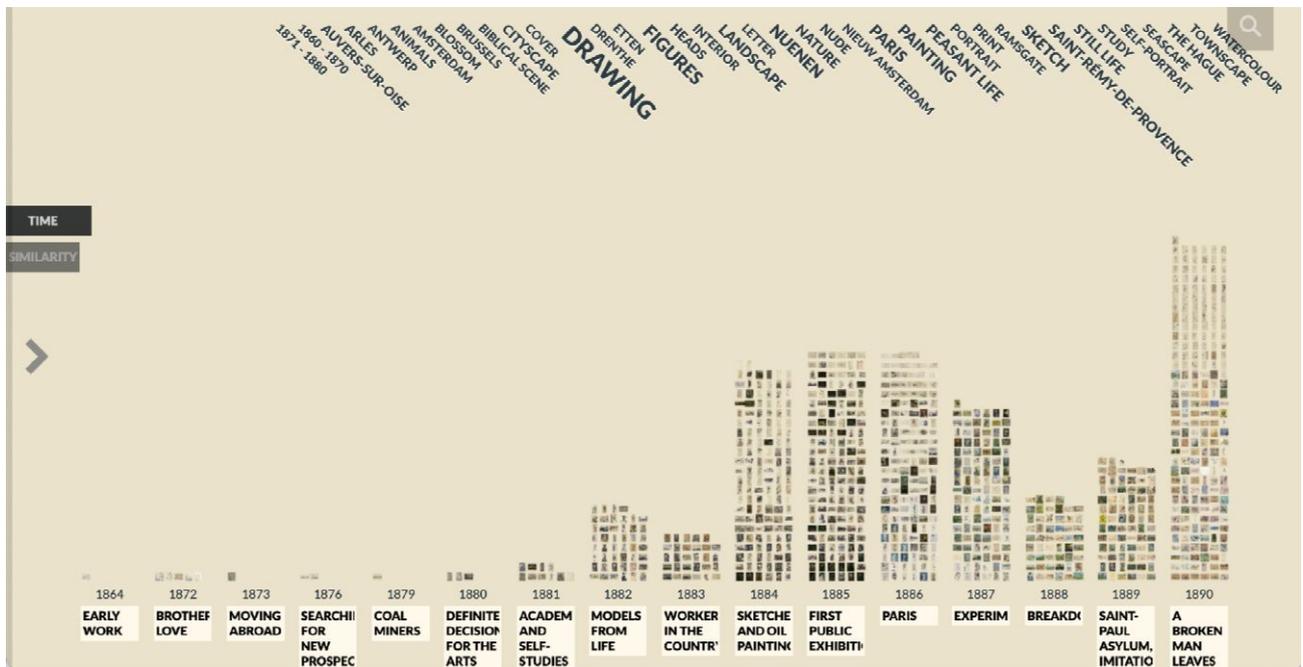


Abb. 15: Tabellarische chronologische Anordnungen nach Jahren und kunsthistorischen Perioden. Sichtbar werden, wie in einem Balkendiagramm, die in der Sammlung vorhandenen Mengen an Bildern mit der größten Anzahl im letzten Lebensjahr (Screenshot, 12.2.2022).

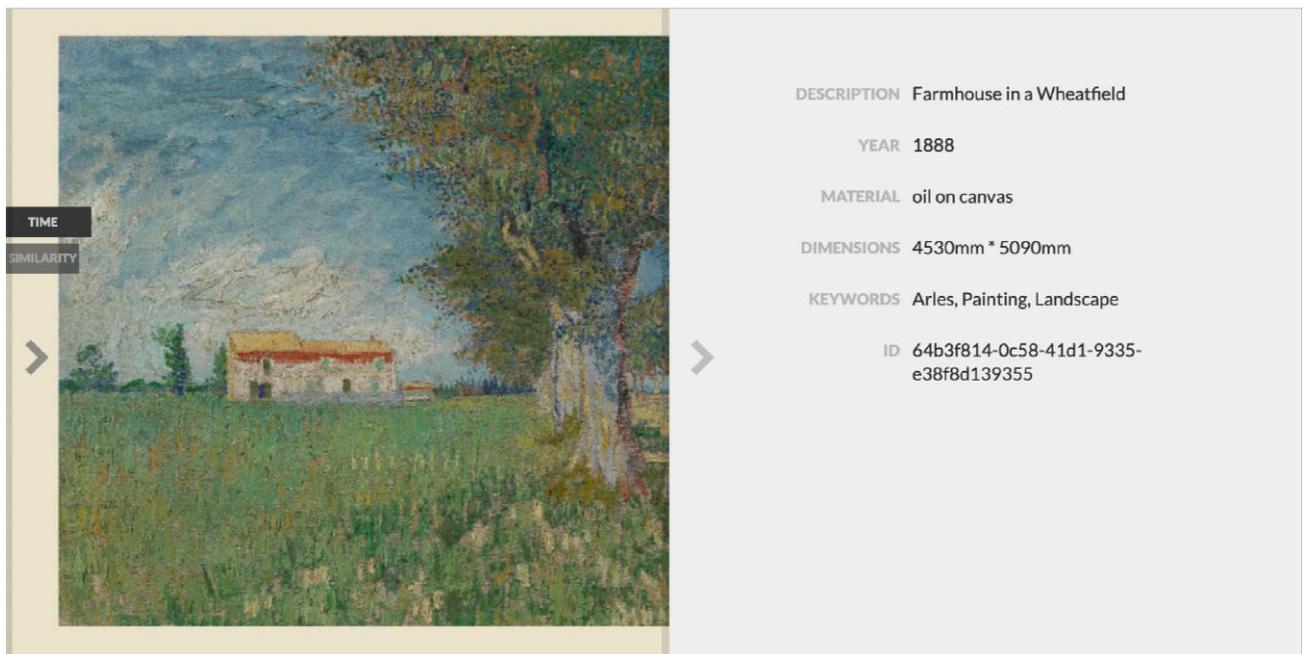


Abb. 16: Detailansicht und Metadaten eines einzelnen Bildes (Screenshot, 12.2.2022).

Die schräg gestellten Stichworte in der oberen Leiste funktionieren als einschließende Filter, sodass über die Jahre hinweg beispielsweise alle Bilder angezeigt werden können, deren Metadaten das Stichwort ‚Landschaft‘ enthalten. Auf ein einzelnes Bild kann geklickt werden, sodass durch Heranzoomen eine Großansicht des Bildes sowie der zugehörigen Metadaten erfolgt (Abb. 16).

Scatterplot

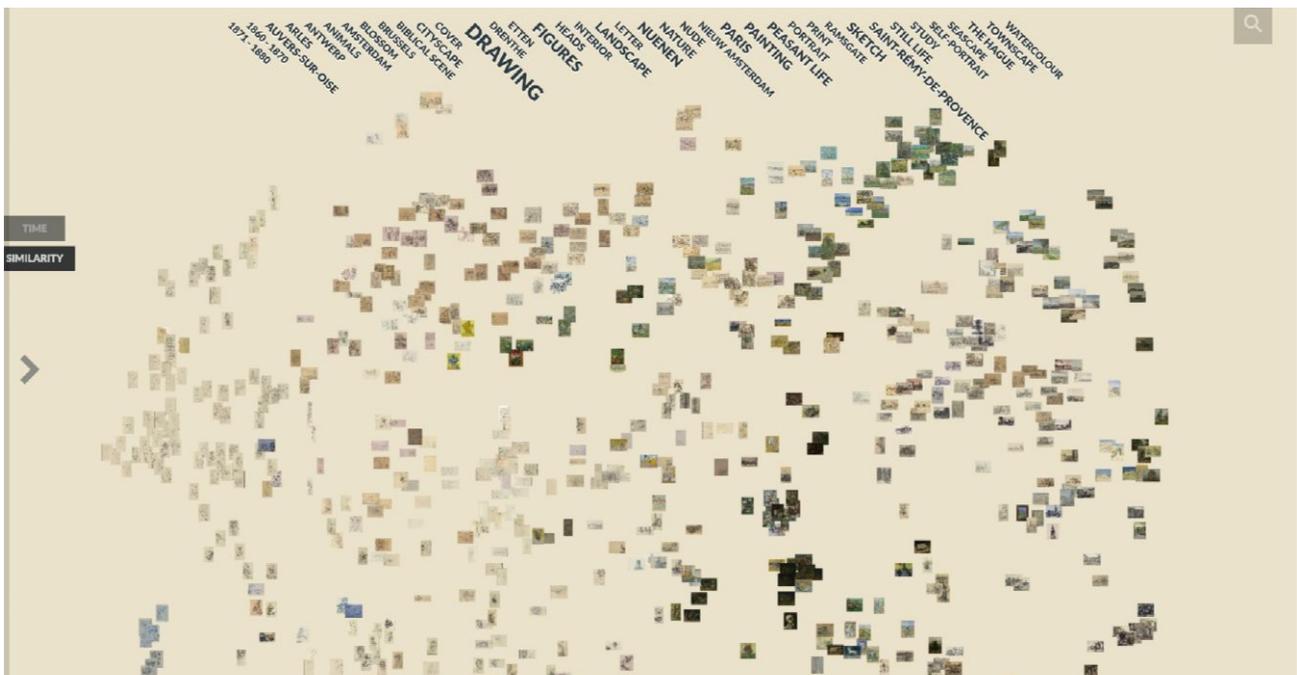


Abb. 17: Scatterplot-Darstellung nach Ähnlichkeit für den Van-Gogh-Datenkorpus (Screenshot, 12.2.2022).

Ein erster Blick auf den Scatterplot mit Gemälden und Zeichnungen Van Goghs zeigt eine durchaus sichtbare Verteilung auf, doch sehen wir sie bereits zu viel, zu klein, als dass sich genauere Aussagen treffen ließen (siehe Abb. 18). Durch Heranzoomen einzelner Regionen schält sich ein Cluster mit Skizzen – sich von links unten ausbreitend – heraus, welches nach oben hin zunehmend dunklere Tönung des Papiers aufweist. Ab der Mitte treten zunehmend Malereien auf, rechts oben ein Cluster von Landschaften. Diese werden oben rechts durch detaillierte Landschaftszeichnungen, die wesentlich ausgearbeiteter sind als die Skizzen, unterbrochen, ehe sich rechts ganz unten eine Reihe gemalter Portraits versammelt. Dieser erste Blick informiert die Betrachter*innen auch, dass sich in der Sammlung des Van Gogh Museums, Amsterdam, in der Mehrzahl Zeichnungen und Skizzen befinden und dass die Anzahl der farbigen Gemälde in der Unterzahl ist. Eine chronologische Ordnung findet nicht statt, wobei sich teilweise durch stilistische Ähnlichkeiten chronologische Cluster sammeln, zum Beispiel bei den umrisshaften Skizzen aus Van Goghs später Schaffensphase, ganz links. Die Bilddaten sind mittels des Aktivierungslayers des ImageNet-trainierten gewichteten Netzes Mobile-Net nach Ähnlichkeit verteilt.¹² Daher ist davon auszugehen, dass viele der

¹² Dieses Verfahren ähnelt dem ersten Prototypen von Training the Archive (Bönisch 2021).

Problematiken, die in dem Paper *"Why so many windows?" – Wie die Bilddatensammlung ImageNet die automatisierte Bilderkennung historischer Bilder beeinflusst* (Hunger 2021a) aufgezeigt wurden, zutreffend sind, und zwar in Bezug auf die in ImageNet vortrainierten Formen (welche Kunst an sich nicht kennen), die Ahistorizität von ImageNet und den von Geirhos et al. 2019 beschriebenen Texturbias.

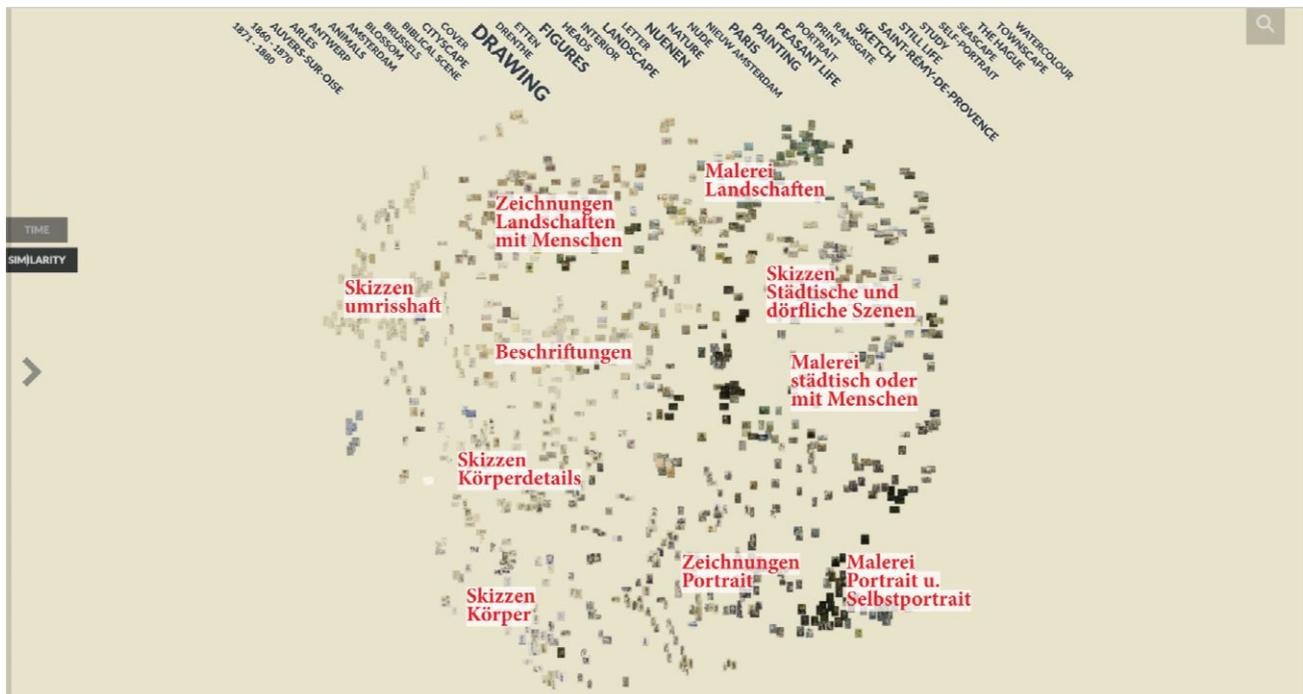


Abb. 18: Clustering der Van-Gogh-Sammlung mittels Vikus-Viewer (Die rote Beschriftung erfolgte durch den Autor, Screenshot, ebd.).

Geirhos et al. hatten für ImageNet-basierte gewichteten Netze ermittelt, dass die mathematischen Optimierungen bevorzugt auf Texturen reagieren. So wird beispielsweise ein mit Elefantenhaut überlagerte Katze als Elefant klassifiziert (Geirhos u. a. 2019, 1). Kurz gesagt: Textur, als Modus maschinistisch-statistischer Ähnlichkeit und nicht Umriss – wie in der menschlichen Wahrnehmung vorrangig – beeinflussen das Clustering. Das Clustering erfolgte basierend auf den aus ImageNet/Mobilenet extrahierten Aktivierungen mit Hilfe des t-SNE Algorithmus. Im Anschluss, um die Dichte des Interface zu verringern, fügte das Forschungsteam zusätzliche Abstände zwischen den Bildern hinzu, damit diese großzügiger verteilt sind und weniger visuelle Überlagerungen erzeugen. Zusätzlich zum maschinell-automatisierten Clustern der Bilddaten anhand statistischer Pixel-Ähnlichkeiten bietet Vikus Viewer den Modus der Überlagerung: Ähnlich den tabellarischen Darstellungen kann der Scatterplot mit kategorialen, von Menschen kuratierten Informationen angereichert werden. Bei Auswahl eines der kategorialen Stichworte aus der oberen Filterleiste werden die unzutreffenden Bilder im Scatterplot ausgeblendet (und in einem äußeren Kreis reorganisiert). Übrig bleiben am selben Ort die durch das Stichwort von Menschen kategorisierten Bilder, sodass das Verhältnis aus Clustering und Stichwort sichtbar wird. Diese Ansicht wirft Fragen der Ontologie der Stichwörter auf, zeigt aber auch, dass das maschinell-statistische Clustering zwar neue Ordnungen schafft, aber nicht notwendigerweise nach humanen Kriterien.

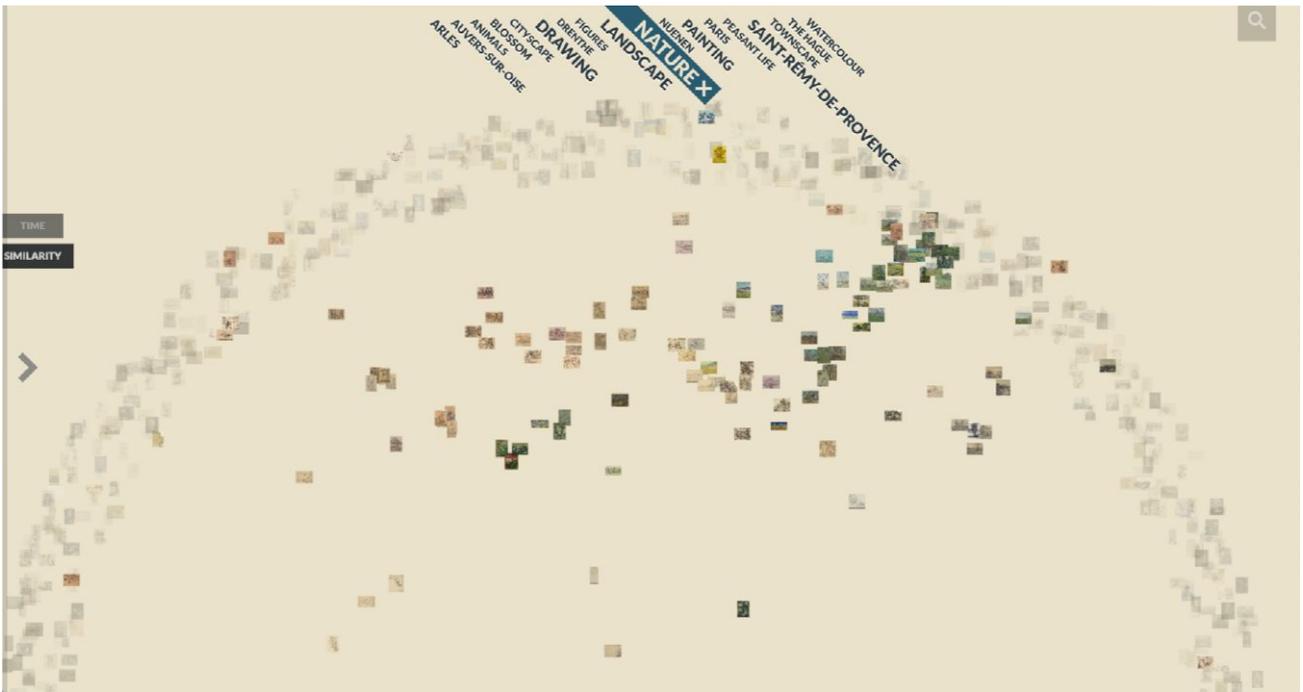


Abb. 19: Filter-Auswahl aller Bilder, die mit dem Schlüsselwort ‚nature‘ versehen wurden, in der nach Ähnlichkeit angeordneten Clusteransicht (Screenshot, 12.2.2022).

Eines der Potentiale, welches dieses frühe Projekt hat und welches wir beispielsweise in *Imgs.ai* realisiert sehen, ist eine Veränderung des Clustering anhand der Auswahl von negativen oder positiven Bildern. Ein anderes Potential wäre das manuelle Gruppieren bestimmter Bilder per Drag und Drop (ein Ändern ihrer Gewichte) und die maschinell-statistische Reaktion durch Neuberechnung aller anderen Gewichte und Positionen im Cluster.¹³

2.5 X Degrees of Separation als tabellarischer K-Nearest-Neighbor-Pfad

X Degrees of Separation (<https://artsexperiments.withgoogle.com/xdegrees/>) ist ein visuelles Experiment auf der Google-Arts-Plattform. Es ist die Kooperation des Künstlers Mario Klingemann mit Simon Doury (Google Cultural Institute), welche circa 250.000 Bilddatenobjekte aus Museums-Sammlungen (Stand 2017) als Teil des Google Arts & Culture Projekts, in Form von Big Data und Machine Learning verfügbar macht. Das Projekt ruft eine weitere Darstellungsform auf: den K-Nearest-Neighbor-Pfad. Im Unterschied zum Modus des Überblicks in Tabellen und Scatterplots verspricht diese Visualisierungsform ein ‚Close Reading‘ des Bildmaterials. Ziel ist es, mit Mitteln der Automatisierung eine Narration zu entfalten,

¹³ Für den Abschnitt des Vikus-Viewer gelten auch viele Aussagen, die im entsprechenden Abschnitt zu *iArts* Scatterplot und Clustering bereits vorgenommen wurden. Diese wurden an dieser Stelle nicht erneut ausgeführt.

welche die Betrachter*innen durch das umfangreiche und kaum zu überblickende Material führt und unerwartete Zusammenhänge sichtbar macht.¹⁴

Das Interface bietet eine Reihe von Bildern und Objektfotografien zur Auswahl, wobei eine Abbildung als Start und eine als Ziel ausgewählt werden kann, damit zwischen beiden ein Pfad der jeweils nächsten Nachbarn erstellt wird.

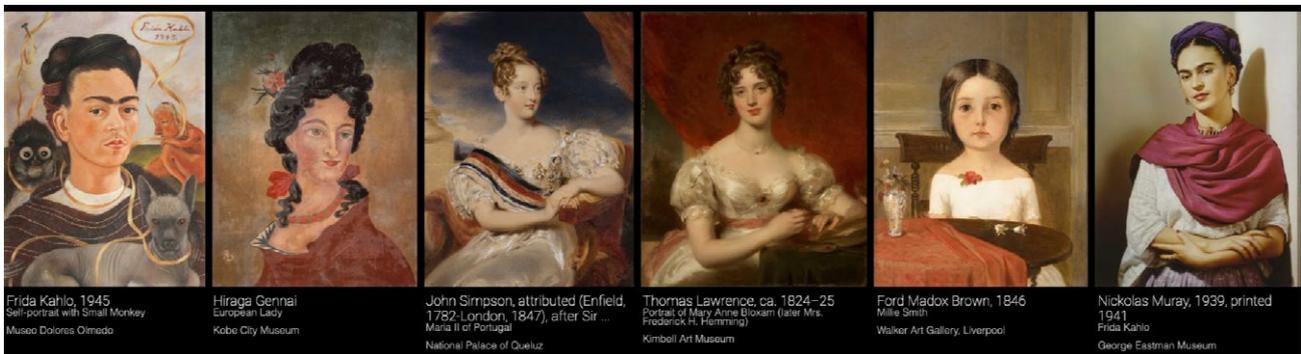


Abb. 20: Durchgang von einem Selbstbildnis Frida Kahlos von 1945 als Startpunkt zu einer Portraitfotografie der Künstlerin, 1941 (Screenshot, 12.2.2022).

In der tabellarischen Darstellung von X Degrees of Separation werden die Bildhöhen einander angepasst, sodass die ursprünglichen Größenverhältnisse verloren gehen. Der Algorithmus orientiert sich an Formenzusammenhängen wie Haar und Gesicht, Stoff mit Faltenwurf und relativ monochromatischen Hintergründen und liefert einen wilden Ritt durch die Kunstgeschichte, der ahistorisch und zusammenhanglos bleibt (Abb. 20). Kunsthistorisch wären Frida Kahlos Arbeiten beispielsweise in Zusammenhang oder in Abgrenzung zu den Mexikanischen Muralisten oder zum Surrealismus zu positionieren (Deffebach 2015), die hier jedoch überhaupt nicht auftreten. Jenseits dieser Feststellung könnte der Mehrwert in der zufälligen (Wieder-)Entdeckung korrespondierender künstlerischer Positionen liegen, doch müsste diese These an einem konkreten kuratorischen Vorhaben überprüft werden.



Abb. 21: Durchgang von einer Skulptur Tilman Riemenschneiders aus dem 15. Jahrhundert zu einem Frauenschuh des 20. Jahrhunderts (Screenshot, 12.2.2022).

¹⁴ Eine Implementierung mit PyTorch ist hier beschrieben: <https://d3lab.github.io/dl/2019/02/02/X-Degrees-Separation/>.

In einem zweiten Beispiel fällt als Konstante, welche die Abbildungen vermutlich zu K-Nearest-Neighbors macht, der helle, monochrome Hintergrund, der zur Freistellung der Objekte dient, auf (Abb. 21). Unter Verweis auf die Wirksamkeit von Textur-Bias (Geirhos u. a. 2019) ist zu vermuten, dass der Umriss wenig an der Verteilung der Bilder im latenten Raum der gewichteten Netze beteiligt ist. Dies fällt beispielsweise auf, wenn im vorletzten Bild der Schuh (als Umriss gedacht) nach rechts zeigt, im Zielbild hingegen nach links. Die Skulpturen zeichnen sich durch Faltenwürfe aus, die auch den Männerstiefel von Eugene C. Miller prägen – eine mögliche Erklärung für den Übergang. Es fällt auf, dass die Skulptur von Henry Glicenstein im Umriss ähnlich ausgerichtet ist, wie der darauf folgende Stiefel von Miller. Kunsthistorische Zusammenhänge sind nicht erkennbar.

Diese beiden Tests geben einen Einblick in die Potentiale und Grenzen eines K-Nearest-Neighbor-Experiments.¹⁵ Dieses ist sicherlich für publikumswirksame interaktive Projekte in Museumssammlungen geeignet, um weniger bekannte Aspekte einer Sammlung zugänglich zu machen. Eine Abwägung gilt es letztlich im Auge zu behalten: Explorationsprojekte müssen nicht zwangsläufig auf dem ressourcen-intensiven Machine Learning basieren. Eine technisch einfachere Möglichkeit zeigt das Projekt Never been Seen des Science Museum, London: Auf der Website <https://thesciencemuseum.github.io/never-been-seen/> wird per Zufallsgenerator das Digitalisat eines Sammlungsobjektes ausgewählt, welches zuvor noch nie online von einer Person angesehen wurde.

¹⁵ Ein weiterer K-Nearest-Neighbor-Ansatz, die Kooperation der FH Potsdam und der Staatlichen Museen Berlin *Visuelle Exploration zweier musealer Sammlungen* (<https://visualisierung.smb.museum>) ist ausführlich in (Brügge-mann u. a. 2022) beschrieben. Die hier am Beispiel von X Degrees of Separation aufgeworfenen Fragestellungen treffen auf die Pfaddarstellung in *Visuelle Exploration zweier Musealer Sammlungen* in ähnlicher Weise zu.

3 Schlussfolgerungen

3.1. Erkenntnis durch Anschauung

Interaktiv werden die Medienvisualisierungen, weil das menschliche Schauen-als-Denken fortwährend sinnerzeugende Ordnungen schafft, die, so die Medienphilosophin Sybille Krämer, räumlich ausgerichtet sind: „Indem in der diagrammatischen Einschreibung die Fläche als Ordnungs- und Anordnungsraum dient, werden topologische Verhältnisse wie oben/unten, links/rechts, zentral/randständig zu bedeutungskonstituierenden Komponenten. Es gibt keine Schrift ohne Lese- und Schreibrichtung, kein Diagramm ohne eine Ausrichtung [...]“ (Krämer 2010, 36). Zur Wirkungsweise diagrammatischer Anordnung zählt, dass die zweidimensionale Fläche eine simultane Präsenz ermöglicht, welche die Informationsobjekte homogenisiert. Krämer argumentiert, dass Menschen, ausgehend von ihrer eigenen Körperlichkeit, mittels Gerichtetheit Erkenntnisse gewinnen, also Sinn erzeugen durch Ordnungen, Anordnungen und Muster.

Der menschliche Erkennungsvorgang basiert wesentlich auf visuellen Ähnlichkeiten. Dafür erkennen Auge und Gehirn die Kanten von Objekten und die Komponenten, und zwar umso besser, je mehr Kontext vorhanden ist, der die Erkennung auch zuvor ungesehener Objekte erleichtert (Biederman 1987). Im Sehen-Erkennen-Prozess streift der Blick über die Fläche und Unterscheidungen oder Ähnlichkeiten können anhand von Nähe und Entfernung getroffen werden. Dabei wirken sich Ähnlichkeit und Nähe additiv auf die menschliche Erkennungsleistung aus (Kubovy und van den Berg 2008).

Die neuere Forschung geht von einem mehrstufigen Sehen-Erkennen aus, welches Momente des Encodierens, der Auswahl und des Decodierens visueller Informationen umfasst. Zentral ist, dass in der Auswahlstufe ein Großteil des Bildkontextes ausgeblendet wird – zugunsten einzelner Komponenten, auf welche sich das Gehirn in der Folge für das Decodieren konzentriert (Zhaoping 2019).

Zusammenfassend ist für Medienvisualisierungen zu konstatieren, dass der umherstreifende menschliche Blick in der Lage ist, mittels visueller Ähnlichkeiten und räumlicher Zuordnungen neues Wissen zu generieren und dass dieser kulturtechnische Vollzug des Schauens medial, historisch und kulturell geprägt ist. Daher ist das User-Interface von The Curator's Machine als Wissenswerkzeug, vergleichbar Latours ‚Labor‘ (Latour 1987, 63–101), an der Generierung neuen Wissens mitbeteiligt. Es bedarf einer gesonderten Aufmerksamkeit für den Versuchsaufbau des ‚Labors‘, also für die Gestaltung des User-Interfaces.

3.2 Die Grenzen des Clustering

Für den Autor überraschend deutlich treffen die hier im Material entwickelten Fragestellungen auf ein grundlegendes Problem von Big-Data und Künstlicher ‚Intelligenz‘: „Wissenschaftler*innen und andere verdrängen mit der Behauptung, dass Daten Entscheidungen trafen, unterschiedlichste Formen ideologisch beeinflusster Subjektivität, die stark an der Kuration von Datensätzen, der Auswahl verfügbarer Codes, Algorithmen, Parameter und

dem Labeln bekannter Daten[objekte] beteiligt sind“ (Dobson 2019). Je weiter die Untersuchung der Fallbeispiele voranschritt, desto mehr stellte sich die Frage, ob Scatterplots überhaupt in der Lage sind, etwas über die zugrundeliegenden Daten auszusagen.

Die Tests zeigten, dass zwar das Clustering anhand von Ähnlichkeit teilweise interpretierbar war, dass aber andere Informationen, beispielsweise Trends, die in klassischen Scatterplots durch lineare Regressionslinien und Beschriften der Skalen sichtbar werden, nicht visualisierbar werden. Daher bleiben die hier untersuchten Scatterplots häufig rätselhaft und sind schwer interpretierbar. Dieses Problem zieht sich grundsätzlich durch die von mir gesichteten Fallbeispiele und bedarf weiterer Forschung.

Diese Feststellung soll weiter zugespitzt werden. Der Informationsdesign-Theoretiker Edward Tufte vermerkte 1983 in seinem Standardwerk *The Visual Display of Quantitative Information* mit sprachlicher Dringlichkeit: „Grafiken dürfen Daten nicht ohne Kontext zitieren“ (Tufte 1983, 74). Doch die Frage, was der Kontext von Scatterplots ohne Skalenbeschriftung sei, bleibt offen, wie der Informationsdesigner Michael Correll in einem Blogpost polemisch anmerkt: „Niemand [...] weiß, was die Achsen des sich ergebenden Diagramms bedeuten, abgesehen von einem sehr einfachen vulgär-algorithmischen ‚diese Punkte liegen nahe beieinander, also nehme ich an, dass sie ähnlich sind‘, und man hätte diese Information direkt mit hierarchischem Clustering oder einer Nachbarschaftsmatrix oder etwas anderem als der obskuren Menge linearer Algebra oder dunkler Magie des maschinellen Lernens vermitteln können, die man anwenden musste, um dieses seltsame, uninterpretierbare Streudiagramm zu erstellen“ (Correll 2022). Tuftes Diktum und die Kritik Corrells lenken besondere Aufmerksamkeit auf die Frage, welchen grafischen Kontext heutige Visualisierungen gewichteter Netze überhaupt liefern.

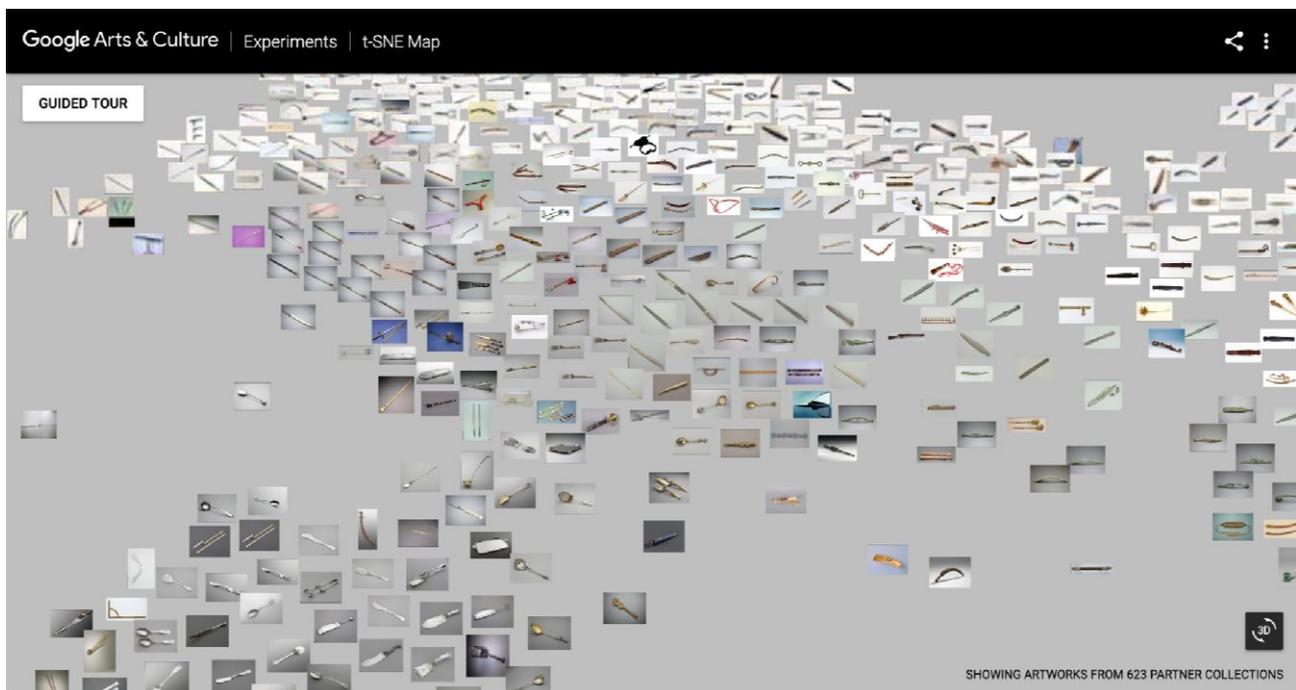


Abb. 22: Die willkürliche Ähnlichkeit des Clusterings zeigt sich am Beispiel des t-SNE-Map-Experiments, denn die fotografierten Gegenstände aus verschiedenen Sammlungen (Löffel, Werkzeuge, Instrument) werden nach der Farbe des Bildhintergrunds geordnet: oben weiß, in der Mitte grau-grünlich, unten grau-dunkelgrau (Screenshot Google Arts & Culture, t-SNE Map).

Das Clustering ohne Skalen gibt keine Aussage über die zwei Dimensionen, nach denen die Punkte des Clusters dargestellt werden (Sarıkaya und Gleicher 2018, 5). Es kann daher auf zwei Weisen gelesen werden: 1.) Intuitiv als eine Aussage über die Zueinandergehörigkeit der Datenpunkte und 2.) als Aussage über die innere Konfiguration der gewichteten Netze. Daraus ergibt sich ein Interpretationsproblem: Da keine Parameter angegeben werden, wissen die Betrachter*innen nicht, was sie überhaupt diagrammatisch präsentiert bekommen: Sind es die Zusammenhänge von Datenpunkten? Oder ist es der innere Aufbau der zugrundeliegenden gewichteten Netzwerke, der Convolutional ‚Neural‘ Networks?

Trotz dieser grundlegenden Zweifel ist nicht alles verloren. Denn Scatterplots dienen ganz hervorragend als Explorations-Interfaces. Sie vermögen in gewissem Maße den Nachteil tabellarischer Darstellungen auszugleichen, in welchen Zueinandergehörigkeit durch Reihenfolge ausgedrückt wird, welche wiederum durch willkürliche Zeilenumbrüche verfälschend dargestellt wird. Die Verteilungen von Scatterplots erlauben – ausgehend von einem Zentrum – räumliche Orientierungen in zahlreiche Leserichtungen und somit einen höheren visuellen Freiheitsgrad als die Orientierung in der Tabellenzeile, die in genau eine Richtung zeigt. Vorteilhaft zeigt sich die nahezu zufällige Anordnung von Ähnlichkeitsverteilungen in Scatterplots, wenn sie in Verbindungen mit kategorialen Ordnungsprinzipien (Filter, Suchfunktionen etc.) und visuellen Zoom-Möglichkeiten auftritt. Hier gelingt es, den inhärenten visuellen Problemen tabellarischer Darstellung, vor allem fehlerhaften visuellen Hierarchien, zu begegnen.

Nur bringen Scatterplots ihre ganz eigenen Probleme mit sich, denn es bleibt, ebenso wie in den zugrundeliegenden gewichteten Netzen (den Convolutional Neural Networks) intransparent, wie sich in ihnen ‚Ähnlichkeit‘ tatsächlich konstituiert. Der Erkenntnisgewinn aus ‚Ähnlichkeitsverteilungen‘ ist daher begrenzt. Andere visuelle Metaphern oder Ordnungsprinzipien sollten weiter erforscht werden. Diese müssten sich vermutlich auch stärker auf die kategorialen Ontologien beziehen, welche Sammlungsgegenständen zwar in der Digitalisierung eingeschrieben werden (als Metadaten), aber im multi-dimensionalen Raum gewichteter Netze (vor allem der Convolutional Neural Networks) ausgelöscht werden.

3.3 Praktische Folgen für die weitere Arbeit von Training the Archive

Die durch Training the Archive zu entwickelnde Curator’s Machine wird nur einem Teil der aufgeworfenen Probleme durch neue Lösungsansätze begegnen können. Für grundlegendere neue visuelle Lösungen bedarf es eines anderen und deutlich ausgedehnteren Forschungsdesigns. Im Folgenden sollen einige praktische Schlussfolgerungen und Vorschläge für Lösungsansätze identifiziert werden:

Konzepte: Die User-Interfaces, welche nicht-triviale Konzepte des Machine Learnings jenseits tabellarischer (und damit eingeübter) Darstellungen visualisieren, verwenden Konzepte wie

„K-Means-Clustering“, „Global Weights“¹⁶ oder „Distance: Angular/Manhattan“. Diese neuen Ordnungskonzepte werden in den User-Interfaces bisher nicht ausreichend erklärt und dokumentiert. Ihre Wirkung kann bestenfalls durch Ausprobieren ermittelt werden. Daraus folgt die Notwendigkeit einer besseren Dokumentation oder Erklärung dieser Optionen.

Diskriminierung: Bei einigen Projekten fiel auf, dass diskriminierende Sprache, so zum Beispiel die rassistischen Begriffe „Neger“, „Nigger“ oder sexistische Zuschreibungen wie „Slut“ bzw. „Schlampe“ einschlägige Treffer erzeugten. Dies traf sowohl Projekte mit historischen Ontologien, als auch Sprach-Embeddings mit Hilfe von CLIP. Aus den untersuchten Fallbeispielen waren allein die inhaltlich eng umrissenen und kuratierten Projekte des Vikus-Viewers vor diesen diskriminierenden Sprach- und Bildpolitiken gefeit.

Sprachembedding: Der Wechsel vom Englischen ins Deutsche erzeugt, wie aufgezeigt wurde, Medienwechsel, zum Beispiel von Malereien von Gebirgen zu Postkartenansichten von Gebirgen. Die Nutzer*innen sollten daher deutlicher über die Arbeitssprache der Embeddings informiert werden. Alternativ kann es auch sinnvoll sein, die Sprachen auf eine einzelne Sprache zu limitieren.

Korrelationslinie: Innerhalb der Cluster würde eine sichtbare Korrelationslinie Orientierung geben, wie stark die einzelnen Datenpunkte sich zum (statistischen) Mittel orientieren.

Ausreißer: Eine Filteransicht, die nur Ausreißer anzeigt, also Datenpunkte, die weit entfernt vom Mittel eines Clusters liegen, wäre hilfreich, damit sie einzeln inspiziert werden können.

Maßstab: Die Darstellungen der Bilder untereinander erfolgt nicht maßstabsgerecht, da alle Gegenstände an ein einheitliches Maß angepasst werden. Eine mögliche Lösung wäre, den Grad der Verkleinerung prozentual in den Bildern (als Overlay) anzugeben.

Drag & Drop und Gruppierung: Eine Funktion, die aus User*innen-Sicht in der Verteilungen-Darstellung wünschenswert wäre, ist die interaktive Sortierung und Gruppierung durch Drag-and-Drop. Dies entspräche in etwa der visuellen Suchfunktion des Projektes *Imgs.ai*, in welchem durch die positive oder negative Auswahl von Bildern diese eine Neuordnung des latenten Raums verursachen. Die Drag-and-Drop-Möglichkeit soll für *The Curator's Machine* nach derzeitigem Stand interaktiv mit der Neuberechnung visueller Ähnlichkeiten verschaltet werden und beispielsweise visuelle Gruppierungen ermöglichen. Darin würde sich die Funktionsweise von *The Curator's Machine* von anderen Projekten unterscheiden.

Hilfe/Maintenance: Die aufgeführten Projekte verfügen nur zum Teil über einen institutionalisierten technischen Support. Bei jenen Projekten, die sich noch im Experimentierstadium befanden, waren die Projektverantwortlichen zu erreichen, die je nach Verfügbarkeit antworteten. Daraus folgt für *Training the Archive* die Notwendigkeit, in Zusammenarbeit mit der RWTH eine dauerhafte Hostinglösung und institutionell verankerte technische Ansprechpartner*innen zu initiieren.

Interaktive Medienvisualisierungen wie die hier dargestellten ermöglichen ein Wechselspiel zwischen Wissensoperationen, indem die Sammlungsgegenstände als Datenobjekte (erste Schicht), operational (zweite Schicht) und über ein Graphical User Interface (dritte Schicht) durch Nutzer*inneneingriffe diagrammatisch interpretiert werden. Im Umkehrschluss lassen

¹⁶ Erklärt in der UI als „You can change weighting of plug-ins and thus adjust the order of results.“

sich User-Interfaces daraufhin ‚lesen‘, welche Wissensoperationen sie anbieten, um Relationen herzustellen: Auswählen, Filtern, Sortieren, Ordnen, Gruppieren, Summieren, Ähnlichkeiten vergleichen.

Tabellen, Scatterplots und K-Nearest-Neighbors – ihnen gemein ist die räumliche Verteilung als visuelles Argument, welches Wissen erzeugen soll. Gleichzeitig wurde deutlich, in welchem Grade es sich bei den Visualisierungen um Reduktionen und Rekonstruktionen von Wirklichkeit handelt. Diese führen bei Verwendung von gewichteten Netzen zu einem Interpretationsproblem, da beim Betrachten von skalenlosen Diagrammen unklar bleibt, welche Parameter überhaupt zueinander ins Verhältnis gesetzt wurden. Daher bleibt unklar, ob das, was die Betrachter*innen sehen, tatsächlich Ähnlichkeiten von Datenpunkten sind oder Interpretationen der Konfigurationen jener gewichteten Netzwerke, die zur Berechnung verwendet wurden.

Der vorliegende Text hat anhand zahlreicher Fallbeispiele ein kritisches Close-Reading von User-Interfaces durchgeführt. Die tabellarische Darstellung und die verteilten Medienvisualisierungen in Scatterplots standen im Zentrum dieser Diskussion. Aus den vorliegenden Überlegungen ergibt sich ein verstärkter Forschungsbedarf für die Visualisierung der Ergebnisse sogenannter Künstlicher Intelligenz.

Vielen Dank an Inke Arns, Dominik Bönisch, Roland Meyer und Kim Albrecht für die wertvollen Kommentare und Hinweise. Ein besonderer Dank gilt Andreas Broeckmann, der eine frühe Version kommentierte.

LITERATURVERZEICHNIS

- Biederman, Irving. 1987. Recognition-by-components: A theory of human image understanding. *Psychological Review* 94, Nr. 2 (April): 115–147. [doi:10/c9dqh2](https://doi.org/10.1037/0033-2909.94.2.115).
- Bönisch, Dominik. 2021. The Curator's Machine. Clustering von musealen Sammlungsdaten durch Annotieren verdeckter Beziehungsmuster zwischen Kunstwerken. *Training the Archive – Working Paper*, Aachen/Dortmund. [doi:10.5281/ZENODO.4604880](https://doi.org/10.5281/ZENODO.4604880).
- Bowker, Geoffrey C. 2014. The Theory/Data Thing Commentary. *International Journal of Communication* 8: 1795–1799.
- Bowker, Geoffrey C. und Susan Leigh Star. 1999. *Sorting Things Out – Classification and its Consequences*. Cambridge, MA: MIT Press.
- Brüggemann, Viktoria, Mark-Jan Bludau, Christopher Pietsch und Marian Dörk. 2022. Von der Wolke zum Pfad – Visuelle und assoziative Exploration zweier kultureller Sammlungen. *Zenodo*. [doi:10.5281/ZENODO.6327939](https://doi.org/10.5281/ZENODO.6327939).
- Bry, François, Clemens Schefels und Corina Schemainda. 2018. Eine qualitative Analyse der ARTigo-Annotationen: Serious Games, Games With A Purpose, Suchmaschine für Kunstwerke, Kunstgeschichte, Human Computation, semantische Tiefe. Hg. von Piotr Kuroczyński, Peter Bell, und Lisa Dieckmann. *Computing Art Reader*: 96–114. [doi:10.11588/ARTHISTORICUM.413.C5771](https://doi.org/10.11588/ARTHISTORICUM.413.C5771).
- Chun, Wendy Hui Kyong. 2021. *Discriminating Data – Correlation, Neighborhoods, and the new Politics of Recognition*. Cambridge, MA: MIT Press.
- Correll, Michael. 2022. Bad Ideas In Visualization. *Medium*. 6. Januar. <https://mcorrell.medium.com/bad-ideas-in-visualization-77d378148d35> (zugegriffen: 24. März 2022).
- D. Haggerty, Kevin und Richard V. Ericson. 2000. The surveillant assemblage. *British Journal of Sociology* 51, Nr. 4 (1. Dezember): 605–622. [doi:10.1080/00071310020015280](https://doi.org/10.1080/00071310020015280).
- Deffebach, Nancy. 2015. *María Izquierdo and Frida Kahlo: Challenging visions in modern Mexican art*. Latin American and Caribbean arts and culture. Austin, TX: University of Texas Press.
- Dobson, James E. 2019. *Critical Digital Humanities – The Search for a Methodology*. Topics in the digital humanities. Urbana, Illinois: University of Illinois Press.
- Few, Stephen. 2005. *Effectively Communicating Numbers – Selecting the Best Means and Manner of Display*. Proclarity White Paper. https://www.perceptualedge.com/articles/Whitepapers/Communicating_Numbers.pdf.
- . 2012. *Show me the numbers*. Berkeley, CA: Analytics Press.
- Friendly, Michael und Daniel Denis. 2005. The early origins and development of the scatterplot. *Journal of the History of the Behavioral Sciences* 41, Nr. 2: 103–130. [doi:10.1002/jhbs.20078](https://doi.org/10.1002/jhbs.20078).
- Galloway, Alexander R. 2012. *The Interface Effect*. Cambridge: Polity Press.
- Geirhos, Robert, Patricia Rubisch, Claudio Michaelis, Matthias Bethge, Felix A. Wichmann und Wieland Brendel. 2019. ImageNet-trained CNNs are biased towards texture – increasing shape bias improves accuracy and robustness. *arXiv:1811.12231* (14. Januar). <http://arxiv.org/abs/1811.12231>.
- Gitelman, Lisa. 2013. *Raw data is an Oxymoron*. Cambridge, MA: MIT Press.
- Glinka, Katrin, Christopher Pietsch und Marian Dörk. 2017. Von sammlungsspezifischen Visualisierungen zu nachnutzbaren Werkzeugen. In: *Books of Abstracts*, hg. von Michael Stolz und Patrick Helling, 204–207. Bern. <https://uclab.fh-potsdam.de/wp/wp-content/uploads/dhd2017-aus-programm.pdf> (zugegriffen: 18. März 2022).
- Hunger, Francis. 2021a. „Why so many windows?“ – Wie die Bilddatensammlung ImageNet die automatisierte Bilderkennung historischer Bilder beeinflusst. *Training the Archive – Working Paper 2*, Aachen/Dortmund (1. Juni). [doi:10.5281/ZENODO.4742621](https://doi.org/10.5281/ZENODO.4742621).
- . 2021b. Curation and its Statistical Automation by means of Artificial „Intelligence“. *Training the Archive – Working Paper*, Aachen/Dortmund (16. November). [doi:10/gn8dn7](https://doi.org/10.5281/ZENODO.4742621).
- . 2022. *Die Form der Datenbank – Genealogien, Operationalitäten und Praxeologien relationaler Datenbanken in Ost und West*. Dissertation, Weimar: Bauhaus Universität Weimar [in Vorbereitung].
- König, Dénes. 1936. *Theorie der endlichen und unendlichen Graphen*. Hg. von Horst Sachs und H. Sachs. 1936 Reprint. Teubner-Archiv zur Mathematik 6. Leipzig: Teubner.

- Krämer, Sybille. 2010. Notationen Schemata und Diagramme – »Räumlichkeit« als Darstellungsprinzip. In: Notationen und choreographisches Denken., hg. von Gabriele Brandstetter, Franck Hofmann, und Kirsten Maar, 29–45. Freiburg/Berlin/Wien: Rombach Verlag.
- Krämer, Sybille und Horst Bredekamp. 2003. Kultur, Technik, Kulturtechnik – Wider die Diskursivierung der Kultur. In: Bild, Schrift, Zahl., hg. von Sybille Krämer und Horst Bredekamp, 11–22. Kulturtechnik. München: Fink.
- Kubovy, Michael und Martin van den Berg. 2008. The whole is equal to the sum of its parts – A probabilistic model of grouping by proximity and similarity in regular patterns. *Psychological Review* 115, Nr. 1 (Januar): 131–154. [doi:10/b2vw3s](https://doi.org/10.1037/a0012333).
- Latour, Bruno. 1987. *Science in Action – How to follow Scientists and Engineers through Society*. Cambridge, MA: Harvard University Press.
- Manovich, Lev. 2020. *Cultural analytics*. Cambridge, MA: The MIT Press.
- McInnes, Leland, John Healy und James Melville. 2020. UMAP – Uniform Manifold Approximation and Projection for Dimension Reduction. arXiv:1802.03426 (17. September). [http://arxiv.org/abs/1802.03426](https://arxiv.org/abs/1802.03426).
- Radford, Alec, Jong Wook Kim, Chris Hallacy, Aditya Ramesh, Gabriel Goh, Sandhini Agarwal, Girish Sastry, u. a. 2021. Learning Transferable Visual Models From Natural Language Supervision. arXiv, 26. Februar. [http://arxiv.org/abs/2103.00020](https://arxiv.org/abs/2103.00020).
- Russell, Henry Norris. 1914. Relations between the Spectra and Other Characteristics of Stars. In: *Popular Astronomy*, 22:275–294. Mai.
- Sarikaya, Alper und Michael Gleicher. 2018. Scatterplots – Tasks, Data, and Designs. *IEEE Transactions on Visualization and Computer Graphics* 24, Nr. 1 (Januar): 402–412. [doi:10.1109/TVCG.2017.2744184](https://doi.org/10.1109/TVCG.2017.2744184).
- Schneider, Stefanie und Hubertus Kohle. 2021. Bilderlesemaschinen – Einsatzszenarien maschineller Lernverfahren im musealen Raum. 28. Mai, *Museum and the Internet (MAI) 2021 Tagung*. <https://www.youtube.com/watch?v=lsSpzjtS-rM> (zugegriffen: 5. Januar 2022).
- Simonyan, Karen und Andrew Zisserman. 2015. Very Deep Convolutional Networks for Large-Scale Image Recognition. arXiv:1409.1556 [cs] (10. April). [http://arxiv.org/abs/1409.1556](https://arxiv.org/abs/1409.1556).
- Springstein, Matthias. 2021. AW: iArt Fragen. E-Mail an den Autoren, 14. Dezember.
- Springstein, Matthias, Stefanie Schneider, Javad Rahnama, Eyke Hüllermeier, Hubertus Kohle und Ralph Ewerth. 2021. iART – A Search Engine for Art-Historical Images to Support Research in the Humanities. arXiv:2108.01542 (3. August). [http://arxiv.org/abs/2108.01542](https://arxiv.org/abs/2108.01542).
- Taylor, Frederick Winslow. 1903. *Shop Management*. London, New York, NY: Harper & Brothers.
- Toshev, Alexander und Christian Szegedy. 2014. DeepPose – Human Pose Estimation via Deep Neural Networks. 2014 IEEE Conference on Computer Vision and Pattern Recognition (Juni): 1653–1660. [doi:10/ggtmv3](https://doi.org/10.1109/cvpr.2014.223).
- Tufte, Edward R. 1983. *The Visual Display of Quantitative Information*. Cheshire: Graphics Press.
- Wilke, C. 2019. *Fundamentals of Data Visualization – A Primer on making Informative and Compelling Figures*. Sebastopol, CA: O'Reilly Media.
- Zhaoping, Li. 2019. A new framework for understanding vision from the perspective of the primary visual cortex. *Current Opinion in Neurobiology* 58 (Oktober): 1–10. [doi:10/ghndbt](https://doi.org/10.1016/j.conb.2019.08.001).