THE CURATOR'S MACHINE: CLUSTERING OF MUSEUM COLLECTION DATA THROUGH ANNOTATION OF HIDDEN CONNECTION PATTERNS BETWEEN ARTWORKS

DOMINIK BÖNISCH

ABSTRACT | The discoverability of items within digitized museum holdings is an ongoing problem for collection managers, who strive to make objects maximally visible to both researchers and the public. Here it is not enough to simply limit the search in databases to narrowly defined keywords. Rather, specific interfaces and visualizations should allow the user to explore an online inventory as well as to 'stroll' through the digital collection. Artificial intelligence can support the systematic and structured processing of the mass of data in the museum. Machine learning can also reveal connections and links between artworks that otherwise might not have been fully legible. The text presents the research project "Training the Archive" as well as a tool to investigate the machine-aided, explorative (re)discovery of connections within the museum's collection.

KEYWORDS | collections, digital/digitized, feature extraction, GLAM institutions, machine learning

Introduction

The recent reconstruction of Aby Warburg's "Mnemosyne Atlas", displayed with the initial material for the first time since his death in 1929 for the exhibition "The Original" at the Haus der Kulturen der Welt, presents an opportunity for a thought experiment. Assume that the 971 images of the "Atlas," which were compiled for the show by the curators Roberto Ohrt and Axel Heil from the Photographic Collection of 400,000 objects[1] of the Warburg Institute London, would form a digitized data set. Several procedures could be carried out in order to analyze this data: For a start, the data could be complemented by the information about which individual image belongs to which of the 63 completed panels,[2] that Warburg had covered with black cloth to place his selected reproductions of artworks from European antiquity, the Middle East, and the Renaissance on them.[3] A next step could be, adding annotations from the so-called "Mnemosyne Pathways"[4], which experts of the Warburg Institute developed to offer an interpretation of how and in which order the panels are to be understood. Finally, the thematic as well as art historical references[5] of the different panels could be interlinked to open knowledge databases (e.g. of Wikimedia) to establish semantic queries. Based on the compiled images and the collected meta-information, an artificial neural network (ANN)[6] could eventually be trained that would recognize, generalize, and reproduce the connection patterns within the "Atlas." Thus, using machine learning (ML),[7] statistical predictions would be possible on how the unfinished panels would be composed of the images in the Warburg Institute's Photographic Collection. Admittedly, this example, which will remain hypothetical[8] for the foreseeable future, does not do justice to the achievements of the art historian Warburg, who created a complex visual reference system in the form of the "Mnemosyne Atlas."[9] It shows, however, to what extent artificial intelligence (AI) can contribute to refreshing the view on existing knowledge in digital collections and archives or to gaining previously unknown insights from available data.

The digitization in art museums[10] promises extended access to the objects of the collection both for scientific purposes and for an interested public, and this preferably online—independent of location and at any time.[11] Here it is not enough to simply transfer the traditional archive logic onto the digital realm and limit the search in databases to narrowly defined keywords.[12] Rather, special interfaces and visualizations[13] should allow the user to explore the digital inventory and to 'stroll' through the online collection without necessarily having to follow a given search term.[14] AI can help to process the mass of collection data in the museum in a systematic and structured way, and ML can reveal connections and links between artworks that are either difficult for humans to perceive or only incompletely accessible.[15] The research project "Training the Archive" at the Ludwig Forum Aachen, in cooperation with the HMKV Hartware MedienKunstVerein, Dortmund and the RWTH Aachen University, aims at investigating the machine-aided, explorative (re)discovery of various connections between the artworks within the collection of the museum.

Connecting AI with the Museum Collection Data

"Training the Archive" is funded by the Digital Culture Program of the German Federal Cultural Foundation (Kulturstiftung des Bundes). The project investigates the potential of ML methods to visualize patterns, connections as well as associations between objects within digital archives. The goal is to structure information and data of museum collections and to make them accessible to curators in an exploratory way. "Training the Archive" approaches the object of research by iterative development of prototypes.

A Prototype for Clustering Museum Collection Data

A prototype was developed within the context of the "AI school"[16] of the Foundation of Lower Saxony (Stiftung Niedersachsen) in cooperation with the tutors Jan Sölter and Thomas Rost.[17] The prototype utilized pre-trained network architectures, which were originally developed for image classification[18] and are now publicly available as models, to cluster digitized artworks, i.e. to divide them into different groups, and to sort the museum collection hierarchically. For this purpose, image data are fed into an ANN.[19] For each imported image, features can be extracted on the basis of visual as well as technical characteristics, either as color values, structures, textures, or as shapes and objects on the image. An ANN can cluster the imported image data according to the model-specific features. This allows the intuitive and associative discovery of one's collection since images with similar features can be aggregated into scalable groupings.

Peter Bell and Björn Ommer point to a "semantic gap"[20] between the mere representation of an artwork as a digital image and its actual image content, which can often only be read with the appropriate prior knowledge:

> *As a result, semantic, stylistic, aesthetic, technical, and historical connections between images and image areas are not sufficiently established. The entire area of the reception between artworks is hardly ever digitally mediated. Thus, the intermediality of art and the transformation of image discourses are barely tangible, even in the abundance of material.*[21]

"Training the Archive" would like to narrow this gap.[22] Based on a prototype, it investigates whether the clusters of artworks can be expanded to include elusive aspects, such as hidden connection patterns—as described by Bell and Ommer—or personal intuition and the subjective taste of an expert, by training an ANN to adopt a 'curatorial gaze.' Although the models used for image classification are already capable of a machine sorting (unsupervised learning)[23] of artworks, an effective process of human-machine interaction[24] must be established (supervised learning)[25]. This takes into account Bell and Ommer's claim that although the computer makes suggestions, the conclusions should still be drawn by the human, based on their historical, stylistic, and object-based contextual knowledge. As a result, juxtapositions could occur that match to such a degree they could be equated with findings that merely needed to be confirmed.[26]

In the following, it will be clarified whether it is possible to change unsupervised formed clusters of a data set by a supervised training with man-made annotations on the relations between artworks. The conducted experiment, which is discussed in the context of this text, is thereby considered as a Proof of Concept (PoC).[27] The concept is regarded as proven if the training has been successfully implemented (objective criterion) as well as if changes in the sorting of the formed clusters become evident (subjective criterion).

Formation and Visualization of Clusters

Applications that operate with ML or Computer Vision (CV) are to a great extent dependent on high-quality data sets, e.g. with regard to the existing image resolution, data variance, and amount of meta-information. For the development of the prototype, the digital collection of the Danish National Gallery Statens Museum for Kunst (SMK), which has been publicly accessible since 2019, was used. The museum's collection promises a wide range of both classical and contemporary artworks. It also covers a variety of different media, such as painting, prints, sculpture, installation, and applied arts.

The museum makes the data available via an API[28] (Application Programming Interface). According to the SMK, meta-information on more than 70,000 digitized items are available.[29]

MODEL	SOURCE
InceptionV3	https://keras.io/api/applications/inceptionv3/
Xception	https://keras.io/api/applications/xception/
ResNet152V2	https://keras.io/api/applications/resnet/#resnet152v2-function
InceptionResNetV2	https://keras.io/api/applications/inceptionresnetv2/
EfficientNetB7	https://keras.io/api/applications/efficientnet/#efficientnetb7-function
VGG19	https://keras.io/api/applications/vgg/#vgg19-function

Table 1: The table shows the models used for the PoC in Keras.

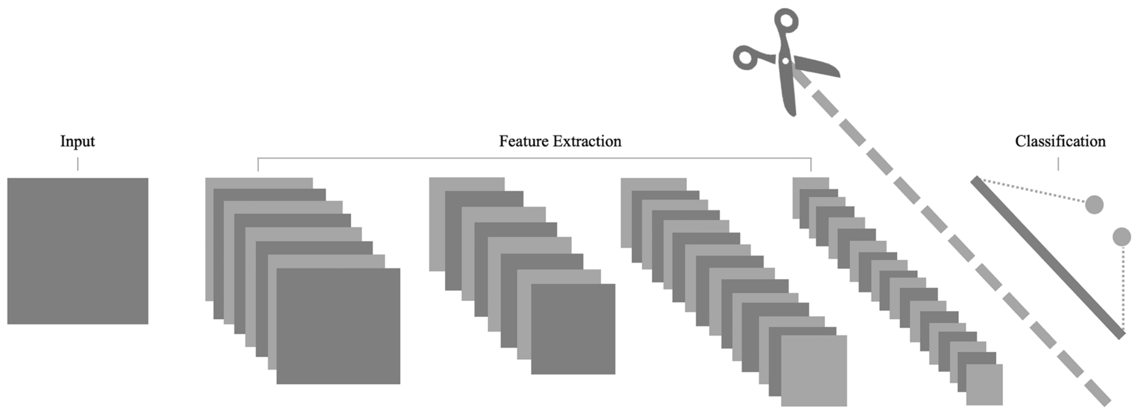

Figure 1: Illustration of omitting the classification.

In fact, at the stage of prototype development (March 2020) the data of 79,002 objects could be accessed. Only those data sets were scraped—meaning retrieved—which contained an image file (44,154 images).[30] In addition, a query was made as to whether the scraped work is in the public domain or whether it is copyrighted (32,411 images without copyright). This information should help to ensure that the image data was handled responsibly from the outset and to rule out any copyright infringement.

After careful analysis of the imported images, limitations to the usability of the data set became apparent, which could influence the training or object recognition and thus lead to false assignment. For example, numerous images that visibly contained a print control strip; in other cases, only black-and-white images of color paintings were available. On the other hand, the same works had been photographed from several perspectives and were filed twice as a verso. To further complicate matters, colored outlines were detected on various images, and it was difficult to differentiate whether they were part of the artwork or had been drawn in by the SMK's restoration department for the purpose of documentation. During the later standardization of the data—the preprocessing—the aspect ratio of the images was not retained. This can lead to changes in the source material. Despite the limitations, we stuck to the digital collection of the SMK, because the high data volume allowed us to sufficiently verify the PoC.

There are a large number of ANNs available in Keras and TensorFlow, which can be used for deep learning and which already contain pre-trained mathematical weights. These models can be applied to solve different problems, such as statistical predictions, feature extraction, or fine-tuning. The complex training of the weights is based on the visual data set ImageNet.[31] This allows the training with 1.3 to 14 million images, which were collected from the Internet and manually sorted into different classes and annotated.[32] The approach of instantiating a fully trained network from one of the publicly available libraries, in order to feed it with the collected image data, is known as transfer learning. Here, features based on the solution of one problem are applied to a different but similar problem.[33] This is advantageous for research because the respective models have already been trained to have a fundamental 'understanding' of the human world in regard to the general structure and content of images, and this knowledge does not have to be taught from scratch.

In addition to the Keras network architectures mentioned in Table 1, Big Transfer (BiT) by Google in the medium version (BiT/m-r152x4)[34] was instantiated using TensorFlow.

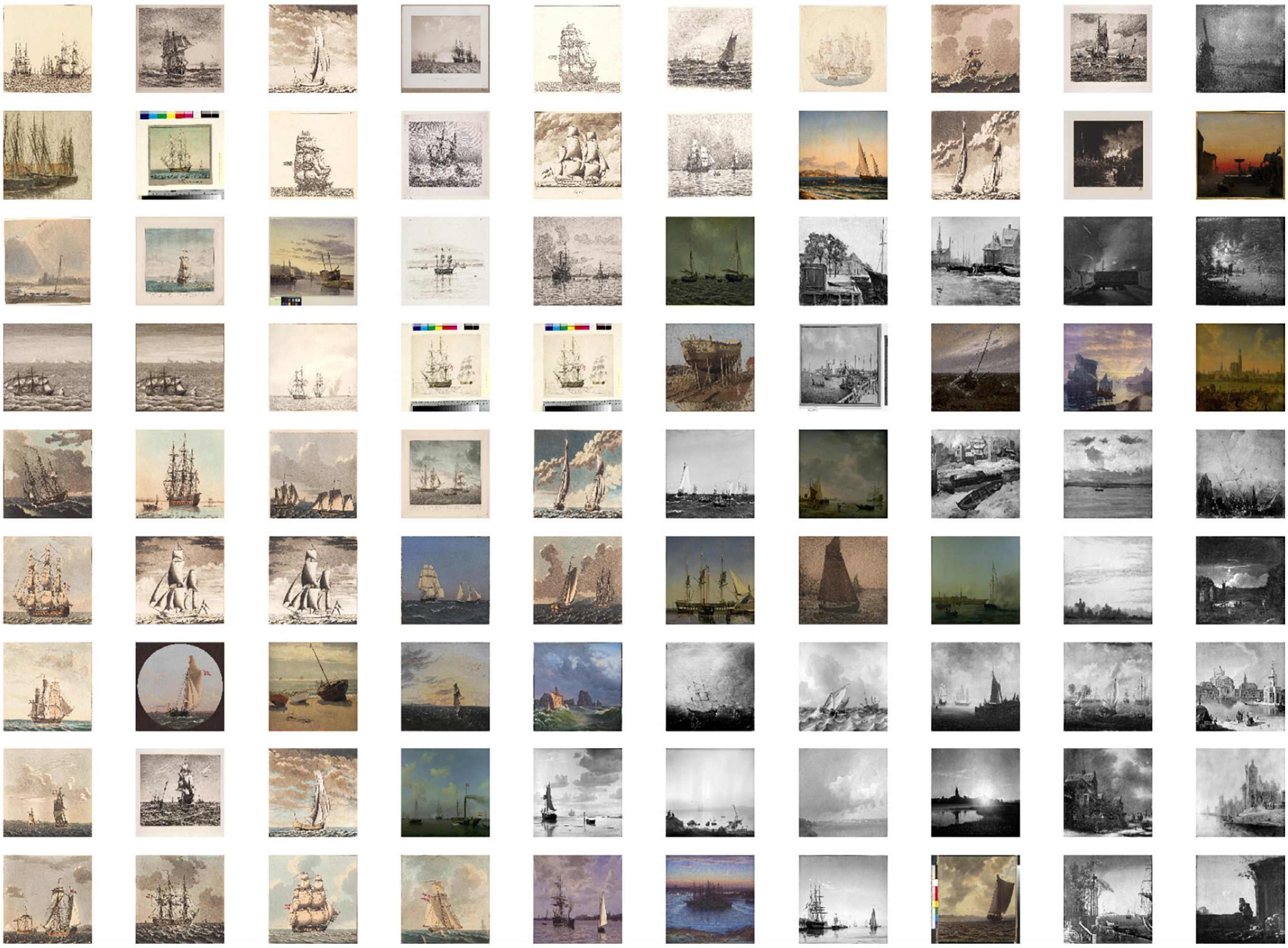

Figure 2: Gridplot of a cluster, which brings together images with maritime scenes of boats and sailing ships. All images are open-source data from the SMK.

The network VGG19 was of particular interest, as research has shown that it is possible to separate the representation from the content and style of an image.[35] Therefore, information on color texture or the specifications of brush strokes in paintings can be processed using specific layers of VGG19. As a result, the 'style' of an image and not only its content can be viewed by the ANN. The prototype used two of these layers to enhance the desired clustering of the scraped images by this style information.

From a technical perspective, the models mentioned above were used for the so-called feature extraction. In Keras or TensorFlow an ANN, pre-trained for image classification, has been used as a basis for building our own. The final layer—as a classifier—was omitted, which in its original state is responsible for the prediction of the respective categories of the ImageNet (see Figure 1).

What was needed and used, however, was the penultimate layer and its weights, which were propagated to the model with the help of the previously completed training. For this purpose, the features of each network architecture were extracted for every image of the scraped data set. This resulted in an 'interpretation' of the image files for all artworks of the SMK according to numerous feature-parameters of the ANN. An output of 44,154 feature vectors for each model were returned in a compressed archive file, which were used in a further step for the construction of a high-dimensional latent space and the clustering of the compiled image data set. Anna Großwendt et al. concludes the clustering approach as follows:

> *Clustering is a fundamental tool in machine learning. As an unsupervised learning method, it provides an easy way to gain insight into the structure of data without the need for expert knowledge to start with.*[36]

In order to initially structure the scraped image data without the addition of expert knowledge, a hierarchical cluster analysis was performed. Here, an agglomerative method was applied, which means that each image in the narrowest grouping forms exactly one single-element cluster, and those clusters are successively combined more roughly until each image belongs to a group sorted according to similarity.[37] Similarity refers thereby to the smallest distance between two relevant images within the latent space matrix.[38] To obtain

Figure 3: Scatterplot of a cluster, which combines images with different animal species. All images are open-source data from the SMK.

Figure 4: Illustration of a gridplot of nearest neighbors. All images are open-source data from the SMK.

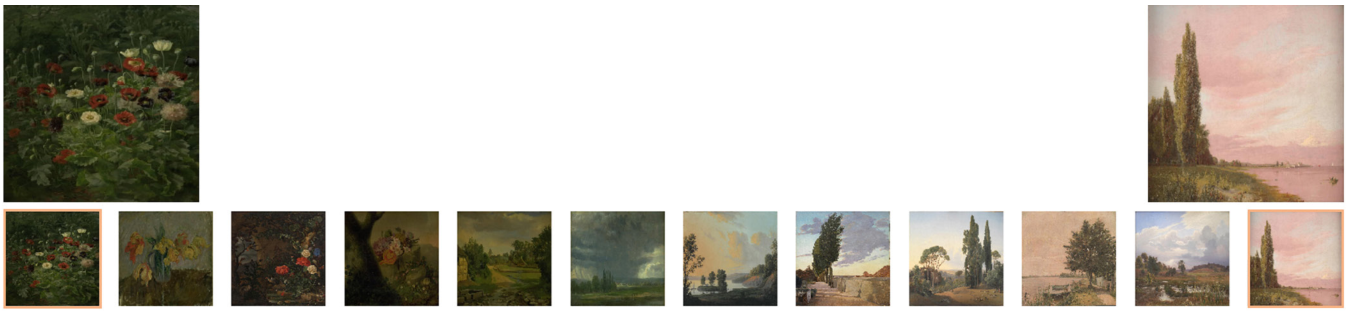

Figure 5: Illustration of a gridplot of the walk through the high-dimensional space from one artwork to another. All images are open-source data from the SMK.

useful clusters and to optimize the required computing power, the dimensionality of the latent space was reduced to 150 components by using the principal component analysis (PCA).[39] The number of clusters can be defined according to a threshold, for which there is no 'correct' size.[40] For the PoC, a threshold was chosen that combines about 95 clusters, which allowed an insight into the image data. If clusters contain too many images, they can also be divided into sub-clusters to maintain clarity.

To further explore the scraped image data, the feature vectors of the different models can be combined; in this case the BiT/m-r152x4 and the InceptionV3 were combined, and enhanced with the style information from the VGG19. The images that were assigned to the various clusters can be visualized in several ways. The prototype supports a gridplot (Figure 2) as well as a scatterplot (Figure 3).

To get an impression of the connections between the individual artworks and their corresponding positioning in the high-dimensional latent space, links can be visualized in different ways. In doing so, for example, the output of so-called nearest neighbors (Figure 4) is possible, which combines works that are as similar as possible (or vice versa, as farthest neighbors). Even a path in the latent space, which connects a starting point and a distant destination by way of specific artworks, can be navigated through a grid- or scatter-plot (see Figures 5 and 6).[41]

The Triplet Loss

The aim of the PoC was to investigate whether clusters, formed using agglomerative hierarchical clustering, can be modified by means of a training that used manual annotations on how artworks belong together. Therefore, a procedure had to be implemented that would enable curators to connect digital collection objects with one another according to specific—for the CV 'hidden'—criteria for context, aesthetics, iconography, and art historical references. To generate the annotations, the logic of the triplet loss was exploited.[42] Triplet loss is a loss function where a basis (anchor) is compared to a positive and a negative input. The aim is to embed the positive in the latent space proportionally closer to the anchor than the negative, which, in contrast, moves further away.[43] The distances of the positive and the negative are not absolute but relative, and therefore—in principle—both could be distant from the anchor, the negative being always further away than the positive. Thus, triplet loss models have the potential to characterize 'fuzzy' similarity connections

Figure 6: Scatterplot of the walk through the high-dimensional space from one artwork to another, the direction of the path is illustrated by arrows. All images are open-source data from the SMK.

between objects in the latent space. Regarding the prototype, this means that artworks with an annotated connection to a chosen reference work should be close to each other and those without a connection should be farther apart. By way of a subsequent clustering, connected artworks could then be sorted together or unconnected artworks could be separated. To obtain the annotations necessary for the triplet loss, an experimental setup as shown in Figure 7 was established.

Per iteration nine randomly chosen images from the data set of the SMK were presented. For each selection, an anchor had to be defined. This is done by connecting one artwork with another matching object. By entering the digits the chosen positive became linked to the respective anchor, and the works of art were thus related to each other. All remaining image files could be classified as negatives. For the PoC 3,000 annotations were created in this way and saved by the author.

Training Using the Generated Annotations

To optimize the models for our purposes, the different feature vectors must be trained using the generated annotations. In the future, information from curators about which artworks are closer to each other and which are not, can be generalized by the ANN after training. Consequently, it will be possible to adjust the formed clusters more closely to the specifications or knowledge of the annotators. A training is feasible with the help of the so-called cosine similarity, which can determine the similarity between objects. Mathematically speaking, the cosine similarity measures the cosine of the angle between two vectors projected in a multi-dimensional space, whereby the distance from each other becomes measurable.[44] A training should change the distances of the feature vectors in such a way that the connection patterns between the artworks according to the annotations are considered.

In the PoC, a training was performed on the author's 3,000 annotations in a total of 600 epochs.[45] 75 % of the annotations (2,250) were used as training data and 25 % (750) were held back as test data for the validation with unseen annotations. After training, the modified feature vectors were saved to a new archive file and could later be used for clustering. To verify the learning success, the meanrank was calculated as an average value. An instrument to evaluate whether the model has learned successfully: A control procedure that assesses the accuracy of the model's prediction compared to the images from the training set that were originally tagged as connected, or from an unseen set of evaluation data.

To determine this value, it had to be clarified at which position in a series of the nine randomly selected image files the (un)trained ANN selects the artwork that had been labeled as positive to the defined anchor, according to the annotation. For this purpose, the images of each iteration were sorted once more into positives and negatives, according to their relation to the anchor. After that, the position of the positive was determined and compared with the initial order. If the positive was in the first position—thus directly assigned to

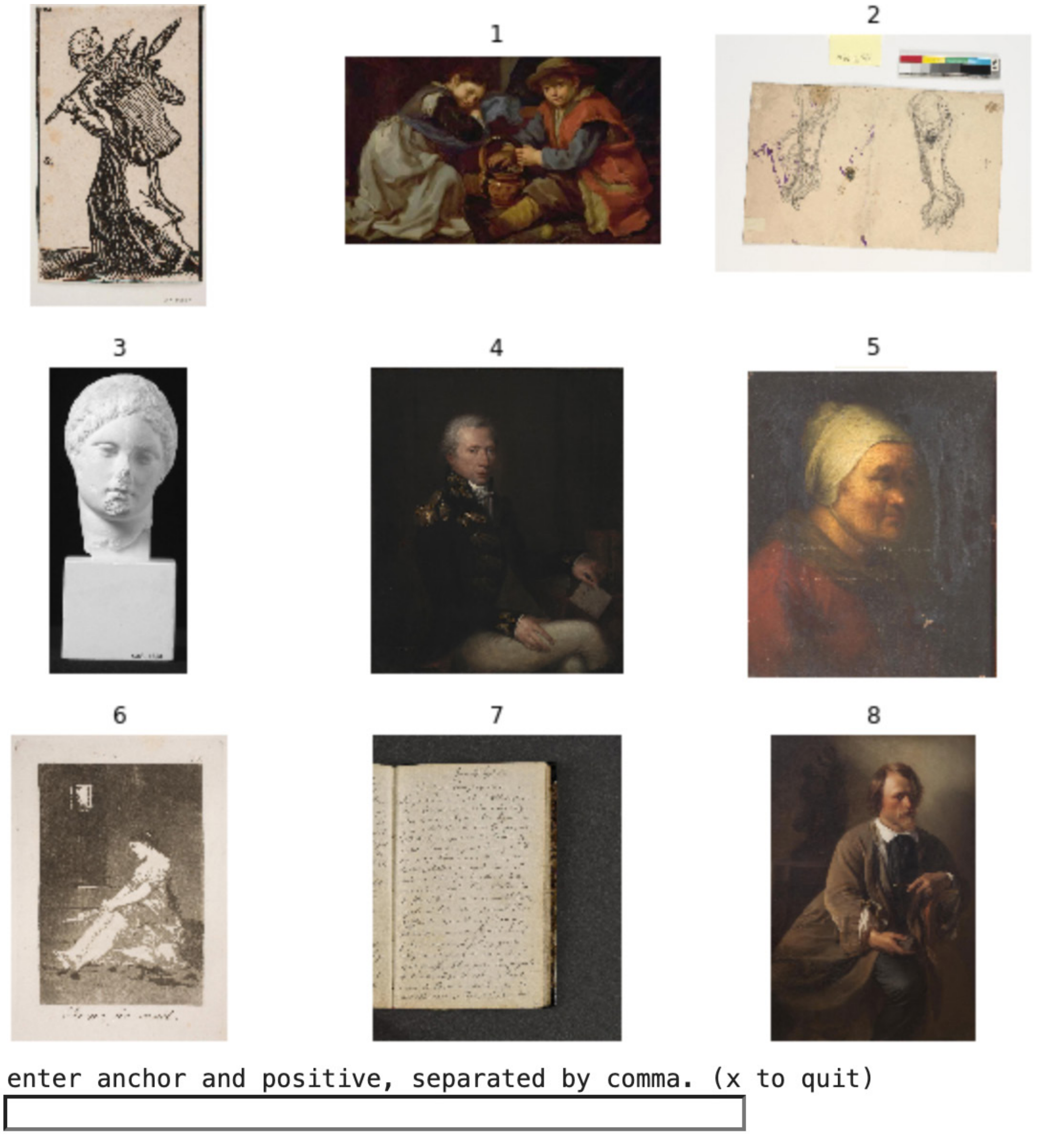

Figure 7: Process of an iteration to obtaining annotations. All images are open-source data from the SMK.

MODEL/MEANRANK	BEFORE TRAINING (3.000 ANNOT.)[a]	TRAINING (3.000 ANNOT., 600 EPOCHS)[a]	VERIFICATION (500 ANNOT.)[b]
BiT/m-152x4	3,79	2,03	1,93
InceptionV3	4,33	2,78	2,90
Xception	4,38	2,83	2,72
ResNet152V2	4,56	2,83	3,08
InceptionResNetV2	4,46	3,27	3,20
EfficientNetB7	4,54	3,67	3,78

Table 2: Meanranks for different models before and after training
[a] These 3.000 annotations were those also used in training.
[b] 500 annotations were held back and given unseen to the trained models for evaluation.

the anchor—the calculated meanrank was 1.0. If the correct selection of the following image file occurred only at the second position, the meanrank decreased to 1.5 on average. A result of 4.5 can be classified as randomly sorted. All values below this rank indicate a successful training. The meanrank was determined initially before training for the respective model, and after training according to the annotations that were used. For additional verification, the meanrank for 500 completely unseen annotations was ascertained in a verification step. The results were documented in Table 2.

The additive combination of the feature vectors of different models even improved the calculated meanrank. The combination of a trained BiT/m-152x4 and InceptionV3 resulted in a meanrank of 2.01 using the annotations from the training and 1.89 respectively, using unseen annotations. To be able to classify the results of the meanrank, a separate meanrank was calculated for the annotating person for comparison. This meanrank should not only be as low as possible, but it also limits the maximum value up to which the ANN can be trained. To determine his own meanrank, the author was given the task of resorting iterations from the annotation set. After 100 repetitions, the author's self-meanrank was 1.73. A roughly equal value was also reached by the trained nets.

Conclusion of the PoC

Based on the meanrank—the objective criterion—the training was a success, in ML terms. At present, the optimal combination of feature vectors consists of the trained BiT/m-152x4 combined with the trained InceptionV3 and supplemented by the style information of the VGG19. Hence, it is basically possible to model some of the knowledge and understanding of curators with the machine-based clustering of image data from museum collections. After training the annotations, the clusters became—the subjective criterion—more dynamic and more oriented to the author's specifications.[46] In the experiment, abstract paintings or drawings were annotated with sculptures and installations, which became apparent in later clusters. In the further course of the "Training the Archive" project, it will be necessary to check whether the test design used to generate the annotations shall be maintained in this form. Often connections between artworks had to be constructed or 'forced,' which could interfere with the training. An improved method shall be developed in dialogue with curators and data scientists for the next prototype.

The Curator's Machine

The confirmation, that machines could generalize the specific knowledge of curators of the collection of a museum invites us to consider a productive thought experiment: The Curator's Machine.[47] It is technically possible to store the annotations on the hidden connection patterns between individual artworks in an ANN as a separate model, so that it can be continuously retrained with new expert knowledge, without losing the specific findings from the annotation work of the individual experts. This could result in an application that, like using a sliding controller, incorporates the evaluations of various curators and thus influences the clustering of digitized museum collections.[48] It follows, therefore, that there is the potential—at least in theory—to emphasize views of groups that are underrepresented in the canon of cultural studies through the ANN, and to renew or redirect a predominant view of art: The Unbiased Curator's Machine.

"Training the Archive" would like to take up and examine this thought experiment further in the course of the project. In addition, text information shall be added to the prototype in a future development stage. Possible semantic connections can be deduced from artwork meta-data, Linked Data, or from keywords that can be derived from the museum or gathered from other research projects as "ARTigo."[49] The area of text-image-interlacing opens up additional possibilities of visualization, such as knowledge-graphs and will be considered in one of the next project phases of "Training the Archive."

Today, art and AI can be intertwined in many different ways. In the art museum, the involvement with this technology mainly—or exclusively—happened through exhibitions on the subject.[50] But AI is also suitable for the automated processing of the collections. ANNs are already able to systematically evaluate knowledge about artworks and their connections between each other and can thus, visualized appealingly, be used as a meaningful tool to support curators.

NOTES

[1] "Aby Warburg: Bilderatlas Mnemosyne," Haus der Kulturen der Welt (HKW), without date, accessed August 18, 2020, https://www.hkw.de/de/programm/projekte/2020/aby_warburg/bilderatlas_mnemosyne_start.php.

[2] Warburg planned to produce at least 79, possibly even up to 200 panels, which eluded him until his death. See Christopher Johnson, "About the Mnemosyne Atlas," Cornell University, without date, accessed August 18, 2020, https://live-warburglibrarycornelledu.pantheonsite.io/about.

[3] "Aby Warburg: Bilderatlas Mnemosyne. Publikation," Haus der Kulturen der Welt (HKW), without date, accessed August 18, 2020, https://www.hkw.de/de/programm/projekte/2020/aby_warburg/publikation_aby_warburg_bilderatlas_mnemosyne/detail.php.

[4] "Mnemosyne pathways," Cornell University, without date, accessed August 18, 2020, https://live-warburglibrarycornelledu.pantheonsite.io/about/mnemosyne-pathways.

[5] Available at https://live-warburglibrarycornelledu.pantheonsite.io/about/mnemosyne-themes, accessed August 20, 2020.

[6] An overview of the structure and functionality of ANNs is given by Klaus Jung, "Semantische Segmentierung mit Hilfe neuronaler Netzwerke zur effizienten Verarbeitung digitalisierter Dokumente," in *Proceedings of 2016 EVA Conference on Electronic Media & Art, Culture and History, Berlin, November 9-11, 2016* (Berlin: EVA, 2016), 56-59, accessed August 19, 2020, https://doi.org/10.11588/arthistoricum.256.338.

[7] See i.a. Fraunhofer-Gesellschaft zur Förderung der angewandten Forschung, *Maschinelles Lernen. Eine Analyse zu Kompetenzen, Forschung und Anwendung* (Munich: Fraunhofer-Gesellschaft, 2018), 8, accessed September 1, 2020, https://www.bigdata.fraunhofer.de/content/dam/bigdata/de/documents/Publikationen/Fraunhofer_Studie_ML_201809.pdf.

[8] Lisa Dieckmann and Martin Warnke stress that the "Mnemosyne Atlas" has a variable structure that is not limited to a final order, which would make the translation into a binary, mathematical logic difficult or even impossible. See Lisa Dieckmann and Martin Warnke, "Meta-Image und die Prinzipien des Digitalen im Mnemosyne-Atlas Aby Warburgs," in *Computing Art Reader: Einführung in die digitale Kunstgeschichte*, ed. Piotr Kuroczynski et al. (Heidelberg: arthistoricum.net, 2018), 83, accessed August 19, 2020, https://doi.org/10.11588/arthistoricum.413.

[9] "Aby Warburg: Bilderatlas Mnemosyne. Publikation," Haus der Kulturen der Welt (HKW), without date, accessed August 18, 2020, https://www.hkw.de/de/programm/projekte/2020/aby_warburg/publikation_aby_warburg_bilderatlas_mnemosyne/detail.php.

[10] In accordance with his professional expertise, the author refers in his remarks primarily to art museums, although an adaptation to other museums is conceivable. A comprehensive insight into the progress of digitization in European museums is provided by the empirical study: Network of European Museum Organisations. *Final report. Digitization and IPR in European Museums* (Berlin: Network of European Museum Working Group on Digitalisation and Intellectual Property Rights, 2020), accessed August 20, 2020, https://www.ne-mo.org/fileadmin/Dateien/public/Publications/NEMO_Final_Report_Digitisation_and_IPR_in_European_Museums_WG_07.2020.pdf.

[11] Katrin Glinka and Marian Dörk, "Zwischen Repräsentation und Rezeption – Visualisierung als Facette von Analyse und Argumentation in der Kunstgeschichte," in *Computing Art Reader: Einführung in die digitale Kunstgeschichte*, ed. Piotr Kuroczynski et al. (Heidelberg: arthistoricum.net, 2018), 236, accessed August 19, 2020, https://doi.org/10.11588/arthistoricum.413.

[12] Mitchell Whitelaw, "Generous interfaces for digital cultural collections," *digital humanities quarterly* 9, No. 1 (2015), accessed August 19, 2020, http://www.digitalhumanities.org/dhq/vol/9/1/000205/000205.html.

[13] The research project "VIKUS: Visualization of Cultural Collections" at the Potsdam University of Applied Sciences investigates graphical user interfaces for the visual exploration of cultural objects. See Marian Dörk and Katrin Glinka, "Der Sammlung gerecht werden: Kritisch-generative Methoden zur Konzeption experimenteller Visualisierungen," in *Book of Abstracts of 2018 DHd Conference on critique of digital reason, Cologne, February 26-March 2, 2018* (Cologne: DHd, 2018), 162, accessed August 21, 2020, http://dhd2018.uni-koeln.de/wp-content/uploads/boa-DHd2018-web-ISBN.pdf. See also the article on new access models for digital collections: Enrico Bertacchini and Federico Morando, "The Future of Museums in the Digital Age: New Models of Access and Use of Digital Collections," *International Journal of Arts Management*, working paper No. 5 (2011), accessed August 21, 2020, https://www.researchgate.net/publication/254455846.

[14] Viktoria Brüggemann et al., "Museale Bestände im Web: Eine Untersuchung von acht digitalen Sammlungen," in *Proceedings of 2016 EVA Conference on Electronic Media & Art, Culture and History, Berlin, November 9-11, 2016* (Berlin: EVA, 2016), 227, accessed August 18, 2020, https://doi.org/10.11588/arthistoricum.256.338.

[15] Cp. Peter Bell and Björn Ommer, "Visuelle Erschließung. Computer Vision als Arbeits- und Vermittlungstool," in *Proceedings of 2016 EVA Conference on Electronic Media & Art, Culture and History, Berlin, November 9-11, 2016* (Berlin: EVA, 2016), 68, accessed August 19, 2020, https://doi.org/10.11588/arthistoricum.256.338.

[16] Stiftung Niedersachsen, KI-Schule, Hannover: Stiftung Niedersachsen, 2018, accessed April 27, 2020, https://www.link-niedersachsen.de/ki_schule.

[17] Available at https://github.com/DominikBoenisch/Training-the-Archive, accessed October 5, 2020.

[18] Currenly, 26 different models for image classification are online available as Keras frameworks at https://keras.io/applications/#available-models, accessed May 8, 2020. The architectures of the models differ in size, parameters and topological depth.

[19] To improve readability, ANN is used as a generic term in the article, even if Convolutional Neural Network (CNN) would be technically more precise.

[20] See i.a. Taylor Arnold and Lauren Tilton, "Distant viewing: analyzing large visual corpora," *Digital Scholarship in the Humanities* 0, No. 0 (2019): 3, accessed April 27, 2020, https://doi.org/10.1093/digitalsh/fqz013.

[21] German original: „Dadurch werden semantische, stilistische, ästhetische, technische und historische Bezüge zwischen Bildern und Bildpartien nur schwach hergestellt. Der ganze Bereich der Rezeption zwischen Kunstwerken wird kaum digital vermittelt. Somit werden die Intermedialität von Kunst und der Wandel von Bilddiskursen auch in der Fülle des Materials nur wenig greifbar." Peter Bell and Björn Ommer, "Visuelle Erschließung. Computer Vision als Arbeits- und Vermittlungstool,"

in *Proceedings of 2016 EVA Conference on Electronic Media & Art, Culture and History, Berlin, November 9-11, 2016* (Berlin: EVA, 2016), 68, accessed August 19, 2020, https://doi.org/10.11588/arthistoricum.256.338.

22 Just as other projects e.g. "Replica." See Benoît Seguin et al., "Visual Link Retrieval in a Database of Paintings," in *ECCV 2016 Workshops*, ed. G. Hua and H. Jégou (Basel: Springer International Publishing, 2016), 753-767, accessed September 7, 2020, https://doi.org/10.1007/978-3-319-46604-0_52.

23 An explanation of the method is provided i.a. by Michaela Tiedemann, "So funktioniert Unsupervised Machine Learning," [at], Last modified Mai 22, 2018, accessed April 27, 2020, https://www.alexanderthamm.com/de/blog/erklaert-so-funktioniert-unsupervised-machine-learning/.

24 Principle of the "human-in-the-loop": Mothi Venkatesh, "What is Human-in-the-Loop for Machine Learning?" Hackernoon, Last modified July 17, 2018, accessed April 27, 2020, https://hackernoon.com/what-is-human-in-the-loop-for-machine-learning-2c2152b6dfbb. These systems are designed with humans as an integral part. See Meredith Broussard, "Letting Go of Technochauvinism," Public Books, Last modified June 17, 2019, accessed April 27, 2020, https://www.publicbooks.org/letting-go-of-technochauvinism/.

25 Michaela Tiedemann, "Supervised Machine Learning – So funktioniert 'Überwachtes Maschinelles Lernen'," [at], Last modified Mai 16, 2018, accessed April 27, 2020, https://www.alexanderthamm.com/de/blog/supervised-machine-learning/.

26 Peter Bell and Björn Ommer, "Computer Vision und Kunstgeschichte – Dialog zweier Bildwissenschaften," in *Computing Art Reader: Einführung in die digitale Kunstgeschichte*, ed. Piotr Kuroczynski et al. (Heidelberg: arthistoricum.net, 2018), 74, accessed August 20, 2020, https://doi.org/10.11588/arthistoricum.413.

27 The term is understood as it is defined in project management, where a PoC is used to prove the basic feasibility of a project, e.g. by using a prototype. Starting from this milestone, the project can be developed further. See e.g. Stefan Riedl, "Was ist ein Proof of Concept?" IT-Business, Last modified January 2, 2017, accessed April 27, 2020, https://www.it-business.de/was-ist-ein-proof-of-concept-a-666615/.

28 A concise explanation of how an API works is given in the video by MuleSoft: "What is an API?" YouTube, Last modified June 19, 2015, accessed May 6, 2020, https://www.youtube.com/watch?v=s7wmiS2mSXY.

29 "The SMK API (beta version): How to access SMK's collection data," Statens Museum for Kunst (SMK), without date, accessed May 6, 2020, https://www.smk.dk/en/article/smk-api/.

30 For this purpose, all identification numbers of the artworks were read via the standardized IIIF (International Image Interoperability Framework) interface. Subsequently, it was checked for each individual number whether an associated thumbnail was available and if applicable, downloaded.

31 Cf. Adrian Rosebrock, "ImageNet classification with Python and Keras," Pyimagesearch, Last modified August 10, 2016, accessed July 15, 2020, https://www.pyimagesearch.com/2016/08/10/imagenet-classification-with-python-and-keras/.

32 Based on the so-called WordNet hierarchy, predictions can be generated by ANNs e.g. for 1,000 (ILS-VRC-2012) or even 21,843 categories (ImageNet-21k), such as "bird", "flower" or "piece of furniture." See: "About ImageNet," IMAGENET, Last modified 2016, accessed July 15, 2020, http://image-net.org/about. In conjunction with Alexander Kolesnikov et al., "Big Transfer (BiT): General Visual Representation Learning," arXiv:1912.11370v3, 2020, 3, accessed July 15, 2020, https://arxiv.org/pdf/1912.11370.pdf.

33 François Chollet, "Transfer learning & fine-tuning," Keras, Last modified May 12, 2020, accessed July 15, 2020, https://keras.io/guides/transfer_learning/.

34 The term is derived from the architecture of the network, which uses a quadruple ResNet-152 (r152x4) and was trained as m-version on the ImageNet-21k data set: "bit/m-r152x4," Google, Last modified August 14, 2020, accessed October 1, 2020, https://tfhub.dev/google/bit/m-r152x4/1.

35 Cp. Leon Gatys et al., "A Neural Algorithm of Artistic Style," arXiv:1508.06576v2, 2015, 4, accessed July 15, 2020, https://arxiv.org/pdf/1508.06576.pdf.

36 Anna Großwendt et al., "Analysis of Ward's Method," arXiv:1907.05094v1, 2019, 0, accessed July 16, 2020, https://arxiv.org/pdf/1907.05094.pdf.

37 In detail, Ward's method is used, which minimizes possible losses of homogeneity when clusters are merged and is therefore considered the most efficient of the agglomerative methods: Petra Stein and Sven Vollnhals, "Grundlagen clusteranalytischer Verfahren" (Duisburg-Essen: Teaching material, 2011), 37, accessed July 16, 2020, https://www.uni-due.de/imperia/md/content/soziologie/stein/skript_clusteranlyse_sose2011.pdf.

38 Reinhold Kosfeld, "Hierarchische Klassifikationsverfahren," teaching material, without date, 1 f., accessed July 16, 2020, https://www.uni-kassel.de/fb07/fileadmin/datas/fb07/5-Institute/IVWL/Kosfeld/lehre/multivariate/Multivariate12_Clusteranalyse2_.pdf.

39 See i.a. Zakaria Jaadi, "A Step by Step Explanation of Principal Component Analysis," builtin, Last modified August 5, 2020, accessed October 8, 2020, https://builtin.com/data-science/step-step-explanation-principal-component-analysis.

40 Cp. Cosma Shalizi, "Distances between Clustering, Hierarchical Clustering," lecture: Data Mining, 2009, 4–8, accessed July 16, 2020, https://www.stat.cmu.edu/ffcshalizi/350/lectures/08/lecture-08.pdf.

41 This kind of visualization is known in particular through the project "X Degrees of Separation" by the artist: Mario Klingemann, "X Degrees of Separation," artist's homepage, Last modified November 5, 2016, accessed July 22, 2020, http://underdestruction.com/ 2016/11/05/x-degrees-of-separation-2016/.

42 Triplet loss is used especially for face recognition models. Cp. Florian Schroff et al., "FaceNet: A Unified Embedding for Face Recognition and Clustering," arXiv:1503.03832v3, 2015, accessed July 17, 2020, https://arxiv.org/pdf/1503.03832.pdf.

43 Olivier Moindrot, "Triplet Loss and Online Triplet Mining in TensorFlow," Olivier Moindrot blog, Last modified March 19, 2018, accessed July 17, 2020, https://omoindrot.github.io/triplet-loss.

44 Selva Prabhakaran, "Cosine Similarity – Understanding the math and how it works (with python codes)," ML+, Last modified October 22, 2018, accessed July 22, 2020, https://www.machinelearningplus.com/nlp/cosine-similarity/.

45 An epoch refers to a complete run of the fed input data through the ANN.

46 Here, the ANN's basic 'understanding' of image composition and content is not overridden, but the additional information about which artworks have a connection between each other is added moderately.

[47] Refers to “The Artist’s Machine” (2018) by Tillman Ohm and his AI “Arcu.”

[48] With the help of so-called user embeddings. Cf. Philipp Blandfort et al., “Fusion Strategies for Learning User Embeddings with Neural Networks,” arXiv:1901.02322v1, 2019, accessed September 1, 2020, https://arxiv.org/pdf/1901.02322.pdf.

[49] “ARTigo” is a research project of the Ludwig-Maximilians-University of Munich, which aims to tag artworks with keywords by means of crowdsourcing via an online game: http://www.artigo.org, accessed July 23, 2020.

[50] In 2018 the Frankfurter Kunstverein presented the group exhibition “I am here to learn: On Machinic Interpretations of the World.” In 2019 the House of Electronic Arts Basel (HeK) presented an exhibition entitled “Entangled Realities – Living with Artificial Intelligence.” In the same year the Kunstverein Hannover explored the question to which extent artists use the technical possibilities of AI in their practice in the exhibition “Artistic Intelligence.” The ZKM | Center for Art and Media Karlsruhe hosted a conference in 2019 themed “Art and Artificial Intelligence.” The Deutsches Hygiene-Museum, Dresden is planning a comprehensive show on the topic for 2021.

BIBLIOGRAPHY

Arnold, Taylor and Lauren Tilton. “Distant viewing: analyzing large visual corpora.” *Digital Scholarship in the Humanities* 0, No. 0 (2019): 1-14. https://doi.org/10.1093/digitalsh/fqz013.

Bell, Peter and Björn Ommer. “Computer Vision und Kunstgeschichte – Dialog zweier Bildwissenschaften.” In *Computing Art Reader: Einführung in die digitale Kunstgeschichte*, edited by Piotr Kuroczynski, Peter Bell, and Lisa Dieckmann, 60-75. Heidelberg: arthistoricum.net, 2018 (Computing in Art and Architecture, Vol. 1). https://doi.org/10.11588/arthistoricum.413.

Bell, Peter and Björn Ommer. “Visuelle Erschließung. Computer Vision als Arbeits- und Vermittlungstool.” In *Proceedings of 2016 EVA Conference on Electronic Media & Art, Culture and History, Berlin, November 9-11, 2016*, edited by Andreas Bienert, Anko Börner, Eva Emenlauer-Blömers, and James Hemsley, 67-73. Heidelberg: arthistoricum.net, 2017 (2016) (EVA Berlin, Vol. 23). https://doi.org/10.11588/arthistoricum.256.338.

Bertacchini, Enrico and Federico Morando. “The Future of Museums in the Digital Age: New Models of Access and Use of Digital Collections.” *International Journal of Arts Management*. Working paper No. 5 (2011): 1-14. Accessed August 21, 2020. https://www.researchgate.net/publication/254455846.

Blandfort, Philipp, Tushar Karayil, Federico Raue, Jorn Hees, and Andreas Dengel. “Fusion Strategies for Learning User Embeddings with Neural Networks.” arXiv:1901.02322v1, 2019. Accessed September 1, 2020. https://arxiv.org/pdf/1901.02322.pdf.

Broussard, Meredith. “Letting Go of Technochauvinism.” Public Books. Last modified June 17, 2019. Accessed April 27, 2020. https://www.publicbooks.org/letting-go-of-technochauvinism/.

Brüggemann, Viktoria, Sarah Kreiseler, and Marian Dörk. “Museale Bestände im Web: Eine Untersuchung von acht digitalen Sammlungen.” In *Proceedings of 2016 EVA Conference on Electronic Media & Art, Culture and History, Berlin, November 9-11, 2016*, edited by Andreas Bienert, Anko Börner, Eva Emenlauer-Blömers, and James Hemsley, 227-236. Heidelberg: arthistoricum.net, 2017 (2016) (EVA Berlin, Vol. 23). https://doi.org/10.11588/arthistoricum.256.338.

Chollet, François. “Transfer learning & fine-tuning.” Keras. Last modified May 12, 2020. Accessed July 15, 2020. https://keras.io/guides/transfer_learning/.

Dieckmann, Lisa and Martin Warnke. “Meta-Image und die Prinzipien des Digitalen im Mnemosyne-Atlas Aby Warburgs.” In *Computing Art Reader: Einführung in die digitale Kunstgeschichte*, edited by Piotr Kuroczynski, Peter Bell, and Lisa Dieckmann, 78-93. Heidelberg: arthistoricum.net, 2018 (Computing in Art and Architecture, Vol. 1). https://doi.org/10.11588/arthistoricum.413.

Dörk, Marian and Katrin Glinka. “Der Sammlung gerecht werden: Kritisch-generative Methoden zur Konzeption experimenteller Visualisierungen.“ In *Book of Abstracts of 2018 DHd Conference on critique of digital reason, Cologne, February 26-March 2, 2018. Cologne: DHd, 2018*: 162-166. Accessed August 21, 2020. http://dhd2018.uni-koeln.de/wp-content/uploads/boa-DHd2018-web-ISBN.pdf.

Fraunhofer-Gesellschaft zur Förderung der angewandten Forschung. *Maschinelles Lernen. Eine Analyse zu Kompetenzen, Forschung und Anwendung*. Munich: Fraunhofer-Gesellschaft, 2018. Accessed September 1, 2020. https://www.bigdata.fraunhofer.de/content/dam/bigdata/de/documents/Publikationen/Fraunhofer_Studie_ML_201809.pdf.

Gatys, Leon A., Alexander S. Ecker, and Matthias Bethge. “A Neural Algorithm of Artistic Style.” arXiv:1508.06576v2, 2015. Accessed July 15, 2020. https://arxiv.org/pdf/1508.06576.pdf.

Glinka, Katrin and Marian Dörk. “Zwischen Repräsentation und Rezeption – Visualisierung als Facette von Analyse und Argumentation in der Kunstgeschichte.” In *Computing Art Reader: Einführung in die digitale Kunstgeschichte*, edited by Piotr Kuroczynski, Peter Bell, and Lisa Dieckmann, 234-250. Heidelberg: arthistoricum.net, 2018 (Computing in Art and Architecture, Vol. 1). https://doi.org/10.11588/arthistoricum.413.

Großwendt, Anna, Heiko Röglin, and Melanie Schmidt. “Analysis of Ward’s Method.” arXiv:190 7.05094v1, 2019. Accessed July 16, 2020. https://arxiv.org/pdf/1907.05094.pdf.

Jaadi, Zakaria. “A Step by Step Explanation of Principal Component Analysis.” builtin. Last modified August 5, 2020. Accessed October 8, 2020. https://builtin.com/data-science/step-step-explanation-principal-component-analysis.

Johnson, Christopher D. “About the Mnemosyne Atlas.” Cornell University. Without date. Accessed August 18, 2020. https://live-warburglibrarycornelledu.pantheonsite.io/about.

Jung, Klaus. “Semantische Segmentierung mit Hilfe neuronaler Netzwerke zur effizienten Verarbeitung digitalisierter Dokumente.” In *Proceedings of 2016 EVA Conference on Electronic Media & Art, Culture and History, Berlin, November 9-11*, 2016, edited by Andreas Bienert, Anko Börner, Eva Emenlauer-Blömers, and James Hemsley, 54-61. Heidelberg: arthistoricum.net, 2017 (2016) (EVA Berlin, Vol. 23). https://doi.org/10.11588/arthistoricum.256.338.

Kolesnikov, Alexander, Lucas Beyer, Xiaohua Zhai, Joan Puigcerver, Jessica Yung, Sylvain Gelly, and Neil Houlsby. “Big Transfer (BiT): General Visual Representation Learning.” arXiv:1912.11370v3, 2020. Accessed July 15, 2020. https://arxiv.org/pdf/1912.11370.pdf.

Kosfeld, Reinhold. “Hierarchische Klassifikationsverfahren.” Teaching material. Without date. Accessed July 16, 2020. https://www.uni-kassel.de/fb07/fileadmin/datas/fb07/5-Institute/IVWL/Kosfeld/lehre/multivariate/Multivariate12_Clusteranalyse2_.pdf.

Moindrot, Olivier. “Triplet Loss and Online Triplet Mining in TensorFlow.” Olivier Moindrot blog. Last modified March 19, 2018. Accessed July 17, 2020. https://omoindrot.github.io/triplet-loss.

Network of European Museum Organisations. *Final report. Digitization and IPR in European Museums*. Berlin: Network of European Museum Working Group on Digitalisation and Intellectual Property Rights, 2020. Accessed August 20, 2020. https://www.ne-mo.org/fileadmin/Dateien/public/Publications/NEMO_Final_Report_Digitisation_and_IPR_in_European_Museums_WG_07.2020.pdf.

Prabhakaran, Selva. “Cosine Similarity – Understanding the math and how it works (with python codes).” ML+. Last modified October 22, 2018. Accessed July 22, 2020. https://www.machinelearningplus.com/nlp/cosine-similarity/.

Riedl, Stefan. “Was ist ein Proof of Concept?” IT-Business. Last modified January 2, 2017. Accessed April 27, 2020. https://www.it-business.de/was-ist-ein-proof-of-concept-a-666615/.

Rosebrock, Adrian. “ImageNet classification with Python and Keras.” Pyimagesearch. Last modified August 10, 2016. Accessed July 15, 2020. https://www.pyimagesearch.com/2016/08/10/imagenet-classification-with-python-and-keras/.

Schroff, Florian, Dmitry Kalenichenko, and James Philbin. “FaceNet: A Unified Embedding for Face Recognition and Clustering.” arXiv:1503.03832v3, 2015. Accessed July 17, 2020. https://arxiv.org/pdf/1503.03832.pdf.

Seguin, Benoît, Carlotta Striolo, Isabella diLenardo, and Frederic Kaplan. “Visual Link Retrieval in a Database of Paintings.” In: *ECCV 2016 Workshops*, edited by G. Hua and H. Jégou, 753-767. LNCS 9913, Basel: Springer International Publishing. https://doi.org/10.1007/978-3-319-46604-0_52.

Shalizi, Cosma. “Distances between Clustering, Hierarchical Clustering.” Lecture: Data Mining, 2009. Accessed July 16, 2020. https://www.stat.cmu.edu/ffcshalizi/350/lectures/08/lecture-08.pdf.

Stein, Petra and Sven Vollnhals. “Grundlagen clusteranalytischer Verfahren.” Duisburg-Essen: Teaching material, 2011. Accessed July 16, 2020. https://www.uni-due.de/imperia/md/content/soziologie/stein/skript_clusteranlyse_sose2011.pdf.

Tiedemann, Michaela. “So funktioniert Unsupervised Machine Learning.” [at]. Last modified Mai 22, 2018. Accessed April 27, 2020. https://www.alexanderthamm.com/de/blog/erklaert-so-funktioniert-unsupervised-machine-learning/.

Tiedemann, Michaela. “Supervised Machine Learning – So funktioniert ‘Überwachtes Maschinelles Lernen’.” [at]. Last modified Mai 16, 2018. Accessed April 27, 2020. https://www.alexanderthamm.com/de/blog/supervised-machine-learning/.

Venkatesh, Mothi. “What is Human-in-the-Loop for Machine Learning?” Hackernoon. Last modified July 17, 2018. Accessed April 27, 2020. https://hackernoon.com/what-is-human-in-the-loop-for-machine-learning-2c2152b6dfbb.

Whitelaw, Mitchell. “Generous interfaces for digital cultural collections.” *digital humanities quarterly* 9, No. 1 (2015). Accessed August 19, 2020. http://www.digitalhumanities.org/dhq/vol/9/1/000205/000205.html.

DOMINIK BÖNISCH studied Cultural Studies and Aesthetic Communication in the Master's Program "Cultural Mediation" at the University of Hildesheim and the Moholy-Nagy University of Art and Design Budapest. Currently, he is working as the scientific project manager of the research project "Training the Archive" at the Ludwig Forum Aachen, investigating the connection between artificial intelligence (AI) and museum collections. As a curatorial assistant, Bönisch dealt with Virtual Reality (VR) in the exhibition "Thrill of Deception. From Ancient Art to Virtual Reality", amongst others. His research interest focuses particularly on the effects of AI and VR on art and the collection as well as exhibition activities of the museum. Bönisch, who is also holding seminars on the topic at the University of Applied Sciences Düsseldorf (Hochschule Düsseldorf), lives and works in Aachen and Düsseldorf, Germany.

Correspondence e-mail: dominik.boenisch@mail.aachen.de