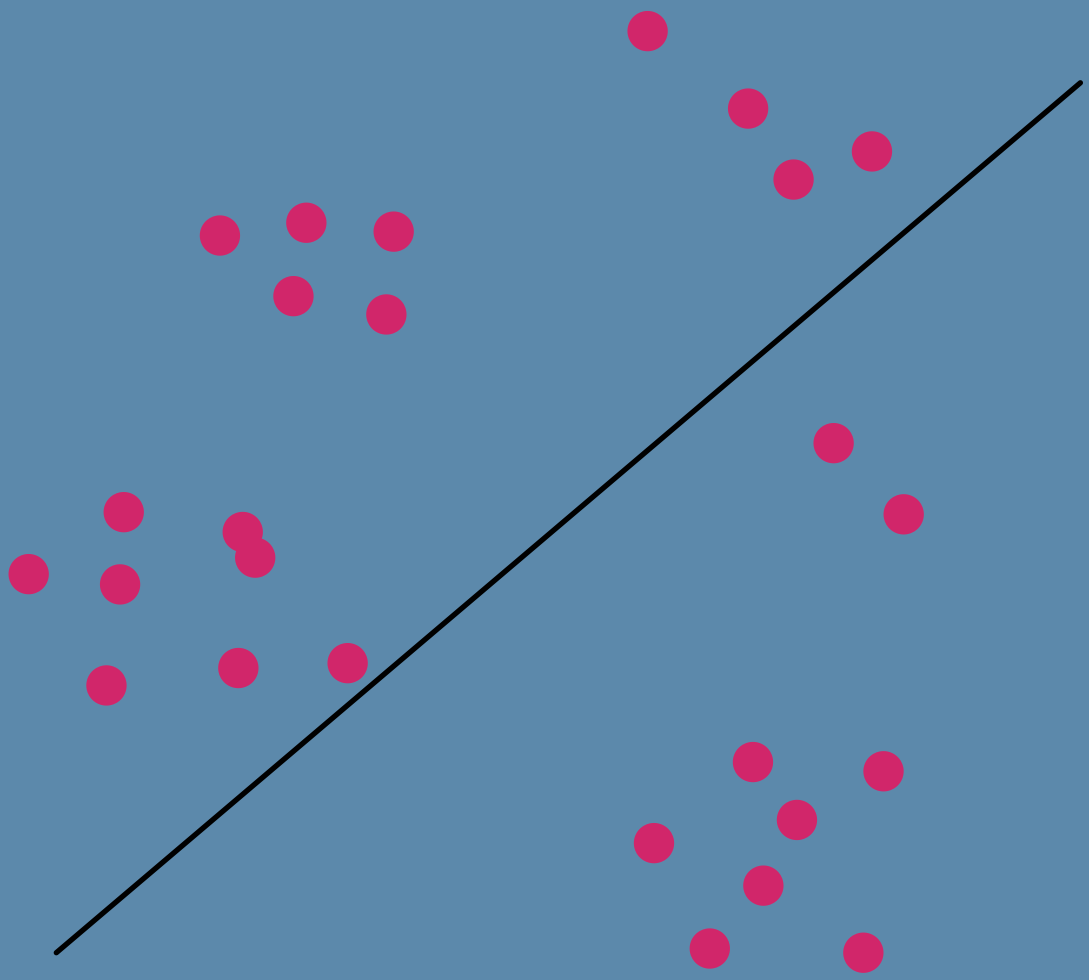
Training the Archive – Working Paper 5



Francis Hunger

Point Clouds. Scatterplots and Tables as User Interfaces of Artificial 'Intelligence'

Training the Archive – Working Paper Series 5
Francis Hunger – Point Clouds. Scatterplots and Tables as User Interfaces of Artificial 'Intelligence'.
Training the Archive (ed.), Aachen/Dortmund, January 2023
DOI: 10.5281/zenodo.7554939

Training the Archive

Ludwig Forum for International Art Aachen
Jülicher Str. 97-109, 52070 Aachen
https://ludwigforum.de/

Ludwig Forum für Internationale Kunst Aachen

Hartware MedienKunstVerein
Büro: Hoher Wall 15, 44137 Dortmund
www.hmkv.de

HMKV Hartware MedienKunstVerein

Funded by the *Digital Culture* programme of the German Federal Cultural Foundation

Kultur Digital KULTURSTIFTUNG DES BUNDES

Funded by the Federal Government Commissioner for Culture and Media

Die Beauftragte der Bundesregierung für Kultur und Medien

Working Paper 5: Point Clouds. Scatterplots and Tables as User Interfaces of Artificial 'Intelligence'.

> "Categories are central to being in the world. Big data does not do away with categories at all" (Bowker 2014, 1797).

> "Cultural techniques are promoting the achievements of intelligence through the senses and the externalizing operationalization of thought processes" (Krämer and Bredekamp 2003, 18).

Abstract

Scatterplots and tabular structures are the main graphical user interfaces for the diagrammatic depiction of large image data collections, which have been processed with visual recognition tools. This study discusses various media visualisations as case examples: ARTigo (LMU Munich), Imgs.ai (UC Santa Barbara/Bildarchiv Foto Marburg), iArt (Universities of Munich and Hannover), Vikus Viewer (FH Potsdam), and X Degrees of Separation (Google). It further investigates how these projects employ visualisation algorithms like PCA, t-SNE and UMAP in relation to artificial weighted networks such as VGG19 or CLIP and provides practical proposals how to deal with the identified problems.

---

1 Introduction

This working paper evaluates visualisation concepts for The Curator's Machine, a machine learning software module that is being developed in the scope of the research project Training the Archive. The Curator's Machine is aimed at automating curatorial decisions for large image sets with the help of pattern recognition and artificial intelligence (Bönisch 2021; Hunger 2021a, 2021b).

The results of machine learning processes are arranged in user interfaces based on diagrammatic principles. Building on the thesis that spatial relationships dominate knowledge formation in diagrams, the task of this text is to investigate visualisations used in machine learning. Thus, the text develops a syntax of visualisations that identifies several diagrammatic types essential for Training the Archive: the table, the scale diagram and scatterplot, and the k-nearest-neighbours method. A further section describes three layers of reduction and reconstruction of reality in visualisations that stem from weighted networks (convolutional neural networks).[1] The three layers include 1.) data collection, 2.) statistics/computation and 3.) visualisation. These introductory considerations are followed by a central section with case studies that I have chosen due to their proximity to our project in terms of content and technology. The user interface illustrations examine ARTigo, an art-historical search engine (LMU Munich), the experimental art-historical image search Imgs.ai (UC Santa Barbara/ Bildarchiv Foto Marburg), the interactive tool for analysing image datasets iArt (Universities of Munich and Hannover), the project Vikus Viewer with curated datasets (FH Potsdam), and finally the path-oriented image experiment X Degrees of Separation (Google). I then draw conclusions for the prototype of Training the Archive, which can be generalised for similar user interfaces. Before the text explores the syntax of media visualisations, two relevant concepts must be introduced and located: diagrams and media visualisations.

1.) Diagrams constitue visual arguments. They serve to visualise facts by placing graphic elements in spatial relationships that can be 'read'. Diagrams are often hybrids of textual and graphical elements. The strength of text-graphic hybrids is that they can present numbers and categories, that is quantitative and categorical information objects from different knowledge domains, combined in one visualisation. This unification in two-dimensional spatiality is one of humanity's most important tools of cognition (Krämer 2010). The goals of the interfaces are to validate or falsify hypotheses and to provoke new research questions. Observing how the design itself – the interface and diagrammatic operation – influences these goals is therefore worthwhile.

Human reading, seeing and understanding of diagrams is based on visual hierarchies. In a visual hierarchy, I interpret 'top' as better compared to 'bottom'. First comes before second, so I read from the top down. I also read left before right because the eye finds its hold on the left text border. The hierarchies listed here are examples from the author's own linguistic sphere. In other cultural and historical contexts, the hierarchies are different, such as in Arabic or Hebrew, which are read from right to left. The cultural techniques of comprehending reading and writing however, have one thing in common: the spatial organisation of what is recorded and the often unspoken, culturally and technically learned hierarchies and orders permeate reading and what is read.

In the following paper, I examine those diagrammatic representations that make image information accessible as a visual argument using examples from the field of artificial 'intelligence', computer vision and data visualisation. This study therefore is positioned as part of interface criticism, which understands interfaces not only as a transition between 'the inside' and 'the outside', but also as "autonomous zones of activity" (Galloway 2012, vii).

2.) The visual methods discussed here are "media visualisations" (Manovich 2020, 215), because they explore visual patterns in large amounts of data. The spatial arrangements of data objects create visual arguments. But most importantly, media visualisations do not represent

[1] To avoid the anthropomorphising term 'neural network', I use 'weighted network' as an alternative. For more on this and in general for the technological procedures of machine learning in 'Convolutional Neural Networks' see (Hunger 2021a, 2–7, 9 and FN 1).

data objects in an abstracted way, such as by points, graphemes or numerical representation, but rather as icons, that is as reduced, visual references to the original data objects. Media visualisations are also characterised by interactivity and the availability of different scales. In many user interfaces, it is possible to zoom into the image collections or individual images. This constant oscillation between detail and overview allows new knowledge to emerge in ever new visual configurations.
User interfaces that display image collections, on websites or AI software for example, form the image datasets algorithmically into ordered arrangements. In these sets, the individual image vanishes in favour of the overview. The referentiality of the individual image is no longer of interest, but rather the synopsis of the relations between the images. However, the display of mass images, which now appear as operative images, comes with its own problems. These will be explored in the case studies.

1.1 Syntax of media visualisations

The following section discusses the visualisation types 'tables', 'scale diagrams and scatter-plot' and 'k-nearest neighbour', which are often used for media visualisations. We can look into them to identify certain strengths and weaknesses of the visualisations for generating knowledge. A number of other visualisations, such as network diagrams, maps, infographics and others are not discussed here because they do not apply to the visualisations in the framework of The Curator's Machine.

Tables

Tables provide an overview and allow the entries to be recorded, looked up and operationalised. By entering the data in rows and columns and through the operations of grouping, sorting and summing, the data is brought into ever new constellations of knowledge, so that it serves people as a cognition tool for producing new knowledge. We can distinguish between statistical, mathematical, transactional/process-related, and knowledge-building tables.
Statistical tables owe their rise to the development of nation states from the 17th century onwards, in which state structures were systematically underpinned by recording population, production and the budget. For this purpose, data was intentionally produced by collecting observations and transferring them into tabular grids.
Mathematical tables comprise, on the one hand, data obtained from empirical observation, for example the orbit of celestial bodies, and, on the other, calculated series of numbers that could be used for further calculations, such as square numbers, interest and compound interest, logarithms and trigonometric functions. The development of calculating machines and computers replaced mathematical tables. They enable these calculations to be made ad hoc or stored in spreadsheets and databases.
Knowledge-building tables assemble pieces of knowledge in the form of an overview. They allow knowledge on topics such as art-historical, secular-historical or political events to be arranged chronologically or categorically and served on the one hand as a tool for rote learn-

ing and on the other for generating new knowledge. They tie in with the new knowledge order of the emerging museums, academies and encyclopaedias from the 16th century onwards, which established our present-day science and knowledge ordering systems.
Transactional tables have become fundamental management and process tools (see Fig. 1 for an example). They represent current statuses as 'transactions' and were historically used for activities such as changing account balances at banks, organising internal work in companies, and the coordinated processing of services, such as the combination of flight and passenger data at airlines. The transaction data recorded in them describe the respective current status of an object or procedure and serve to ensure the controlled processing of a transaction from its defined beginning to its defined end.

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 machined							
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

Fig. 1: Transactional table for the production of steel tyres for vehicles in the Midvale Steel Works 1883. The respective transaction can be addressed via the order no. The first column shows the workflow of the transaction from beginning to end (Taylor 1903, 86).

Tables are among the cultural techniques of knowledge production: their use must be learned.

The following visual arguments characterise the table: Firstly, the tabular grid creates a construct that makes entering data possible at all. Secondly, the reading eye jumps between rows and columns to gain an overview of the individual information objects (the row) according to different categories (per column). Thirdly, empty space indicates missing data and data yet to be generated, and fourthly, the spatial relationships within the table change through rearrangement (filtering, ordering, grouping) of the data in the grid (Hunger 2022, 76–128). Tables thus enable a unifying, common thinking space us and allow us to combine both categorical/qualitative and quantitative data in a uniform grid.

Scale diagrams and scatterplots

Scale diagrams, such as bar charts and scatterplots, consist of graphic elements (lines, rectangles or bars, points) distributed in a two-dimensional space, which is spanned by two scales with quantitative or qualitative values (X-axis and Y-axis). They facilitate the representation of time series, rankings, percentage distribution, deviations in relation to the zero value and distributions of the total values (e.g. bell curve as normal distribution). The scales usually have even spacing so that a grid is created (Wilke 2019, 13–25).
The visual arguments of scale diagrams are oriented towards the alternating viewing of the scales and the values represented by graphic elements. The graphic elements allow a number of patterns to be identified as visual relations: focal points, distributions, trends and tendencies, comparisons. Elements can be grouped by colour and thus additionally dimensioned beyond the two-dimensional distribution (Tufte 1983, 28–39; Few 2005, 2012, 87–136). The experiential knowledge of graphic design has led to numerous ways of using diagrams. Quantities, for instance, are most easily compared by means of bar diagrams. If not only numerical but also categorical qualities need to be compared, then we can use grouped and stacked bar charts. For visualising distributions, we prefer histograms, density plots or violin plots. For proportions that can be indicated by percentages, we use pie charts and bar charts. Relations can be displayed as scatterplots, bubble charts and slope graphs, as matrixes or correlograms. For geographical data, maps and also geographical heat maps are suitable (Wilke 2019, 37–44).
In principle, scale diagrams also include scatterplots, which show distributions within a dataset. Why in principle? As we will see later, the user interfaces examined here dispense with the scales. This is accompanied by a loss of information, as meanings that were previously contained in the scales become less clear. Before we delve into this, however, we turn to 'classic' scatterplots, i.e. those with scales.
Scatterplots highlight correlations, clusters, patterns, trends and outliers in a set of data points by organising related points into aggregations, spatially close to one another. "Central to many of these techniques is preserving the meaning of distance between objects as an indicator of similarity" (Sarikaya and Gleicher 2018, 2). In this context, similarity means that several data objects are alike in at least one characteristic, whereby on the one hand an overall similarity can be addressed, and on the other a partial similarity in which data objects are similar in certain individual aspects. However, the visual features of scatterplots are not complete 'proofs' in the mathematical sense, but rather serve as proxies for correlations. Scatter-

plots have been defined as “a plot of two variables, x and y, measured independently to produce bivariate pairs (xi, y), and displayed as individual points on a coordinate grid typically defined by horizontal and vertical axes, where there is no necessary functional relation between x and y” (Friendly and Denis 2005, 105). In other words, the relationship between data is clearly visible in a table or bar chart when x and y are linearly correlated. However, if – as in the practical examples discussed below – the function between x and y is mathematically complex, or if x is correlated to several other parameters, then this relationship is difficult to ‘read’ in a tabular display. Scatterplots have an advantage here because their data points can be evaluated as proxies for correlations by viewing them and can lead to new insights.

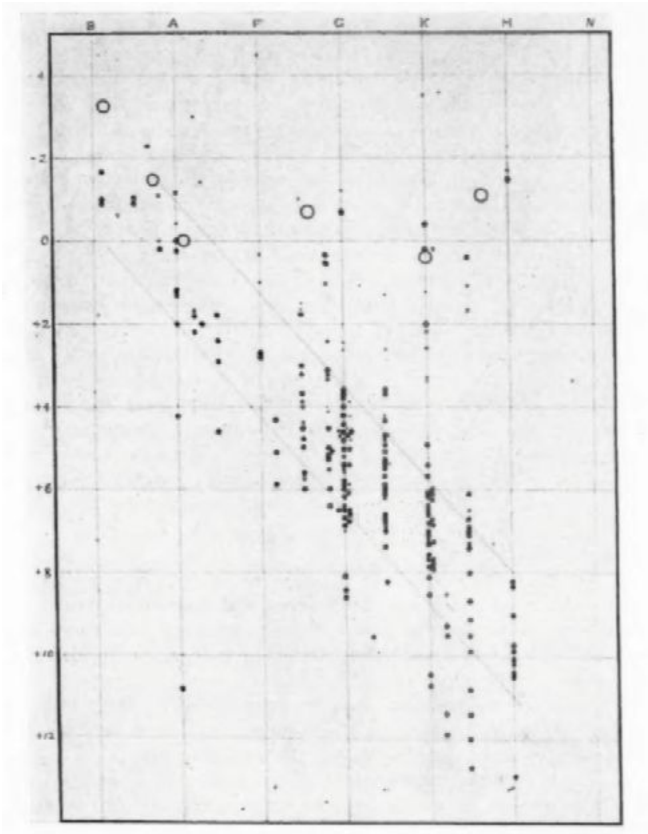


Fig. 2: Historical scatterplot: “The spectral class [B, A, F, G, K, M, N – relating to the color spectrum of the perceived light, F.H.] appears as the horizontal coordinate, while the vertical one is the […] visual magnitude [measure of light energy, F.H.] which each star would appear to have if it should be brought up to a standard distance, corresponding to a parallax of 0.1“ (Russell 1914, 285). Giant stars are at the top and dwarf stars are more likely to be at the bottom right.

Genealogies trace the development of scatterplots back to the naturalist and eugenicist Francis Galton, who, starting in the 1880s, demonstrated the phenomena of regression and statistical correlation through tabular visualisation (Friendly and Denis 2005, 109–113).[2] While it would be going too far to list the many contributions of other statisticians and graphic artists, we should at least mention the astronomer John Frederick W. Herschel,

[2] Media theorist Wendy Chun draws attention to the problematic genealogy of correlation and linear regression as eugenic methods developed by Francis Galton (see Chun 2021, 59–66).

who used a scatterplot, i.e. a graphical solution, to predict the elliptical orbits of double stars despite incomplete data. The terms ‘scatter diagram’ and ‘scatter plot’ came into increasing use between 1906 and 1920, when the technique was described in textbooks on statistics and data visualisation (Friendly and Denis 2005, 116–119).

The density of the respective clusters shows the degree of correlation within the respective clusters. More widely scattered data points indicate a lower correlation between the data points. Correlation lines (regression analysis) allow researchers to see trends, use them to make predictions when appropriate, and visualise deviations from these trends. “As opposed to other graphic forms – pie charts, line graphs, and bar charts – the scatterplot offered a unique advantage: the possibility to discover regularity in empirical data (shown as points) by adding smoothed lines or curves designed to pass “not through, but among them,” so as to pass from raw data to a theory-based description, analysis, and understanding” (Friendly and Denis 2005, 128).

The scatter plot above by astronomer Henry Norris Russel, for instance, shows from left to right that the majority of stars in classes A and B (low spectral colour) are very bright. Most stars in classes K and M are dimmer and in the reddish spectrum (higher spectral colour). In the middle of the diagram are red stars of high brightness. Blue and red super giants and red and white dwarf stars are outliers.

Aggregations thus create a visual order similar to grouping, although the delineation of groups is not as sharp as it is in the tabular grid where columns and rows delineate the data. Grouping allows a summary of data as ‘AND’ or ‘OR’, so that visual cognitive operation places the knowledge objects in groups based on similarities, but not necessarily congruency. In this way, focal points, distributions, tendencies, comparisons and outliers can be displayed.

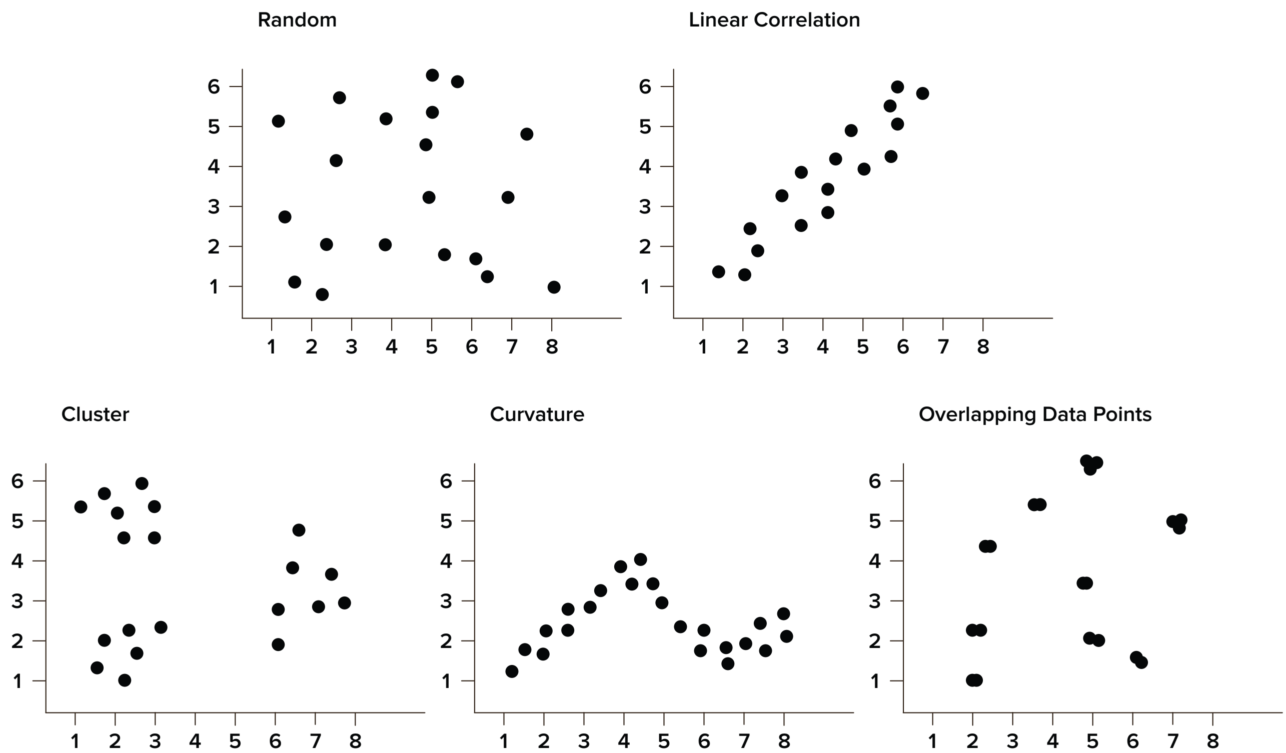


Fig. 3: Five typical distributions in scatterplots (quoted after Sarikaya and Gleicher 2018, 5).

Computer visualisation experts Alper Sarikaya and Michael Gleicher have identified a number of cognitive operations with scatterplots: the identification and location of an object, the comparison of objects (belonging to object classes, for example 'furnishing object: desk', or 'art historical epoch: Impressionism'), an exploration of the neighbourhood and the data, the search for a pattern (cluster, correlation) and its characterisation (dense, widely scattered, etc.), and the identification of anomalies and outliers (Sarikaya and Gleicher 2018, 3). They also show five typified distributions in scatterplots, which are commonly encountered: 1. random distribution, 2. linear correlation, 3. cluster formation, 4. curvature (e.g. in the form of a sine curve), and 5. overlapping data points (See Fig. 3).

K-nearest neighbours

Based on scatterplots, clustering and path displays are playing an increasing role in visual representations of (art) collections by means of machine learning. In the process, the user interface makes the affiliation of data objects visible through common colouring or network lines connecting individual data points.[3] The k-nearest neighbours result from mathematical similarity, which is made visible through the proximity of data objects to each other. K indicates how many classificatory subsets the total data are divided into, i.e. K=3 results in three clusters (aggregations).
The most common uses of the k-nearest neighbour algorithm are currently clustering and path representation:

1.) Clustering: Clustering is in turn divided into two different methods. For one, media visualisation can be achieved via discretised, 'structured' data, i.e. images labelled with keywords. In this process, previously labelled known images are used as the basis for locating unknown images in the multi-dimensional latent space of weighted ('neural') networks. Depending on which feature space the unknown images are closest to, they also receive the corresponding label. In clustering, i.e. classifying images into three clusters for example, one of the labels 'cat', 'dog' or 'horse' is spatially assigned.
In addition to discrete assignments, we can use k-nearest neighbour algorithms for 'continuous regression' in similarity calculations. In this case, rather than assigning a discrete label, a continuous number space, a graph, is spanned based on the comparison data points. The location of a data point is determined in relation to the k-nearest neighbours, i.e. when considering k=3, from the ratio, namely the mean value, to the three nearest neighbours. The result is a numerical value that indicates the degree of similarity in relation to the neighbours, which in turn is used for the distribution in a scatterplot (Fig. 13).

2.) Path: The principle of the nearest neighbour in the multidimensional space of weighted networks is used to construct a path leading from start to destination between these image data. Historically, this procedure goes back to the bridge problem, which Leonhard Euler

[3] James E. Dobson develops an insightful genealogy of the k-nearest neighbours algorithm in the chapter *The Cultural Significance of k-NN* (Dobson 2019). He describes k-nearest neighbours as a means that allowed classification to be derived directly from data, in apparent elimination of subjective human categorisation.

first described in 1736. The algorithm gradually developed from this problem provides a mathematical solution for the shortest path between two points in a city while crossing a series of bridges (See Euler in: König 1936, 290–301). In visual computing projects, the path is used for the visual traversal of latent space, as illustrated in the practical example *X Degrees of Separation*. In this example, the viewers select an initial image and a target image and in several steps, i.e. intermediate images, the algorithm establishes a connection between the start and the destination based on the nearest neighbours. I will discuss how clustering and pathway behave in concrete terms in greater detail in one of the case studies.

Tables, scatterplots and k-nearest neighbours – what they all have in common is spatial distribution as a visual argument for the purpose of generating knowledge. Having covered the syntax of diagrams in media visualisations by means of weighted networks, we now turn to discussing the reduction and reconstruction of reality as an additional aspect.

1.2 Three layers of reduction and reconstruction of reality

Visualisations are reductions and reconstructions of reality. But they are only the last and most visible layer. The second layer is computation and statistics, and the first layer is data collection. This is relevant because each layer of reductive and constructive information processing introduces its own political and ethical values, which remain invisible to the observer. As we can see in the practical examples, reduction and reconstruction stand in an operative relationship to one other. The reality reduced in the data space undergoes its reconstruction in diagrams and only becomes operational through these steps.

First layer 'data collection': As early as the data collection phase, an information model is used to decide which parts of reality are to be collected as data and which are literally left splitt off and excluded. These exclusions are not 'bad' per se, because generating new knowledge requires concentration and clarity. This leads to a spatial, temporal or spatio-temporal segmentation of the world, as Bowker/Star (1999, 10) note. The data collection already establishes standards, and in the course of standardisation reality is reduced and at the same time also reconstructed (compare Gitelman 2013; Bowker 2014, 1797). But reduction does not occur solely as exclusion, as the example of 'data doubles' illustrates (D. Haggerty and Ericson 2000). Data Doubles are representations of a person or an object at a specific point in time. They are always reconstructions. Since a person's data can never be fully captured, missing information is compensated for – by correlation with comparison groups, for instance. Data from different sources are often combined with each other, so that here, too, we can speak of a reconstruction. The reconstruction of reality in data occurs on the basis of an information model, in other words, on the basis of an understanding of which part of reality is to be included in the processing and which is to be excluded. In the header of a table, for instance, the column labels define which data can be entered and this constitutes the information model then. Since persons and objects change over time, data doubles per se are never identical to the 'original'. They are historically incorrect, incomplete, fragmented and reduced. Reduction thus occurs both as exclusion and through temporal divergence.

Second layer ‘statistical computation’: The weighted networks of artificial ‘intelligence’ are reduction machines. They reduce large pixel sets, such as image sizes of 5000×3000 pixels, to computable dimensions as in the case of the pre-trained weighted network ResNet 50 to dimensions of 229×229 pixels. In addition, because many modules in convoluted weighted (‘neural’) networks expand the number of data points (features), the multi-dimensional latent space between the modules must be reduced again and again in order to guarantee computing performance. Algorithms such as principal component analysis (PCA) are used for this purpose. If, for example, the dimensions are to be reduced from 1024 to 256, the algorithm filters out those dimensions with the greatest variance and then makes further use of them. To the PCA method, content-related reasons concerning the question of which dimensions should be reduced are irrelevant. The algorithm carries out the reduction solely on the basis of maximum variance, not on the basis of content criteria. In addition to the reduction, a reconstruction also takes place through interpolation and merging of datasets, especially in those modules of convolutional ‘neural’ networks that expand the number of features.

Third layer ‘visualisation’: The already reduced data in these underlying layers are then reduced again in the visualisation through graphic representation. Data, i.e. socially mediated references to facts, become points in a scale diagram, a node and relation in a network diagram or a cell entry in a table or colour coding. Algorithms such as PCA, MDS, PCS, t-SNE or UMAP dramatically reduce the high dimensional spaces of weighted networks for visual output to only two dimensions, the X and Y axes in scatterplot diagrams.
The reference to the originally collected fact becomes weaker, almost homeopathic (Fig. 2).[4]

Layer	Reduction	Reconstruction
Data collection	Spatial, temporal or spatio-temporal segmentation of the world	Information model
Statistical computation in weighted networks	Algorithmic reduction of data volumes in favour of computing speed	Merging and extrapolation
Visualisation	Reduction of the dimensions of weighted networks for two-dimensional representation Symbols, graphs, points	Image representations (icons)

Fig. 4: Reduction and reconstruction of reality in visualisations based on weighted networks (convolutional neural networks).

[4] This analysis also applies to three-dimensional representations, which in principle resemble two-dimensional reductions, and merely introduce the third dimension as a further parameter in reduction algorithms such as PCA, MDS, PCS, t-SNE and UMAP.

But a new type of visualisation adds new visual enrichment after these many reductions: “Rather than representing text, images, videos, or other media though new visual signs such as points, rectangles, and lines, media visualizations build new representations out of the original media. Images remain images; text remains text” (Manovich 2020, 197f.). Even if one may not completely follow Manovich in this statement (“images remain images, text remains text”), he nevertheless brings up an important point. The operationality of ‘classical’ diagrams is more abstract through symbolic representation than is the case in media diagrams, which depict what is represented by means of icons – iconically. Due to the high capacity of computing power, media representations that were difficult to achieve before the 2000s are now possible in real time and enable new diagrammatic solutions for visual knowledge. In this way, as will be explained in the practical examples, the table and scatterplot can be filled not with abstract points, lines and the like, but with image representations and, alongside the spatial distribution, create an additional depth of information that permits new visual interpretations (Fig. 12). A ‘reconstruction’ of reality occurs here in diagrammatic representations. Each graphic construction creates a new space into which the data objects are transferred, based on the grid that this space provides. A dimensional reduction of the original images is carried out to either the same height or the same width. Along with these scalings comes a whole series of interpretational problems as we will see in the case studies.
The first part of this study discussed the syntax of various media visualisations, with a focus on tables and diagrams, especially scatterplots and clustering using k-nearest neighbour procedures. While the visual tools of knowledge mentioned above have spatial references in common as a knowledge-forming operation, there are differences in the way they are read. In a further step, we identified three layers of reduction and reconstruction of reality: 1.) in the course of data collection, which transfers the objects from reality into data structures, 2.) through statistical computation in weighted networks and the dimensionality reductions there using PCA, and 3.) in the course of visualisation, during which the high dimensional spaces are reduced to two dimensions X and Y for scatterplot diagrams using algorithms such as PCA, MDS, PCS, t-SNE and UMAP. Finally, the paper outlined which knowledge operations are relevant to The Curator's Machine and how spatial arrangement becomes operationalisable in each case.

2 Practical examples of AI visualisation

The following projects were selected because of their proximity to Training the Archive in terms of dealing with large image collections. The examples selected were reviewed in the spirit of constructive criticism. They were not chosen to diminish their achievements or characteristics, but to highlight the need for detailed expertise when making decisions on information, visualisation and user interfaces. In the following section, we will discuss the projects ARTigo, Imgs.ai, iART, Vikus Viewer and X Degrees of Separation before formulating a further course of action in the summary.

2.1 ARTigo – the art history search engine

ARTigo (https://www.artigo.org) is a semantic image search engine that makes the databases of the Diathek of the Institute of Art History at LMU Munich searchable. ARTigo does not use artificial intelligence algorithms, but rather invites the users of the website to provide the images with structured metadata in the scope of a crowd intelligence process (Bry, Schefels, and Schemainda 2018). The following section discusses the user interface of the search engine. No machine learning procedures have been used thus far, but ARTigo processes collections of art objects in a similar way to The Curator's Machine and is therefore relevant to our project. The strength of ARTigo, in contrast to many artificial intelligence projects, lies in its focus on the semantic aspects of the images – the content and meaning of each individual image has been annotated and verified. As a result, the referentiality of each individual image in the database remains verifiably preserved. LMU Munich has thus far refrained from transferring semantic meaning to other images by means of machine learning, but the dataset lends itself to such an approach.

Tabular grid

In the standard view, the search results are arranged in a table, with one row marking each search entry (Fig. 5). The columns consist of a tag cloud belonging to the picture, an image and the classic art historical metadata of artist, title, location and date. This listing thus follows the index card logic that we often find in relational databases. The most relevant result to the query is arranged at the top according to the visual top-bottom logic, and the individual entries are separated from each other by white spaces. This view focuses on the scrolling display of individual entries to allow a selection of one or a more entries through activating the filter options.

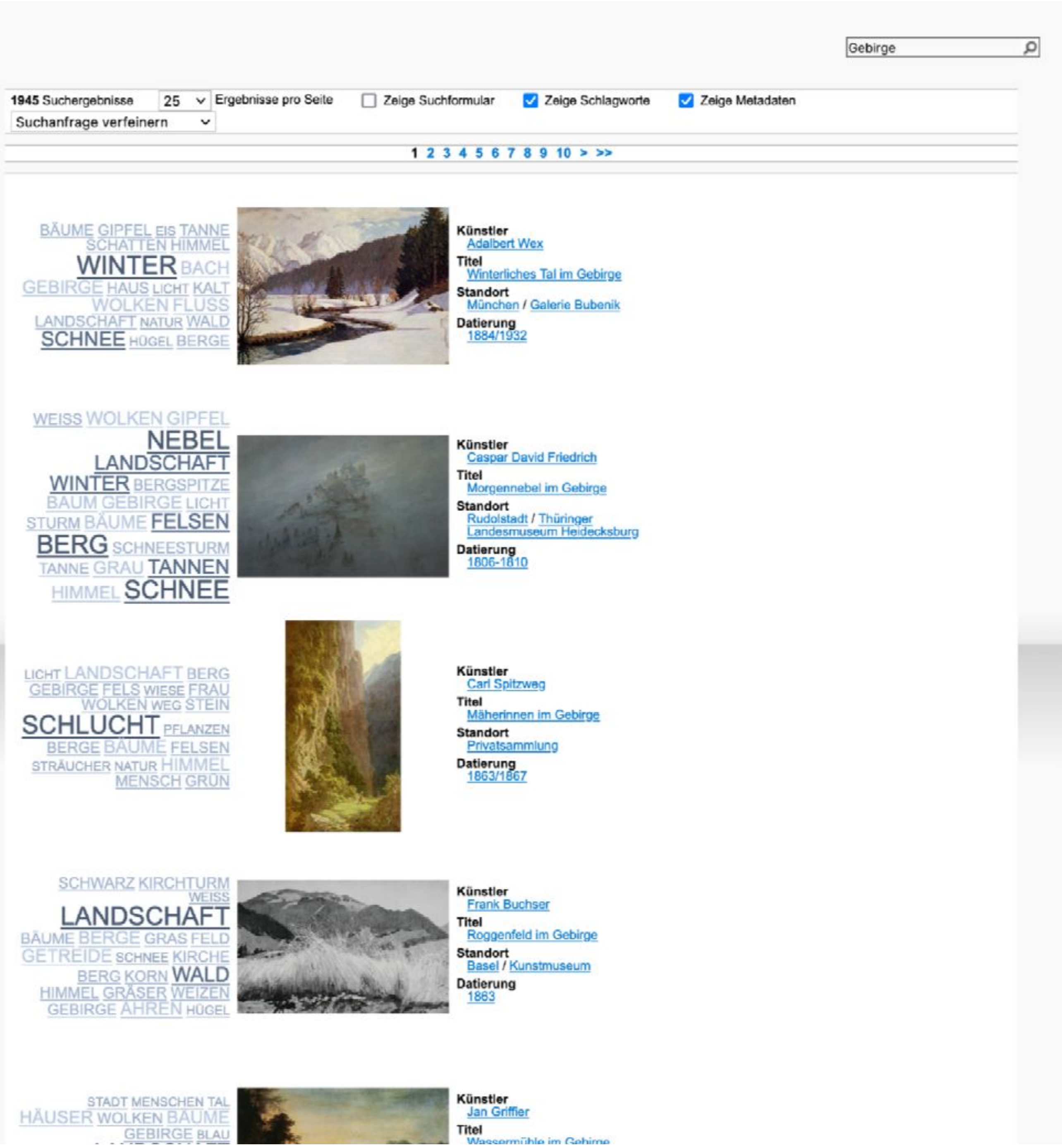


Fig. 5: ARTigo results display in tabular form with three columns and one entry per row (screenshot, 2/12/2022).

The search result can be refined by clicking on a term in the tag cloud. The terms act as filters. In this example, the search term has been ‘mountains’ (Fig. 5). By clicking on ‘trees’ in the tag cloud, the search set is limited to those images annotated as ‘mountains’ and ‘trees’ in the tags. Similarly, clicking on the metadata positioned to the right of the image allows searches for paintings with ‘mountains’ present in the location ‘Munich’.

The search operates according to the principle of the closed-world assumption. The search can find only what is contained in the database. Nothing else can be found, i.e. the search engine is not linked to the outside world – through open data access or links with bibliographic aggregators, for instance.[5] No separate visualisation exists to indicate this circumstance; the search list simply comes to its end. The predefined categories determine the search process. You can search by title, artist, location, date and keywords.

Matrix view

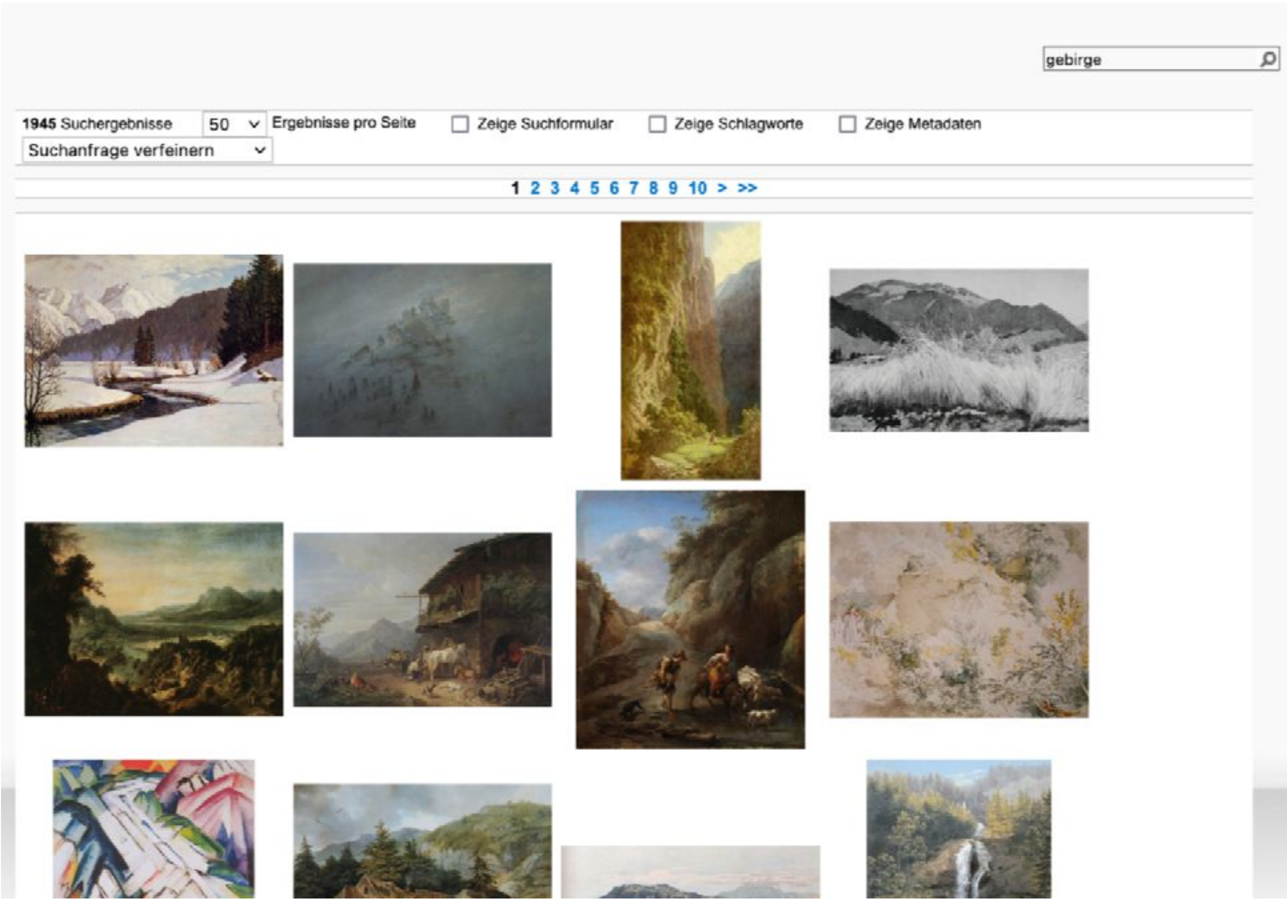


Fig. 6: ARTigo interface in matrix format (screenshot, 2/12/2022).

If the options "show metadata" and "show description" are deselected in the ARTigo user interface, the display format changes to a matrix. However, the reading orientation of this matrix is unclear and is not visually supported (e.g. by spacing that would create rows). The interface does not explicitly communicate the mode change from the three-column grid of the table to a new ordering of images. This view fundamentally changes the order in which the results are displayed. Rather than being displayed from top to bottom as before, they are now displayed from left to right. The most relevant result for the query is at the top left and the subsequent results are now sorted from left to right until the page margin. The various image sizes are limited to a maximum of 200px in height and width, whereby the images are reduced in size, but not to scale, uncropped and displayed in the correct aspect ratio. As a result, many portrait formats create the impression that the results are to be read from top to

---

[5] A bibliographic aggregator is a "service that collects, unifies, manages, maintains and shares metadata from cultural and knowledge institutions" (https://pro.deutsche-digitale-bibliothek.de/glossar/aggregator).

bottom and many landscape formats create the impression that the grid runs from left to right.

Overall, ARTigo is a valuable step towards making historical art collections accessible through a search engine function. However, my description clearly shows that cooperation with user interface experts would have contributed additional expertise to the project.

2.2 Imgs.ai – an experimental art historical image search

Imgs.ai (https://imgs.ai) is an experimental platform for exploring image collections through machine learning. It is based at the University of California, Santa Barbara and the German Documentation Centre for Art History – Bildarchiv Foto Marburg. The software technology CLIP (Radford et al. 2021) enables a content-based search for unclassified images and image content by means of text queries, known as prompts. In addition, images can be set as positive or negative templates for further searches, and uploading your own images allows you to compare them with the visual orders of museum and image collections. The images are sorted according to the k-nearest neighbour principle, which displays similar images from the collection closer to each other. Imgs.ai gives curators and researchers the option of using their own images or keywords to search for similar images in the available museum collections even if the collections are not annotated.

Tabular grid

Imgs.ai's user interface consists of a section with numerous filtering options, such as the dataset used (Metropolitan Museum, Rijksmuseum), the weighted network used (VGG 19, raw, CLIP-vit, Poses),[6] an upload option for one's own images as search terms and a selection for the column width of the display, 124 pixels wide for example, and for the number of neighbouring images between 10 and 100. The image data is arranged horizontally in columns. By clicking on a picture, you can find its URL on the Internet or go directly to the high-resolution picture file on the websites of the institutions of origin.

The various image sizes are limited to a column width of 32px to 224px, selectable via the menu item 'size'. All images are reduced in size accordingly and displayed uncropped and in the correct aspect ratio. The number of columns is determined by the available screen width in relation to the column width selected.

---

[6] Pose can identify human poses, especially the relationship of arms, legs to the body (Toshev and Szegedy 2014); VGG 19 is a weighted network of 19 layers pre-trained on ImageNet (Simonyan and Zisserman 2015); CLIP is trained on combined image data and word embeddings (image captions and metadata), e.g. by training with annotated Flickr images (Radford et al. 2021).

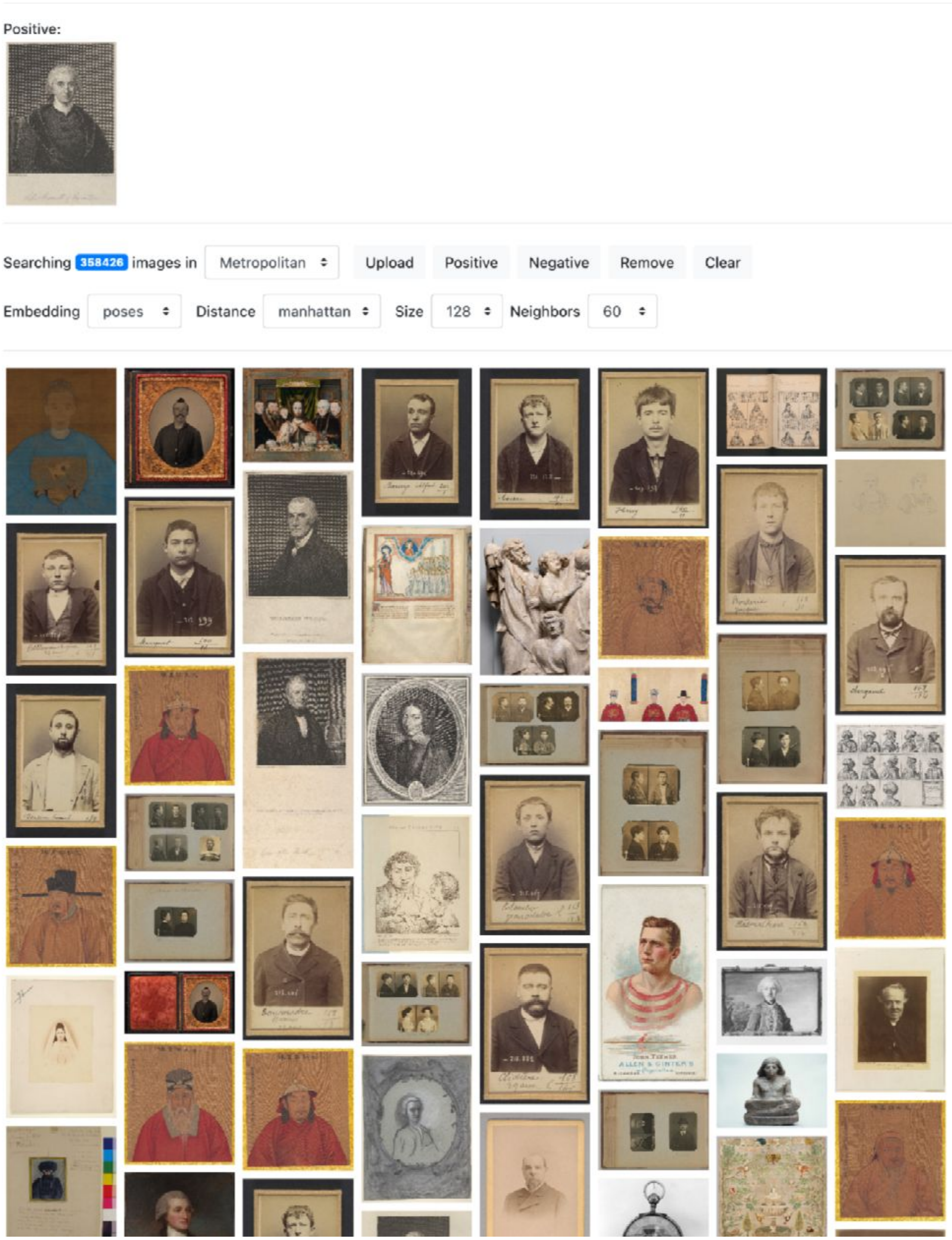


Fig. 7: Imgs.ai interface with the etching "Charles Carrol of Carrolton" (ca. 1835) from the collection of the Metropolitan Museum New York as a prompt. 'Poses' was chosen as the embedding here, so that in principle similar ratios of head, shoulders and arms are shown (screenshot, 2/12/2022).

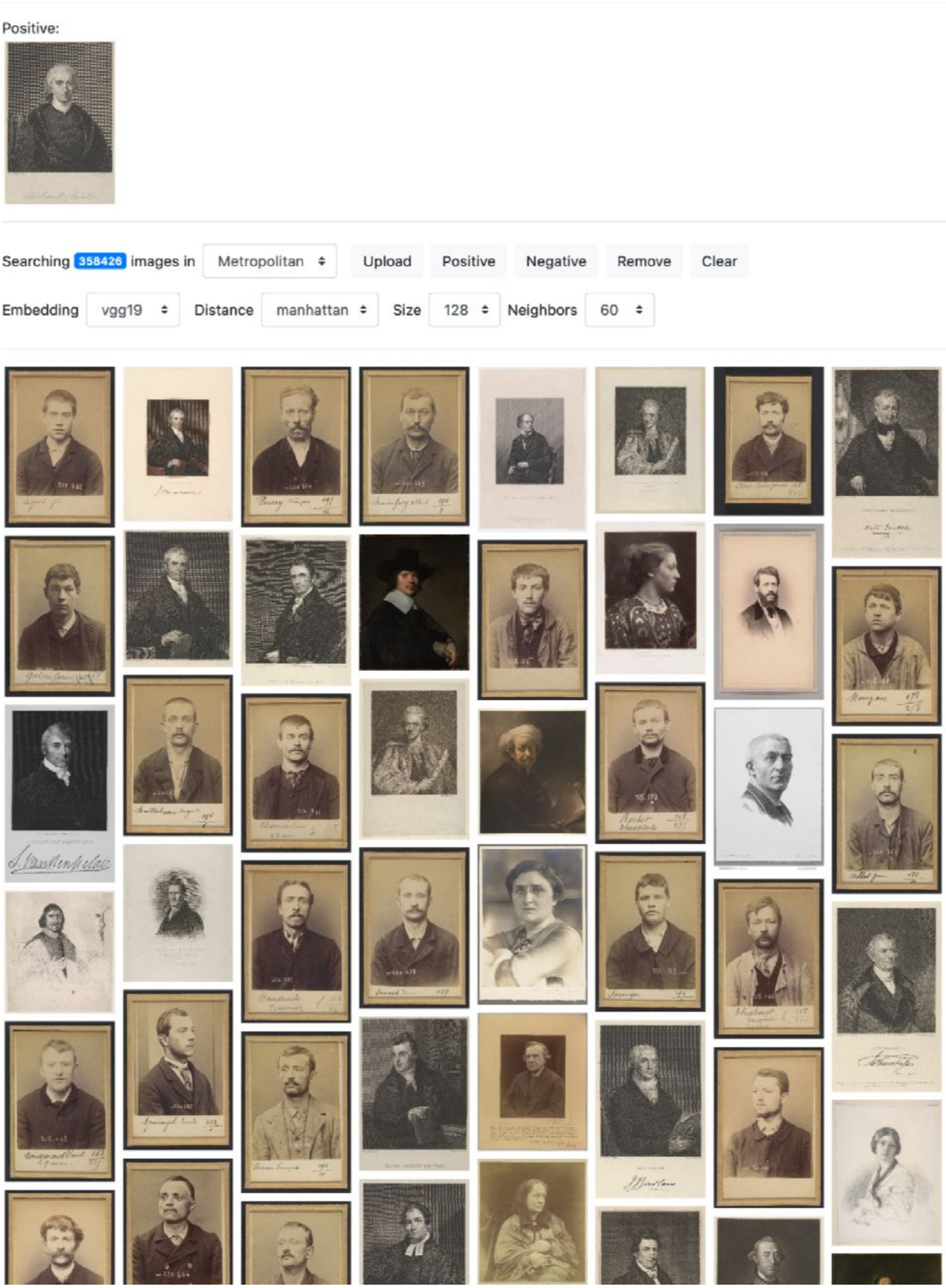


Fig. 8: Same search parameters as in the previous figure, but the k-nearest neighbours are not calculated using the Poses network but rather using VGG-19 (screenshot, 2/12/2022).

In Fig. 7, horizontal orientation results solely from the image formats and does not receive any further visual support. In this vertical visual logic, the first and closest image to the search query is placed at the top left of column 1. Further images follow in the column with descending 'similarity' from top to bottom. At the eighth image, the eye jumps to the top of column two.
According to the k-nearest neighbour principle, this image is further away from the first image than all the images in column 1, but in this graphical solution it is directly next to the first image. This reveals a problem with matrix ordering in vertical columns. They generate a specific reading order. The display in a scale-oriented grid is fundamentally problematic for the underlying knowledge task because proximity and distance relationships are inadequately represented due to the grid.
When images with very similar aspect ratios appear in this column-oriented matrix, visual order is further disrupted, because it is unclear to the reading eye how the image arrangements are oriented: from top to bottom or from left to right?
Using the img.ai interface example, we see that for k-nearest neighbours a matrix-shaped arrangement is rather unsuitable, as a divergence arises from calculated statistics and visual mapping.

2.3 iART – an interactive tool for analysing image datasets

iART (https://labs.tib.eu/iart) is an interactive analysis and retrieval tool for image collections focusing on art history. The DFG-funded research project of the LMU Munich, the University Library of Hannover (TIB) and the University of Paderborn arranges the image data according to weights as k-nearest neighbour.
It aims to combine three different knowledge-generating modes: pattern recognition and classification through deep learning, similarity analysis through clustering and k-nearest neighbour analysis, and application of user preferences through filtering tools. According to the project developers, the interface is oriented towards the Google user interface design, as it is familiar and users are accustomed to it (Schneider and Kohle 2021, 6:50 min).
The project used data from collections provided and classified by third parties. Therefore, it could dispense with conducting its own time-consuming classification work. The approximately 1 million images stem from the numismatic collection KENOM, the Wikimedia Commons category "Art_by_Subject", the Rijksmusem Amsterdam and other smaller datasets with approximately 60,000 images, including ARTigo (Springstein 2021). The criteria for the appearance of certain artistic positions are based on how often an artist is included in the datasets. The German-American landscape painter Albert Bierstadt (1830–1902), for instance, dominates the search query for 'mountain', as his images with the keyword 'mountain' are well represented in the Wikimedia Commons collection. It follows that romantic depictions of mountains by Bierstadt are represented in large numbers in the search results, and this could give the impression that his works were of outstanding artistic importance (in reality, they were only well keyworded).
The visual search is not only based on the metadata itself, but also on embeddings. Embeddings are mathematical translations of words expressed as vectors. These vectors are related

to each other mathematically-spatially in order to model meaning contexts, whereby words with similar vectors (should) result in a similar meaning context. This makes it possible to specify, for instance, that the word 'painting' has a similarity to 'sculpture' of 0.845, to 'painter' of 0.842 and to 'watercolour' of 0.841 on a scale of 1, while the similarity to 'create' is only 0.359.[7]

Because certain word combinations of the metadata lie close to each other in weighted networks, the images distributed in the latent space are positioned correspondingly close to each other. In doing so, the authors refer to classical principles of order: "Ordering criteria that were already common in the early modern Wunderkammer, such as colour, material, or function, can be applied as well as more iconographically based classification principles that, e.g., examine objects for biblical motifs or Christian themes" (Springstein et al. 2021, 1). The interface allows for the selection of different embeddings of pre-trained weighted networks: CLIP, Wikimedia and ImageNet, whereby the more precise designations of the embeddings (e.g. versions, datasets) are not stored in the interface. This makes the underlying genealogies and biases of the training datasets non-transparent for researchers.

In the following section, we will discuss the two main viewing modes, image grid and distribution, with reference to iART.

Tabular grid

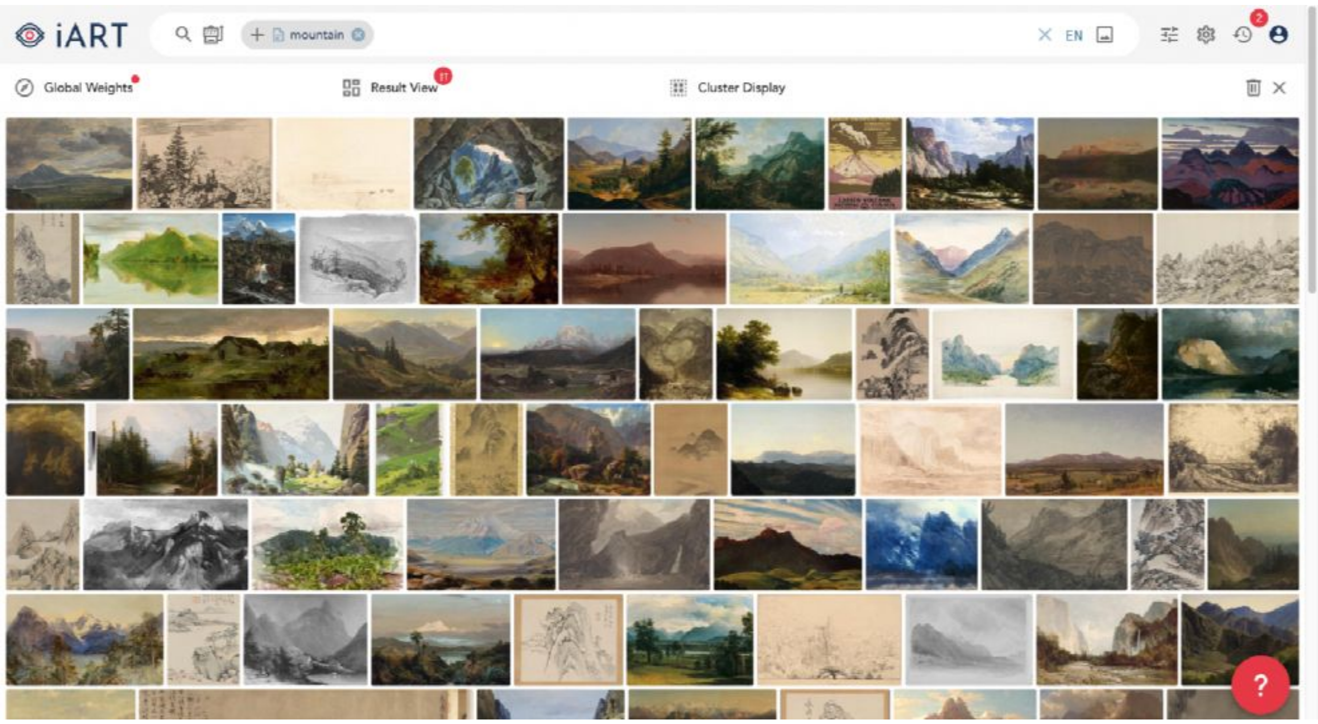


Fig. 9: iART, Search term 'mountain' without grouping/clustering (screenshot, 2/12/2022).

In the grid, results are displayed in order of relevance. When hovering over an individual image, additional information, such as metadata and a link to a high-resolution version, appears

[7] For an example, see the display at https://embeddings.sketchengine.eu.

(Fig. 9). Since the images are adjusted to the same height, the size relationships between the images are lost. The images are displayed as a continuous set of data, with the first image representing the most ideal mountain and then organised horizontally, row by row, the degree of the mountain’s calculated relevance sinking in a decreasing ranking. Thus the user's gaze sweeps from left to right and then jumps to the next line. The order results from the word embeddings in the multi-dimensional space of the weighted network based on the CLIP logic (Springstein et al. 2021, 2). This relevance view can be sorted by image title or chronologically, by year.

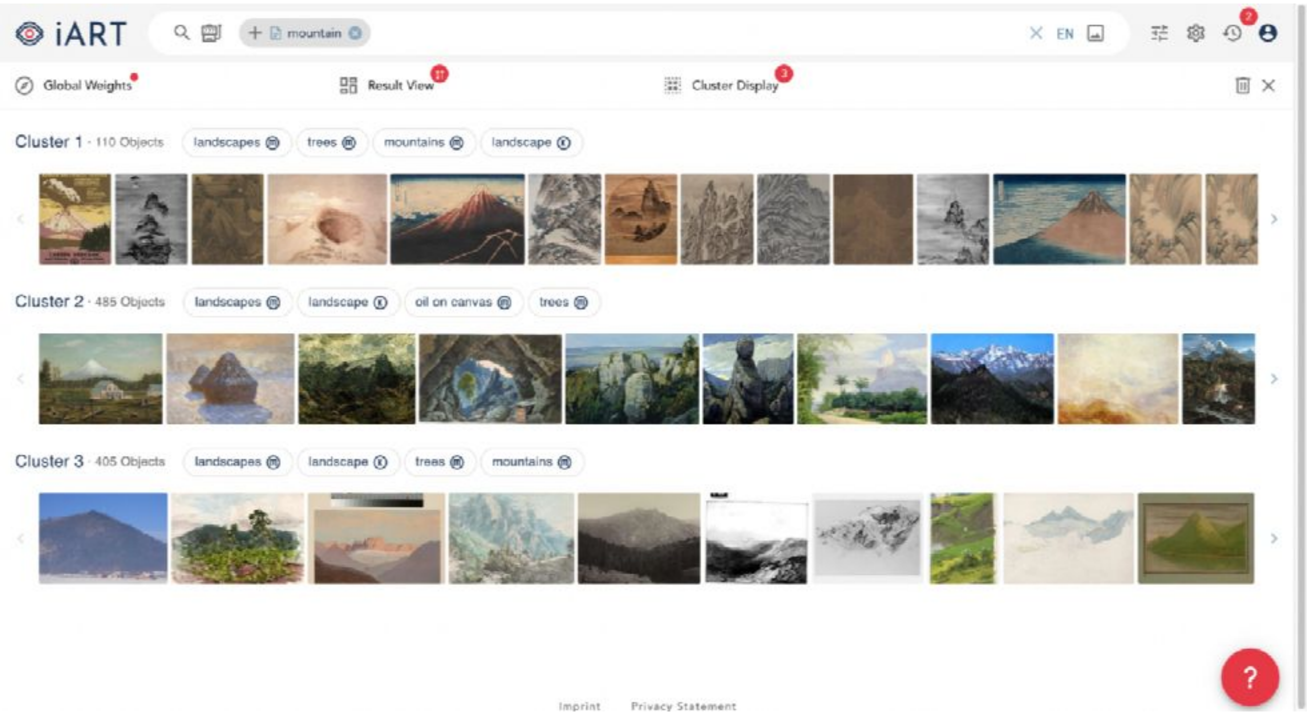


Fig. 10: The search term ‘mountain’ divided into three clusters (screenshot, 2/12/2022).

An additional mode groups the grid view into clusters. Each cluster formed is given its own line, which scrolls infinitely to the right. This display has the advantage that the eye actually sweeps through a continuous dataset from left to right and does not have to jump to the next line at the page margin, as discussed in the example of the Imgs.ai interface.
The clusters are formed automatically – in this example with the search term ‘mountain’. The clusters are identified by very similar keywords, some of which only differ in order:

Cluster 1: 110 images – landscapes, trees, mountains, landscape
Cluster 2: 485 images – landscapes, landscape, oil on canvas, trees
Cluster 3: 405 images – landscapes, landscape, trees, mountains

Upon inspection, we notice that cluster 1 includes many drawings, predominantly from Japanese culture. These images nearly always contain Japanese characters. Cluster 2 mainly shows oil paintings and differs from the other two clusters in its brilliant, high-contrast colouring. Cluster 3 contains watercolour paintings, ink drawings, chalk drawings and some photographs, but also engravings with rather subdued and restrained colours. Colour saturation could therefore be one of the criteria according to which the clusters were formed. However, other factors may also play a role in the clustering, such as the sensitivity of the underlying language embedding CLIP for specific word groups, or the tendency of weighted networks, discussed in Working Paper 2, to distinguish image data based on textures rather than outlines as humans would, the phenomenon known as texture bias.[8] If the Japanese drawings, ink drawings and prints gathered in cluster 1 did not contain Japanese characters, the algorithm may have clustered them differently. We would have to establish this through experimentation.

In the search entry, the abbreviation EN indicates that an English-language entry is expected. However, entries in other languages are possible and produce surprising results (Fig. 11).

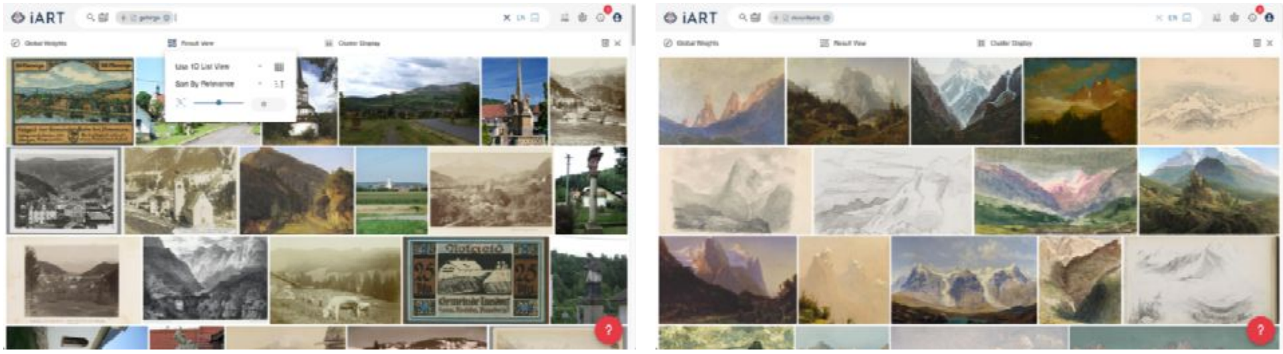


Fig. 11: Left: Results of the German search query ‘Gebirge‘, with mostly postcard illustrations; Right: Results for the English search query ‘mountains’ with mainly paintings (screenshot, 2/12/2022).

Scatterplot

The iART interface allows another view of the same dataset by using a scatterplot. The Scatterplot reduces the dimensions of the latent data space using the Uniform Manifold Approximation and Projection algorithm (UMAP) (McInnes, Healy, and Melville 2020), a refinement of the t-SNE algorithm (Springstein et al. 2021, 2).[9] UMAP reduces the high dimensional output of weighted networks along the variances of individual dimensions, while maintaining the relative distances of the images in the respective dimension as much as possible.[10] “UMAP is derived from the axiom that local distance is of more importance than long range distances. […] Multi-dimensional scaling specifically seeks to preserve the full distance matrix of the data, and as such is a good candidate when all scales of structure are of equal importance” (McInnes, Healy, and Melville 2020, 45). In this context, it is crucial to note that

[8] Cf. (Geirhos u. a. 2019).

[9] The authors of UMAP point out that UMAP lacks the strong interpretability of the reduction algorithm PCA (Principal Component Analysis), which means that one cannot make such strong statements about how the inside of the algorithm comes to decisions. “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] For novices wishing to gain a better understanding, the following YouTube tutorials are recommended: *Principal Component Analysis (PCA), Step-by-Step*, https://www.youtube.com/watch?v=FgakZw6K1QQ and *t-SNE, Clearly Explained*, https://www.youtube.com/watch?v=NEaUSP4YerM.

all dimensional scales are similarly important in UMAP. This is different from PCA and t-SNW, which use the two scales with the largest variance, then yield the X and Y axes.

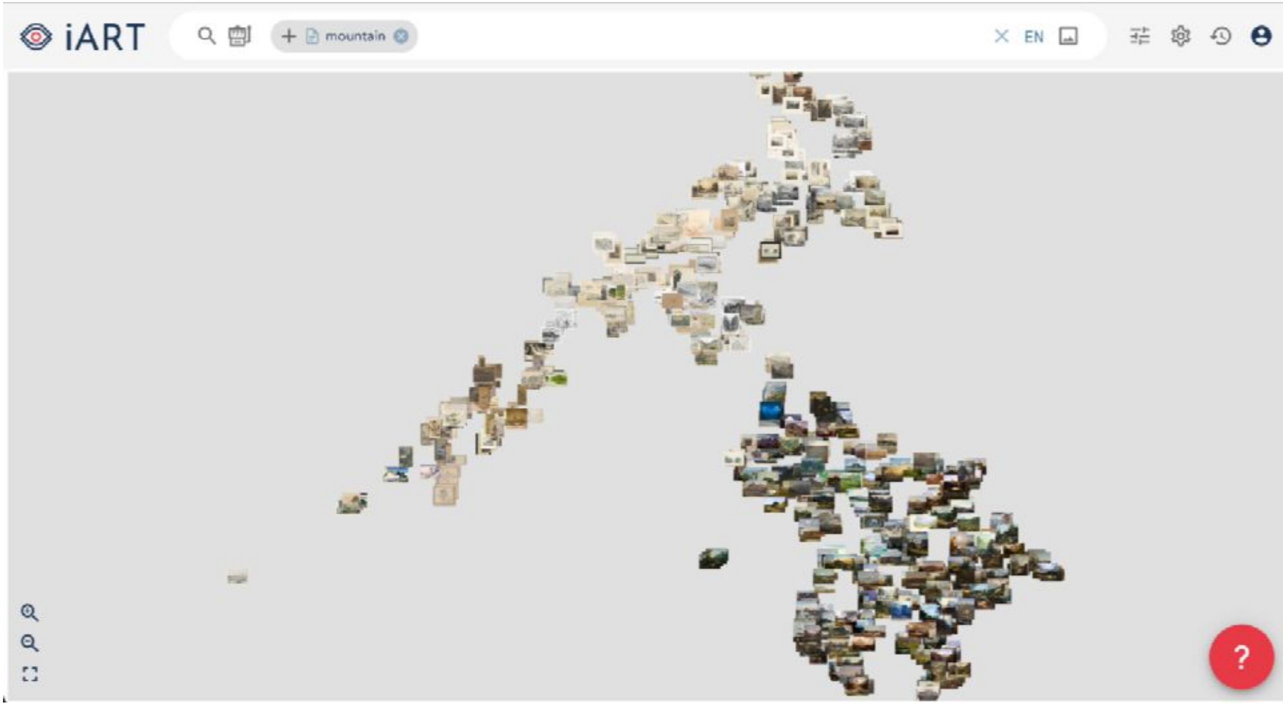


Fig. 12: Scatterplot for the term 'mountain' with three clusters (screenshot, 2/12/2022).

A visual comparison of different dimensionality reduction algorithms used for scatterplots, such as PCA, t-SNE, UMAP, by the McInnes/Healy research team shows strikingly different visual results for the same dataset. Thus, the dimension reductions of t-SNE and UMAP differ in that UMAP's reduced dimensions make clearer delimitations of the groupings among themselves than those of t-SNE (McInnes, Healy and Melville 2020, Fig. 4). In practice, explaining the exact differences to users of the user interface would be going a step too far. However, some hints, through explanatory videos for instance, would be helpful because most people are not as experienced in reading scatterplots as they are in reading charts such as tables (compare Fig. 3).

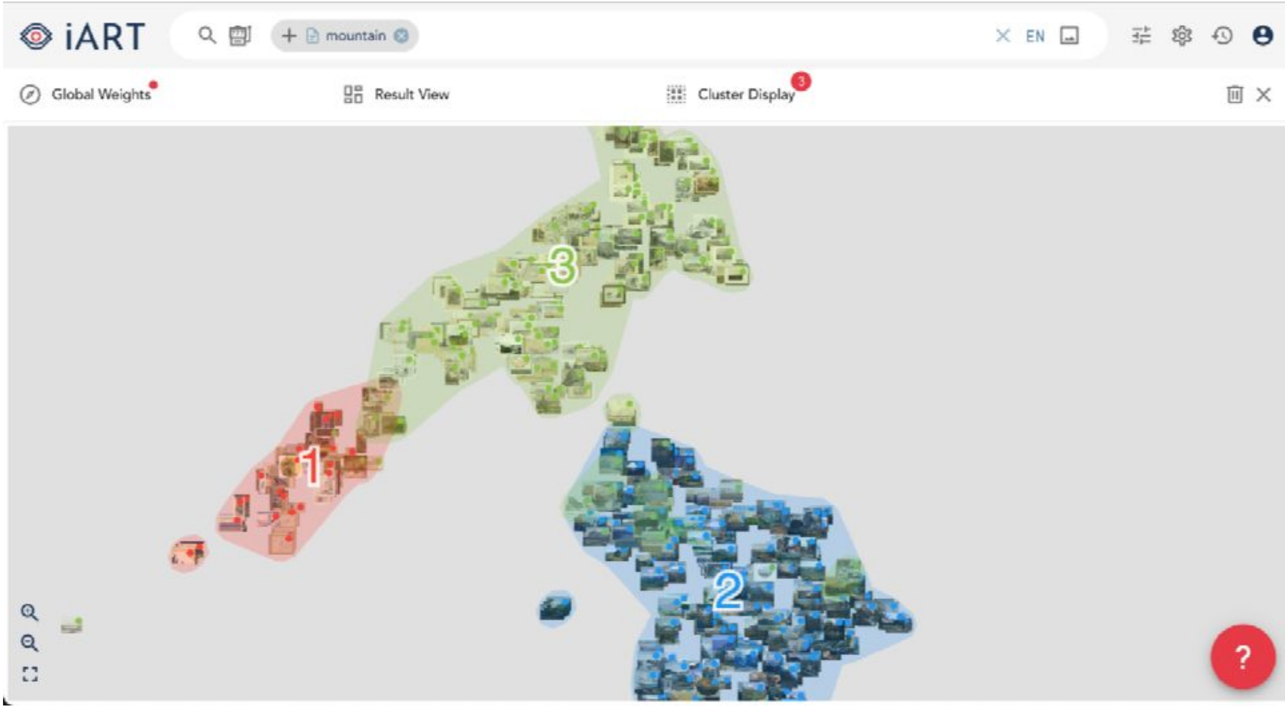


Fig. 13: The view of the clusters can be additionally clarified with coloured dots and colour overlays. Cluster 2 also contains pictures with green dots, which therefore belong to cluster 3 (screenshot, 2/12/2022).

In order to make the distributions and affiliations to the respective clusters visible, the images of the scatterplot are additionally colour-coded with a small, coloured dot in iART. This makes clusters visually recognisable, but more importantly it clarifies which outliers belong to which cluster.

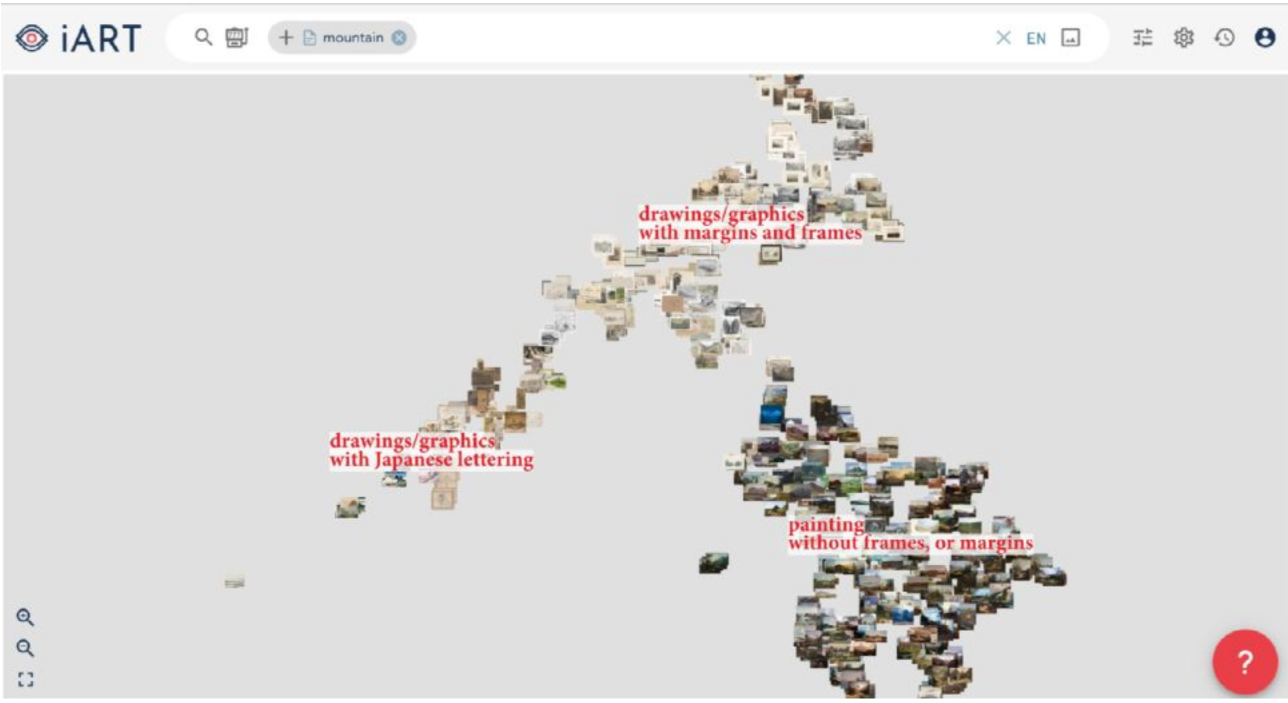


Fig. 14: Three clusters are formed by means of CLIP-embedding (screenshot, 2/12/2022, red captions by the author).

The advantage of the scatterplot is that in principle all images present in the resulting set can be represented. In the grid view, by contrast, the respective cluster set is not completely represented because the screen width limits it. You can only see what fits on the screen, without the possibility of zooming in (Fig 11).
iART offers numerous other options. In principle, users can import their own image inventories, combine the cluster display with keyword filters and search terms, and in particular, make use of an image search ('Search by Image') similar to the Imgs.ai project. However, a detailed discussion on this topic reaches beyond the scope of the present paper.

iART is a complex research tool that requires more thorough documentation, such as video tutorials. iART is an open analysis tool for exploring large image sets and offers numerous usage options, which are only briefly presented here. iART needs to provide more guidance and explanation on the user face's extensive range of organisational tools.

2.4 Vikus Viewer

Vikus Viewer (https://vikusviewer.fh-potsdam.de) is part of a research project on the display of large image sets on websites, based at the Urban Complexity Lab of the University of Applied Sciences Potsdam 2014–2017.[11] Christopher Pietsch originally designed and programmed Vikus Viewer for the research project Past Visions (Katrin Glinka, Christopher Pietsch and Marian Dörk). His implementation builds on Nikhil Thorat's Teachable Machine, based on the machine-learning framework Tensorflow. The project, which was released in 2017 and has been under development since 2015, is likely to be one of the earliest such approaches not only in Germany but worldwide.
In the course of workshops, lectures and conferences, the research group investigated interfaces and their generalisability. In doing so, they resorted to two 'classic' methods for ordering knowledge: 1.) temporal ordering and 2.) keywords in objects' metadata.
The following section describes another iteration of the Past Visions project, entitled Vikus Viewer – Vincent van Gogh. It is based on 986 drawings and paintings by Vincent van Gogh (1853–90) from the collection of the Van Gogh Museum, Amsterdam. Like the other collections Vikus used, this one was curated, meaning that project collaborator Viktoria Brüggemann additionally curated and corrected the data from the Van Gogh Museum. In the course of user interface experiments in the summer of 2014, designer and programmer Christopher Pietsch introduced another method of ordering knowledge – in a scatterplot. Pietsch drew on his work with generative design, in particular Kyle McDonald's Module in open framework, which also enabled similarity analysis in the C++ language as part of creative coding practices.
Particularly noteworthy is the approach of publishing the underlying web-based software as open source on the Gitlab platform. The sustainability considerations are also reflected in a

---

[11] This case refers to the *Vikus Viewer – van Gogh*, one of many visual research projects of this working group. See also https://uclab.fh-potsdam.de/projects and in: *Von der Wolke zum Pfad – Visuelle und assoziative Exploration zweier kultureller Sammlungen* (Brüggemann et. al. 2022).

conference paper by the authors, who list some minimum requirements: “This primarily includes chronological classification and keywording based on a controlled vocabulary. In addition, the collection should be digitised in sufficiently good quality (and be available as a jpg) so that full use can be made of the zoom function (from overview to detail). The required collection data must be available in standardised CSV format” (Glinka, Pietsch and Dörk 2017, 205).

Tabular grid

In the tabular view, the entire dataset is displayed by year. An additional line contextualises the years by noting the artistic periodisation. Above the years are the icons of the corresponding images in the column (Fig. 15). Within any one year, the images are in random order. An order based on similarity, on light-dark values or alphabetically by title would also have been possible. In fact, in the default setting, the picture icons are so tiny that colourfulness and light-dark contrast are the most visible characteristics. However, a stepless zoom feature allows the user to select other image sizes that reveal more detail.

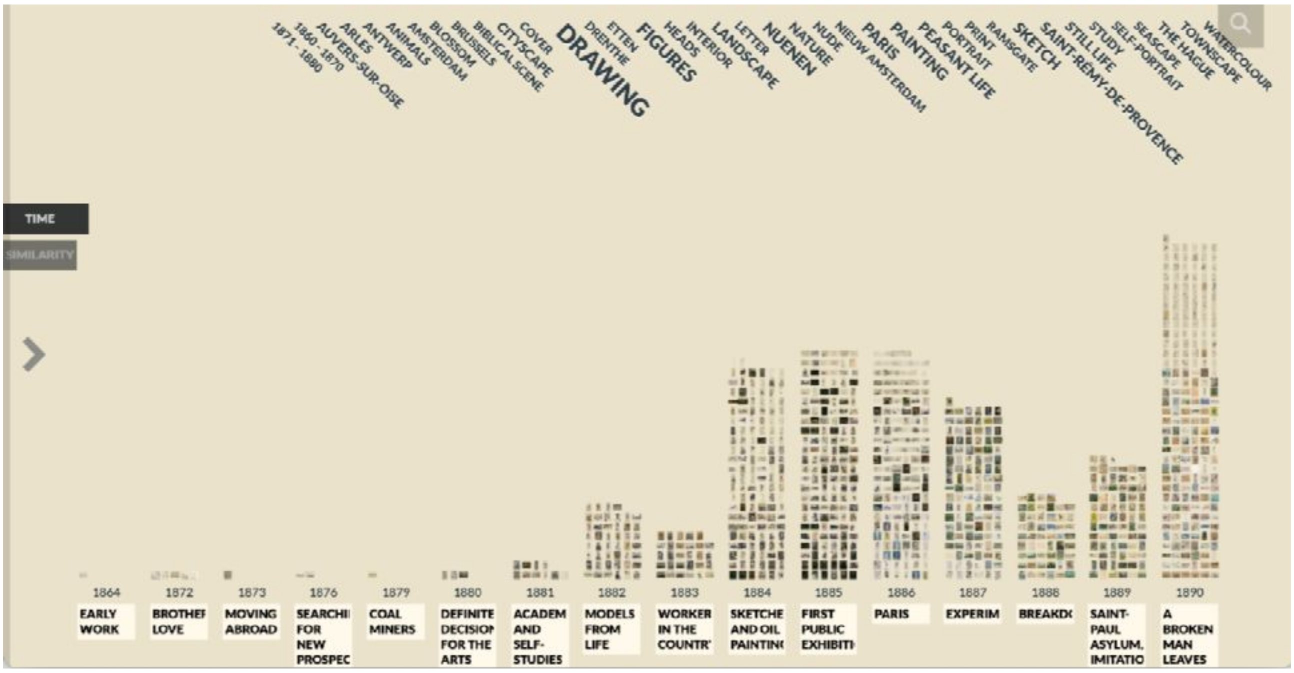


Fig. 15: Tabular chronological ordering by year and creative period. As in a bar chart, the quantities of paintings present in the collection are visible, with the largest number in his final year of life (screenshot, 2/12/2022).

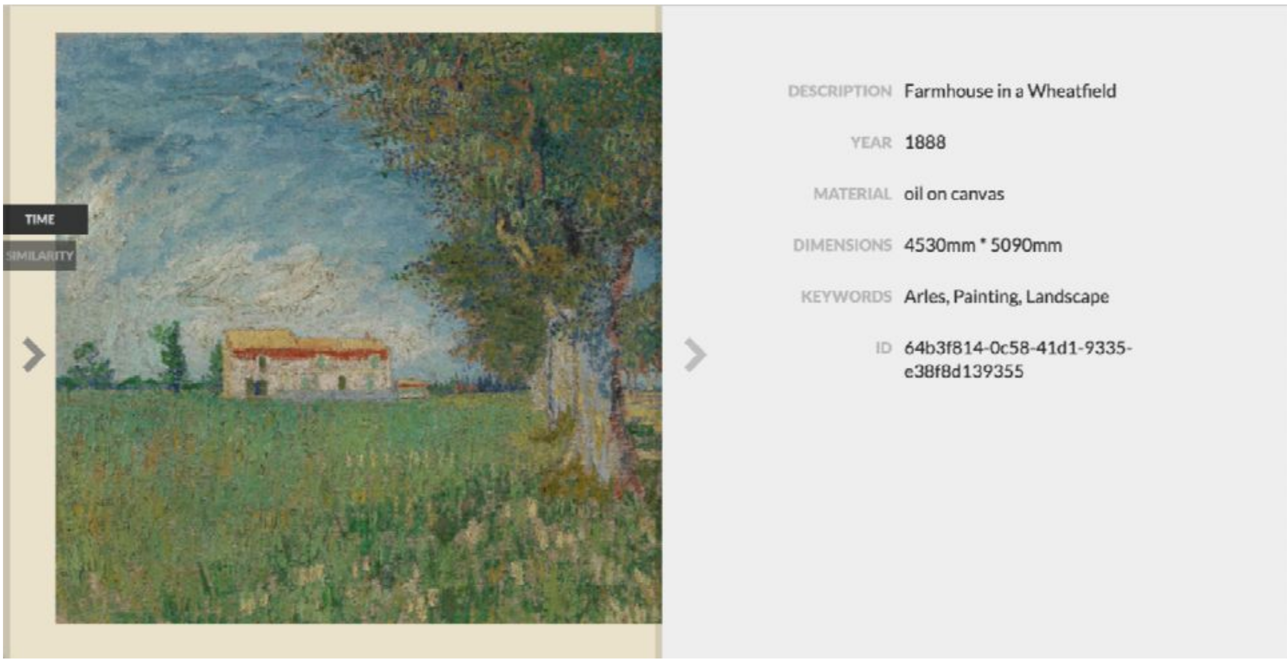


Fig. 16: Detail view and metadata of an individual image (screenshot, 2/12/2022).

The slanted keywords in the top bar function as inclusive filters, so that a user can select to display all the images throughout the years that contain metadata with a given keyword, such as ‘landscape’. Clicking on an individual images zooms in for a larger view of the image and its metadata (Fig. 16).

Scatterplot

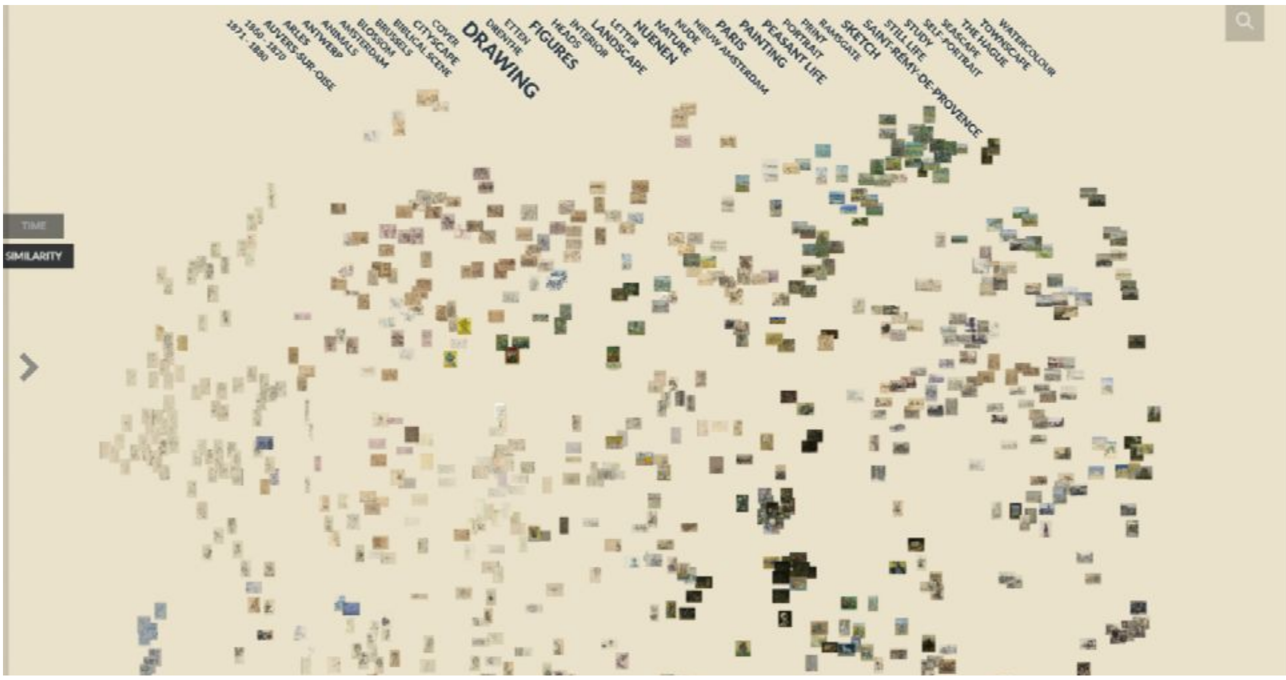


Fig. 17: Scatterplot display by similarity for the Van Gogh data corpus (screenshot, 2/12/2022, red captions the author).

Even at first glance, the scatterplot of van Gogh's paintings and drawings reveals a clearly visible distribution, but we see too much at once and all is too small to make more precise statements (see Fig. 18). Zooming in on individual regions reveals a cluster of sketches – spreading out from the lower left – with the paper becoming increasingly darker towards the top. Paintings increasingly appear from the middle onwards, with a cluster of landscapes in the upper right. These are interrupted in the upper right by detailed landscape drawings, which are much more elaborate than the sketches, before a series of painted portraits gathers in the lower right. This first glance also informs the viewer that the collection of the Van Gogh Museum, Amsterdam, consists mainly of drawings and sketches and that the number of colour paintings is in the minority. The images do not appear in any form of chronological order, although stylistic similarities sometimes result in chronological clusters, such as in the outline-like sketches from van Gogh's later creative phase, on the far left.

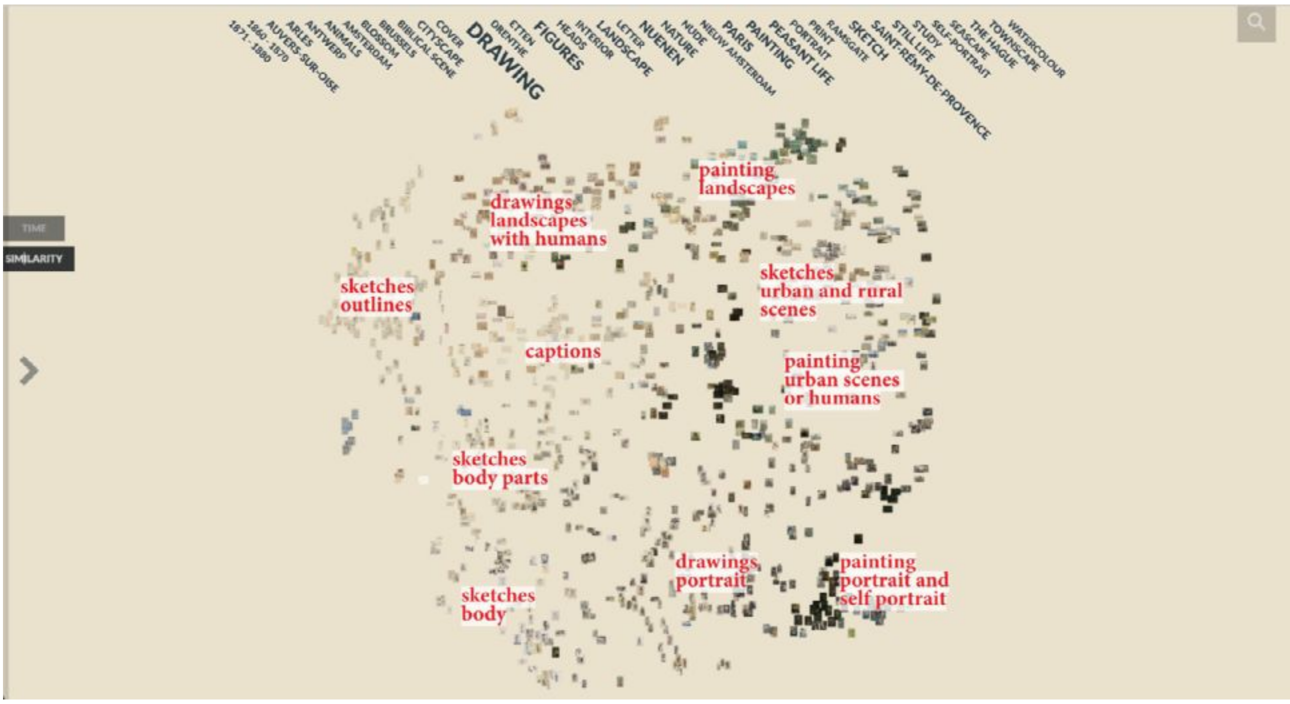


Fig. 18: Clustering of the Van Gogh collection using Vikus Viewer (screenshot, ibid.).

The image data is distributed by similarity using the activation layer of the ImageNet-trained weighted network Mobile-Net.[12] We can therefore assume that many of the issues raised in the paper *"Why so many windows?" – How the ImageNet image database influences automated image recognition of historical images* (Hunger 2023) are applicable, namely in relation to the pre-trained shapes in ImageNet (which are inherently unaware of the concept of art), the ahistoricity of ImageNet, and the texture bias described by Geirhos et al. 2019. Geirhos et al. had determined that for ImageNet-based weighted networks, the mathematical optimisations respond preferentially to textures. For example, a cat overlaid with elephant skin is clas-

[12] This procedure is similar to the one described in Bönisch 2021, which was also used for the first prototype of Training the Archive's The Curator's Machine.

sified as an elephant (Geirhos et. al. 2019, 1). In short: texture, as a mode of machinistic-statistical similarity, influences clustering more than outline, which is the primary mode of human perception.
Clustering was based on the activations extracted from ImageNet/MobileNet using the t-SNE algorithm. In order to reduce the density of the interface, the research team then added additional spacing between the images so that they are more widely distributed and create less visual overlap.

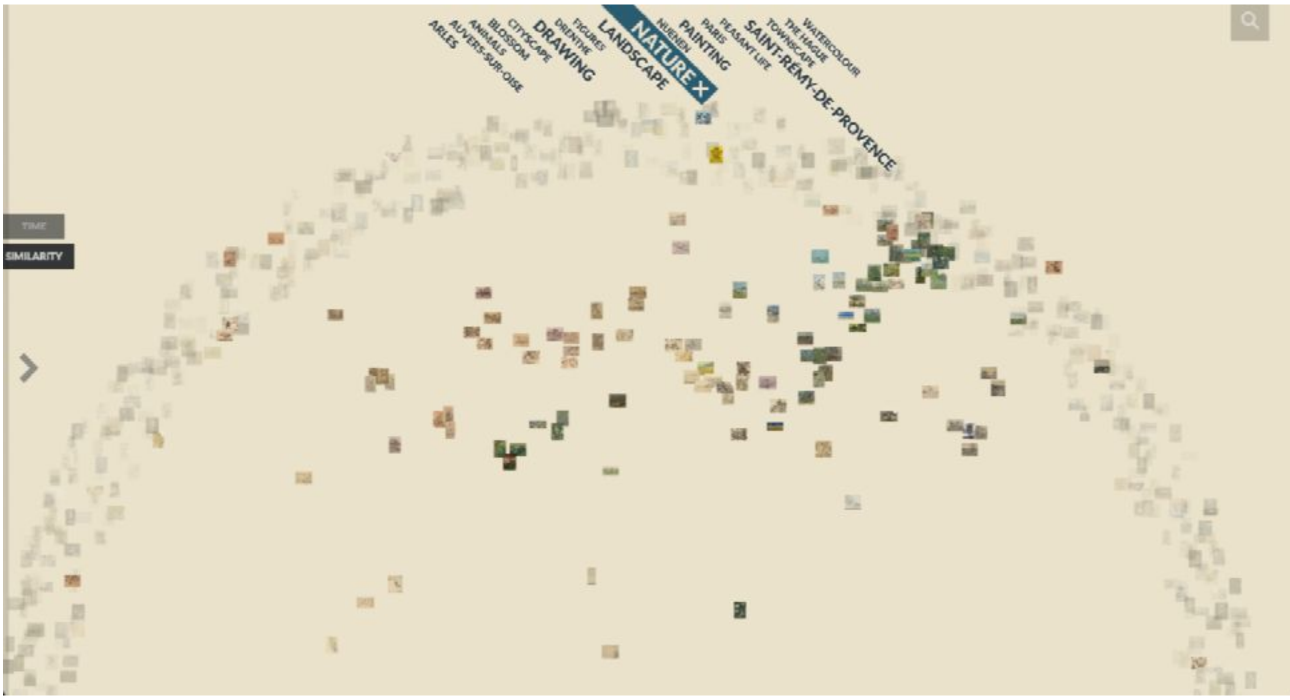


Fig. 19: Selection of all images filtered for the keyword 'nature' arranged by similarity in the cluster view (screenshot, 2/12/2022).

In addition to the machine-automated clustering of image data based on statistical pixel similarities, Vikus Viewer offers an overlay mode. As with tabular displays, the scatterplot can be supplemented with categorical, human-curated information. Selecting one of the categorical keywords from the upper filter bar hides the inapplicable images in the scatterplot (and reorganises them in an outer circle). Images categorised by the human-curated keyword remain in the same place, so that the relationship between clustering and keyword is visible. This view raises questions about the ontology of keywords, but also shows that machine-statistical clustering creates new orderings, but not necessarily according to humane criteria.
One potential options, which we don't see realised in this early project, but in Imgs.ai for instance, is changing the clustering based upon the selection of negative or positive images. Another potential option would be manually grouping certain images via drag and drop (changing their weights) and the machine-statistical reaction through recalculating all other weights and positions in the cluster.[13]

---

[13] Many statements already made in the corresponding section on iArt's scatterplot and clustering also apply to the Vikus Viewer section. I have not repeated them here.

2.5 X Degrees of separation as tabular k-nearest neighbour path

X Degrees of Separation (https://artsexperiments.withgoogle.com/xdegrees/) is a visual experiment on the Google Arts platform. It is a collaboration between artist Mario Klingemann and Simon Doury (Google Cultural Institute), which makes about 250,000 image data objects from museum collections available as part of the Google Arts & Culture project (status 2017), in the form of Big Data and machine learning. The project invokes another display format: the k-nearest neighbour path. In contrast to the overview mode in tables and scatterplots, this form of visualisation promises a 'close reading' of the visual material. The aim is to unfold a narrative using automation, thereby guiding the viewer through the extensive and almost unmanageable material and making unexpected connections visible.[14]
The interface offers a range of images, from paintings to photographs of objects, to choose from. The user selects one image as the starting point and another as the destination, resulting in the creation of a nearest neighbour path between the two.

Fig. 20: Pathway connecting a self-portrait of Frida Kahlo from 1945 as the starting point to a portrait photograph of the artist from 1941 (screenshot, 2/12/2022).

In the tabular display of X Degrees of Separation, the image heights are adjusted to each other so that the original proportions are lost. The algorithm is oriented towards correlation in form, such as hair and face, fabric with drapery and relatively monochromatic backgrounds. It takes users on a wild ride through art history that remains both ahistorical and incoherent (Fig. 20). In terms of art history, Frida Kahlo's works could be positioned in relation to or in distinction from the Mexican Muralists or Surrealism, for instance (Deffebach 2015). However, these do not appear here at all. Outside this observation, added value could lie in the accidental (re)discovery of corresponding artistic positions, but this thesis would have to be tested in the scope of a concrete curatorial project.

[14] An implementation with PyTorch: https://dzlab.github.io/dl/2019/02/02/X-Degrees-Separation/.

Fig. 21: Pathway leading from a 15th century sculpture by Tilman Riemenschneider to a woman's 20th century slipper (screenshot, 2/12/2022).

In a second example, the bright, monochrome background that serves to highlight the objects stands out as a constant that presumably helps to make the images k-nearest neighbours (Fig. 21). Referring to the effectiveness of texture bias (Geirhos et. al. 2019), we can assume that the outline plays a very minor role in the distribution of images in the latent space of the weighted networks. This is apparent, when in the penultimate picture, for instance, the shoe (intended as an outline) points to the right, whereas in the target picture it points to the left. The sculptures are characterised by folds, which also characterise Eugene C. Miller's men's boot – a possible explanation for the transition. It is striking that the sculpture by Henry Glicenstein is similarly aligned in outline to the boot by Miller that follows it. No art-historical connections are discernible.

These two tests provide insight into the potential and limitations of a k-nearest neighbour experiment.[15] It is certainly suitable for interactive projects in museum collections to make lesser-known aspects of a collection accessible to the public. Ultimately, there is one consideration to keep in mind: exploration projects do not necessarily have to be based on resource-intensive machine learning. The Science Museum of London's Never been Seen project, for instance, presents a technically simpler possibility. On the website https://thesciencemuseum.github.io/never-been-seen/, a random generator selects the digital copy of an object in the collection that has never been viewed online by a person before.

[15] Another k-nearest neighbour approach, the cooperation between FH Potsdam and the Staatliche Museen Berlin *A Visual Exploration of Two Museum Collections* (https://visualisierung.smb.museum) is described in detail in (Brüggemann et. al. 2022). The questions raised here by the example of X Degrees of Separation apply in a similar way to the pathway display in *A Visual Exploration of Two Museum Collections.*

3 Conclusions

3.1. Knowledge through observation

Media visualisations become interactive because the human gazing-as-thinking continually creates orders of meaning-generating, which according to media philosopher Sybille Krämer, are spatially oriented: “In the diagrammatic inscription, the surface serves as a space of order and arrangement, and topological relationships such as above/below, left/right, central/peripheral become components that constitute meaning. There is no writing without a reading and writing direction, no diagram without an orientation [...]” (Krämer 2010, 36). One of the effects of diagrammatic arrangement is that the two-dimensional surface enables a simultaneous presence that homogenises the information objects. Krämer argues that humans, starting from their own physicality, gain knowledge by means of directionality, i.e. they create meaning through orders, arrangements and patterns.

The process of human recognition is essentially based on visual similarities. The eye and brain recognise the edges of objects and the components, and the more context there is, the better, facilitating the recognition of even previously unseen objects (Biederman 1987). In the process of seeing-recognising, the eye sweeps over the surface and identifies distinctions or similarities based on proximity and distance. During this process, similarity and proximity have an additive effect on human recognition performance (Kubovy and van den Berg 2008). Recent research supposes a multistage vision-cognition that includes moments of encoding, selection and decoding of visual information. Central to this process is that the brain blocks out a large part of the image context – in favour of individual components, on which it subsequently concentrates for decoding (Zhaoping 2019).

In summary, we can state for media visualisations that the wandering human gaze is capable of generating new knowledge by means of visual similarities and spatial allocations and that this cultural-technical process of observation is shaped by media, history and culture. Hence, the user interface of The Curator's Machine, as a knowledge tool comparable to Latour's ‘laboratory’ (Latour 1987, 63–101), is involved in the generation of new knowledge. Special attention must be paid to the experimental set-up of the ‘lab’, i.e. the design of the user interface.

3.2 The limits of clustering

What is surprisingly obvious is that the issues developed here in the material coincide with a more fundamental problem of big data and artificial ‘intelligence’: “In claiming that data make decisions, scientists and others displace multiple forms of ideologically influenced subjectivity that are heavily involved in the curation of datasets, selection of available codes, algorithms, parameters, and the labeling of known data” (Dobson 2019). The more the investigation of the case studies progressed, the more the question arose as to whether scatterplots are capable of saying anything at all about the underlying data.

The tests showed that while clustering based on similarity was partially interpretable, other information, such as trends, which become visible in classic scatterplots through linear regression lines and labelling of the scales, were not available. Thus, the scatterplots examined here often remain enigmatic and are difficult to interpret. This problem runs through all the case studies I have examined and requires further research.
We should take this observation a step further. The information design theorist Edward Tufte noted with urgency in 1983 in his standard work *The Visual Display of Quantitative Information*: “Graphics must not quote data out of context” (Tufte 1983, 74). But the question as to what the context of scatterplots without scale labelling is remains open, as information designer Michael Correll polemically notes in a blog post: “Nobody with any actual connection to your data knows what the axes of the resulting chart mean beyond a very basic folk-algorithmic ‘these points are close together so I guess they are similar,’ and you could have communicated that information directly with hierarchical clustering or an adjacency matrix or something other than whatever unique set of linear algebra or machine learning dark magics you had to do to make your weird, uninterpretable, scatterplot” (Correll 2022).
Tufte’s dictum and Correll’s critique draw particular attention to the absence of graphic context (which would provide meaning) in today's visualisations from weighted networks.

Fig. 22: In this example we can see clustering’s arbitrary similarity in the example of the t-SNE map experiment. The photographed objects from various collections (spoons, tools, instruments) have been ordered according to the colour of the image background: white at the top, greenish-grey in the middle, grey dark-grey at the bottom. (Screenshot Google Arts & Culture, t-SNE Map).

Clustering without scales does not yield any information about the two dimensions according to which the points of the cluster are represented (Sarikaya and Gleicher 2018, 5). We can therefore read it in two ways: 1.) Intuitively as a statement about the correlation between the data points and 2.) as a statement about the inner configuration of the weighted networks. This results in an interpretational problem. Since no parameters are given, viewers have no idea what the diagram is presenting to them. Is it the correlation of data points? Or is it the inner structure of the underlying weighted networks, the convolutional 'neural' networks?

Despite these fundamental doubts, all is not lost. After all, scatterplots serve as excellent exploration interfaces. To a certain extent, they are able to compensate for the disadvantage of tabular displays, in which correlation is expressed by sequence, which in turn is distorted by arbitrary line breaks. The distributions of scatterplots allow – starting from a centre – spatial orientations in numerous reading directions and thus a higher degree of visual freedom than the orientation in the table row, which points in only one direction. The almost random arrangement of similarity distributions in scatterplots proves advantageous when it occurs in conjunction with categorical ordering principles (filters, search functions, etc.) and visual zoom options. In this case, it is possible to counter the inherent visual problems of tabular display, especially faulty visual hierarchies.
However, scatterplots come with their own set of problems, because, as in the underlying weighted networks (the convolutional neural networks), how 'similarity' is actually constituted remains intransparent. The knowledge gained from 'similarity distributions' is therefore limited. Other visual metaphors or ordering principles should be further explored. These would presumably also have to relate more strongly to the categorial ontologies, which are inscribed in collection objects in digitisation (as metadata) but are eliminated in the multidimensional space of weighted networks (especially the convolutional neural networks).

3.3 Practical implications for the further work of Training the Archive

The Curator's Machine being developed in the scope of Training the Archive will only be able to address some of the issues raised through new approaches. For more fundamental new visual solutions, a different and much more extended research design is needed. The following section identifies some practical conclusions and suggestions:

Concepts: The user interfaces, which visualise non-trivial concepts of machine learning beyond tabular (and thus practised) displays, use concepts such as 'k-means clustering', 'global weights', or 'distance: angular/Manhattan'. These new ordering concepts have not yet been adequately explained and documented in the user interfaces. Their effect can at best be determined by trial and error – hence the need for better documentation and/or explanation of these options.
Discrimination: In some projects, discriminatory language, such as the racist terms 'negro', or sexist descriptions like 'slut' or 'bitch', generated relevant hits. This highly problematic behaviour affected both projects with historical ontologies, and projects with word-embeddings such as CLIP. Of the case studies examined, only the narrowly defined and curated projects of the Vikus viewer were immune to these discriminatory language and image politics.
Language Embedding: Switching language from English to German, as has been shown, creates changes in media – from paintings of mountains to postcard views of mountains, for

instance. Users should therefore be better informed about the working language of the embeddings. Alternatively, it may be useful to limit the languages to a single language.
Correlation line: A visible correlation line within the clusters would provide orientation as to how strongly the individual data points are oriented towards the (statistical) mean.
Outliers: A filter view that only shows outliers, i.e. data points that are far away from the mean of a cluster, would be helpful so that they can be inspected individually.
Scale: Displays of the images in relation to each other are not to scale, as all objects are adjusted to a uniform dimension. A possible solution would be to indicate the degree of reduction as a percentage in the images (as an overlay).
Drag & Drop and Grouping: One display function that would be desirable from the user's point of view is interactive sorting via drag-and-drop in order to group items. This would be similar to the Imgs.ai project's visual search function, in which the positive or negative selection of images results in a reordering of the latent space. According to the current goals, the drag-and-drop option for The Curator's Machine is to be interactively interconnected with the recalculation of visual similarities and enable further options, such as visual groupings.
Help/maintenance: Only some of the projects listed have institutionalised technical support. We were able to reach the project leaders of the projects that were still in the experimental stage, and they responded when availability allowed. As a result, Training the Archive needs to initiate a permanent hosting solution and institutionally based technical contact persons in cooperation with the RWTH.

Interactive media visualisations like the ones presented here enable an interplay between cognitive operations. They do so by interpreting the collection objects diagrammatically as data objects (first layer), operationally (second layer), and via a graphical user interface (third layer) through user interventions. Conversely, user interfaces can be 'read' in terms of which cognitive operations they offer to establish relations: selecting, filtering, sorting, ordering, grouping, summing, comparing similarities.
Tables, scatterplots and k-nearest neighbours – what they have in common is spatial distribution as a visual argument that aims at generating knowledge. At the same time, it became obvious to what extent the visualisations are reductions and reconstructions of reality. These lead to an interpretation problem when using convolutional weighted networks, since it remains unclear when looking at scale-free diagrams which parameters have been set in relation to each other. Therefore, it also remains unclear whether the viewers actually see similarities of data points or rather interpretations of the configurations of those convolutional weighted networks that were used for the calculation.
This text has conducted a close, critical reading of user interfaces based on numerous case studies. The tabular display and the distributed media visualisations in scatterplots were at the centre of this discussion. From the preceding considerations, we require more research on visualising the results of artificial 'intelligence'.

Many thanks to Inke Arns, Dominik Bönisch, Roland Meyer and Kim Albrecht for their valuable comments and advice. A special thanks goes out to Andreas Broeckmann who provided beneficial comments on an early draft.

BIBLIOGRAPHY

Biederman, Irving. 1987. "Recognition-by-Components: A Theory of Human Image Understanding." *Psychological Review* 94 (2): 115–47. doi:10/c9dqh2.

Bönisch, Dominik. 2021. "The Curator's Machine. Clustering von Musealen Sammlungsdaten Durch Annotieren Verdeckter Beziehungsmuster Zwischen Kunstwerken." *Training the Archive – Working Paper, Aachen/Dortmund*, May. doi:10.5281/ZENODO.4604880.

Bowker, Geoffrey C. 2014. "The Theory/Data Thing Commentary." *International Journal of Communication* 8: 1795–99.

Bowker, Geoffrey C., and Susan Leigh Star. 1999. *Sorting Things Out – Classification and Its Consequences*. Cambridge, MA: MIT Press.

Brüggemann, Viktoria, Mark-Jan Bludau, Christopher Pietsch, and Marian Dörk. 2022. "Von Der Wolke Zum Pfad – Visuelle Und Assoziative Exploration Zweier Kultureller Sammlungen." In *Zenodo*. doi: 10.5281/ZENODO.6327939.

Bry, François, Clemens Schefels, and Corina Schemainda. 2018. "Eine Qualitative Analyse Der ARTigo-Annotationen: Serious Games, Games With A Purpose, Suchmaschine Für Kunstwerke, Kunstgeschichte, Human Computation, Semantische Tiefe." Edited by Piotr Kuroczyński, Peter Bell, and Lisa Dieckmann. *Computing Art Reader*. University Library Heidelberg, 96–114. doi: 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*. January 6. https://mcorrell.medium.com/bad-ideas-in-visualization-77d378148d35.

D. Haggerty, Kevin, and Richard V. Ericson. 2000. "The Surveillant Assemblage." *British Journal of Sociology* 51 (4): 605–22. doi: 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, and Daniel Denis. 2005. "The Early Origins and Development of the Scatterplot." *Journal of the History of the Behavioral Sciences* 41 (2): 103–30. doi: 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, and Wieland Brendel. 2019. "ImageNet-Trained CNNs Are Biased towards Texture – Increasing Shape Bias Improves Accuracy and Robustness." *ArXiv:1811.12231*, January. http://arxiv.org/abs/1811.12231.

Gitelman, Lisa. 2013. *Raw Data Is an Oxymoron*. Cambridge, MA: MIT Press.

Glinka, Katrin, Christopher Pietsch, and Marian Dörk. 2017. "Von Sammlungsspezifischen Visualisierungen Zu Nachnutzbaren Werkzeugen." In *Books of Abstracts*, edited by Michael Stolz and Patrick Helling, 204–7. Bern. https://uclab.fh-potsdam.de/wp/wp-content/uploads/dhd2017-aus-programm.pdf.

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, June. doi: 10.5281/ZENODO.4742621.

———. 2021b. "Curation and Its Statistical Automation by Means of Artificial 'Intelligence.'" *Training the Archive – Working Paper 3,* Aachen/Dortmund, November. doi: 10/gn8dn7.

———. 2022. "Die Form Der Datenbank – Genealogien, Operationalitäten Und Praxeologien Relationaler Datenbanken in Ost Und West." Dissertation, Weimar: Bauhaus Universität Weimar [in preparation].

———. 2023. "'Why so Many Windows?' – How the ImageNet Image Database Influences Automated Image Recognition of Historical Images." *International Journal for Digital Art History*, [pre-print].

König, Dénes. 1936. *Theorie Der Endlichen Und Unendlichen Graphen*. Edited by Horst Sachs and 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.*, edited by Gabriele Brandstetter, Franck Hofmann, and Kirsten Maar, 29–45. Freiburg/Berlin/Wien: Rombach Verlag.

Krämer, Sybille, and Horst Bredekamp. 2003. "Kultur, Technik, Kulturtechnik – Wider Die Diskursivierung Der Kultur." In *Bild, Schrift, Zahl.*, edited by Sybille Krämer and Horst Bredekamp, 11–22. Kulturtechnik. München: Fink.

Kubovy, Michael, and 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 (1): 131–54. doi: 10/b2vw3s.

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, and James Melville. 2020. "UMAP – Uniform Manifold Approximation and Projection for Dimension Reduction." *ArXiv:1802.03426 [Cs, Stat]*, September. http://arxiv.org/abs/1802.03426.

Radford, Alec, Jong Wook Kim, Chris Hallacy, Aditya Ramesh, Gabriel Goh, Sandhini Agarwal, Girish Sastry, et. al. 2021. "Learning Transferable Visual Models From Natural Language Supervision." arXiv. http://arxiv.org/abs/2103.00020.

Russell, Henry Norris. 1914. "Relations between the Spectra and Other Characteristics of Stars." In *Popular Astronomy*, 22:275–94. https://ia802300.us.archive.org/16/items/sim_popular-astronomy_1914-05_22_5/sim_popular-astronomy_1914-05_22_5.pdf.

Sarikaya, Alper, and Michael Gleicher. 2018. "Scatterplots – Tasks, Data, and Designs." *IEEE Transactions on Visualization and Computer Graphics* 24 (1): 402–12. doi: 10.1109/TVCG.2017.2744184.

Schneider, Stefanie, and Hubertus Kohle. 2021. "Bilderlesemaschinen – Einsatzszenarien Maschineller Lernverfahren Im Musealen Raum." Museum and the Internet (MAI) 2021 Tagung, May 28. https://www.youtube.com/watch?v=lsSpzjtS-rM.

Simonyan, Karen, and Andrew Zisserman. 2015. "Very Deep Convolutional Networks for Large-Scale Image Recognition." *ArXiv:1409.1556 [Cs]*, April. http://arxiv.org/abs/1409.1556.

Springstein, Matthias. 2021. "AW: IArt Fragen." Email to the Author.

Springstein, Matthias, Stefanie Schneider, Javad Rahnama, Eyke Hüllermeier, Hubertus Kohle, and Ralph Ewerth. 2021. "IART – A Search Engine for Art-Historical Images to Support Research in the Humanities." *ArXiv:2108.01542 [Cs]*, August. http://arxiv.org/abs/2108.01542.

Taylor, Frederick Winslow. 1903. *Shop Management*. London, New York, NY: Harper & Brothers.

Toshev, Alexander, and Christian Szegedy. 2014. "DeepPose – Human Pose Estimation via Deep Neural Networks." *2014 IEEE Conference on Computer Vision and Pattern Recognition*, June, 1653–60. doi: 10/ggtmv3.

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 (October): 1–10. doi: 10/ghndbt.