Artl@s Bulletin

Volume 6
Issue 3 *Visualizing Networks: Approaches to Network Analysis in Art History* Article 6

2017

The Computer as Filter Machine: A Clustering Approach to Categorize Artworks Based on a Social Tagging Network

Stefanie Schneider
Ludwig-Maximilians-Universität München, stefanie.schneider@itg.uni-muenchen.de

Hubertus Kohle
Ludwig-Maximilians-Universität München, hubertus.kohle@gmail.com

The Computer as Filter Machine: A Clustering Approach to Categorize Artworks Based on a Social Tagging Network

Stefanie Schneider *
Hubertus Kohle **
Ludwig-Maximilians-Universität

Abstract

Image catalogs containing several million reproductions of artworks still pose a costly or computationally intensive challenge if one tries to categorize them adequately, either in a manual or automatic way. Using crowdsourced annotations assigned by laypersons, this article proposes the application of a clustering algorithm to segment artworks into groups. It is shown that the resulting clusters allow for a consistent reclassification extending the traditional categories (history, genre, portrait, still life, landscape), and thus enable a finely-grained differentiation which can be used to search in and filter image inventories, among other things.

Auszug

Es stellt noch immer eine kosten- oder rechenintensive Herausforderung dar, Bildkataloge mit mehreren Millionen Reproduktionen von Kunstwerken händisch oder automatisch zu kategorisieren. Dieser Aufsatz schlägt die Anwendung eines Clustering-Algorithmus auf crowdgesourcte Annotationen von Laien vor, um Kunstwerke in Gruppen zu segmentieren. Es zeigt sich, dass die resultierenden Cluster eine konsistente Reklassifizierung ermöglichen, die von den traditionellen Gattungskategorien ausgehen (Historie, Genre, Porträt, Stillleben, Landschaft), diese aber auch transzendieren. Dadurch wird eine feinkörnige Differenzierung erreicht, die unter anderem zur Suche in und Filterung von Bildinventaren genutzt werden kann.

** Stefanie Schneider is a M.Sc. student in statistics for economics and social sciences at Ludwig-Maximilians-Universität München. Her research interests center on techniques which combine methods of text mining and computer vision for application in the digital humanities, in particular art history.*
*** Hubertus Kohle is a professor of art history at Ludwig-Maximilians-Universität München with interests in 18th to 20th century German and French art and digital art history. He published books on French 18th to 19th social art history, Adolph Menzel, overviews on German mid-19th century art, the computer in art history, and Denis Diderot's art theory.*

The term "network" refers to a system which places items of any kind (*nodes*) in relation to one another in terms of *edges*. In art history, if even applicable, the unavoidable spatialization that occurs in visualizing such a system has led to the analysis of spatial relations as networks, in particular within the context of the methodically modern "spatial turn."[1] Networks, though, do not generally refer to spatial relationships. Even a library catalog can be seen as a network. But although a spatial element is addressed in the relationship between author and place of publication, the edge between author and book title also forms a relation, without resulting in a spatial one.

For art history, image catalogs are even more important than library catalogs, having emerged in many places over the past years, some containing several million reproductions of artworks.[2] Normally such catalogs are used as *containers*, from which individual, already-known works are extracted. If need be, one searches through them based on keywords, for example parts of the title or human-assigned classifying terms, in order to identify a thematically restricted image inventory, which was formerly unknown. In any case this method is reminiscent of analog precursors, as if one were to sift through different boxes of card indexes.

The *clustering method* proposed hereinafter is based on the network structure of the image catalog and attempts to take the organizing potential of the computer into account more strongly, thus enabling it to determine the ordering principles itself. These principles will not be completely different from traditional ones though, as they continue to be based on labels generated by humans. Our approach differs from similar efforts to categorize artworks with the help of computational tools insofar as it does not include formal aspects of a pixel by pixel image addressing.[3] Instead it relies solely on *crowd* annotations, namely those from ARTigo.

About the Corpus

ARTigo[4] is two things: Firstly, it is an internet platform in which digital reproductions of artworks are presented to an audience with unknown qualifications, who then annotate these artworks in a playful and competitive way. As there is no obligation to register or to provide socio-demographic characteristics such as gender, age, family background, or level of education, the diversity of the users cannot be defined in detail. However, due to the fact that the project was developed at a university, one can assume that many students (in particular of art history) are among the over 30,000 users who have played the game until now. Secondly, ARTigo is a semantic search engine which can master large image sets based on these crowdsourced annotations (*tags*) without having to rely on the expensive *manpower* of specialists—or even on artificial intelligence from the field of computer vision. The resulting corpus is used to search for works whose identity cannot be determined by identifying the author and title, which are available as metadata in traditional image archives.

Since 2007, we at the Institute of Art History in cooperation with the Institute of Computer Science at Ludwig-Maximilians-Universität München have gathered 9.3 million German, English, and French

[1] See also Martin Papenbrock and Joachim Scharloth, "Datengeleitete Analyse kunsthistorischer Daten am Beispiel von Ausstellungskatalogen aus der NS-Zeit: Musteridentifizierung und Visualisierung," in *Kunstgeschichte. Open Peer Reviewed Journal*, 2011, accessed January 7, 2017, http://www.kunstgeschichte-ejournal.net/248/.

[2] The leading German databases are that of the Bildarchiv Foto Marburg, "Bildindex Kunst und Architektur" (http://www.fotomarburg.de/forschung/datenbanken/bildindex accessed January 7, 2017) as well as the Prometheus Bildarchiv (http://prometheus-bildarchiv.de/, accessed January 7, 2017).

[3] See for instance Jana Zujovic et al., "Classifying Paintings by Artistic Genre: An Analysis of Features & Classifiers," in *Proceedings of the 16th International Workshop on Multimedia Signal Processing*, Rio de Janeiro, Brazil, 2014, accessed January 7, 2017, http://infolab.northwestern.edu/static/papers/classifying-paintings-by-artistic-genre-an.pdf; Babak Saleh and Ahmed Elgammal, "Large-scale Classification of Fine-Art Paintings: Learning The Right Metric on The Right Feature," 2015, accessed January 7, 2017, https://arxiv.org/pdf/1505.00855v1.pdf.

[4] http://www.artigo.org/ (accessed January 7, 2017). For a more detailed description of the game, see Hubertus Kohle, "Kunstgeschichte goes Social Media. Laien optimieren eine Bilddatenbank – mit einem digitalen Spiel," in *Aviso : Zeitschrift für Wissenschaft und Kunst in Bayern* 3 (2011), 38–43; Hubertus Kohle, "Artigo. Social image tagging pour les œuvres d'art," in *L'art et la mesure. Histoire de l'art et méthodes quantitative*, ed. Béatrice Joyeux-Prunel (Paris: Ed. Rue d'Ulm, 2009), 153–164.

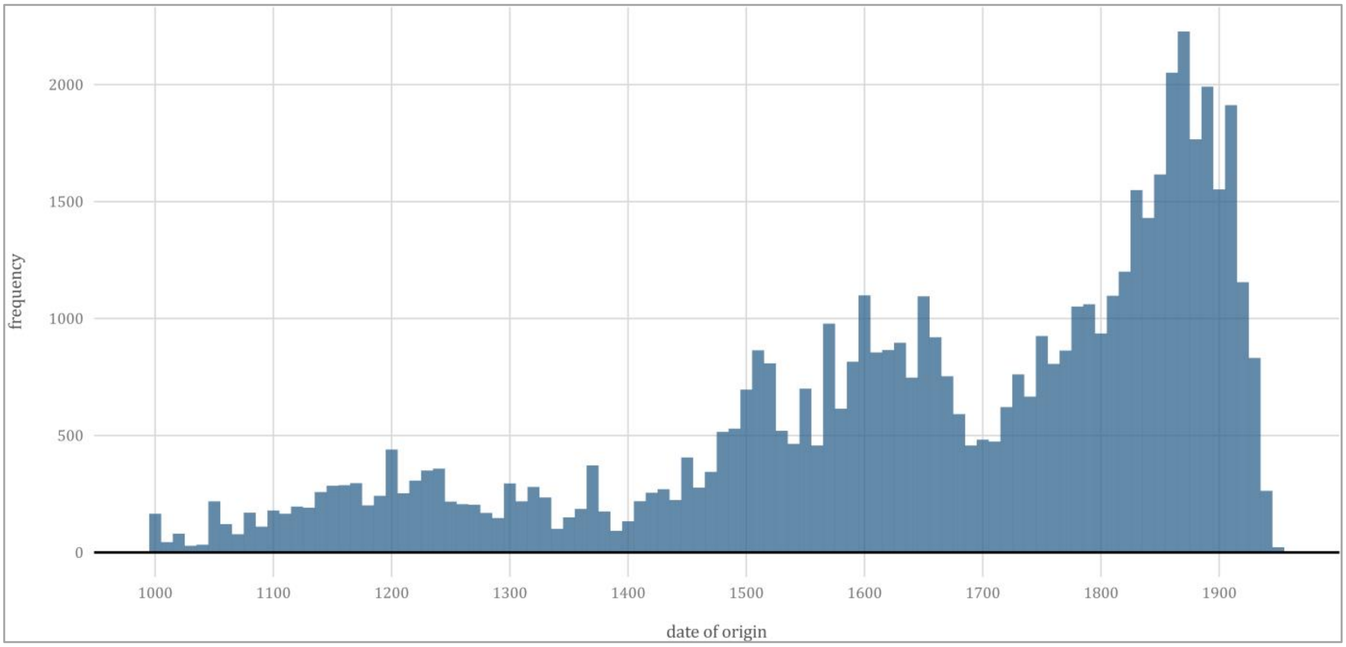

Figure 1. Distribution of the date of origin in ARTigo. Ranges were approximated by their arithmetic mean. Given that the database contains only 153 digital reproductions prior to 1000, which furthermore spread over 25 centuries, those cases were excluded in the visualization. The database solely consists of images in the public domain and therefore includes reproductions by artists who died before 1946. Created with R and the package ggplot2 (Hadley Wickham, *ggplot2: Elegant Graphics for Data Analysis*, New York: Springer, 2009).

language *taggings*[5] for over 55,000 artworks[6] through this "ecosystem."[7] The image repository, which has been expanded over time,[8] encompasses digital reproductions from the 15th century B.C. (Thutmose III, "Karnak, the Temple of Amun") to Modernism (Franz Marc, "Fighting Forms"). The database, however, was not created according to systematic criteria; instead, it was orientated towards the Europe-centered research and teaching interests of the faculty at the time of its inception. This is accompanied by a focus on 19th century art, which makes up 28.9 percent of the corpus and which shows itself as a peak in Figure 1. A second, though less pronounced peak can be found with an amount of 14.3 percent in the 17th century.

The triangular relationship between the crowdworker, the resource which is to be annotated, and the annotation itself results in a tripartite network. In contrast to hierarchical classification structures, social tagging services rely on an indexing process which introduces no authority in order to verify which tags are considered suitable and which are not. On the one hand, this system offers an advantage: the players do not have to stick to a predetermined vocabulary; instead, they can give annotations which represent "their own voice," without having been influenced by any presettings. This results in a set of dynamic, heterogeneous tags. On the other hand, semantic relations between words also pose a challenge to computational methods: Without algorithmic "tuning" in the backend, an image that has been given the word "horse" would be recognized as dissimilar to an image that was tagged with the plural form "horses" or synonymous expressions like "stallion" or "pony"—and thus would not be listed in the results of a search query.

[5] Tagging refers to the user-generated process of annotating a resource with a tag.
[6] These figures, as well as the following statistics, are based on a database dump from May 3, 2016.
[7] For further details, see Christoph Wieser et al., "ARTigo: Building an Artwork Search Engine With Games and Higher-Order Latent Semantic Analysis," in *Proceedings of Disco 2013, Workshop on Human Computation and Machine Learning in Games at HComp*, Palm Springs, CA, USA, 2013, accessed January 7, 2017, http://www.en.pms.ifi.lmu.de/publications/PMS-FB/PMS-FB-2013-3/PMS-FB-2013-3-paper.pdf.
[8] From the inventories of the Kunsthalle Karlsruhe, the Rijksmuseum Amsterdam, the Albertina in Vienna, and the Mead Art Museum of Amherst College.

In order to avoid misuse, the program has a validation mechanism in effect. Therefore, a tag is only considered valid, and thus rewarded with points according to the game's rules, when at least two users have entered it, whether in the current game session or in a previous one. This integrated competitive spirit does not decrease the versatility gained through collaboration. Rather, it encourages the crowd to describe an image with contextual aspects, which are quite easy to come up with at first glance,[9] instead of focusing on more complex formal criteria.[10] After all, this increases the chance that an already assigned annotation will be entered and—not insignificant for reasons of ambition and for the joy of playing—that one's high-score will be improved.

It is still disputed among experts how reliable crowdsourced information is on art historical artefacts.[11] Therefore, it seems reasonable to question the gathered annotations based on their relevance, that is, to determine to what extent the descriptions comply with professional criteria. On the other hand, it could turn out that non-expert annotations about artworks could be particularly significant. The assumption that insights remaining underexposed in professional discourses could be hiding in laymen evaluations is put aside for now, though. Even if such assessments should be examined based on their added value, which becomes apparent particularly when traditional knowledge is questioned (at least in regard to its rationality), they only play a minor role in the network analysis that will be covered in the following sections.

Methodology

The "wisdom of crowds" serves as a vehicle for connecting self-organizing, iterative methods of *unsupervised learning*, which initially leave out specialized academic considerations. Instead, the computer is entrusted with the task of mathematically detecting *non-random* patterns in the crowdsourced tags. As a result, the observed regularities allow the inventory to be segmented into preferably homogenous, disjointed groups. In contrast to categorization and classification approaches in *supervised learning*, unsupervised methods operate without assistance from a human authority assigning concrete allocations *a priori* to which an algorithm can orient itself.[12] The machine decides autonomously which criteria are used to create partitions. In the end, it is again up to humans to discuss such computational recom--mendations, particularly to the extent that they can quantitatively confirm traditional patterns as well as raise new questions.

Our procedure rests on two mathematical pillars. Guiding these is a term-document matrix with *tf-idf weighting*[13] whose two-dimensional structure depicts tags (*terms*) in rows and resources (*documents*) in columns. First, a *Partial Singular Value Decomposition* reduced this matrix to ten principal components.[14] For this purpose, we used the Lanczos algorithm according to Baglama and Reichel,[15] implemented in R[16] in the package irlba,[17] to uncover correlations existing between tags and transfer them to a lower-dimensional feature space, which no longer focuses on individual annotations but rather on latent concepts. Because they have been annotated together more frequently than random in the resources, words with

[9] For example, Turner's "The Burning of the Houses of the Lords and Commons, October 16, 1834" lists as the most prominent annotations (in descending order): "Bridge," "Sky," "Fire," "Water," "Clouds" (in German: "Brücke," "Himmel," "Feuer," "Wasser," "Wolken").

[10] François Bry and Christoph Wieser, "Squaring and Scripting the ESP Game: Trimming a GWAP to Deep Semantics," in *Proceedings of the International Conference on Serious Games Development and Applications*, Bremen, Germany, 2012, accessed January 2, 2017, http://www.en.pms.ifi.lmu.de/publications/PMS-FB/PMS-FB-2012-10/PMS-FB-2012-10-paper.pdf.

[11] Cf. Clay Shirky, *Here Comes Everybody: The Power of Organizing Without Organizations* (London: Penguin, 2008).

[12] Ludwig Fahrmeir, Alfred Hamerle and Gerhard Tutz, *Multivariate statistische Verfahren*, 2nd revised edition (Berlin: De Gruyter, 1996), 437–439.

[13] Tf-idf is a statistical measure. It refers to the frequency with which a tag is annotated in one resource (*term frequency*) in relation to the frequency with which this tag is annotated in all resources (*inverse document frequency*).

[14] Ten principal components retain 34.3 percent of the data's variance.

[15] James Baglama and Lothar Reichel, "Augmented implicitly restarted Lanczos bidiagonalization methods," in *SIAM Journal of Scientific Computing*, 27(1), 2005, 19–42, accessed January 2, 2017, http://www.math.kent.edu/~reichel/publications/auglbd.pdf.

[16] R is a programming language and an *open-source-software* which is ideal for statistical issues, data analysis, and data visualization and which functionality can be expanded through packages, see R Core Team, *R: A language and environment for statistical computing*, R Foundation for Statistical Computing, Vienna, Austria, 2016, accessed January 3, 2017, https://www.r-project.org/.

[17] Jim Baglama and Lothar Reichel, *irlba: Fast Truncated SVD, PCA and Symmetric Eigendecomposition for Large Dense and Sparse Matrices*, R package version 2.1.2, 2016.

obviously similar meaning ("water," "river," "lake") are "pulled together" just like descriptions which seem rather disparate to human understanding of language ("dog," "rocks," "curtain"). The goal was to obtain an economical form that condenses the corpus with minimal loss of information and that can also function as an appropriate basis for a graphical representation. Furthermore, infrequent--ly annotated artworks can definitely be a part of further study under a concept-based model, even if the number of tags is so sparse that alternative mathematical approaches are no longer capable of useful classification.

Afterwards, the clustering algorithm *Partitioning Around Medoids*, which is initialized with the R-package cluster,[18] segmented the dimensionally reduced matrix into groups using the angular distance[19] and a previously specified start configuration, i.e. the number of clusters that should be formed. The approach developed by Kaufman and Rousseeuw[20] finds representative centers (*medoids*) and assigns objects to them which are, in the mathematical sense, close. A medoid is approximated by a concrete resource and is not only defined by a key figure. For instance, Rembrandt's "The Stoning of Saint Stephen" takes the role of a medoid in Cluster 5 (by partitioning into nine groups), which can be seen in the following section. Certain "alliances" develop: the extent to which they are considered to be "neighbors" is based on similar feature constellations, here on the abstract level of concepts. Compared to *k-means*, another non-hierarchical clustering method, the afore--mentioned algorithm is more robust against extreme observations (*outliers*) which are difficult to categorize due to their somewhat particular composition, whatever kind that may be. Thus, these observations should be interpreted very carefully.

Results

In order to evaluate the applied methods, we took a stratified random sample of 7,105 resources, which maintains the distribution of the total inventory according to the date of origin. Five and nine clusters were used as starting configurations of the clustering algorithm. A division into five clusters was deemed plausible from an art historical perspective because it corresponds to the way artworks are traditionally categorized in genres; a division into nine was carried out based on the highest average silhouette coefficient.[21] Figures 2 and 3 show the detected medoids for five and nine clusters, respectively, including information about the corresponding group sizes. It is clear upon considering the results that not every cluster size leads to similarly satisfying results, whereby "satisfying" should initially be understood as the proximity to traditional classifications. In the context of art historical data, these classifications particularly refer to different categories, usually represented by the classifiers "history," "genre," "landscape," "portrait," and "still life." One could also add architecture: in terms of characteristics, architecture stands contrary to the classifications of the pictorial arts; however, it is equally extensively documented in our database.

The results of a classification into five clusters were only satisfying to some extent, although the identity of the cluster and category numbers promise the greatest odds for the unbiased viewer.[22] The first group could be defined as landscape cluster—which can be seen in the homogeneous aggregation in Figure 4, first graphic on the left—yet it also includes many artworks that do not emphasize landscapes. These are, however, predominantly history paintings or cityscapes with pronounced landscape components (Heinrich Gentz, "Plan of a Royal Summer Palace"), though there are also themes in which landscape does not even play a

[18] Martin Maechler et al., *cluster: Cluster Analysis Basics and Extensions*, R package version 2.0.5, 2016.
[19] In contrast to the transformation of the cosine similarity common in information retrieval, the angular distance is a proper distance metric, and was therefore preferred.
[20] Leonard Kaufman and Peter J. Rousseeuw, *Finding Groups in Data: An Introduction to Cluster Analysis* (New York: John Wiley and Sons, 1990).
[21] The silhouette coefficient is a measure to assess the goodness of clustering. It calculates the ratio of an objects distance to all other objects in its cluster to its distance to all objects in its nearest neighboring cluster. See Peter J. Rousseeuw, "A Graphical Aid to the Interpretation and Validation of Cluster Analysis," in *Journal of Computational and Applied Mathematics* 20 (1978), 53–65.
[22] If one disregards the fact that architecture falls out of the category scheme but generates its own group in both cluster models, as will be seen later on.

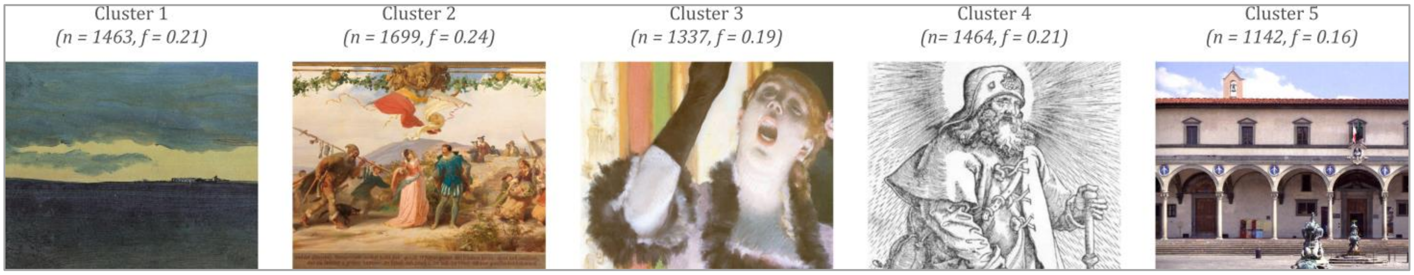

Figure 2. Obtained medoids and group sizes by partitioning into five clusters. The following artworks serve as medoids: Karl Blechen, "Evening Sky over an Italian Plain with Aqueduct" (Cluster 1), Adolf Schrödter, "Triumph of King Wine" (Cluster 2), Edgar Degas, "Singer with a Glove" (Cluster 3), Hans Baldung, "Apostle James the Great" (Cluster 4), Filippo Brunelleschi, "Hospital of the Innocents" (Cluster 5). Absolute and relative frequencies of the artworks assigned to the respective cluster are represented in brackets. All images are in the public domain.

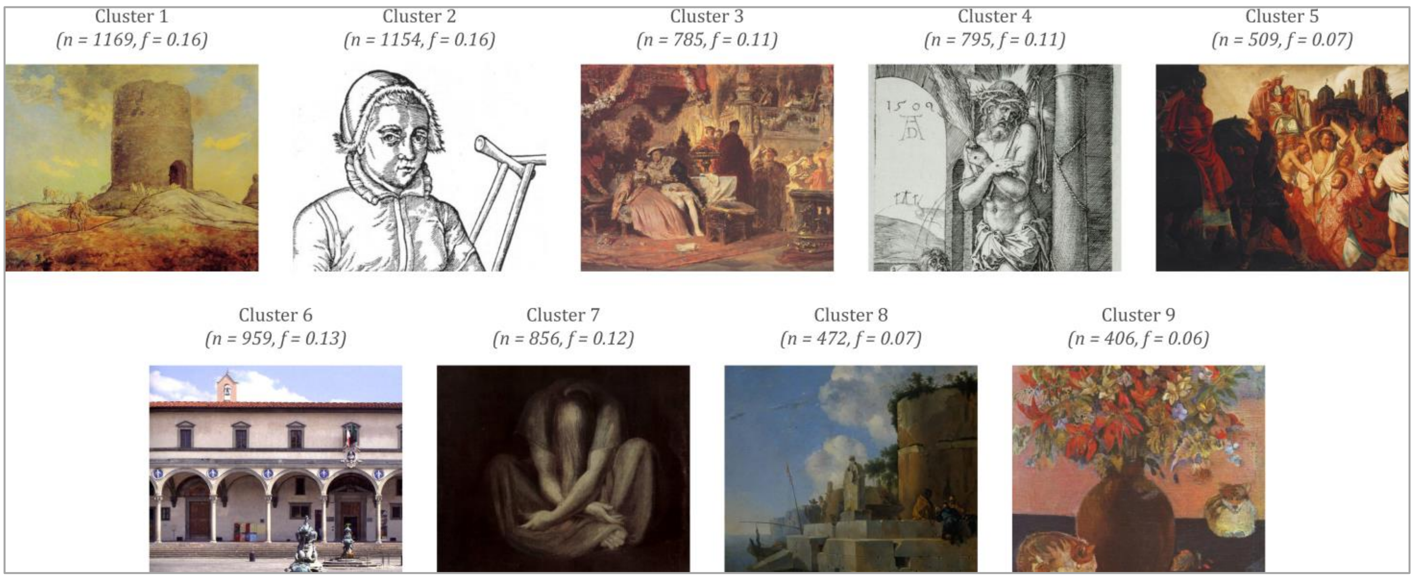

Figure 3. Obtained medoids and group sizes by partitioning into nine clusters. The following artworks serve as medoids: Jean-François Millet, "The Tower of Chailly" (Cluster 1), Johann Weyer, "Twelve-year-old Girl from Unna, Cured Cripple" (Cluster 2), Karl Theodor von Piloty, "Henry VIII and Anne Boleyn" (Cluster 3), Albrecht Dürer, "The Man of Sorrows Standing by the Column" (Cluster 4), Rembrandt, "The Stoning of Saint Stephen" (Cluster 5), Filippo Brunelleschi, "Hospital of the Innocents" (Cluster 6), Henry Fuseli, "Silence" (Cluster 7), Jan Asselyn, "A Coastal Ruin in Italy" (Cluster 8), Paul Gauguin, "Flowers and Cats" (Cluster 9). Absolute and relative frequencies of the artworks assigned to the respective cluster are represented in brackets. All images are in the public domain.

secondary role (Giovanni Battista Tiepolo, "The Reception of Henry III"). Several abstract images are also included (Vladimir Stenberg, "Color Construction," 1918). Similar observations can be made in the other clusters.

Better results can be obtained if one divides the works to be analyzed into nine clusters, even if some things seem to be inexplicable here as well. Cluster 1 is again landscape-oriented. At larger distances from the medoid (over 12 percent), examples from other categories that contain landscape elements are added. The trend strengthens by a distance of over 20 percent. Cluster 2 is allocated to portraits. Images with questionable portrait features first appear with a distance value of over 35 percent and should be further examined based on their individual tags. To give just one example, Melchiore della Bella's "Glove from the Sarcophagus of Henry VI" appears in this cluster but is so strongly associated with a figure wearing this glove that it was at least partially given annotations also corresponding to a person, which explains why it appears in the portrait cluster. Cluster 3 pertains primarily to genre paintings, although it is apparent that the ones which are closest to the medoid almost all

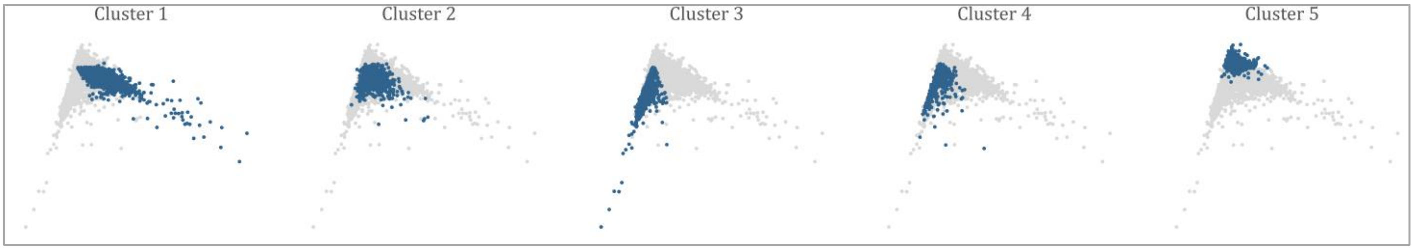

Figure 4. Obtained clusters by partitioning into five clusters, projected onto a two-dimensional space. The x-coordinate represents the first, the y-coordinate the second principal component. Created with R and the package ggplot2 (Wickham, *ggplot2: Elegant Graphics for Data Analysis*).

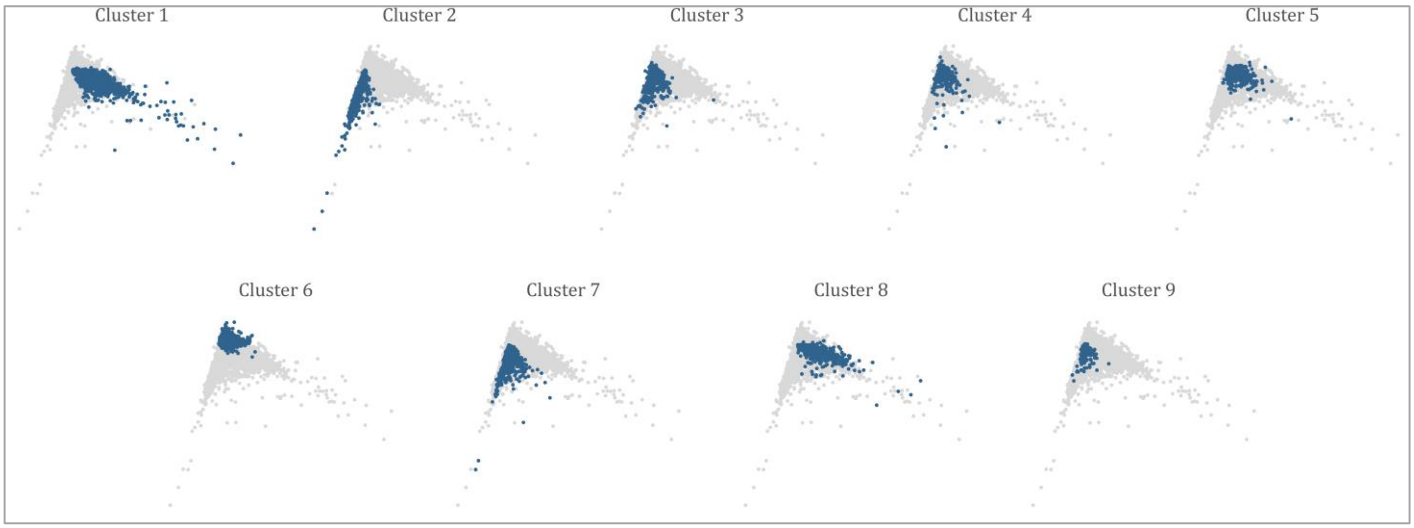

Figure 5. Obtained clusters by partitioning into nine clusters, projected onto a two-dimensional space. The x-coordinate represents the first, the y-coordinate the second principal component. Created with R and the package ggplot2 (Wickham, *ggplot2: Elegant Graphics for Data Analysis*).

belong to the historical genre (Cristoforo de Predis, "Maximilian Sforza at the Table with his Nurses"). General genre paintings are added later; portraits can be found repeatedly starting from distance values of over 20 percent. Cluster 4 contains classical history paintings. Surprisingly, many book illustrations stand out at distances around 30 percent (Lieven van Lathem, "Book of Hours of Mary of Burgundy"). The clear overlap between Clusters 3 and 4 (Figure 5) should be examined more precisely to determine if this has to do with the presence of historicizing phenomena in both areas or if the overlapping area refers to the history genre. Cluster 5 denotes a group which primarily contains portrayals of fights: episodes of war, battles, and hunting scenes. Starting at a distance value of 20 percent, the points become more widely distributed, though representations of horses as well as human confrontations indicate the main theme of conflict before that. Cluster 6 is uniformly related to architecture—unsurprising due to the distinctiveness of architectural phenomenology. Abstract subject matter first appears at a high distance value (30 percent and over); their composition, however, is determined by a specific architectural element (for example, the geometric forms of Sol LeWitt or the geometrics found in Dürer's illustrations of perspective theory, which turn up in this cluster).

Figure 6. Excerpt of the similarity network for Edouard Manet's "Still Life with Lilac and Roses". To emphasize the network character only edges originating from the center were drawn. All images are in the public domain.

We encounter an unusual group in Cluster 7, i.e. many individual figures and figures in groups; mostly, these figures are not portrayed in a narrative context (Lovis Corinth, "Susanna and the Elders"). Nudes are predominant in this cluster, and thus there are far more women than men. Cluster 8 is also landscape-oriented, but distinguishing it from Cluster 1 is virtually impossible if one does not want to take into account that genre-like elements and architecture, and above all, water-related scenes, are more strongly present as a differentiating criterion. That there exists a broad spatial consistency between the clusters may confirm the close connection. One could question, however, if the computer only designated two clusters at this point because we forced it to this high differentiation. Cluster 9 is in turn more definite and primarily contains still lifes (see Figure 6). It is only at a large distance (over 20 percent) that highly abstract objects appear.

Conclusions

Taking into account that similar results occur when a different sample of the database is clustered, one can conjecture that the approach has certain universality. We see three outcomes in particular: first, crowdsourced annotations can be considered quite valuable for research, in that they can be used for classification and filtering tasks. Because specialized categorizations are not necessary for such purposes, relatively simple descriptions of the subject are sufficient. Second, a subtler classification of the image inventory can be made by clustering into nine groups, surpassing a differentiation into five genres. It is crucial that a consistent reclassification arises, incorporating the entire material into a convincing grid, which comes along with similar plausibility as the traditional category rasterization does. Third, the resulting clusters can be used to train a *Convolutional Neural Network* and therefore open the possibility of automatically classifying digital reproductions of artworks that have not been pre-annotated and may even lack any metadata.[23]

Original methods are also conceivable. Thus, it would be feasible to isolate a geometricizing abstraction from a more biomorphic-organic one by searching in Cluster 6, according to evidence provided by the network analysis performed here. In addition, battle and combat scenes, whose subject matters waver between history paintings and landscapes corresponding to the classic hierarchy of genres, are set aside in its own cluster (5) and can be addressed accordingly. Nudes, which certainly cannot always be identified by their given title, are found especially in the group that the computer located in Cluster 7. Even though nowhere near all portrayals in this cluster are nudes, the proportion of them is decidedly higher than in a general selection of images.

In practice, the clustering becomes relevant when crowdsourced data can be used to consistently filter an image inventory. It is left to the work of a more sophisticated study to find out if, and if so how, an alternative clustering would allow further possibilities. One example would be the application of a soft classification method instead of the hard one used here.[24] A resource would then no longer be assigned to a single cluster; rather, the exact rate of affiliation to a group's center would be calculated in order to determine that the specified resource belongs, for example, 60 percent to Cluster 1, 10 percent to Cluster 2, and 30 percent to Cluster 5. Queries would thus be enabled to extract images, e.g. through a slide control implemented in the search interface, which thematically focus on landscapes but also contain a certain amount of architecture. In addition, it is plausible that classifying the data into 30 to 40 partitions would produce an even more fine-grained differentiation according to content-related criteria. Equally fine-grained bins can be achieved if one takes a specific group and applies the proposed clustering method once or several times more, constructing a hierarchical tree-like structure whose branching further segment the inventory. This whole process would be of great interest in preparing research processes, not in replacing them.

These kinds of clustering methods demonstrate one thing which frequently stands out in quantitative analyses and which causes some humanities scholars to make ironic observations: often, the result is to be expected and largely corresponds to what was already known. This point, however, is not the focus of our work; one could rather say it is desired. In this way, a mathematical method can be applied in connection with common art historical questions to render large, and otherwise hardly ascertainable, quantities of data usable. Particularly if one combines the clustering with other computer-based approaches, for example *Content-based Image Retrieval*,[25] collections of images can be created which are even more tailored to a particular interest. It is unnecessary to point out that other problems can be worked on with the collected data. Though the categorization of artworks is the focus of this approach, using the annotating behavior as a source of information about the players is also possible. This behavior is probably not identical across different ages, genders, and especially interesting, cultural background; it is therefore plausible that deviations in annotating behavior can be defined more precisely.

[23] In simple terms, this technique learns to detect features in images, whether these are edges, shapes or higher-level patterns, by stacking up layers, with each layer further trying to extract more complex characteristics; see Ian Goodfellow, Yoshua Bengio and Aaron Courville, *Deep Learning* (Cambridge, MA: MIT Press, 2016).

[24] For an application-oriented overview of so-called fuzzy clustering algorithms, see Sadaaki Miyamoto, Hidetomo Sadaaki and Katsuhiro Honda, *Algorithms for Fuzzy Clustering* (Berlin: Springer, 2008).

[25] The term might be misleading at first sight. *Content* refers to the visual attributes of an image: its color, texture, shape and edges.