Through the Eyes of the Machine: Exploring Historical Photo Collections with Convolutional Neural Networks

Max Frischknecht *

Abstract: The article explores Convolutional Neural Networks to cluster historical photo collections by visual similarity to aid exploration and mediation. It examines potentials and challenges of human-machine collaboration by juxtaposing human and machine ways of seeing. Clustering 48 000 negatives from the collection Ernst Brunner the analysis reveals how sociotechnical imaginaries in infrastructure act as an epistemological Trojan horse and emphasizes the need for thematic data sets to utilize machine-learning approaches for visual data analysis.

Keywords: Digital humanities, convolutional neural networks, data visualization, critical data studies, Ernst Brunner

À travers les yeux de la machine : exploration de collections de photos historiques à l'aide de Convolutional Neural Networks

Résumé : L'article explore l'utilisation des Convolutional Neural Networks pour regrouper des collections de photos historiques par similarité visuelle. Il examine les défis et les potentialités de la collaboration homme-machine en opposant leurs manières de voir. En regroupant 48 000 négatifs de la collection Ernst Brunner, l'analyse montre comment les imaginaires sociotechniques agissent comme un cheval de Troie épistémologique et souligne l'importance de données thématiques pour l'analyse visuelle via l'apprentissage automatique.

Mots-clés : Humanités numériques, convolutional neural networks, visualisation de données, études critiques des données, Ernst Brunner

Durch die Augen der Maschine: Zur Untersuchung historischer Fotosammlungen mittels Convolutional Neural Networks

Zusammenfassung: Der Artikel untersucht Convolutional Neural Networks für das visuelle Clustern historischer Fotosammlungen. Er untersucht die Möglichkeiten und Herausforderungen einer Mensch-Maschine-Kollaboration durch die Gegenüberstellung derer Sichtweisen. Anhand von 48 000 Negativen der Sammlung Ernst Brunner wird aufgezeigt wie soziotechnische Vorstellungen der Infrastruktur als epistemologisches Trojanisches Pferd agieren. Die Notwendigkeit thematischer Datensätze, um das Potenzial von Machine-Learning für die Datenanalyse auszuschöpfen, wird unterstrichen.
Schlüsselwörter: Digitale Geisteswissenschaften, convolutional neural networks, Datenvisualisierungen, kritische Datenstudien, Ernst Brunner

* University of Bern, CH-3012 Bern, Institute of Design Research, Bern University of the Arts HKB, CH- 3027 Bern, max.frischknecht@hkb.bfh.ch.

DOI 10.26034/cm.sjs.2025.6907

1 Introduction[1]

Since the beginning of the 21st Century, Swiss archives have been increasingly digitizing their materials. One of these is the photo archive of Cultural Anthropology Switzerland (CAS) (formerly Swiss Society for Folklore Studies SSFS) with over 105 000 photographs from 20 collections (CAS 2023a). The digitization of such (photographic) collections pursues the goal of preservation on the one hand and greater accessibility on the other. Digital technology and the advent of the WWW gave novel emphasis to the dissemination of, and accessibility to, archival records (Stevenson, 2008, p. 89; Terras, 2011, p. 9). Especially smaller archival institutions see the potential to reach a larger audience and raise the profile of less known collections (Stevenson, 2008). Correspondingly the demands from users (including researchers) to access, search, and retrieve archival information, for example through a website, have increased, leading GLAM[2] institutions to take a more user-centered perspective and look for novel ways of searching and exploring collections.

In this context, the article examines the application of machine learning, specifically Convolutional Neural Networks (CNNs), to cluster photographic collections based on visual similarity. The article is guided by the assumption that visual similarity can be fruitful for the exploration and examination of large collections, *if the CNN clusters the photographs along the central topics and narratives inherent in the collection*, such as, for example, the photographer's interest in landscape, rural live, or labor. The primary aim of this article is not to use a CNN to discover unknown or never-before-seen things, but rather about investigating whether the algorithm is able to cluster the central aspects of a collection that one would also find in the relevant literature. If this were the case, it would show that such a machine-learning component could meet the archivists' demands, for example, on the mediation of the collection towards an interested public. It would further provide an interesting alternative way for researchers to explore digital collections beyond classical text-search based interfaces to develop novel research questions. At the same time, it is obvious that the machine will see differently to some degree and the question arises where this supposedly unique perspective comes from, how it is shaped, and how it can influence, in turn, the human interpretation of the collection. In essence it is about the examination of a human-machine interpretation interplay and its possible epistemological implications.

The main research questions are: 1) Can a CNN recognize the central (visual) topics and narratives of a historical photographic collection? 2) To what extent does this "machine way of seeing" (Cox, 2022, p. 103) align or differ from a human perspective on the collection? 3) Does this difference allow for interesting modes

1 I would like to thank the anonymous reviewers and the editor, Sebastian W. Hoggenmüller, for their critical review. Without their help this article would not be what it is.

2 GLAM stands for Galleries, Libraries, Archives, and Museums.

of human-machine collaboration? 4) And what epistemological implications arise from such a collaboration?

To answer these questions the article defines the human perspective on the collection through a literature review that identifies the prevalent narratives and contrasts these with the clusters created by a CNN. The collection Ernst Brunner (CAS, 2023b) serves as an example for this study. It is an extremely varied photographic collection comprising around 48 000 negatives that document large parts of rural and urban everyday life in Switzerland approximately between 1935 and 1970 – whereby the majority of photos were taken between the 30s and 50s with only very few from the 60s and 70s. The collection contains portraits, village views, landscapes, private and public events, handcrafted objects as well as work processes and is of high visual and documentary quality. Due to its size and level of indexing, the collection is ideally suited for machine-learning supported exploration. For clustering, the PixPlot application is used. It was developed by Yale Digital Humanities Lab (2017) and uses a CNN pre-trained on the popular ImageNet (Stanford Vision Lab, 2011) data set. Despite its age, ImageNet is one of the most important publicly accessible data sets and has been used to train a variety of applications.[3] Thus, some of the findings of this study can, to a certain extent, be applied to other scenarios involving ImageNet. It shows how crucial the underlying training data is and how it influences PixPlot's ability to see and cluster photographs from the Brunner collection.

The following section 2 situates the article in the broader theoretical discussions around big visual data in sociology and digital humanities and proposes a multidisciplinary theoretical framework to explore the epistemological implications of a human-machine interpretation interplay. Section 3 sets out the methodological approach by introducing the collection Ernst Brunner, explaining the mechanics of PixPlot / the CNN and discussing the ImageNet data set and its biases. A combination of close and distant reading is proposed to compare the PixPlot clustering with the views on Ernst Brunner's work found in the literature. Section 4 then analyses how the algorithm sees by systematically comparing the PixPlot clusters with human perspectives, using the ImageNet dataset as a reference. Section 5 consolidates the findings and provides a critical reflection based on a recourse to the theoretical framework. It outlines a possible pathway for the critical application of CNNs in exploration and mediation of big visual data while paying attention to epistemological challenges.

2 Theoretical Framework

Since the appearance of big (visual) data in the early 2000s, a lively discussion arose around challenges and potentials to study sociological phenomena through such

3 For a long time, ImageNet was one of the few publicly accessible image data sets. Today, however, new and larger datasets are increasingly being published such as LAION-5B, cf. https://laion.ai/blog/laion-5b/ (15. 6. 2024).

data (Savage & Burrows, 2007; Burrows & Savage, 2014; Frade, 2016). In the digital humanities, advocates argue that big data analysis facilitates and augments research (Manovich, 2011) while critics describe the results as superficial and reductionist (Kitchin, 2014, p. 142 with reference to Trumpener 2009). Either way, big (visual) data holds the potential to reframe the epistemology of social science and humanities and the chosen methodological approaches must be thought through accordingly (Kitchin, 2014). This article attempts to make a humble contribution to these methodological discussions through an empirical study that outlines a possible human-machine collaboration. It draws on a multidisciplinary approach combining concepts from digital humanities, media theory, science, and technology studies (STS), and critical data studies to explore how a "machine way of seeing" (Cox, 2022, p. 103) is shaped by its underlying infrastructure, and how this infrastructure, in turn, is shaped by sociotechnical imaginaries. In doing so, it provides a critical framework for how knowledge is created through big visual data and shaped by technology.

2.1 Machine Ways of Seeing

In his seminal essay *Ways of Machine Seeing as a Problem of Invisual Literacy* Geoff Cox (2022) outlines the characteristics of how machines see in reference to John Berger's *Ways of Seeing* (Berger, 1972). Seeing and naming things is a question of literacy. Literacy can be defined as "competence or knowledge of practices that allow users to maintain and build social imaginaries", it is the ability to "read, write and program" (Cox, 2022, p. 105–106). It is a form of power and authority when certain ways of describing and naming things are held over others. This is the case with ImageNet and how it manifests what PixPlot can see through a defined set of images and categories.

Building on Geoff Cox we can identify three aspects of a machine way of seeing. 1) *Machines don't see, they relate.* Seeing for a machine is no longer singular or indexical, but rather distributed and multimodal (Cox, 2022, p. 110). At one point the image is digitized or created, at another it is given a label, at a third, it is viewed. The meaning that an image has for a machine is not based on its indexicality but rather on its relation to its categories and other images and their categories. 2) *Machines don't see, they read.* Machines don't see an image, but rather read it according to the model of the world they know (Cox, 2022, p. 108). What lies outside of this model can't be recognized. This also applies in a technical sense as an algorithm reads an image pixel by pixel to interpret the correlation of a pixel with its neighbors. 3) *Machines don't see, they calculate probabilities.* With reference to Crawford and Paglen (2019), Cox argues that seeing for a machine is a calculative practice, where the algorithm calculates the probability that, for example, the image shows a totem pole rather than a cactus. These models of probability are "built upon inherent human prejudices related to class, gender, and race" (Cox, 2022, p. 109; with reference to Crawford & Paglen, 2019). In summary, a machine way of seeing can be described as based on *relational perception, algorithmic reading, and probabilistic interpretation.* To understand how the machine sees the world is to understand how the humans understand the machine and how they see and teach the machine to see the world (see also Hoggenmüller & Klinke in this Special Issue).

2.2 Data as Infrastructure

Infrastructure study broadens our view of the various components that are at play when a machine sees. PixPlot and its CNN are built upon existing information infrastructure, specifically cyberinfrastructure. Cyberinfrastructures are "those layers that sit between base technology (a computer science concern) and discipline-specific science" (Bowker et al., 2010, p. 100). This applies to PixPlot as it builds on existing base technologies such as Tensorflow[4] or Keras[5] and was specifically developed for a humanities context. Susan Leigh Star introduced us to the idea that infrastructures are relational to social practices and knowledge (Star, 1999). Bowker et al. (2010, p. 102) further proposed to investigate cyberinfrastructures as a set of distributed activities along a technical/social and a local/global axis: "The key question is not whether a problem is a 'social' problem or a 'technical' one. […] The question is whether we choose, for any given problem, a primarily social or a technical solution, or some combination." If the CNNs training data doesn't recognize certain aspects of the collection Brunner, we could define it as a technical problem. But at the same time, we can frame it as a social problem if we ask why certain motives appear in ImageNet and others do not.

2.3 Sociotechnical Imaginaries

Technical systems are inseparable from the social contexts from which they emerge and operate. The exploration of infrastructures such as ImageNet leads us to the *sociotechnical imaginaries* (Jasanoff, 2015) embedded within such technological systems, influencing the development of computer vision and their applications and implications to explore big visual data. I understand a CNN and its training data as collectively held and institutionally stabilized reflections of imagined forms of social life and order, which, in the context of this study, are used to explore yet another imagined form of social life and order embedded within the collection Brunner. The concept of sociotechnical imaginaries allows us to recognize and investigate the multilayered levels of meaning embedded in using CNNs for historical big visual data exploration. It highlights the intertwined nature of technical systems and social contexts. The CNNs way of seeing is built upon an archive (ImageNet) and is used

4 Tensorflow is an open-source software library for machine learning applications developed and maintained by Google, https://www.tensorflow.org/ (14.6.2024).

5 Keras is an open-source deep learning library written in the Python programming language, https://keras.io/ (14. 6. 2024).

to explore an archive (the collection Brunner). Both of these imagined forms of social life, lead, as I will try to show, to a potential clash of meanings.

3 Object of Study, Data, and Methods

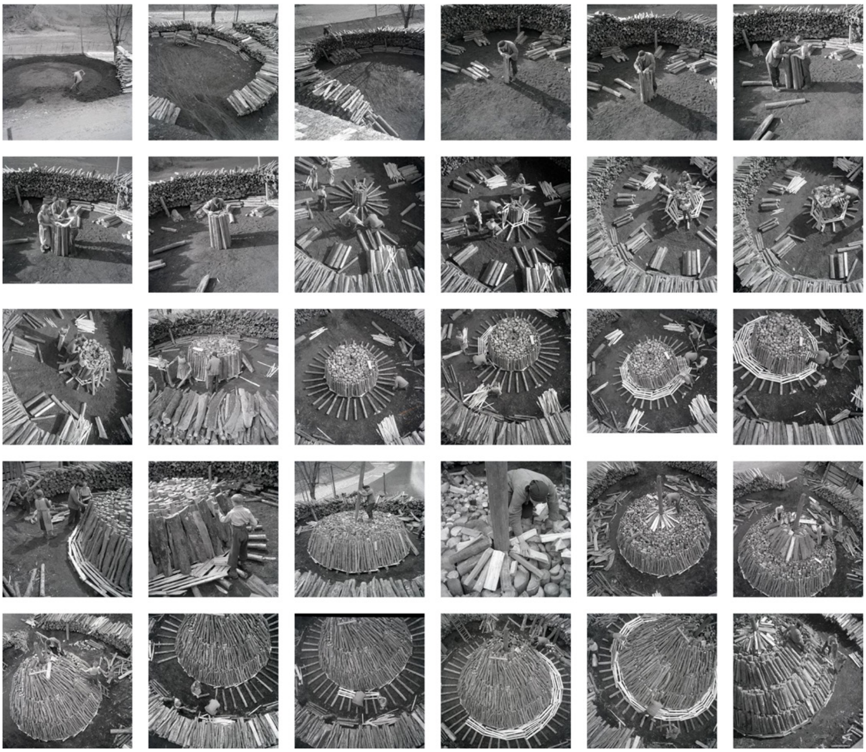
This section sets out the methodological approach of this study. 3.1) Examines the human way of seeing the collection through a literature-based historical recontextualization of the photographer Ernst Brunner and his work. 3.2) Briefly introduces the digitization of the collection and presents the Brunner data set that was clustered with PixPlot. 3.3) Explains in detail the underlying mechanics of the PixPlot application and the ImageNet data set that was used to train the CNN. 3.4) Describes the analytical approach for the interpretation of the clusters through a combination of close and distant reading.

3.1 The Collection Ernst Brunner

Ernst Brunner (1901–1979) attended a carpentry apprenticeship in his father's company in Mettmenstetten, Switzerland. After two semesters at the carpentry college in Nürnberg and studying interior design at the Kunstgewerbeschule Zurich, Brunner moved to Lucerne and worked as an interior designer. During the great depression (1929–1939) Brunner lost his job and attended a public employment program where he worked on an inventory of historical monuments. He taught himself photography autodidactically and presented his first pictures to Zurich Publisher *Regina* around 1936. He quickly began photographing for magazines such as *Das Schweizer Heim* and *Die Schweizer Familie*. Starting in the 1940s, photographs were repeatedly published in the Swiss fine art magazine *Du* (Lüthi & Frei, 2024). In 1955 Brunner was part of the famous exhibition *Family of Man* by Edward Steichen at MoMA New York (Steiger, 1998). From the mid-fifties till his death in 1979 Ernst Brunner shifted focus and became part of the *Aktion Bauernhausforschung in der Schweiz* (Farmhouse Research Campaign in Switzerland) initiated by the CAS (2023c) between 1919 and 1960. He documented the distinct architecture of farmhouses in Lucerne and published a corresponding book in 1977 (Brunner, 1977).[6] It was not until the 1990s that Brunner's photographic work was rediscovered by a broader public (Lüthi & Frei, 2024). Most notably through Peter Pfrunder's monograph *Ernst Brunner: Photographien, 1937–1962* (1995) and an accompanying travelling exhibition (*Verlorene Welten. Ernst Brunner Photographien 1937–1962*). It prominently features Brunner's work to illustrate the historical timber industry, farming, milling, or soil cultivation.

6 Brunner's photographic material on farmhouses in the canton of Lucerne is not part of the CAS collection. The many farmhouses from other cantons found in the CAS collection resemble Brunner's general interest and are not directly part of this research.

Figure 1 First Part of a Longer Photo Series on the Production of Charcoal (SGV_12N_04301 to SGV_12N_04330)



Source: Collection Ernst Brunner, photo archive of Cultural Anthropology Switzerland, https://archiv.sgv-sstp.ch.

Till today, the monograph has decisively shaped the perception of Brunner as the photographer who documented the vanishing rural world (Özvegyi, 2020, p. 26).

While Brunner's work has been published in various magazines, the available academic literature on Ernst Brunner is yet very limited. To the best of my knowledge, only two articles contextualize his work so far with a focus on his oeuvre (Steiger, 1998; Özvegyi, 2020). At the moment, the first dissertation on Brunner's collection is being written at the University of Basel (Lüthi, 2024). Although a large body of Brunner's work is concerned with agriculture and craftsmanship documenting the everyday lives of farmers in rural Switzerland, the academic literature also highlights the diversity of the collection, including photographs about city life, industry, construction projects, or military service (Özvegyi, 2020, p. 28). Due to Brunner's serial approach and rigid cataloguing, his work has further been described as "systematic" and "intended as objective documentations" (Steiger, 1998, pp. 26, 36). Brunner

Figure 2 Ernst Brunners Multifaceted Portrayal of Swiss Soldier During World War II (From Left to Right, Top to Bottom: SGV_12N_03504, SGV_12N_03563, SGV_12N_05317, SGV_12N_04682, SGV_12N_20301, SGV_12N_04603)

Source: Collection Ernst Brunner, photo archive of Cultural Anthropology Switzerland, https://archiv.sgv-sstp.ch.

created an extensive series on work processes, for example, timber processing or charcoal making (cf. Fig. 1). The latter is a prominent example of his photographic work that aimed at visually preserving knowledge that could potentially be lost. This effort to preserve can also be recognized in his involvement in the Swiss farmhouse research movement.

Ernst Brunner's work must further be understood in the political context of its time. As Steiger illustrates, Brunner's photographs contributed to the construction of a public image of Switzerland as a country of strong and free people living in an alpine landscape during World War II (Steiger, 1998, p. 33). Brunner's images were removed from their serial context and shown as collages of national unity in popular Swiss magazines such as *Das Schweizer Heim*. In the succeeding decades the photographs were published several times and generally in a way "which emphasized their formal and artistic character rather than their documentary purpose" to support the construction of a national myth of "wise, but seemingly uncomplicated farmers" that perform "real" work (Steiger, 1998, p. 47). During the war, Switzerland became politically and militarily isolated causing the desire to ensure one's own identity.

Steiger and Özvegyi both point out that Brunner knew how to create photographs that could be sold to magazines in the context of WWII. Ernst Brunner's depictions of farmers as free, independent, and hard-working people and of brave heroic soldiers were appreciated visual material in the effort for an intellectual national defence (*Geistige Landesverteidigung*). However, it would be short-sighted to impute a propagandistic intention to Brunner's work. As Özvegyi (2020, p. 41) shows, Brunner's photographs of the Swiss military not only included portrayals of soldiers fit for service, but also sober recordings of their daily lives (cf. Fig. 2).

To summarize, the way of seeing the collection driven by public discourse, magazines, and exhibitions, focuses on the depiction of *rural life and craftsmanship.* The academic perspective complements this view by highlighting the *documentary and rigid photographic approach,* the *complexity and diversity* of the collection and its specific historical context. For example, its use for the construction of a *nation's image of brave soldiers and independent farmers during WWII* while providing a far more nuanced view. These observations will guide the following examination of the PixPlot clusters. Will the machine see the same?

3.2 Digitization of the Collection and Dataset

After the death of Ernst Brunner in 1979 the collection was handed over to the CAS photo archive which is responsible for its archiving and digitization. The physical collection contains approximately 48 000 black-and-white negatives in medium format, 20 000 prints on index cards organized in a corresponding file system and additional material such as handwritten indexes, several hundred historical prints and specimen copies of published photos. Between 2014 and 2018 CAS conserved, restored, digitized, and, for the most part, also indexed the black-and-white negatives within a larger digitalization effort. Since 2021 and in the context of the SNSF research project *Participatory Knowledge Practices in Analogue and Digital Image Archives* (PIA, 2023a)[7] this is also being done for the additional material. The collection is further transferred into a new data model and base and an extended cataloguing is carried out. These efforts are being made not least to make the collection more accessible and understandable. As shown, the collection has so far received little academic attention despite its size and significance.

The Brunner data set used for clustering in this article has been created by accessing the PIA metadata API (PIA, 2023b) and IIIF API (PIA, 2023c). A Python script collected ID (e.g. SGV_12N_20301) and image title (e.g. "Soldaten beim Sport") for each available digital object in the API. At the time of the collection, the script collected a total of 47 837 ID and title pairs which were saved in a CSV file. A second Python script was used to download each image as a JPG file based on its

7 SNSF Grant Number 193788, cf. https://data.snf.ch/grants/grant/193788 (14. 6. 2024).

Table 1 Construction of the Image Set for Clustering

Type	Count
Available Objects in the Metadata API	47 837
Available Files on the Image Server	47 020
Unavailable Files on the Image Server	817
Total Files used in PixPlot to cluster	47 020

Source: Frischknecht, November 2023.

ID from the PIA image server. Interestingly, this produced a total of 47 020 image files, 817 files less than there are objects in the metadata API. One example of such a missing image file is SGV_12N_27618 ("Häuser auf einer Alp"): While the physical negative exists in the archive and the metadata API returns information on the object, the image server responds with an internal error that the file is temporarily not available. These interruptions in file availability are due to the complexity of the infrastructure and the fact that development is ongoing. They beautifully echo Susan Leigh Stars statement that infrastructure becomes visible upon breakdown (Star, 1999). The historical photographic collection, which we perceive as a stable entity, develops a certain dynamic in the digital. The PixPlot clusters do not represent the collection in itself, but the collection in a specific state at a specific point in time. As PixPlot requires an image file to see, the 817 digital objects with no available image file were excluded from the data set and a total of 47 020 images (98.3%) were used for clustering (cf. Table 1).

3.3 PixPlot and Convolutional Neural Networks (CNN)

PixPlot is a free, open-source application developed by Yale's Digital Humanities Lab in 2017 (see also Herms & Lehmann in this Special Issue). It has been used to cluster big visual data sets such as the collections of the Yale Center for British Art (Duhaime, 2017) or the Harvard Art Museum (Rodighiero et al., 2022). Cyberinfrastructures like PixPlot are not stand-alone software packages but built upon (and dependent on) existing code libraries.[8] PixPlots convolutional neural network (CNN) was pre-trained on ImageNet (Stanford Vision Lab, 2011) to detect key features from images, such as edges, shapes, textures, and patterns, and calculate the probability that the image shows a certain object such as a "tree" or "house". Based on these identified characteristics PixPlot clusters images according to their visual similarity. For each image, a featurization space of 2 048 dimensions is created, meaning that the CNN produces a vector with 2 048 values, each of which corresponds to a specific feature of the image. A vector is basically a list of numbers

8 For a complete list of libraries see the PixPlot code repository on Github (Duhaime, 2017).

that serve as a kind of coordinate that situates an image and its features in relation to other images. The featurization space holds the information on how similar an image is to another one. These features are not necessarily visible or understandable for humans but rather created by the algorithm through an iterative process of guess and check. Finally, to plot the images in two-dimensional space (on an X and Y axis) the 2048 image features need to be reduced to two. This is achieved using a dimensionality reduction algorithm, specifically UMAP (Uniform Manifold Approximation and Projection) (McInnes et al., 2020), that aims at reducing the dimensions while retaining as much of the relevant information as possible (see also Hoggenmüller & Klinke in this Special Issue).

To understand the in PixPlot embedded sociotechnical imaginaries we need to analyze its training data set ImageNet. The central characteristics of PixPlot's way of seeing – relational perception, algorithmic reading, and probabilistic interpretation – are essentially derived from this training data. ImageNet, originally created for visual object recognition, was one of the first widely available large-scale image data sets and has been central for the advancement of computer vision and deep learning research. It was developed at Stanford Vision Lab and first presented in 2009 at the IEEE Conference on Computer Vision and Pattern Recognition (Deng et al., 2009). Each year between 2010 and 2017 the data set and its accuracy have been developed further through the ImageNet Large Scale Visual Recognition Challenge (ILSVRC) (Russakovsky et al., 2015). The data set contains 1281167 training images (to learn visual features), 50000 validation images (to validate how well these features can be generalized), 100000 test images (to test what has been learned on unknown data) and 1000 object classes (that specify the labels such as "tree" or "house"). The images were collected from the internet through automated and manual searches. It remains unclear from which particular sources the images come but it seems likely that they are the results of well-known search engines such as Google or Yahoo on the one hand, and popular image websites such as Flickr on the other (cf. Deng et al., 2009; Russakovsky et al., 2015). The images are largely derived from North American amateur photography (Cox, 2022) and were labelled by precarious workers through Amazon's Mechanical Turk (Crawford, 2022). The classes to label the images are based on WordNet, a lexical database of nouns, verbs, adjectives, and adverbs that are grouped into sets of synonyms (Princeton University, 2010). In the subsequent analysis of this article, the openly available ImageNet data set on Kaggle is used for comparison and interpretation of the clustering results (Kaggle, 2020).

As we know, each data set comes with inherent bias. Generally, we can examine three levels of bias that can vary significantly across different data sets (Tommasi et al., 2017). Knowledge of these biases will support our assessment of the embedded sociotechnical imaginaries and whether a certain behavior of PixPlot can be framed as a technical or social problem. The first form of bias, the *capture bias*, relates to the distinct features of the images, such as angle or lightning. ImageNet-trained CNNs tend to have a bias towards identifying images based on texture rather than shape, contrary to humans where the case is the opposite (Geirhos et al., 2018). The ImageNet data set holds primarily images taken in well-lighted situations contributing to its texture-bias (Hermann et al., 2020). In contrast, humans are trained to see in diverse lighting conditions which is why the recognition of shapes takes on greater importance. The second form of bias, *label bias*, relates to the data sets visual semantic categories. According to Yang et al. (2020), WordNet (the lexical database used to create the ImageNet categories) includes words that are offensive in terms of sexuality or race, sensitive terms that can be offensive in a specific context, and terms that are hardly applicable to the description of images (e. g. "vegetarian"). While some of these words have been removed from the ImageNet data set, Yang et al. show that many slipped through the filtering process. Further, it is worth considering the language difference that results from the temporal difference of ImageNet and the collection Ernst Brunner. Lastly, the *negative bias* relates to the limits of the available images and categories and their representation of the world. A negative bias is challenging to address because changing the form of representation doesn't necessarily lead to a broader or more inclusive representation (Tommasi et al., 2017). Addressing the negative bias would lead to a more extensive data set, but it can never be reduced entirely. The negative bias can be understood as the periphery of the machine's eye.

3.4 Analysis Method: Distant and Close Reading

For the following analysis of the PixPlot clusters a combination of distant and close reading is proposed (Jockers, 2013; Moretti, 2016). Distant reading is understood as the "not reading" of (or not looking at) the collection photograph by photograph but rather from afar "to focus on units that are much smaller or much larger" (Moretti, 2016, p. 50).[9] 4.1) Describes the clusters from afar, not the individual photographs are of importance, but rather groups of photographs and how the clusters relate to each other. 4.2) Proposes a close reading of specific groups of photographs to identify what the machine might see, how this relates to the ImageNet data set and how it aligns or differs from the human perspective identified previously under section 3.1.

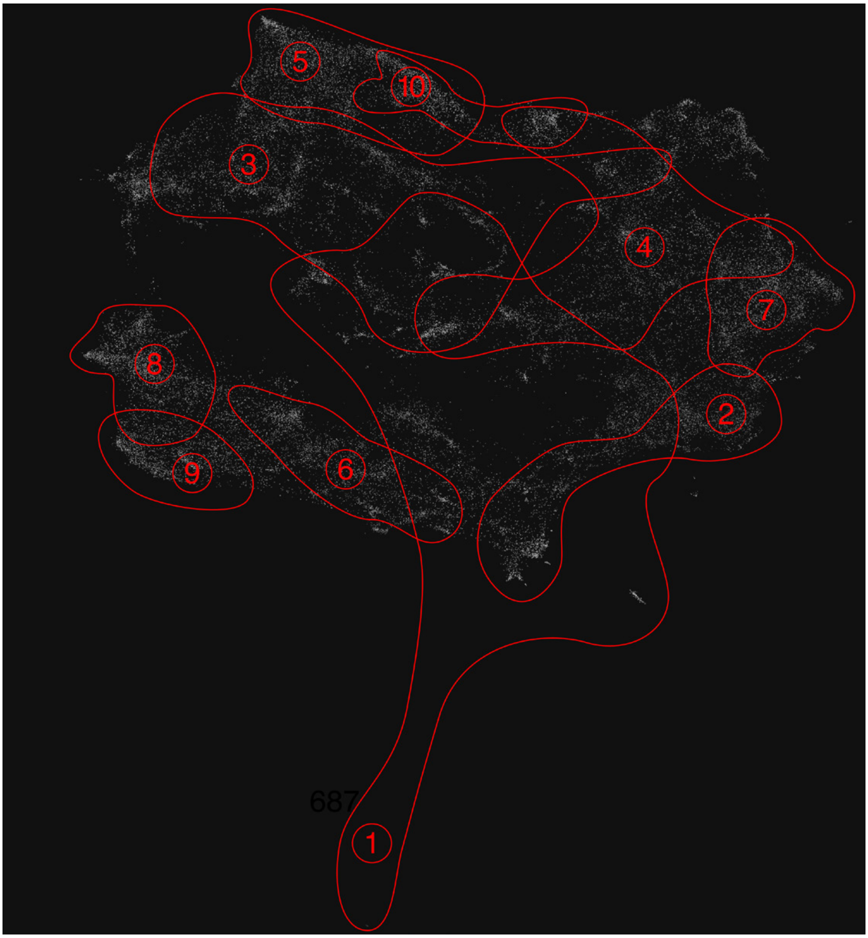
4 Analysis and Results

4.1 Cluster Analysis: A Panorama of Brunner's Work

PixPlot created and numbered ten visual clusters from the collection Brunner that are accessible via side navigation in the application. Figure 3 shows these clusters and

9 While Moretti's distant reading approach was originally developed for literature studies it has been adopted to analyze various media, including images (cf. Arnold & Tilton, 2019).

Figure 3 A Rough Outline of the Ten Clusters Produced by Pixplot



Source: Collection Ernst Brunner clusters, local instance of the PixPlot application [Screenshot, red outlines added, Frischknecht].

visualizes how they overlap. Table 2 provides a brief description of each cluster starting at the top left corner with cluster 3 and then continuing in a clockwise rotation. The ten clusters provide a distant view of the collection and highlight the large diversity of Ernst Brunner's work. Equipped with such an overview, the next section looks at three concrete examples to examine how human and machine ways of seeing might align, differ, or complement.

Table 2 Description of the Ten Image Clusters Created by PixPlot

Cluster	Description
3	Cluster 3 develops from trees and forests on the left, over roads, fields, and lakes in the middle, boats and shores on the bottom, hills in the top area, and rivers, rocks and stone fields on the right side.
5	Cluster 5 develops from rivers and rocky landscapes at the bottom, over mountain trees and forest in the middle, towards mountain valley view's visually similar due to their distinct V shape.
10	Cluster 10 starts with snowy mountains on the left and seems to lose altitude throughout the middle, showing roads, houses, and villages, ending with people skiing, sledging or wood chopping. The dominant visual feature of cluster 10 is the amount of white in the photographs caused by snow.
4	Cluster 4 collects men, women, and children in work settings, such as construction or in the field holding tools like scythes or pitchforks before traversing towards more formal images such as portraits.
7	Cluster 7 holds larger groups of people, for example at folk festivals, parades, or assemblies. The algorithm seems to focus on small repetitive structures such as a dense crowd or soldiers marching in line. However, these structures do not only apply to people. The algorithm also identifies grain sacks, milk cans and skulls in an ossuary.
2	Cluster 2 seems to focus more on images of children, overlapping with clusters 4 and 7. It further includes many photographs of people inside at a desk writing, eating, or working. It also includes a section that unites photographs that focus on the hands of people in working situations, for example, blacksmiths or carpenters.
1	Cluster 1 is the least dense. The center assembles a series of mini clusters showing power lines and cranes, flags, bridges, viaducts, boats (overlapping with cluster 3), locomotives (interestingly next to a cluster of railroads), fences and wood structures, or frontal views of stacked logs. All these motifs carry their own visual structure. At the lower end, it collects a relatively large corpus of photographs of book pages and paintings, an interest of Brunner unknown to me so far and a potential novel research direction. The very bottom holds a mini cluster of faulty scans that happened during the digitization process.
6	The right side of cluster 6 starts with a series of photographs showing decorative religious crosses before moving through gothic archways and interior views of chapels towards doors of all kinds. The center-top holds detailed photos of wooden stables and houses, while the left side shows exterior shots of houses, villages, and city alleys.
9	Cluster 9 collects images of castles, towers, and churches.
8	Cluster 8 is concerned with photographs of farmhouses of all kinds. Together with clusters 6 and 9, this cluster shows Ernst Brunner's interest towards architecture that resulted from his involvement in farmhouse research.

Source: Frischknecht, March 2024.

4.2 Comparing Human and Machine Way of Seeing

Under section 3.1 we found that in public discourse, magazines, and exhibitions, Brunner's work is best known for its depiction of rural life and craftsmanship. Academic perspectives complement this view by highlighting the complexity, diversity, historical context, and the documentary and rigid photographic approach. Based on these observations three questions are proposed to create a juxtaposition of human and machine ways of seeing:

1. Does PixPlot see the prominence of rural live and craftsmanship? Or does it rather focus on complexity and diversity?
2. Does PixPlot see the documentary and rigid photographic approach and cluster the extensive series of work processes?
3. Does PixPlot see the historical context and cluster photographic narratives such as the independent farmer or the strong soldier?

A Timeless View

Peter Pfrunder's monograph decisively shaped public recognition of Brunner for his depictions of rural life and craftsmanship (Pfrunder, 1995). While a book and exhibition involve a necessary selection and curation of photographs, PixPlot on the contrary clusters almost the complete collection (98.3%). A point highlighted by positivist voices from digital humanities which see the advantages of computational approaches in novel scalability of research across a larger volume of resources (cf. Kitchin, 2014). Indeed, PixPlot impressively shows how broad Ernst Brunner's photographic practice was beyond the rural side of Switzerland. For example, cluster 1 reveals the ongoing industrialization of the country, showing power lines, cranes, cable cars, and factories with smoking chimneys, but also large infrastructural projects such as railways, steel bridges, water reservoirs and dams (cf. Fig. 4). Simultaneously, PixPlot echoes public discourse showing farmers and craftsmen such as blacksmiths and carpenters in clusters 2 and 4.

At first, the negative bias – the limits of the representation of the world through a certain set – seems to be bigger in the monograph than within PixPlot. However, while PixPlot illustrates the diversity in the collection, it bears a negative bias of its own. The ImageNet categories contain many industrial and infrastructural terms, such as "steam locomotive", "crane", "steel arch bridge", or "dam". And while the data set includes "hammer" and "lumbermill" it does not include "blacksmith" or "carpenter". Here, ImageNet's focus on objects (not humans) comes to the fore and it performs particularly well on those objects whose appearance hasn't much changed since the 1950s, such as bridges or tools like hammers. For humans, looking at the collection is always a *looking back* at traditions and lifestyles of a different time. For the machine, time doesn't exist. Everything it has learned about the world comes from an internet snapshot made in the 2000s. Although ImageNet likely contains

Figure 4 Clustering of Steel Bridges in Cluster 1

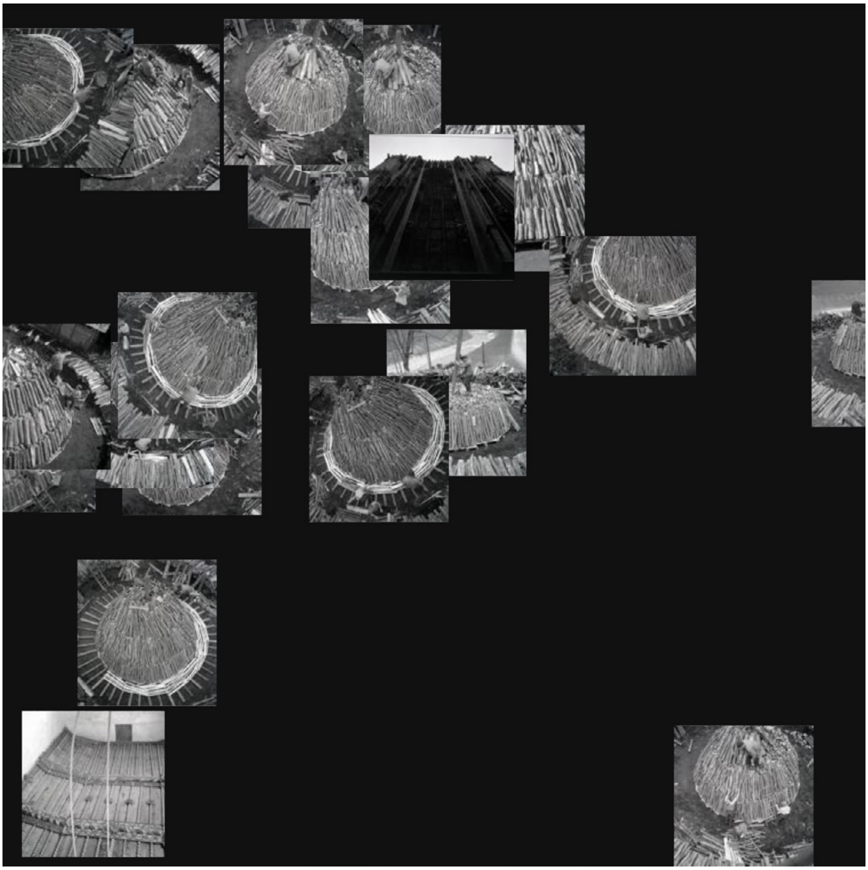
Source: Ernst Brunner clusters, local instance of the PixPlot application [Screenshot, Frischknecht].

historical images, the logic of machine learning classification demands the neutralization of contextual meaning in an image to become part of an aggregated mass of training data (Crawford, 2022, p. 93).

Focus on Texture

Brunner's work is appreciated for his documentary and rigid approach. Steiger (1998, p. 26) notes that Brunner "composed his photos with strong and simple forms" and often from above or below to add descriptive information. One example is the extensive series on charcoal making (cf. Fig. 1). Taken in good lighting conditions

Figure 5 Clustering of the Coal Series



Source: Ernst Brunner clusters, local instance of the PixPlot application [Screenshot, Frischknecht].

and with a birds-eye view, the round wooden pile at the center and the structure of the vertically arranged branches are distinct visual features that seem to jump into the machines eye. Although PixPlot is not able to fully cluster all series, it does make the serial and documentary approach of Brunner comprehensible (cf. Fig. 5). PixPlot's way of seeing aligns with Brunner's emphasis on strongly composed, serial and documentary photography.

Again, ImageNet doesn't include images or categories related to coal making, having a negative bias in this regard. This illustrates the machine's algorithmic reading and probabilistic interpretation. PixPlot iterates through the image pixel by pixel,

Figure 6 Assembling People, Soldiers, Grain Sacks, Milk Cans, and Skulls in the Ossuary

Source: Ernst Brunner clusters, local instance of the PixPlot application [Screenshot, Frischknecht].

line by line, and interprets the correlation of each pixel and its neighbors to calculate the probability of visual similarity. But, because of ImageNet's tendency to identify images based on texture rather than shape (capture bias, cf. 3.3) the photographs are still clustered, even though the machine doesn't know what coal making is. The capture bias, so to say, overcomes the negative bias. While this works well in the case of serial photographs, the texture bias is not always beneficial. In cluster 7, human crowds get paired with marching soldiers, grain sacks, milk cans and skulls, developing a certain morbid comedy (cf. Fig. 6). While there is a certain similarity in texture, a human eye easily recognizes the different contents of the images.

Figure 7 Clustering of Farmers

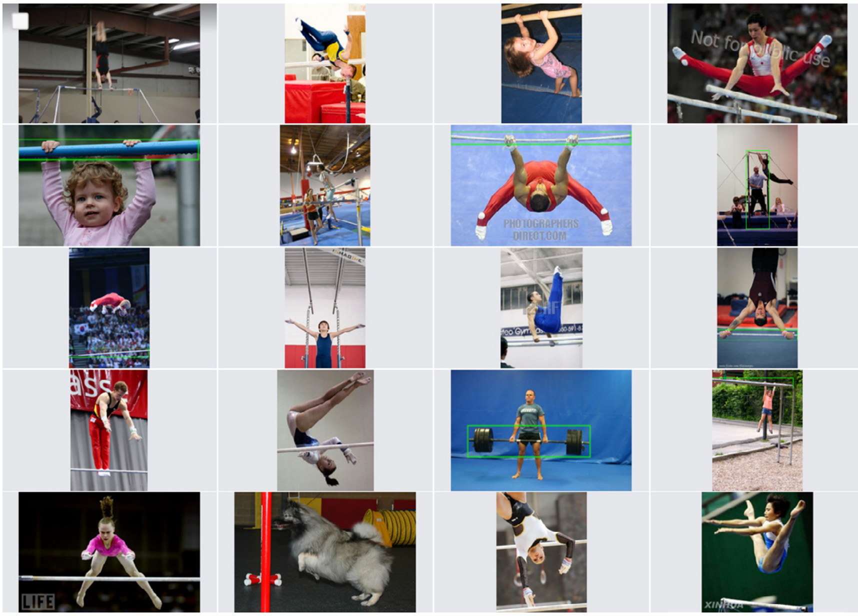
Source: Ernst Brunner clusters, local instance of the PixPlot application [Screenshot, Frischknecht].

Athletes in the Field

When it comes to more complex visual narratives, such as the free, independent farmer or the strong soldier, the human and machine ways of seeing seem to disperse further. Can a machine even detect such a historical narrative?

Farmers are distributed over large areas in clusters 2, 4, 7, and 10. Especially cluster 4 shows many working in the field (cf. Fig. 7). One could almost assume that PixPlot wants to "reassure the people of Switzerland of their roots, and of the meaning of their culture" as Steiger (1998, p. 39) describes the use of Brunner's photographs in *Das Schweizer Heim* in 1940. A closer look at the photographs in cluster 4 reveals

Figure 8 Selection of Imagenet Images Labeled With "Horizontal Bars, Bars"



Source: Navigu, ImageNet Dataset Explorer, https://navigu.net/#imagenet [Screenshot, Frischknecht].

how most farmers appear in conjunction with distinct objects, especially long tools like scythes, shovels, or pitchforks. A look into ImageNet reveals that there are not many categories for specific agricultural tools nor does the category "farmer" exist. The categories "parallel bars, bars" and "horizontal bar, high bar" are probably the closest thing to a pitchfork. Searching for "parallel bars" in the ImageNet data set reveals images of athletes performing high jump (cf. Fig. 8). Comparing them with the photographs in Figure 7 lets us suspect why PixPlot clustered the way it did. The machine doesn't see "wise, but seemingly uncomplicated farmers" doing "real work" (Steiger 1998, 47). The machine's perception is relational and determined by the distinctive horizontal and vertical objects.

Looking at the second historical narrative of the heroic soldier and national defense, PixPlot seems to have a clearer vision. The algorithm clusters photographs with greater visual variance than in the previous example on farmers. Soldiers are shown alone, in groups of different sizes and with different equipment (Fig. 9). Looking at the images makes clear that soldiers introduce more visual consistency due to their helmets and uniforms. Interestingly, in some cases, PixPlot places Soldiers in close proximity to military equipment such as tanks or anti-aircraft guns, which differ greatly visually. This alleged contextual knowledge shows how relational

Figure 9 Clustering of Soldiers in Diverse Situations

Source: Ernst Brunner clusters, local instance of the PixPlot application [Screenshot, Frischknecht].

perception, algorithmic reading, and probabilistic interpretation influence each other when concepts often co-occur in the training data.

It must be noticed that ImageNet is highly militarized containing many categories such as "rifle", "assault rifle", "tank, armored combat vehicle", "warplane", or "cannon".[10] Again, with a focus on objects as the data set includes "military uniform" but not "soldier". This militarized view overlooks the versatility of Brunner's work and his portrayal of soldiers as playful and vulnerable humans playing soccer or sleeping (cf. Fig. 2). On the contrary, the machine falls into an almost propagandistic mode.

10 Here it would be interesting to examine more closely how this speaks for North American society, from which large parts of the ImageNet images originate.

5 Conclusion

The article started with the assumption that visual similarity can be fruitful for the exploration and examination of large collections if the CNN clusters the photographs along the central topics and narratives inherent in the collection. The study juxtaposed a human perspective on the collection derived from literature with a machine perspective through PixPlot and its underlying infrastructure, particularly ImageNet. The main focus of the study was the examination of the epistemological implications of such a human-machine interpretation interplay, and it was conducted along four research questions: 1) Can a CNN recognize the central (visual) topics and narratives of a historical photographic collection? 2) To what extent does this machine way of seeing align or differ from a human perspective on the collection? 3) Does this difference allow for interesting modes of human-machine collaboration? 4) And what epistemological implications arise from such a collaboration?

The central topics and narratives prevalent in public discourse, magazines, and exhibitions, are Brunner's depiction of rural life and craftsmanship. Academic perspectives complement this view by highlighting the collection's complexity, diversity, historical context, and the documentary and rigid photographic approach. The CNN recognized some, but not all, of these central topics and narratives. The clusters make the important role of rural life and craftsmanship comprehensible while simultaneously showing the impressive size and diversity of the collection. This is a result mainly due to the scalability of the computational approach. The CNNs bias towards textures aligns well with Brunner's documentary approach, in particular the extensive photo series, such as the one on coal making, are clustered together, at least for the most part. In relation to historical context, matters get more complicated. A narrative such as the free, independent farmer becomes partly comprehensible while the soldiers are mostly shown in training situations lacking Brunner's humanistic portrayal of their everyday lives.

Besides accessibility resulting from scalability, I see the potential of an interpretative collaboration above all in the fact that the machine view emphasizes formal aspects of Brunner's photographic language. Building on the texture bias, a further specialized CNN could, for example, support questions in art history regarding image composition. On the contrary, the collaboration must be viewed critically when the machine's view serves as the basis for the development of content-related questions, for example, to inform potential novel research directions. Here, the sociotechnical imaginaries embedded in cyberinfrastructure turn out to be a kind of epistemological trojan horse. Hidden behind layers of interface, software and code lies the ImageNet data set with its own classification of the world from which the machine's way of seeing emerges. This has epistemological consequences insofar as that *the CNN has the ability to shift our attention towards certain narratives and imaginaries that we might assume arise from the collection itself but actually originate* *in the infrastructure.* A clash of meaning arises between a historical photo collection and the CNNs training data set. In particular, the different time dimensions of the collection and the training data create hidden temporal references (examining these in more detail would be an exciting undertaking in itself). To assess the origin of phenomena that one might observe in the collection through the CNNs clustering, it is necessary to cross-reference the machine ways of seeing with its original ontology ImageNet. Here, the study showed that infrastructural and critical data studies prove to be a suitable tool to critically interpret, and where necessary, recontextualize information gained from machine-learning based big visual data exploration.

In conclusion, it can be stated that an improved human-machine interpretation interplay depends on training data sets tailored to the human's interpretation interest. It would be interesting to see how a CNN clusters the collection Brunner if trained with historical data including categories derived from description of work processes or oral history. To further promote the potential of machine-learning-based approaches for big visual data analysis, more diverse and thematically specific data sets must become publicly accessible. Inspired by human ways of seeing we should train many ways of machine seeing instead of a few trying to depict the whole world.

6 References

Arnold, T., & Tilton, L. (2019). Distant Viewing: Analyzing Large Visual Corpora. *Digital Scholarship in the Humanities*, *34*(1), i3–i16. https://doi.org/10.1093/llc/fqz013

Berger, J. (1972). *Ways of seeing: Based on the BBC television series*. Penguin Books.

Bowker, G. C., Baker, K., Millerand, F., & Ribes, D. (2010). Toward Information Infrastructure Studies: Ways of Knowing in a Networked Environment. In J. Hunsinger, L. Klastrup, & M. Allen (Eds.), *International Handbook of Internet Research* (pp. 97–117). Springer Netherlands. https://doi.org/10.1007/978-1-4020-9789-8_5

Brunner, E. (1977). *Die Bauernhäuser im Kanton Luzern. Empirische Kulturwissenschaft Schweiz.*

Burrows, R., & Savage, M. (2014). After the crisis? Big data and the methodological challenges of empirical sociology. *Big Data & Society*, *1*(1), 1–6. https://doi.org/10.1177/2053951714540280

CAS (Cultural Anthropology Switzerland). (2023a). Das Fotoarchiv der EKWS (Empirische Kulturwissenschaft Schweiz). https://archiv.sgv-sstp.ch/ (accessed March 12, 2023).

CAS (Cultural Anthropology Switzerland). (2023b). SGV_12 Ernst Brunner. https://archiv.sgv-sstp.ch/collection/sgv_12/all/1 (accessed January 30, 2023).

CAS (Cultural Anthropology Switzerland). (2023c). Die Bauernhäuser der Schweiz. https://www.volkskunde.ch/sgv/publikationen/reihen/die-bauernhaeuser-der-schweiz/ (accessed January 30, 2023).

Cox, G. (2022). Ways of machine seeing as a problem of invisual literacy. In A. Dewdney & K. Sluis (Eds.), *The networked image in post-digital culture* (pp. 101–113). Routledge.

Crawford, K., & Paglen, T. (2019). *Excavating Ai: The Politics of Images in Machine Learning Training Sets*. https://excavating.ai/

Crawford, K. (2021). *Atlas of AI: Power, politics, and the planetary costs of artificial intelligence*. Yale University Press.

Deng, J., Dong, W., Socher, R., Li, L., Li, K., & Fei-Fei, L. (2009). ImageNet: A large-scale hierarchical image database. In *2009 IEEE Conference on Computer Vision and Pattern Recognition* (pp. 248–255). https://doi.org/10.1109/CVPR.2009.5206848

Duhaime, D. (2017). PixPlot. https://github.com/YaleDHLab/pix-plot (accessed February 30, 2023).

Frade, C. (2016). Social theory and the politics of big data and method. *Sociology, 50(5), 863–877.* https://doi.org/10.1177/0038038515614186

Geirhos, R., Rubisch, P., Michaelis, C., Bethge, M., Wichmann, F. A., & Brendel, W. (2018). ImageNet-trained CNNs are biased towards texture; increasing shape bias improves accuracy and robustness. *arXiv*. https://arxiv.org/abs/1811.12231 (accessed June 14, 2024).

Hermann, K. L., Chen, T., & Kornblith, S. (2020). The origins and prevalence of texture bias in convolutional neural networks. *Proceedings of the 34th International Conference on Neural Information Processing Systems*, 19000–19015.

Herms, K., Lehmann, J. (2025). Seeing Like a Field? *Schweizerische Zeitschrift für Soziologie, 51*(2), Special Issue hrsg. von S. W. Hoggenmüller, Big Visual Data als neue Form des Wissens: Potenziale, Herausforderungen und Transformationen.

Hoggenmüller, S. W., Klinke, H. (2025). Metabilder als Forschungswerkzeuge: Zur Kontingenz und algorithmischen Bedingtheit ihrer Herstellung. *Schweizerische Zeitschrift für Soziologie, 51*(2), Special Issue hrsg. von S. W. Hoggenmüller, Big Visual Data als neue Form des Wissens: Potenziale, Herausforderungen und Transformationen.

Jasanoff, S. (2015). One. Future Imperfect: Science, Technology, and the Imaginations of Modernity. In *Dreamscapes of Modernity* (pp. 1–33). University of Chicago Press. https://doi.org/10.7208/chicago/9780226276663.001.0001

Jockers, M. L. (2013). *Macroanalysis: Digital methods and literary history*. University of Illinois Press.

Kaggle. (2020). ImageNet object localization challenge. https://kaggle.com/competitions/imagenet-object-localization-challenge (accessed December 11, 2023).

Kitchin, R. (2014). *The Data Revolution: Big Data, Open Data, Data Infrastructures & Their Consequences*. SAGE Publications Ltd. https://doi.org/10.4135/9781473909472

Lüthi, F. (2024). Partizipative Wissenspraktiken in analogen und digitalen Bildarchiven am Beispiel der Sammlung Ernst Brunner (SGV). https://universe.unibas.ch/projects-collaborations/9088 (accessed March 12, 2024).

Lüthi, F., & Frei, F. (2024). Objektbiografien im interdisziplinären Fokus. Ein Werkstattbericht aus der Sammlung Ernst Brunner. Lecture, Seminar für Kulturwissenschaft und Europäische Ethnologie, Universität Basel.

Manovich, L. (2011). Trending: The Promises and the Challenges of Big Social Data. In M. K. Gold (Ed.), *Debates in the Digital Humanities* (pp. 460–475). University of Minnesota Press. https://doi.org/10.5749/minnesota/9780816677948.003.0047

McInnes, L., Healy, J., & Melville, J. (2020). *UMAP: Uniform Manifold Approximation and Projection for Dimension Reduction*. arXiv. https://doi.org/10.48550/arXiv.1802.03426

Moretti, F. (2016). *Distant Reading*. Konstanz University Press.

Özvegyi, A. (2020). Von der heroisierten Inszenierung zur ernüchterten Darstellung? Fotografien von Ernst Brunner aus seiner Militärdienstzeit bei der Fliegerabwehrbatterie 311. *Schweizerisches Archiv für Volkskunde/ Archives suisses des traditions populaires, 116(2), 25–46.* https://doi.org/10.5169/SEALS-913987

PIA (Participatory Image Archives). (2023a). Participatory Image Archives. https://about.participatory-archives.ch/ (accessed January 30, 2023).

PIA (Participatory Image Archives). (2023b). PIA Metadata API. https://opendata.swiss/de/dataset/pia-metadata-api (accessed December 23, 2023).

PIA (Participatory Image Archives). (2023c). PIA IIIF API. https://opendata.swiss/de/dataset/pia-iiif-api (accessed December 23, 2023).

Pfrunder, P. (1995). *Ernst Brunner: Photographien, 1937-1962*. Schweizerische Gesellschaft für Volkskunde.

Princeton University. (2010). WordNet: A lexical database for English. https://wordnet.princeton.edu (accessed April 30, 2023).

Rodighiero, D., Derry, L., Duhaime, D., Kruguer, J., Mueller, M. C., Pietsch, C., Schnapp, J. T., Steward, J., & metaLAB. (2022). Surprise machines: Revealing Harvard Art Museums' image collection. *Information Design Journal*, *27*(1), 21–34. https://doi.org/10.1075/idj.22013.rod

Russakovsky, O., Deng, J., Su, H., Krause, J., Satheesh, S., Ma, S., Huang, Z., Karpathy, A., Khosla, A., Bernstein, M., Berg, A. C., & Fei-Fei, L. (2015). ImageNet Large Scale Visual Recognition Challenge. *International Journal of Computer Vision*, *115*(3), 211–252. https://doi.org/10.1007/s11263-015-0816-y

Savage, M., & Burrows, R. (2007). The Coming Crisis of Empirical Sociology. *Sociology*, *41*(5), 885–899. https://doi.org/10.1177/0038038507080443

Stanford Vision Lab. (2011). ImageNet. https://image-net.org/index.php (accessed May 17, 2023).

Star, S. L. (1999). The Ethnography of Infrastructure. *American Behavioral Scientist*, *43*(3), 377–391. https://doi.org/10.1177/00027649921955326

Steiger, R. (1998). On the uses of documentary: The photography of Ernst Brunner. *Visual Sociology*, *13*(1), 25–47. https://doi.org/10.1080/14725869808583785

Stevenson, J. (2008). The Online Archivist: A Positive Approach to the Digital Information Age. In L. Craven (Ed.), *What are archives? Cultural and theoretical perspectives a reader* (pp. 89–108). Ashgate.

Terras, M. M. (2011). The Rise of Digitization. In R. Rikowski (Ed.), *Digitisation Perspectives* (pp. 3–20). SensePublishers. https://doi.org/10.1007/978-94-6091-299-3_1

Tommasi, T., Patricia, N., Caputo, B., & Tuytelaars, T. (2017). A Deeper Look at Dataset Bias. In G. Csurka (Ed.), *Domain Adaptation in Computer Vision Applications* (pp. 37–55). Springer International Publishing. https://doi.org/10.1007/978-3-319-58347-1_2

Trumpener, K. (2009). Critical Response I. Paratext and Genre System: A Response to Franco Moretti. *Critical Inquiry*, *36*(1), 159–171. https://doi.org/10.1086/606126

Yale Digital Humanities Lab. (2017). Yale DHLab – PixPlot. https://dhlab.yale.edu/projects/pixplot/ (accessed May 13, 2023).

Yang, K., Qinami, K., Fei-Fei, L., Deng, J., & Russakovsky, O. (2020). Towards fairer datasets: Filtering and balancing the distribution of the people subtree in the ImageNet hierarchy. *Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency*, 547–558. https://doi.org/10.1145/3351095.3375709