Renovating the Search Experience of Art Image Databases

by

Silvia Wang

Department of Art, Art History and Visual Studies
Duke University

Date:________________________
Approved:

____
Victoria Szabo, Co-supervisor

____
Augustus Wendell, Co-supervisor

____
Mark Olson

Thesis submitted in partial fulfillment of
the requirements for the degree of Master of Arts
in the Department of Art, Art History and Visual Studies
in the Graduate School
of Duke University

2022

ABSTRACT

Renovating the Search Experience of Art Image Databases

by

Silvia Wang

Department of Art, Art History and Visual Studies
Duke University

Date:_______________________
Approved:

____
Victoria Szabo, Co-supervisor

____
Augustus Wendell, Co-supervisor

____
Mark Olson

An abstract of a thesis submitted in partial
fulfillment of the requirements for the degree
of Master of Arts in the Department of
Art, Art History and Visual Studies in the Graduate School of
Duke University

2022

Abstract

With the emergence of the computer, the digital image has become one of the most prevalent visual mediums in the 21st century. This paper aims at analyzing the current limitations in interacting with digital image databases in art historical research. In response to the limitations of current image database structures, this paper explores how emerging computer vision technologies can be applied to enrich the ways database users interact with art image databases. While current image databases primarily rely on manually-defined metadata and textual descriptions to associate art images, this digital project demonstrates how deep neural networks can add visual connections between art images through feature extraction algorithms. This thesis documents a digital product that demonstrates how deep neural network models can extract images' visual features and connect art images by these visual features.

Although it offers a new approach to the problem, this digital project is not intended to replace the existed metadata structure and text-based search in existing image database system. Metadata and text-based search have developed over time to assist people in managing data and navigating the digital world in the era of big data. As such, this digital project offers to overlay a visual-driven search path upon the existing database structure in order to provide a more diverse search environment.

Contents

List of Figures

Introduction

In 1949, Italian Jesuit priest Roberto Busa, with the help of IBM, used a computer to successfully index the Latin works of St Thomas Aquinas and associates, a corpus of upwards of 11 million words[1]. Since the format of textual data is more consistent with the compilation processes of the computer itself[2], the practice of what would become the digital humanities began to evolve first around the textual medium, and within the next twenty years. Similarly, in the field of computer science, we can also observe a consistent tendency to focus on text, in that text-based research seems to move more rapidly than that based on images. As a result, art history, a discipline that centered around images and visual analysis, was slower to engage in digital humanities[3] approaches to research. Although access to digital art images have been growing since 1960s, rather than being actively used for computational analysis, use of digitized of images was geared toward reproduction and display. This thesis holds the standpoint that the technological advancement in digital image processing, computer graphics and computer vision since late 70s have substantially increased the computability of digital images. Their

---

[1] Steven E. Jones, *Roberto Busa, S. J. , and the Emergence of Humanities Computing: The Priest and the Punched Cards*, vols. (London, UNITED KINGDOM: Taylor & Francis Group, 2016), online, Internet, 9 Mar. 2022. , Available: http://ebookcentral.proquest.com/lib/duke/detail.action?docID=4470572.
[2] Harald Klinke, "Big Image Data within the Big Picture of Art History" *International Journal for Digital Art History*. 2 (2016), online, Internet, 25 Feb. 2022. , Available: https://journals.ub.uni-heidelberg.de/index.php/dah/article/view/33527.
[3] Johanna Drucker, "Is There a 'Digital' Art History?" *Visual Resources*. 29.1–2 (2013): 5–13.

increasing computability means the increase in machine's ability of extracting meaningful features from digital images for computational analysis.

Lev Manovich distinguishes the concept of "metadata" and "feature"[4] in his book *Cultural Analytics.* He describes metadata as "information about objects that already existed and was transferred into a computer"[5]. Reflecting on the current art image databases, common metadata includes creator's name, nationality, date of birth and death, culture, geography, medium, genre, classification and etc. In comparison to metadata that is pre-defined in a dataset or collection, Lev Manovich describes features as new pieces of information that are generated through "algorithmic analysis of the objects"[6]. Based on the observation of current art image databases, the method of describing image data in databases is still dominated by metadata and textual annotations. This approach, however, has limits. While metadata provides a standard for art image data storage, it also simplifies the information carried by image data and solidifies the association between images with conventional classifications and categorizations. With the progression of digitization, art image databases have been expanded, aiming to organize and integrate the information for people to retrieve. This paper addresses the limitations in the existing search methods used by art image

---

[4] "Cultural Analytics | Books Gateway | MIT Press," n.d., online, Internet, 22 Mar. 2022. , Available: https://direct-mit-edu.proxy.lib.duke.edu/books/book/4966/Cultural-Analytics.
[5] Ibid.
[6] "Cultural Analytics | Books Gateway | MIT Press," n.d., online, Internet, 22 Mar. 2022. , Available: https://direct-mit-edu.proxy.lib.duke.edu/books/book/4966/Cultural-Analytics.

databases and aims to explore how computer vision technology can be applied to extract features relating to images' visual content. These features are generated by computer algorithms and encompass more bottom-level information in digital images. With the increase of digital images' computability, the use of digital images is no longer limited to simple proprietary information such as color, color layout, size, orientation, etc. With the establishment of ImageNet and other large-scale image databases for machine learning training and the continuous development of neural network architectures, computer vision is moving in the direction of understanding the content-level features of images. As a result, the use of digital images in the digital humanities is no longer limited to display and reproduction, but can go further to use deeper content-based features to mine connections between digital art images from a large-scale collection. This paper argues that with the development of computer vision, coupled with a significant increase in computing power, digital image data can be more involved in the computational analysis in digital art history to facilitate the excavation of image relationships based directly on visual elements.

This thesis consists of six main chapters. The first chapter begins with an inquiry into the role and significance of image sources in the art history discipline. The second chapter further investigates the principles used by art historians to store their physical collections prior to digitization. The first two chapters intend to set up a historical framework for the subsequent discussion of current digital art image databases. Chapter three discusses how the emergence of digital image and computational tools have reshaped the way art historians work since 70s. This section then further investigates the historical causes of the current text-based search model used in art image databases from an interdisciplinary perspective and addresses the limitations of such search method under the context of art historical research. The fourth chapter proposes a content-based image retrieval prototype using computer vision technologies. This prototype is designed to provide an image-driven search path to explore digital space of artworks. The last chapter documents a digital project that serves as a proof of concept for this thesis. The demo project aims to illustrate a potential way in which computer vision algorithms could be incorporated with art image database in the near future, facilitating art-related research and stimulating new knowledge generation.

1. “Image” in the Eyes of Art History

To design an art image database that adapts to the needs of art historical research, it is essential to comprehend the fundamental theories underlying the interpretation of artworks in the field of art history. These theories, which have had a profound influence on the development of the discipline, can provide clearer insight into the elements that are of true concern for art historical researchers. Such elements will serve as a guide in the data retrieval methodology for the art image database discussions to follow.

1.1 An Overview of Contemporary Art Historical Theories

In the early twentieth century, Aby Warburg presented the world with the Atlas of Mnemosyne, establishing the centrality of images in the study of art history[1]. Under the influence of Warburg, in 1939 the German art historian Erwin Panofsky further defines iconography through three strata: 1. Primary or Natural Subject Matter 2. Secondary or Conventional Subject Matter 3. Intrinsic Meaning or Content[2]. With its focus on the “meaning” behind the image, iconography aims to examine and encapsulate the themes of visual art from an interdisciplinary perspective. It is a

---

[1] “‘Aby Warburg’s Mnemosyne Atlas: On Photography, Archives, and the Afterlife of Images,’ by Molly Kalkstein – Rutgers Art Review,” n.d., online, Internet, 2 Mar. 2022. , Available: https://rar.rutgers.edu/aby-warburgs-mnemosyne-atlas/.

[2] Erwin Panofsky, *Studies in Iconology: Humanistic Themes In the Art of the Renaissance*, vols. (New York: Routledge, 2019).

perspective that involves cultural, historical and sociological interpretation, tightening the connection between art history and other disciplines in the humanities[3].

While iconography focuses more on extrinsic meanings, some scholars have attempted to investigate the intrinsic development of visual art from a more purely visual realm that is not diverted by external circumstances. This perspective in art image study eventually developed into formalism. One of the most prominent founders of this theory was Swiss art historian Heinrich Wölfflin. In 1915, he published *Principles of Art History*, which later provided theoretical support for abstract and contemporary art. In this book, he introduced five pairs of basic concepts: 1. Linear and Painterly 2. Plane and Recession 3. Closed and open form 4. Multiplicity and Unity 5. Clearness and Unclearness[4].These five pairs of characteristics are an attempt to investigate the intrinsic development of visual art from a more purely visual realm that is not diverted by external circumstances.

In the subsequent chapters, this thesis will draw upon the theories and methods of these two art historians, Aby Warburg and Roberto Longhi, because they lend themselves to a computationally-media approach to art image analysis. On the one hand, they are chosen because they have different perspectives when approaching an

---

[3] Allan H. Gilbert and Horst W. Janson, "Erwin Panofsky, Studies in Iconology: Humanistic Themes in the Art of the Renaissance" *The Art Bulletin*. 22.3 (1940): 172–175.
[4] Wölfflin Heinrich, *Principles of Art History: The Problem of the Development of Style in Early Modern Art*, vols., 2015th ed. (Getty Trust Publications, 1915).

image. The former leans more towards iconography, while the latter leans more towards formalism[5]. By looking at both of their methods, the paper intents to design a more encompass search prototype in the chapter five. On the other hand, despite having distinct interests in images, both of them base their research on the examination and analysis of an extensive number of images[6]. As a result, they have both developed unique approaches to storing images, another important aspect of image archives. . Their methods will serve as a guide and reference for later discussions on how to reconstruct digital art image databases.

1.2 Aby Warburg and Mnemosyne Atlas

In the late nineteenth century, the advancement of the camera industry made photography accessible, and Warburg leveraged its reproducibility to launch his exploration of the image[7]. Although Warburg's *Mnemosyne Atlas* remains an unfinished work, it has had a profound impact on the study of images in the field of art history ever since[8]. Therefore, taking a close look at this *Atlas* and how Warburg worked with it enables us to have a deeper understanding of the way images are perceived and used in

---

[5] Rauwel Alain, "Roberto Longhi: the eye and the verb" *Art Critique*. , 12 Dec. 2018, online, Internet, 3 Mar. 2022. , Available: https://www.art-critique.com/en/2018/12/roberto-longhi-the-eye-and-the-verb/.
[6] "Library – Fondazione Roberto Longhi," n.d., online, Internet, 3 Mar. 2022. , Available: https://fondazionelonghi.it/wordpress/en/eng_library/.
[7] Logan Larsen, "On the Library as Foundation: Thinking On and Beyond Aby Warburg's Systematic Approach to Images and Culture"Thesis, 2020, online, Internet, 3 Mar. 2022. , Available: https://repositories.lib.utexas.edu/handle/2152/81441.
[8] "Bilderatlas Mnemosyne" *The Warburg Institute*. , n.d., online, Internet, 8 Mar. 2022. , Available: https://warburg.sas.ac.uk/archive/bilderatlas-mnemosyne.

art historical research, or at least within the iconographic branch.

When Warburg worked on the *Mnemosyne Atlas* project, he used sixty-three wooden panels to arrange over two thousand images[9]. Hristova describes Warburg's methodology in detail in the article *Images as Data: Cultural Analytics and Aby Warburg's Mnemosyne*:

> "The black-and-white photographs of artworks were pinned either directly to the black fabric, or framed over a white mate. They varied in size, proportion of the depicted image, as well as distance from each other. Warburg sometimes photographed the work of art in its entirety. In other instances, he focused on a detail and included only a close-up. Through the use of the close up, he cropped and framed elements that can be then positioned as points of emphasis within the panels. The panels themselves have not survived and what we have left instead are black and white photographs of these panels instead.[10]"

Hristova's description suggests that Warburg's observation of images is not at a fixed distance. Use figure 1, a photo of No.39 panel in the *Mnemosyne Atlas*, as an example: on the one hand, one can consider his interaction with images as a macroscopic distanced reading[11]—a set of visually resonated images are fixed on a single plate to shape a general vision of the theme. On the other hand, Warburg's particular interest to motifs and substructure in the images is also clearly reflected by those close-up shots.

---

[9] Ibid.

[10] Stefka Hristova, "Images as Data: Cultural Analytics and Aby Warburg's Mnemosyne" *International Journal for Digital Art History*. 2 (2016), online, Internet, 3 Mar. 2022. , Available: https://journals.ub.uni-heidelberg.de/index.php/dah/article/view/23489.

[11] Kathryn Brown, ed., *The Routledge Companion to Digital Humanities and Art History*, vols. (New York: Routledge, 2020).

Another aspect of his work worth noting is that Warburg did not restrict his selection of images to a certain type. The image types in the *Atlas* include "photographic reproductions, photos, diagrams and sketches, postcards and various kinds of printed material including adverts and newspaper clippings[12]". As figure 2 shows, the panels are organized with a carefully designed arrangement[13], so that wooden panels don't exist independently, but echo each other to form a coherent narrative.

The *Mnemosyne Atlas* exhibits a discovery-based approach to image analysis that expands on the juxtaposition of large image sets. This approach emphasizes the visual information embedded in the images, and does not rely on text to articulate relationships between images. Visual narratives are constructed on each panel, or between panels, through the visual characteristic of images.

1.3 Roberto Longhi: Untranslatable Visual Representation

Distinct from Warburg's interpretation of the images, formalism is more concerned with the visual qualities inherent in the images[14]. These qualities are considered as internal visual representations that exist independently of external cultural factors and harder to be projected into language space. Under the influence of

[12] "Bilderatlas Mnemosyne | Final version" *The Warburg Institute*. , n.d., online, Internet, 8 Mar. 2022. , Available: https://warburg.sas.ac.uk/archive/bilderatlas-mnemosyne/final-version.

[13] Larsen, "On the Library as Foundation."

[14] Laura Moure Cecchini, "Baroque Futurism: Roberto Longhi, the Seventeenth -Century, and the Avant-Garde" *The Art Bulletin*. 101.2 (2019): 29–53.

the theories of "pure visibility"[15], Longhi focuses his research on the evaluation of "pure" visual features of images, such as line, color, and composition[16].

By reviewing the table of contents of his *Breve ma veridica storia della pittura italiana*[17], which was written as a compendium for his students, one can obtain an overview of his approach. It can be summarized as drawing visual clues by reviewing a large number of images and defining the observed clues by what he called verbal equivalents[18]. Then, these visual clues are used as basis for tracing the development of artists and artistic periods. Examples of Longhi's definition of styles are shown in the figure 3. He used *Stile plastico*, *Stile plastic-lineare*, *Stile prospettico di forma* and *Stile coloristico puro*[19] to describe the four visual styles of paintings. Longhi invested a remarkable amount of effort in establishing his own verbal equivalents, attempting to translate visual information through his superior linguistic talent[20]. Despite great effort, however, Roberto Longhi's work is still quite difficult to comprehend in non-Italian speaking cultures due to his extremely individualistic rhetoric. Bloch describes Longhi's work as virtually untranslatable[21]. He also points out the fact that Longhi's, and other art

---

[15] Ibid.
[16] Ibid.
[17] Roberto Longhi, *Breve ma veridica storia della pittura italiana*, vols. (Abscondita, 2013).
[18] "Opening art with 'art': Roberto Longhi's approach to art historical research and writing Chinese art history with 'language' (Volume one)" *Art History Library*. , n.d.
[19] "An Introduction to the Use of Images in Art Historical Research: Context and Prospects" *Art History Library*. (2020).
[20] Rauwel, "Roberto Longhi."
[21] Vitale Bloch, "Roberto Longhi" *The Burlington Magazine*. 113.823 (1971): 609–612.

historians' work that share similar methodology, were less influential in their foreign translations, thereby limiting their impact on the field[22].

Nonetheless, Roberto Longli and his methodology are highlighted here for two main reasons. First, both his methodology and the external critiques his method received reflects the difficulty in are highly relevant in the discussion about the remapping of visual features in the language space. Second, Longhi's innovative visual feature mining process based on a large number of images has parallels with unsupervised feature learning in today's computer vision domain. To put it into a nutshell, his work reflects the information loss associated with the conversion between vision and language: the first layer of bias is reflected in the subjectivity when describing images; the second level of bias is due to different linguistic and cultural contexts. Although Longhi had superior rhetorical skills, the quality of visual information conveyed in text was limited by the cultural context, resulting in Longhi's depiction of visual features ultimately being less accessible to audiences from non-Italian speaking cultures.

[22] Ibid.

2. Collections and Libraries Prior to Digitalization: “Art" of Storing Art Images

In the previous section, following the work of Warburg and Longhi, this paper analyzed some influential methodologies that the art history field uses to perceive and interpret images. When comparing Warburg’s and Longhi’s methods in section 1.4, it is notable that despite distinctive focal points and approaches, both of their works are established on large image sets. As a result, both Longhi and Warburg own a substantial volume of personal collections. Because portable cameras were becoming more accessible at the time[1], the volume of photos, much like what we experience today, had expanded. This chapter will investigate how art images were physically stored and retrieved at another critical moment in art historical analysis, and prior to the digital revolution. With a particular focus on Warburg’s insightful strategies for how to store art images, the thesis argues that Warburg's library is informative for today's digital image database navigation.

2.1 Longhi’s Photo Library: Image Mining

Longhi started his collection of images and books in the 1890s, and by the time he died, his personal library had obtained 30,000 books and nearly 70,000 photographs sorted in 133 folders under Longhi’s own criteria[2]. The photos are mostly indexed under

---

[1] “Library – Fondazione Roberto Longhi.”
[2] Vitale Bloch, “Roberto Longhi” *The Burlington Magazine*. 113.823 (1971): 609–612.

the names of the artists whose works are reproduced in photographs. In Roberto Longhi Foundation's website, a list of artists' name and school is provided under Longhi's criteria. Although there is a paucity of English documentation describing Longhi's library, from the few descriptions provided by the Roberto Longli Foundation, we can learn that Longhi had the habit of writing notes on the back of photos, "supply(ing) information about the work of art they reproduce"[3]. If we think about today's digital art image collections, a similar approach is still employed to provide context for the images—that is, adding textual descriptions.

2.2 Warburg's Library: The Law of Good Neighbor

In his essay *The Law of Good Neighbor*, Steinburg provides a lucid explanation and commentary on the principle that Warburg's library follows:

> "Each book, and each newly acquired book, should stage a conversation with its neighbor: ask a question, provide an answer. The result is a conversation both infinite and coherent: the envy, perhaps, of today's digitized world and its infinite capacity for information but not for coherence."[4]

As Larson argues in his thesis, Warburg's *Mnemosyne Atlas* was also created upon this ideology[5]. Each panel is arranged in a position that creates a conversation visually with its neighbors. In today's digital databases, such conversations based on visual information are very difficult to be constructed. Users begin a search with abstract text

---

[3] Ibid.
[4] Michael P. Steinberg, "THE LAW OF THE GOOD NEIGHBOR" *Common Knowledge*. 18.1 (2012): 128–133.
[5] Larsen, "On the Library as Foundation."

and get a bunch of images based on word matching algorithms. Analogizing the returned pages to Warburg's wooden panel, our search yields a panel that does not provide us with visual linkage; digital images are selected to be placed in the user's panel simply because they have been assigned a textual description or category that matches the user's text-based query. Therefore, extrapolating to the digital image databases, Steinburg's comment—" infinite capacity for information but not for coherence"[6]—is still applicable.

The figure 4 provides an overview of library's reading room, photographed with panels of the so-called Ovid-exhibition of books and images, part of the Mnemosyne atlas project. One of the most evident features of this space is its image-centered organization. In the article *Approaching Aby Warburg and Digital Art History,* Preez emphasizes the fact that when walking into Warburg's library, images will be the first thing that the visitors encounter. It is from the images, visitors will then be redirected to separate book sections that provides context for the visual information they obtained from the images[7]. This image-driven query path is significantly different from the navigation direction of our mainstream digital image databases today. Most of them rely on the text-based search engine that requires users to have a specific text query and

---

[6] Steinberg, "THE LAW OF THE GOOD NEIGHBOR."

[7] Amanda Du Preez, "Approaching Aby Warburg and Digital Art History: Thinking Through Images" in *The Routledge Companion to Digital Humanities and Art History*. , vols. (Routledge, 2020).

returns textually described images to the users. It is essentially a text-to-image path opposite to Warburg library's image-to-text arrangement.

3. In the Wave of Digitization

3.1 Digitization: From Analog to Digital

The advent of digital computers in second half of twentieth century has driven the digital revolution[1]. The physical world began to project itself into the digital world. Due to the discrete structure and abstract format inherited in human language, text has been at the forefront of both digitization and digitalization from the very beginning of the digital revolution. One year after the world's first electronic computer ENIAC was born, inspired by language decryption and computers, Warren Weaver proposed the idea of machine translation in 1947 and discussed with Norbert Wiener, the father of cybernetics, for the feasibility of machine translation[2]. This implies that while other analogs were still experimenting with legitimate digital structures (i.e. digitization process), text had already set an eye on the task of "understanding".

While we are familiar with the digital image medium today, it has undergone a more convoluted development than the digital text. The maturity of the digital image as a medium could not be achieved without the invention of MOS integrated circuits and early microprocessors in the 1960s and 1970s, as well as advances in computer memory

---

[1] Michael Riordan, Lillian Hoddeson, and Conyers Herring, "The Invention of the Transistor" in *More Things in Heaven and Earth: A Celebration of Physics at the Millennium*. Ed. Benjamin Bederson, vols. (New York, NY: Springer, 1999), 563–578, online, Internet, 10 Mar. 2022. , Available: https://doi.org/10.1007/978-1-4612-1512-7_37.

[2] Warren Weaver, "Translation" in *Proceedings of the Conference on Mechanical Translation*. , vols. (presented at the EarlyMT 1952, Massachusetts Institute of Technology, 1952), online, Internet, 10 Mar. 2022. , Available: https://aclanthology.org/1952.earlymt-1.1.

storage, display technology and data compression algorithms. The image formats that we are familiar with today, such as .jpg, .png, .tif, were not invented until the 1990s[3]. The efforts that went into this process should be acknowledged as part of the history of the digital humanities and its development, and why text-based approaches largely preceded image-based research methods.

3.2 Digital Image Databases

Many museums, galleries, and educational institutions have been digitizing their collections since the 1980s[4]. Today, almost all mainstream museums have their own databases and online websites. ARTstor, MET, MoMA, and the British Museum, for example, have reached digital collections with millions of images. Just as Warburg devised library principles under the context of the growth of physical images, digital art image databases need a methodology for data storage and acquisition that can adapt to the big image era, as art repositories are growing rapidly today. This section will evaluate the current storage and retrieval methods of today's mainstream art image databases.

---

[3] "Digital image" *Wikipedia*. , 15 Feb. 2022, online, Internet, 10 Mar. 2022. , Available: https://en.wikipedia.org/w/index.php?title=Digital_image&oldid=1072072337.

[4] Drucker, "Is There a 'Digital' Art History?"

3.2.1 A Case Study of Current Mainstream Database

The digital medium is not new to humanities research. A variety of digital projects relating to topic modeling, metadata design, discourse analysis, have been developing over the past few decades and have provided new perspectives to the humanities field[5]. Nevertheless, from the above example, it is not difficult to observe that most digital projects are established and developed on a text medium basis. In her article “Is There a ‘Digital’ Art History”, Johanna Drucker points out that this phenomenon derives from the straightforward remediation from text source to code when the digital medium first appeared[6]. Similarly, the preference to working with text data can be widely observed in the discipline of art history as well.

Most of the mainstream art image databases, independent of their purposes, choose text-based search as the navigation scheme for people to approach the digital images of art objects. Figure 5 is an example of how ARTstor uses text-based search and metadata filters as main approach to navigate its users. Users’ retrieval procedure of image data can be briefly summarized as follows:

1. Enter the web page.
2. Input keywords in the search bar.
3. Refine the search results using the filters.
4. Pinpoint the target information in the returned search results.

---

[5] Drucker, Johanna. “Is There a ‘Digital’ Art History?” *Visual Resources*, vol. 29, no. 1-2, 2013, pp. 5–13., https://doi.org/10.1080/01973762.2013.761106.
[6] Drucker, “Is There a ‘Digital’ Art History?”

Nevertheless, there are several issues embedded in this text-core search process. First, text-based search is also known as "concept-based" search. As the name indicates, this data retrieval method requires a conceptual agreement on language and interpretation between the database designers and the database users[7]. To continue using ARTstor as an example: if an art history student becomes interested in this recurring visual icon in Egyptian art, using ARTstor's current search model, the student has to translate this visual element into the textual concept before searching the database. In other words, database users must have the equivalent textual translation of other non-textual digital mediums embedded in their knowledge base before conducting further database searches. In this specific case, for those who want to inquire about the visual element in figure 6, they need to know this visual element is called "djed" and input text query "djed" into the search bar. To retrieve this image, it requires the users to have a preconceived conceptual knowledge on the visual element in English.

Apart from requiring a highly compatible knowledge and language structure between database users and database creators, the feasibility of text-based image retrieval relies substantially on the textual descriptions chosen by the database designer for the digital images. To better illustrate this issue, it is necessary to understand the matching schema of text-based image retrieval. As shown in figure 7, text-based image

[7] Rafferty, Pauline, and Rob Hidderley. "Indexing Multimedia and Creative Works."

retrieval is essentially a match between the text query (user input) and the text description corresponding to an image. A conventional evaluation score for text-pairing in the text-based search engine is TF-IDF (term frequency-inverse document frequency), which measures the importance of a keyword to a document in a collection or corpus. Under this matching mechanism, some images, which are visually correlated with the user’s search query, cannot be retrieved due to the target visual element is not covered in the textual description of the image. For instance, in figure 8, although the image visually contains “djed” (marked in red), the user cannot access this image by textually querying "djed," since this visual element was not selected to be described textually (marked in green) when the image was stored into the database. In the process of manually translating visual art images into textual descriptions, those well-known artworks are more likely to be retrieved because they are more thoroughly described in language. Marginal works, on the other hand, are less likely to be discovered since they tend to have less detailed text descriptions (see chapter 5.3.2 for examples). From the issues addressed above, it is reasonable to contend that such a search pattern may solidify the pre-existing knowledge structure and impede the discovery of new knowledge from digital image databases. This means that art images databases reinforce existing biases through a system that otherwise appears to treat image sources objectively.

3.2.2 Cultural Analytics and Distance Reading

As Lev Manovich points out in *Cultural Analytics:*

> "In all cases, the categories, metadata, and tags were originally entered by the archivists who manage the collections. This process imposes particular orders on the data. As a result, when a user accesses an institutional media collection via its website, she can only move along a fixed number of trajectories defined by the taxonomy of the collection and types of metadata."[8]

In order to go beyond the limited trajectories of conventional digital database structure and generate new knowledge out of the pre-existing cannon, Manovich established the cultural analytics lab in 2007. The lab has developed several projects to explore new methods to explore image data. Using computational tools, they remap the image data in the same space. For instance, in their project *A view from Above: Exploratory Visualizations of MoMA Photography Collection*, 18941 photographs are presented on the same graph (figure 9). With quantitative analysis and data visualization, Manovich provides an innovative way to view image data, which allows people to glance at a large volume of images at once and discovers macro patterns or trends. The images in most of Manovich's projects are considered as data points that used to generate quantitative conclusions. In the field of literature, a similar approach was proposed by Franco Moretti back in 2000. He points out in *Conjectures on World Literature* that close readings is "theological exercise—very solemn treatment of very few texts taken very seriously" and considers distance reading as an approach to break the unequal caused by close

[8] "Cultural Analytics | Books Gateway | MIT Press."

reading.[9] Similarly, each image in the Manovich project is read and described by the program in a standardized way. It means that unlike the unequal attention in traditional close reading, the images have equal level of representation in Manovich's projects. While cultural analytics can take image mining beyond the confines of existing canons and provide a platform for "serendipitous" discoveries, however, it sacrifices the individuality of each image while reading at a distance. Such systems allow for zooming in to see individual images, but do not take full advantage of adjacency to convey meaning.

3.3 Discovering Good Neighbor in Digital Space

Based on the above analysis of current art image databases, it is evident that the digital properties of art image databases force the need for fixed rules and structures in the digital space. The rigid metadata and category rules followed by art image databases restrict the users to search only under a path preset by the database curator. The images returned by such text query are independent from each other and are therefore unable to create meaningful conversation. This is very different from Warburg's Law of good neighbor, which emphasizes the arrangement of items[10]. Moreover, in terms of the methodology of art historical research, whether from Warburg's iconographic or

---

[9] Franco Moretti, "Conjectures on World Literature" *New Left Review*. 1 (2000): 54–68.
[10] Preez, "Approaching Aby Warburg and Digital Art History."

Longhi's formalist perspective, text-based image retrieval sets limits on the art historian's excavation of new knowledge through engagement with the database system. Although the museum is now trying to provide good neighbors to art images by recommending images that are manually tagged under the same content or theme, database users are still constrained under database owners' dictatorship. Reflecting on Warburg's storage principle, this means database users still have very little control in choosing "good neighbors" to assist their personalized probe. Taking Metropolitan Museum's website as an example, figure 10 provides an example of how the museum create connections between art images based on these tags. The tags used to associate artworks can be broadly summarized as same artist, same material, same period and same culture, which are defined by the database owners. These choices implicitly confine users within limited trajectories when exploring the database. The process of database curators marking up images determines not only what characteristics of art images get described, but also what relationships between art images are worth-noting. Manovich argues that this process "imposes particular orders on the data"[11].

While the library journey under Warburg's rules is open-ended, the discovery path in the digital image database is progressively contracted. Whether it is a metadata-guided browse or a text-based search, the user must come with a strong sense of

---

[11] Lev Manovich, "Data Science and Digital Art History" *International Journal for Digital Art History*. 1 (2015), online, Internet, 10 Mar. 2022. , Available: https://journals.ub.uni-heidelberg.de/index.php/dah/article/view/21631.

purpose and an adequate amount of preconceived knowledge to form a set of keywords. Within this framework, the user quickly contracts the view and is ultimately directed to a specific image. Except for the additional comparison tools provided by ARTstor that allows users to add ten images into a self-defined gallery for further investigation, almost every search ends up with a single image. Yet, reflecting back to Warburg's Atlas, it is clear that his research did not start with a single image. Rather, it began with a set of images to create a visual context.[12] Therefore, the thesis argues that in addition to textual context, image databases should also provide users with an intuitive visual context and give users the opportunity to self-define the good neighbors based on their own probes.

3.4 Content-based Image Retrieval

To design a more adaptable system for art image databases, it is essential to take the visual features of art images into consideration. This paper proposes to apply content-based image retrieval in art image databases for a more efficient search engine. Instead of using text input as the search query, content-based image retrieval directly takes image data as the input query. This means that users can acquire information relating to a certain image without preconceived conceptual knowledge of it. The core concern in computer vision area is how to transform images back into knowledge—in

[12] Hristova, "Images as Data."

other words, how to make machines “recognize” and “understand” visual information. Content-based image retrieval has been an ongoing and important topic in the field of computer vision. In 1996, the QBIC (Query by Image Content) project at IBM's Almaden Research Center investigated strategies for querying huge online image databases based on the images' content. The interface of the search engine looks as figure 12. Unlike commonly used image retrieval engines today, the QBIC system at that time did not support images directly as input. It required the user to describe the image content in abstraction first, using the standardized tools provided on the interface. As figure 12 shows, the search demo on the left supports query of color and color ratio, and the search demo on the right provides simple graphic tools to support the query of image layout. IBM defines the image content as “color, color layout, texture, shape, size, orientation, and position of image objects and regions”[13]. Compared to the image contents that are extensively discussed in computer vision today, the image contents the QBIC project focused on were mostly images’ inherent properties that are easier to be extracted and quantified. Intrinsic properties refer to the data that are embedded when the digital image is created, which can be obtained without the assistance of complex algorithms. For example, color of each pixels can be easily remapping into numbers

[13] “IBM Research,” REPLACE, online, Internet, 1 Mar. 2022. , Available: https://dominoweb.draco.res.ibm.com/dominoweb.draco.res.ibm.com/a3ea9019a4ed364985256593006fd727.html.

follow RGB rule or HSV rule because it is how pixels are initially stored. Similarly, shape and size are also inherent properties.

With the development of the field of computer vision, more powerful and convenient content-based search engines, such as Google Image search and TinEye, have emerged. Compared with the visual descriptors used in IBM's QBIC system, these two engines both allow users to directly input images as search queries without textual keywords. Although the specific feature extraction methodologies of these two engines are not open-sourced, from the returned outputs, we can observe that the features extracted by today's content-based image search engines are more sophisticated and returns more accurate results. Consider figure 13 and 14, which present the returns of Google Image and TinEye when querying an art image. Their search results all returned the identical images as the search image and relevant information about the image, such as the name of the work, the artist, and the associated URLs. Yet, while such search engines give considerable attention to the visual information of an image and do not require preconceived knowledge to complete an image retrieval, they can hardly be used to excavate connections between art images. In the context of art history, locating an image that is identical to the query image does not provide new insights for research. For example, if an art historian wants to study the connection between the works of Van Gogh and Hiroshige Kogawa, what he/she needs is that, when querying a Van Gogh's digital art image, computational tools can algorithmically create features capturing the visual features of the image query and return art images sharing visual commonalities from Hiroshige Kogawa's digital art image sets. As Manovich points out, if the current interface navigation method does not change, computational tools will "not allow serendipitous explorations of media collections or making research links."[14]

[14] "Cultural Analytics | Books Gateway | MIT Press," n.d., online, Internet, 22 Mar. 2022. , Available: https://direct-mit-edu.proxy.lib.dukeDrucker.edu/books/book/4966/Cultural-Analytics.

4. Computer Vision & Digital Image Space Redesign

4.1 Technological Background: A Brief History of Computer Vision

In the past decade, the topic of computer vision has garnered an unprecedented level of interest from researchers throughout the world. According to CVPR (the Top Computer Vision and Pattern Recognition Conference), 1600 papers, selected from more than 7000 submissions, are presented in twelve sessions last year[1]. Nevertheless, while some may consider computer vision as an emerging discipline, the quest to build vision systems for machines was carried out in various fields as early as the 1960s. In the field of mathematics, Ulf Grenander published the monograph on *Probabilities on Algebraic Structures* in 1963[2], which laid an essential mathematical foundation for computer vision. His following work, integrated algebra, groups and probability theory, provides a unified representation of patterns found in both the natural and man-made worlds[3]. Approaching the topic from a neurobiological perspective, David Marr first proposed the notion of computational vision at MIT's AI Lab in 1976[4]. In his subsequent book *Vision*, he "integrates insights and data from neuroscience, psychology, and

---

[1] '"CVPR 2021 Conference Report | IEEE Computer Society," n.d., online, Internet, 27 Feb. 2022. , Available: https://www.computer.org/publications/tech-news/events/cvpr-2021-recap/.
[2] Nitis Mukhopadhyay, "A Conversation with Ulf Grenander" *Statistical Science*. 21.3 (2006), online, Internet, 28 Feb. 2022. , Available: http://arxiv.org/abs/math/0701092.
[3] Stuart Geman, "Ulf Grenander, 1923–2016" *Journal of the Royal Statistical Society. Series A (Statistics in Society)*. 179.4 (2016): 1122–1124.
[4] D. Marr and T. Poggio, "From Understanding Computation to Understanding Neural Circuitry" (1976), online, Internet, 1 Mar. 2022. , Available: https://dspace.mit.edu/handle/1721.1/5782.

computation"[5], which became an essential theoretical framework for computer vision in the 1980s. Marr provided systematic framework and prospective direction to the development of the field. Many of his ideas have been corroborated in the succeeding research and are still valuable for reference to this day.

For the general public today, computer vision is no more an abstract term, but rather a wide spectrum of applications that pervade almost every aspect of daily life. If one looks back, however, the large-scale application of computer vision technologies came much later than scholars expected in the 60s, and the path was more convoluted. Although the theoretical exploration in this area started early, it had not been widely applied prior to the twenty first century. This was largely due to the lack of hardware and digital image data basis at that time. Songchun Zhu, the general chair of CVPR 2012 and 2019, recalls in an interview that when he was working on his doctoral dissertation in the 90s, his advisor provided him with the access of SUN, the best workstation of the Harvard Robotics Lab at the time with a memory of 32 megabytes. Their lab also spent $250,000 to build an image acquisition system because there were no digital cameras at the time[6]. This recollection of the lab experience at the time indicates that the limitation in hardware conditions and data acquisition approach. It was not until the mass

---

[5] David Marr, *Vision: A Computational Investigation into the Human Representation and Processing of Visual Information*, vols. (Cambridge, MA, USA: MIT Press, 2010).
[6] Song-Chun Zhu, "A Discussion on the Origin of Computer Vision and Artificial Intelligence," The Vision Seeker, Oct. 2016.

production of digital cameras in late 90s, that the data basis for the application of computer vision established. The spread of digital cameras also brought changes to the field of art history, as will be discussed in later sections.

As stated above, the lack of credible image datasets, and lagging computing power, together bottlenecked the development of computer vision. To explain computer vision's take-off and its success in large-scale application in the twenty first century, it is inevitable to mention the emergence of GPU and the famous dataset ImageNet. NVIDIA first introduced the notion of GPUs in 1999 with the release of its signature product, the GeForce256[7]. Prior to that, most personal computers used CPUs for graphics and image operations. From 1994 onwards, chips for image processing gradually appeared. Nvidia's GeForce256 standardized the various image acceleration chips that emerged during this period, making GPUs officially popular. The emergence of GPU, while greatly reducing the pressure on the CPU, also gave greater computability to the image. Digital images gradually expanded from primitive functions such as storage, image enhancement, and restoration to the level of analyzability.

Feifei Li, the initiator of ImageNet, addresses the significance of data resources in computer vision as an increasing amount of data became available since the dawn of

[7] "NVIDIA LAUNCHES THE WORLD'S FIRST GRAPHICS PROCESSING UNIT: GEFORCE 256" *ResponseSource Press Release Wire*. , n.d., online, Internet, 4 Mar. 2022. , Available: https://pressreleases.responsesource.com/news/3992/nvidia-launches-the-world-s-first-graphics-processing-unit-geforce-256/.

digital age and invention of the internet[8]. Her belief in the potency of large-scale image data prompted her to launch the creation of ImageNet in 2007[9]. ImageNet's massive sorted and labeled image data and GPU's high computing power provide the conditions for neural network models, such as CNN, to flourish. It is still widely used as benchmark in the current researches in the field.

4.2 Prototyping

As previously noted, the standpoint of this thesis is not to criticize or discard the existing metadata and text-based search. Considering the exponentially growing data volume, Warburg's library concept is difficult and inappropriate to be fully reconfigured in the digital image space. Metadata and text-based search have both evolved over a long period of time, following their own logic to help people manage data and navigate digital space in this big data era. Therefore, this prototype proposes to add an image-driven search path on top of the original database structure to create a more diverse search environment.

So now the question arises, how will computer vision help to create a new search path that enables database users to briefly step outside of the established metadata framework, and as Warburg suggests, "formulat[e] a new problem"[10]? This section will

---

[8] Fei-Fei Li, *How we're teaching computers to understand pictures*, vols., 1427123194, online, Internet, 25 Feb. 2022. , Available: https://www.ted.com/talks/fei_fei_li_how_we_re_teaching_computers_to_understand_pictures.
[9] "ImageNet," n.d., online, Internet, 25 Feb. 2022. , Available: https://www.image-net.org/index.php.
[10] Preez, "Approaching Aby Warburg and Digital Art History."

focus on how, given an analysis of current capabilities and boundaries of computer vision, art historical research will leverage technology to navigate the exponentially growing image data space in order to gain new insights upon the existed metadata framework and digital classification systems. In this section, two types of computer vision algorithms based on convolutional neural networks will be analyzed. The digital project will incorporate these algorithms based on different features they can generate when analyzing art image data computationally.

4.2.1 Holistic Level Features

The conventional visual features in this type of measurement include, but are not limited to, color, general composition, luminating effect and patterns. Since these features are more representable in compiler language, they can be relatively easier to extract or quantify. Meanwhile, this set of information is somewhat inherited in the digital format, so that this level of computing could be done even without high demand of computing power and large-scale dataset. As a result, the machine vision exploration at this level started earlier than other more complicated tasks. Extrapolating back to Warburg's image study paradigm, this technique can be analogized to the process of Warburg pulling visually similar images on the wooden panel to create a general vision before digging further into the details[11].

---

[11] Hristova, "Images as Data."

4.2.2 Unsupervised Visual Feature Mining

The second research direction focuses on clustering tasks. The difference between this approach and the previous approach is that it is an unsupervised learning task and does not require a pre-defined classification framework to annotate data. It is a mechanism that maps the features of an image to a higher mathematical space and then centralizes the data points around multiple centroids to generate clusters. The most classical algorithms to fulfill this task are KNN, K-means. These traditional unsupervised clustering algorithms are now also incorporated with convolutional neural networks nowadays. DeepCluster, an unsupervised visual feature discovery model, is a good representative here. As shown in figure 15, such model “iteratively clusters deep features extracted from unlabeled row image data and use the cluster assignments as pseudo-labels to learn the parameters of the convnet”[12].

This mechanism is highly compatible with Longhi's methodology: after scanning a large number of images and clustering visual features in his mind, Longhi then used self-defined, innovative vocabularies to define each visual feature group, which perceived as “style” under art history context. This unsupervised approach is not restricted to an existed textual classification framework and echoes Longhi's formalist perspective.

---

[12] Mathilde Caron et al., “Deep Clustering for Unsupervised Learning of Visual Features” *arXiv:1807.05520 [cs]*. (2019), online, Internet, 3 Mar. 2022. , Available: http://arxiv.org/abs/1807.05520.

5. Digital Implementation Details

This is a chapter that documents the digital component of this thesis. The first subsection of the chapter defines the goal of this project and how does it attribute to the previous chapters. This section also documents the thinking process during the design and implementation of this project, showing how the solution was conceptualized. The second subsection explain the technical selection provides the details of both front-end and back-end implementation.

The design of this digital project is guided by the art historical research paradigm discussed in the previous section, and is inspired by the "Law of Good Neighbor"[1] applied by Warburg in his personal library and his atlases. It is hoped that this database search method will better utilize the computability of the digital image and facilitate the concept of “search for researching"[2]. With the continuous creation of art repositories in this digital era, art history field is confronted with an unprecedented amount of data. Through Warburg's library theory, it is essential to be more aware of the importance of the way we navigate the expanding digital image space and use newly generated computer vision technologies to find "good neighbors" for art image data retrieval in the age of big images.

[1] Steinberg, “THE LAW OF THE GOOD NEIGHBOR.”
[2] Preez, “Approaching Aby Warburg and Digital Art History.”

5.1 Dataset and Framework

One part of the dataset used in this demo project is adopted from Kaggle, named "Chinese Fine Art—Computer Vision: Famous Chinese Paintings"[3]. This dataset contains 395 artworks from 30 Chinese artists, ranging over 200 years. It is provided by Ricky J Li and has a license under CC0: Public Domain. The data is of .jpg format and does not include any labels or segmentations. In addition to the image data, each artist's name, ancestry, gender, and date of birth/death are also included in the dataset. The other image datasets used in the project are Utagawa Hiroshige's artworks in nineteenth century, Utagawa Kuniyoshi's artworks in nineteenth century, Katsushika Hokusai's artworks in nineteenth century, Van Gogh's artwork in nineteenth century and Hongshow Chen's artworks in Late Ming dynasty and etc. The images are downloaded or crawled from the MET's public collection or internet's public domain. Digital images of these artists' artworks are selected in order to test if the digital project can discover visual links between both inter-cultural and inner-cultural artworks.

As figure 11 shows, the digital project framework consists of two main modules. The first one deals with user interface implementation and the second one deals with machine learning tasks requested by the users at the front-end. For the back-end implementation, the project chooses Python as the major programming language as it is

[3] "Chinese Fine Art," n.d., online, Internet, 10 Mar. 2022. , Available: https://kaggle.com/rickyjli/chinese-fine-art.

widely used in the field of machine learning. For the front-end implementation, the project uses JavaScript as programming language and traditional HTML, CSS combo to do visual design. The front-end and back-end is connect using Flask, a micro web framework written in Python.

5.2 User-Interface Design

As previously mentioned, this digital project intends to explore the question of how to enhance the search experience of art image databases from an art historical perspective. Referring back to Warburg's Atlas, the interface design hope is to create a visual space that enables the database users to conduct distance reading on image sets. Therefore, the user interface is designed to have multiple search panels, having similar function as the wooden panel Warburg used for his Atlas. On each search panel, with the assistant of existing database filters, the users are able to self-define "good neighbors" that are pertinent to their research interest.

For example, suppose a researcher or art history student wants to probe the impact of Japan's reopening in the mid-19th century on the Western artist. He/she can use the filters to narrow down to a period of time or a specific artist in originally classified in two very different metadata categories and juxtapose their work to develop an overview (distance reading of images). Figure 16 provides a demo of how the interface looks like if the researcher narrows the search to Van Gogh and Utagawa Hiroshige. (The two artists are selected as examples here because for someone who is interested in Japonisme, he/she might be already aware of the influence of Hiroshige on Van Gogh's work.)

5.3 Model Selection, Algorithmic Logic and Case Analysis

This section will discuss the model selection in the implementation. At the same time, the article intends to reveal how the algorithmic logic behind the models, parameters, data preprocessing and etc. will affect the credibility of the result under an art historical perspective. As Murray points out in the book *Inventing the Medium*, due to today's pervasiveness of digital artifacts, how people think, act, communicate and understand the world is reshaped by these digital artifacts and the design decisions that go into their creation[4]. Therefore, when using computing tools to solve field-specific problems, it is vital to not only concentrate on the outcomes delivered by the technology, but also trace the design philosophy of the digital tools.

5.3.1 Feature Extraction Model

This implementation is corresponding to the prototype of section 4.2.1. In this experiment, the project uses VGG model to extract image features. VGG is convolutional network structure for unsupervised feature extraction[5]. For this project, the VGG 16 is

---

[4] Janet H. Murray, *Inventing the Medium: Principles of Interaction Design As a Cultural Practice*, vols. (Cambridge, UNITED STATES: MIT Press, 2011), online, Internet, 28 Feb. 2022. , Available: http://ebookcentral.proquest.com/lib/duke/detail.action?docID=3339350.
[5] Karen Simonyan and Andrew Zisserman, "Very Deep Convolutional Networks for Large-Scale Image Recognition" *arXiv:1409.1556 [cs]*. (2015), online, Internet, 9 Mar. 2022. , Available: http://arxiv.org/abs/1409.1556.

imported from Keras library[6]. As presented in figure 17, the VGG 16 model consists of sixteen convoluted layers and takes digital image with a 224-pixel * 224-pixel size and 3 color dimensions as input. Since VGG16's initial design includes three fully connected layers, the image has to be resized to the same pixel size. According to the VGG's corresponding paper published at ICLR 2015, the image resizing strategy used on training images is randomly cropped out a 224*224 part from the rescaled training image set. Nevertheless, the resizing of images is essentially a compromise of information. Different resizing strategies introduces different presumptions into out result. Reflecting on this project, resizing sacrificed the visual information embedded in the portions of the images that were cropped off in order to achieve the 224*224 ratio. In addition, for tasks that focusing on a small motif or figure in a high-resolution image (such as images of Haltadefinizione[7]), resizing might be a factor to be carefully considered for potential impact. As figure 18 shows, the features generated by the algorithm is stored in NumPy array with a size of 4096, which is same as the number of channels in the fully connected layer of VGG-16.

Due to the emulative kernel of the neural network, it is less interpretable than traditional symbolic machine learning models. To show its functionality more

---

[6] Keras Team, "Keras documentation: VGG16 and VGG19," n.d., online, Internet, 10 Mar. 2022. , Available: https://keras.io/api/applications/vgg/.

[7] "About us | Haltadefinizione," n.d., online, Internet, 10 Mar. 2022. , Available: https://www.haltadefinizione.com/en/about-us/.

intuitively, a simplified dataset is created to test VGG-16's ability to capture visual features. Figures 19 and 20 documents VGG-16's visual feature extraction schema on a simplified test dataset. The result indicates that using the deep features extracted by VGG-16, images with similar objects, components, or textures can be retrieved.

In addition, figures 21-23 attempt to reveal the black box in which the neural network extracts features. Through the sample figures, we can observe from the visualized feature maps that the perceptual fields of the convolutional network differ from layer to layer. As the number of layers of the network increases, the size of the features that the model focuses on gradually expands as well. This architecture allows image features to be captured more comprehensively.

5.3.2 Application of Neural Network Models in Museums: Supervised or Unsupervised?

As mentioned earlier, the image's visual content is underrepresented in digital art databases. Computer vison models based on neural networks have also recently been applied by the Metropolitan museums to alleviate this issue. The museum uses computer vision algorithms to tag its art collections based on their contents. Elena Villaespesa and Seth Crider have carefully evaluated this method based on the tags generated by three well-known computer vision systems—Google Cloud Vision, Amazon Rekognition and IBM Watson[8]. This approach is essentially an object detection task, which classifies pixels to a set of predefined categories, such as *Person, Bronze, Clothing, Painting and etc.*[9]. As figure 24 and 25 shows, this experimental project demonstrates that supervised learning can provide a standardized mechanism for describing image content with predefined taxonomy (i.e. classification labels), enriching the visual features of image data. With the support of algorithms that can quickly add content-related features to a large number of images, marginalized artworks in the conventional canon can receive the same level of description as the mainstream images. Therefore, a relatively unbiased exploration space is created.

Yet, this supervised learning requires a decision on which categories will be used as training labels. This choice determines what content in the image is worth keeping and discards the rest of the features. Then, who gets to decide which image contents are valuable to be emphasized? Usually it is the database owners or the model creators. The database users, on the other hand, are not involved in the selection process. In addition, most of the categories in these three common systems are designed to describe realistic photographs. As shown in figures 24 and 25, such tags do not seem to fit well with art historical terminology. In the process of associating visual features generated by neural networks with textual labels, we then fall back into the highly generalized, language-

---

[8] Elena Villaespesa and Seth Crider, "A critical comparison analysis between human and machine-generated tags for the Metropolitan Museum of Art's collection" *Journal of Documentation*. 77.4 (2021): 946–964.
[9] Ibid.

dependent, established paradigm of metadata, which is a waste of extracted features. In order to push beyond these limits, this thesis's digital component does not use such supervised classification after feature extraction. Instead, the digital project retains complete vectors outputted by the VGG-16 to keep a more comprehensive description of images' visual features.

5.3.3 Result Analysis

This section will present the features of this digital project and instructions for how it can be used to mine visual connections between art images in the digital space. The first step is to change the directory to the Art_CBIR folder, in which the digital project is implemented. The project is implemented on the local server at this stage, using Visual Studio Code as IDE. The system of the device is macOS Monterey. A problem encountered during the implementation is that system default python version 3.10 does not support installation of PyTorch, an open source machine learning framework. Therefore, since PyTorch will be used in the back-end machine learning tasks, the project is built under the virtual environment with the python 3.8. The virtual environment can provide an isolated space for the project's dependencies so that there will not be contradictions with another project's implementation. Figure 26 is the terminal log of how to setup a virtual environment and the specific dependencies of this project.

The final project is presented as a website. The main feature of this project is to support the users to execute visual-driven interrogation between customized image sets. As figure 27 shows, an art image will be the first thing that the users encountered when entering the digital space, and a new image will be arbitrarily extracted from the image database when being clicked, providing all the images in the image database with the same possibility to be seen at some point. Then, before moving on to visual linkage mining, the system's user can customize the number of panels they required for visual inquiry and the scope of artworks corresponding to each panel (figure 28 and 29). After the users submit a list of the parameters, like artists' name, time range and motifs included, the site will return the image sets that they think will echo with each other and provide insights for their questions (figure 30).

Following the previous multi-panel search example, let's assume that the user is familiar with the well-known fact that Vincent van Gogh copied Hiroshige's *The Plum Garden at Kameido Shrine* in one of his oil painting dated 1877[10] and *Sudden Shower over Shin-Ōhashi bridge and Atake* in 1887[11]. It is reasonable to assume that with such prior knowledge, the user might yield questions like "Did Hiroshige Kogawa exert an

---

[10] Artstor, "The Plum Garden at Kameido Shrine," Artwork, n.d., online, Internet, 10 Mar. 2022. , Available: https://library.artstor.org/public/SS35559_35559_34101454.

[11] "Utagawa Hiroshige | Sudden Shower over Shin-Ōhashi Bridge and Atake (Ōhashi Atake no yūdachi), from the series One Hundred Famous Views of Edo (Meisho Edo hyakkei) | Japan | Edo period (1615–1868)" *The Metropolitan Museum of Art*. , n.d., online, Internet, 10 Mar. 2022. , Available: https://www.metmuseum.org/art/collection/search/55433.

influence on Van Gogh's other works from this time period? If so, how? " Assume the user selected Van Gogh's oil painting *The Harvest*, which created in 1888, as his/her image query, this digital product will return Hiroshige's woodblock print *Shono* and several other works from the series *狂歌入东海道* (figure 32). This system calculates the distance between the feature vector of the query image and the feature vector of the images in the corresponding panels, and returns ten images with the closest distances to the query image features.

Commonly used formulas for calculating vector distances are: Euclidean distance, Cosine distance, Minkowski distance, Hamming distance and etc. The digital project uses the first two functions to measure the distance between image features. The function of Euclidean distance and Cosine distance and their python code are shown in figure 31. The Euclidean distance can be interpreted as the length of a line segment connecting two points. The Euclidean distance formula is straightforward and uses the Pythagorean theorem to calculate the distance from the Cartesian coordinates of these points. The Cosine distance, on the other hand, measures the similarity of two vector's direction. The cosine similarity of two vectors with exactly the same direction is 1, while the cosine similarity of two vectors opposite each other is - 1. (Note that their magnitudes do not matter, because this is a measure in the direction.) In this digital project, since the vectors are normalized before distance measurement, that the returned images under the two matrixes are the same. The returned ranking of the two matrixes can be different, however, when the vectors are not normalized. This situation can happen when the two sets of image vectors are generated by distinctive neural network models with different vector dimensions.

Shono can also be found in MET's public collection as figure 33 shows. Compared to Utagawa's other works that have received more attention in the Western world, such as *Sudden Shower over Shin-Ōhashi bridge and Atake*, there is very little context for this work in the digital collection. It has no descriptive text or relevant literature. It echoes with the issues discussed in chapter 3.1.1 that the text-based search relies on the manually added text descriptions of art images can makes it harder for marginal works to be discovered since they tend to have less detailed text descriptions.

After getting the result, it is the art historian's job to do further analysis on the returned result. In the given sample demo, we can observe farmers, "a half mowed wheatfield, ladders and several carts. A reaper works in the background"[12]. Similar themes can be observed in *Shono* as well; Hiroshige also depicted the peasant life in addition to some tools that farmers uses. *The Harvest* and *Shono* are both depicting peasant life and work on the land. From a formalist perspective, there are parallels that can be observed between the two art images as well. For instance, the space is both flattened in the two artworks.

[12] "The Harvest Vincent van Gogh, 1888" *Van Gogh Museum*. , n.d., online, Internet, 10 Mar. 2022. , Available: https://www.vangoghmuseum.nl/en/collection/s0030V1962.

The previous example illustrates how deep feature extraction can assist in mining visual linkages between inter-cultural artworks that are not linked together under the metadata structure or textual descriptions. The visual connections discovered by deep neural networks can also facilitate visual comparison between intra-cultural artworks. Taking Hokusai and Hiroshige as examples. Both artists are prominent Japanese painters of the Ukiyo-e period, and their paintings share many similarities in both themes and techniques. When searching for their works in MET or ARTstor, however, it is difficult for users to observe an intuitive visual connection between their works. With deep features extracted by VGG-16 however, the users are able to quickly obtain the artworks that share similar visual features between the two artists.

In distinction from the supervised object detection method implemented by Metropolitan museum, this digital project attempts to avoid additional supervised learning that involves labeling and human intervention as much as possible. The project encompasses as many visual features as possible, aiming to minimize the dictatorship of database owners in data presentation and to provide more resilient and comprehensive image content for different entering points. Although the project has preserved the complexity of the image content in the mathematical space as intact as possible, there is still a brutal simplification when it comes to the evaluation of the images' relevance. In art historical research, it is clear that the relevance of images is a flexible indicator. It depends on the research topic or on the divergent ways people read the images. Yet in computational analysis, it is difficult to completely avoid such simplistic quantification because computer models usually require an explicit metric to function. Therefore, it is important to keep in mind that a critical human eye is ultimately required to steer a meaningful computational analysis[13].

[13] Claire Bishop, "Against Digital Art History" *International Journal for Digital Art History*. 3 (2018), online, Internet, 13 Apr. 2022. , Available: https://journals.ub.uni-heidelberg.de/index.php/dah/article/view/49915.

6. Conclusion

This paper addresses the limitations of the current text and metadata-driven navigation approach of art image databases. It argues that when adding textual descriptions to images, decisions on which images should be textually described and which visual features in the images should be emphasized are highly subjective and biased. These system-level decisions influence the accessibility of any given art image. This paper argues that computer vision algorithms can create a more inclusive and equitable digital space for art image databases. A deep neural network is capable of contributing content-related features to digital images systematically, so that each image will receive same level of representation, regardless of the amount of attention it receives from mainstream research.

Another important problem is how to use these algorithm-generated features to facilitate the discovery of new knowledge in art image databases. Drucker argues that while digitization has driven the construction of online art repositories that make art historical corpus more available and accessible, methods for computational analysis have been sluggish to emerge[1]. Yet, algorithms alone do not facilitate the discovery of new knowledge. Reverse image search engines like Google Image and TinEye, for example, also employ computer vision algorithms to extract visual features and support

[1] Miriam Kienle, "Digital Art History 'Beyond the Digitized Slide Library': An Interview with Johanna Drucker and Miriam Posner" 6.3 (2017): 6.

content-based image retrieval, but they will "not allow serendipitous explorations of media collections or making research links."[2] As Manovich suggests, the current interface navigation method is not adequate for forming new problems and uncovering new relationships[3]. Referring back to the research methods of Aby Warburg and Roberto Longhi, this paper proposes an image-driven multi-panel search interface, aiming to assist the formation of research links while browsing an art image database. The digital project attached with this thesis uses VGG-16 model to extract visual features from art image data and to construct a feature database. In conjunction with the database's original metadata framework, users are then able to initialize different image panels according to their individualized research interests, and to use the feature database to mine the visual relationships between the image sets contained in these panels. This approach of overlaying a visual-driven search path upon the existing metadata structure allows art historians to discover the undefined visual relationships between art images, and beyond the database's confined, text-based search trajectories. Meanwhile, preserving the metadata structure enhances the efficiency and precision of the exploration, so that researchers will not get lost in the massive amount of data.

This digital project, however, has undeniable limitations. Compared to manually annotating and characterizing image data using texts, the process of extracting image

---

[2] "Cultural Analytics | Books Gateway | MIT Press," n.d., online, Internet, 22 Mar. 2022. , Available: https://direct-mit-edu.proxy.lib.dukeDrucker.edu/books/book/4966/Cultural-Analytics.
[3] Manovich, "Data Science and Digital Art History."

data's visual features by neural networks is more systematic, efficient, and impartial. Compared to traditional machine learning models, neural networks are more autonomous and require less human intervention in feature extraction, which leads to less human bias. This is because traditional models mostly fall under symbolicism, whereas neural networks fall under connectionism. The former relies on mathematical logic and statistics, and the latter is grounded on neurocognitive science and bionics. Nevertheless, most of the extensively used convolutional neural network models (including the VGG-16 model used in the digital project of this thesis) are pre-trained on ImageNet. Therefore, the performance of neural networks is strongly influenced by the data, meaning that a bias is imposed on the model with the training data itself. In addition, traditional machine learning models, such as decision trees, SVM, Bayesian, etc., have better interpretability than neural network models.

While these limitations remain unresolved, they should not be a reason to forgo computational analysis tools. This thesis holds the standpoint that each method has its own limitations and what truly matters is that when applying computational tools, their limitations and underlying assumptions are revealed and communicated to the users.

Appendix: Illustrations

Figure 1: Panel 39 of Bilderatlas Mnemosyne (1929 Final Version).[1]

[1] "Bilderatlas Mnemosyne | Final version" *The Warburg Institute.* , n.d., online, Internet, 8 Mar. 2022. , Available: https://warburg.sas.ac.uk/archive/bilderatlas-mnemosyne/final-version.

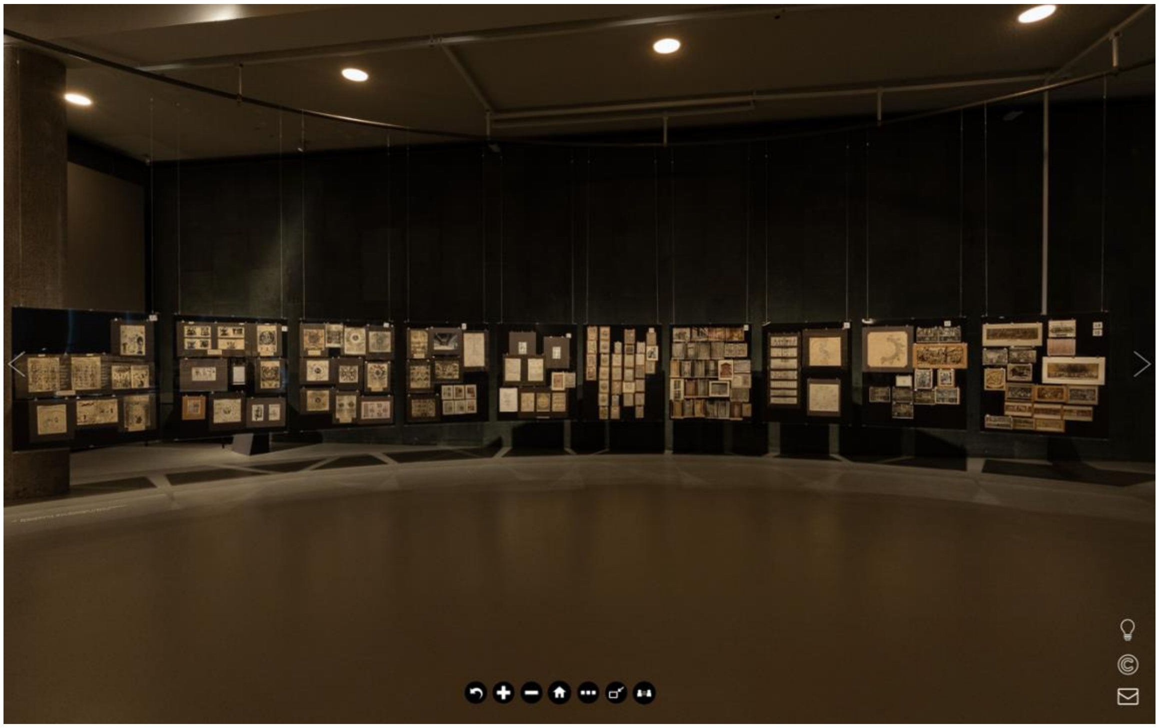


Figure 2: Virtual Tour of Bilderatlas Mnemosyne exhibition at Haus der Kulturen der Welt.[2]

[2] "Virtual Tour - Aby Warburg: Bilderatlas Mnemosyne exhibition at Haus der Kulturen der Welt | The Warburg Institute," n.d., online, Internet, 13 Apr. 2022. , Available: https://warburg.sas.ac.uk/virtual-tour-aby-warburg-bilderatlas-mnemosyne-exhibition-haus-der-kulturen-der-welt.

Figure 3: The four images represent for four different styles (Pittura) under Roberto Longhi's definition.[3]

[3] "An Introduction to the Use of Images in Art Historical Research: Context and Prospects."

Figure 4: Reading Room of the Kulturwissenschaftliche Bibliothek Warburg, February 1927, Warburg Institute Archive, London.[4]

[4] Aby Warburg, Christopher Johnson, and Claudia Wedepohl, "From the Arsenal to the Laboratory" *West 86th: A Journal of Decorative Arts, Design History, and Material Culture*. 19 (2012): 106–124.

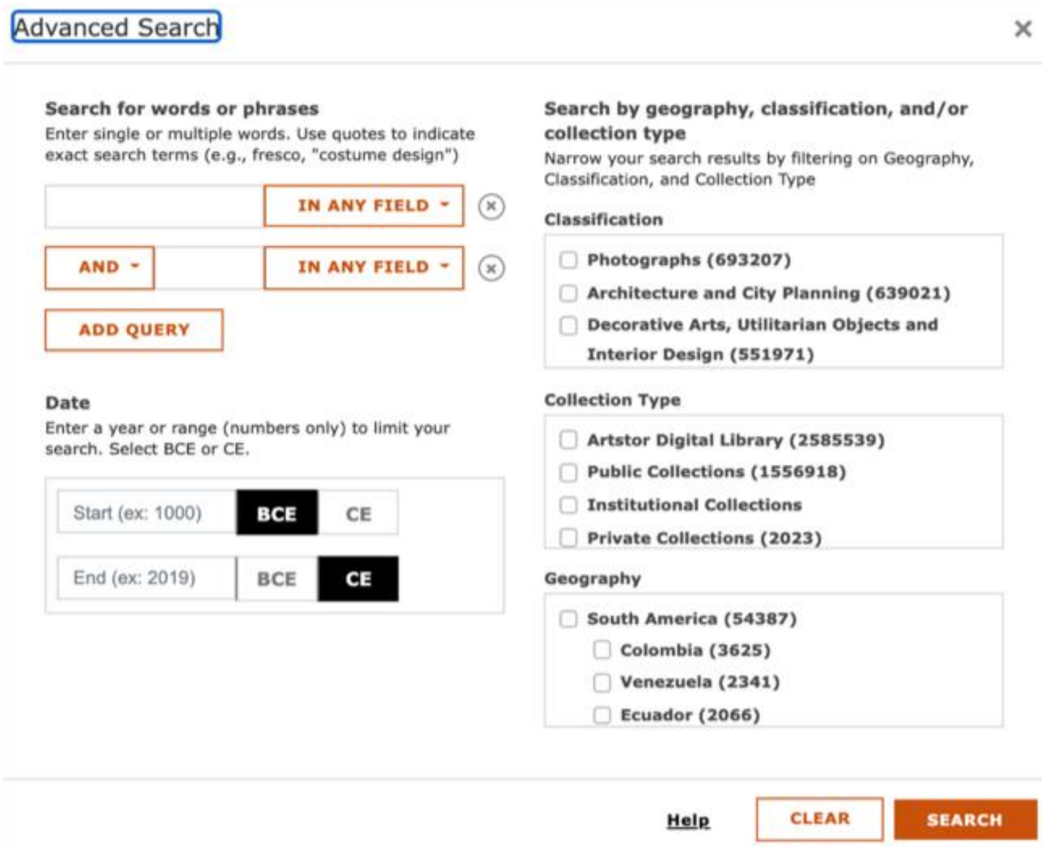


Figure 5: The advanced search interface of ARTstor website.[5]

[5] ARTstor, https://library.artstor.org/#/home

Figure 6: An Egyptian djed pillar made by the blue glass from Metropolitan Museum's collection.[6]

[6] Djed pillar. 525-30 BC. Artstor, library.artstor.org/asset/SS7731421_7731421_11511239

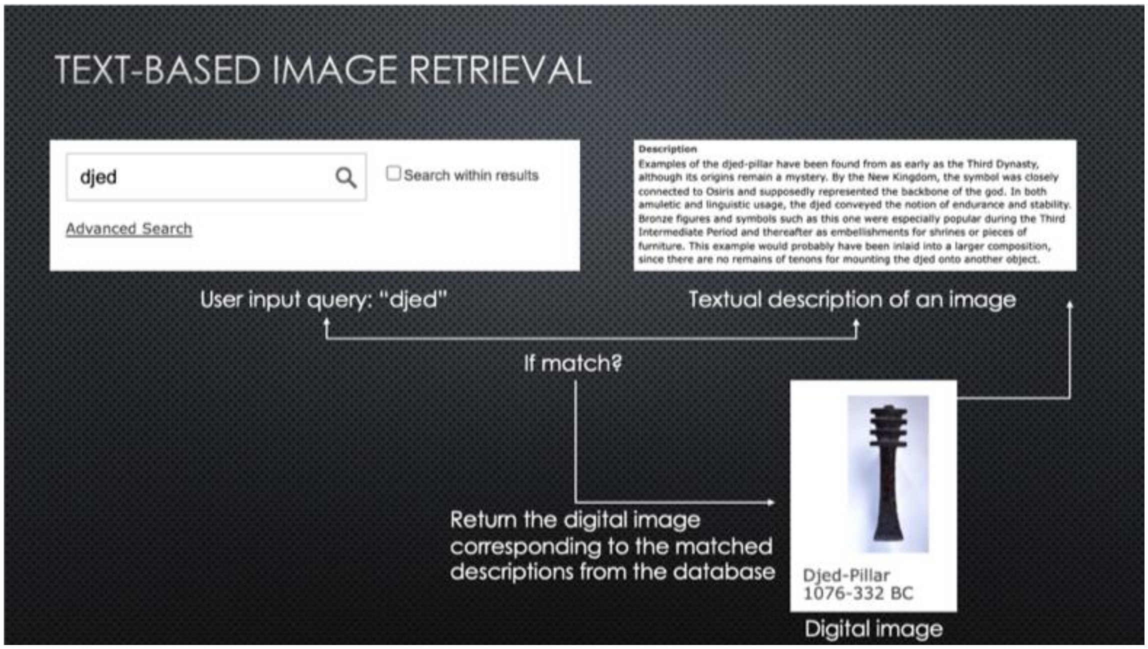


Figure 7: A brief illustration of text-based image retrieval's matching schema.[7]

[7] Djed-Pillar. 1076-332 BC. Artstor, library.artstor.org/asset/SS35538_35538_29884201

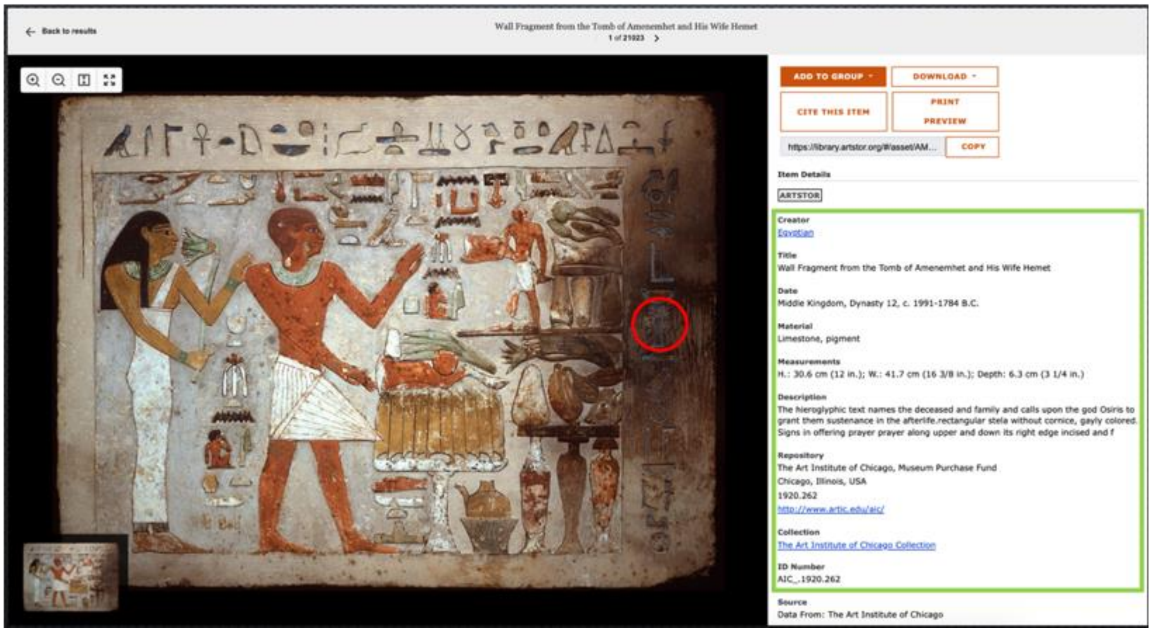


Figure 8: Digital image of the "Wall Fragment from the Tomb of Amenemhet and His Wife Hemet" and its corresponding textual description on ARTstor website.[8]

[8] Egyptian. Wall Fragment from the Tomb of Amenemhet and His Wife Hemet. Middle Kingdom, Dynasty 12, c. 1991-1784 B.C.. Artstor, library.artstor.org/asset/AMICO_CHICAGO_1031149800

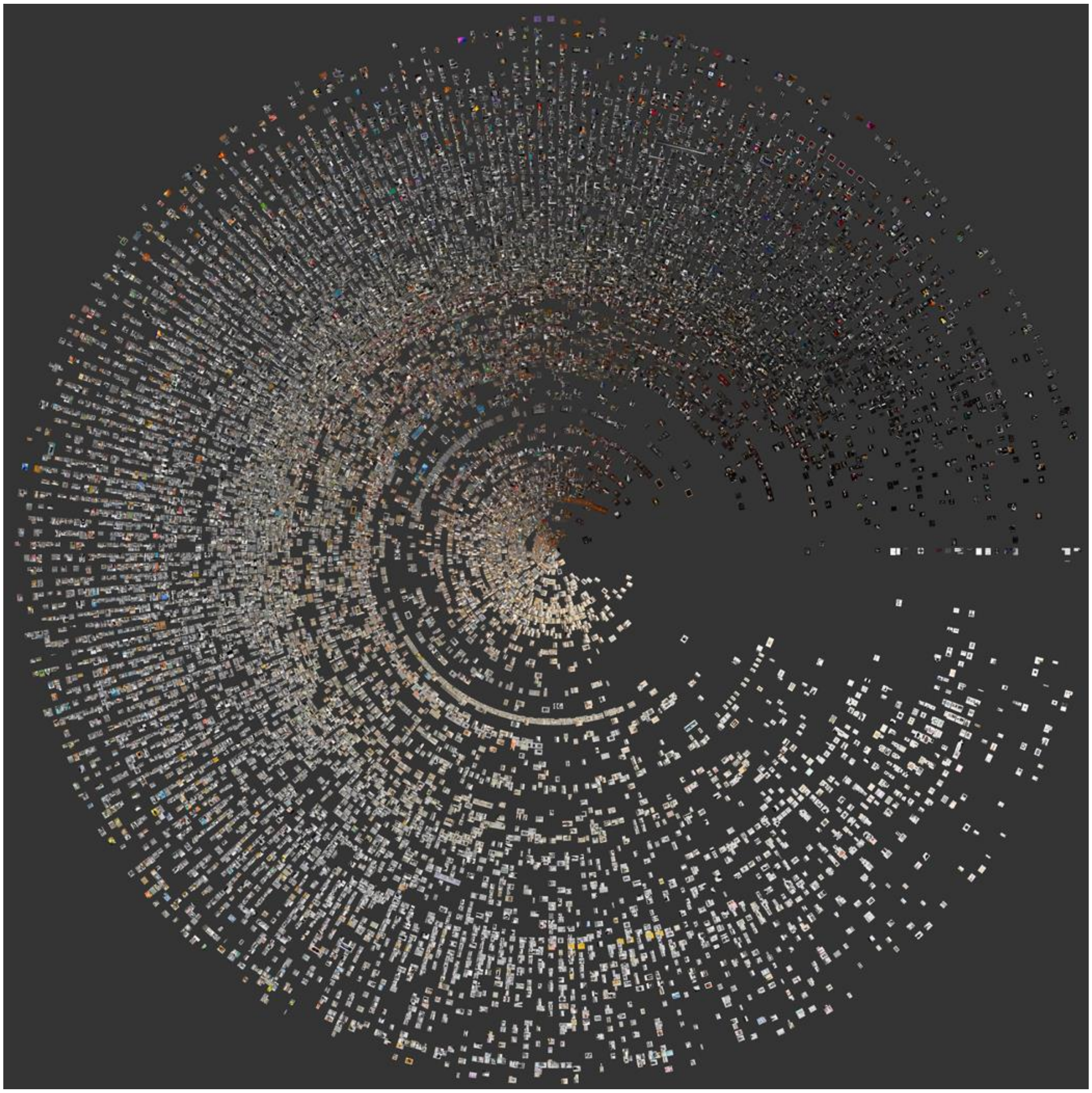


Figure 9: Radius visualization of 18941 photographs from MoMA.[9]

[9] "A View from Above: Exploratory Visualizations of MoMA Photography Collection," n.d., online, Internet, 13 Apr. 2022. , Available: http://lab.culturalanalytics.info/2016/04/exploratory-visualizations-of-thomas.html.

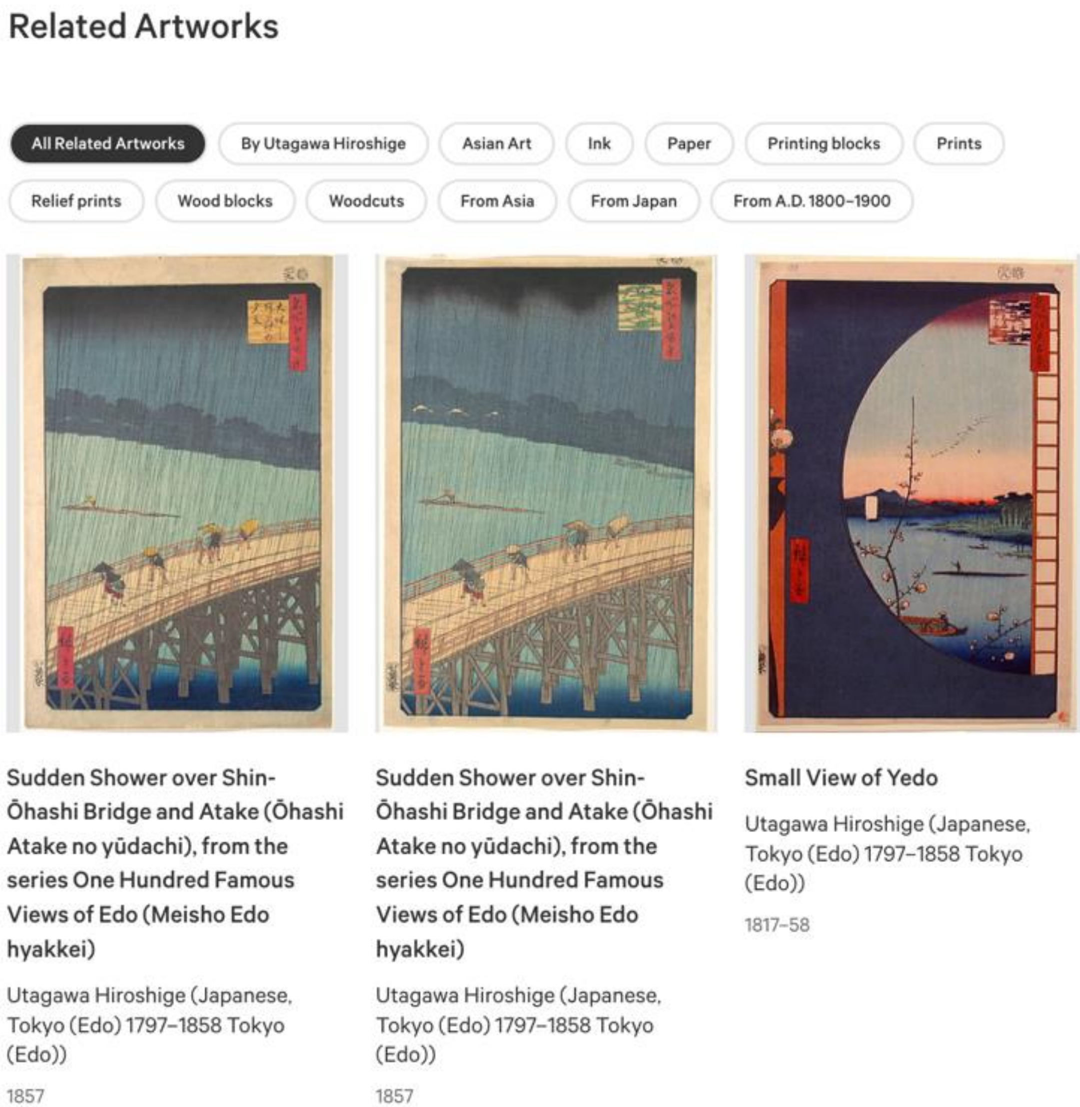


Figure 10: An example of MET's related objects recommendation for Hiroshige's *Sudden Shower over Shin-Ohashi Bridge and Atake*.[10]

[10] "Utagawa Hiroshige | Sudden Shower over Shin-Ōhashi Bridge and Atake (Ōhashi Atake no yūdachi), from the series One Hundred Famous Views of Edo (Meisho Edo hyakkei) | Japan | Edo period (1615–1868)" *The Metropolitan Museum of Art*. , n.d., online, Internet, 10 Mar. 2022. , Available: https://www.metmuseum.org/art/collection/search/55433.

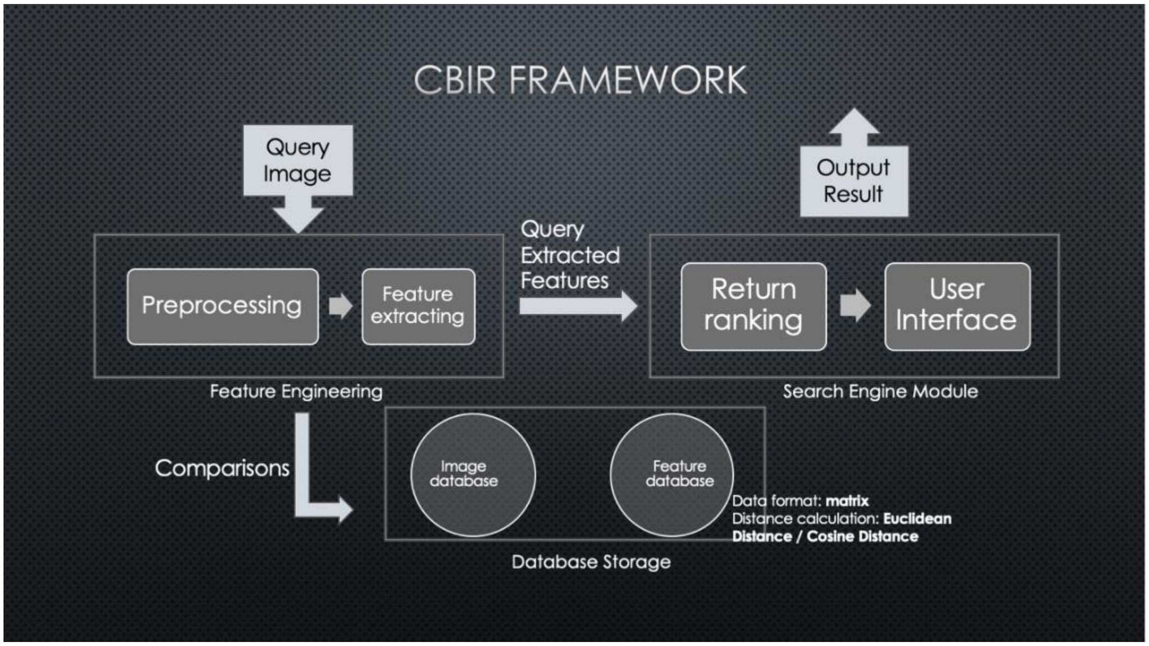


Figure 11: The framework of content-based image retrieval.

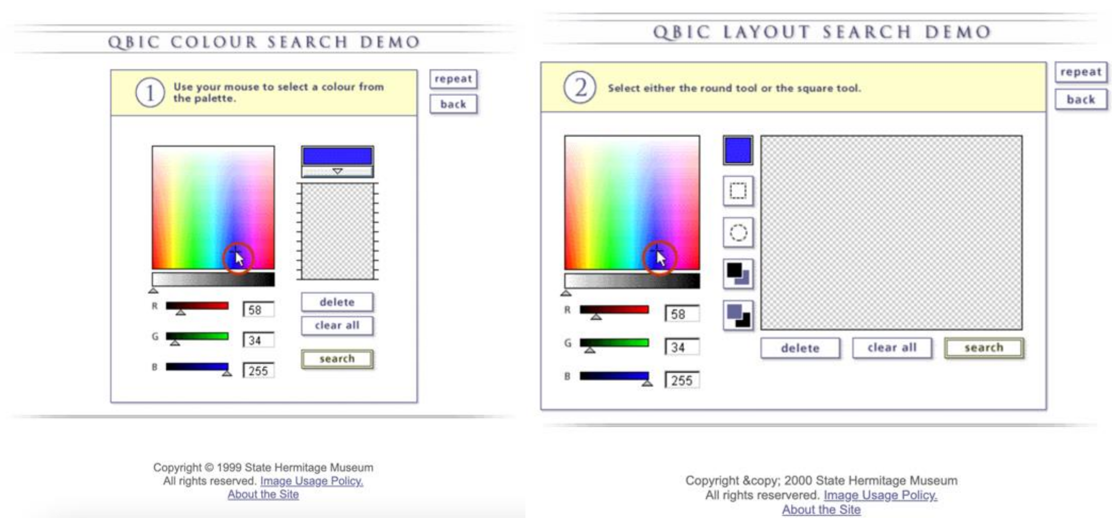


Figure 12: The search interface of IBM's QBIC syste, which was applied on Imagebase of Fine Arts Museums of San Francisco and other museum collections.[11]

---

[11] "The State Hermitage Museum: Digital Collection -- Powered by IBM," n.d., online, Internet, 1 Mar. 2022. , Available: https://web.archive.org/web/20030808170654if_/http://www.hermitagemuseum.org:80/fcgi-bin/db2www/qbicSearch.mac/qbic?selLang=English.

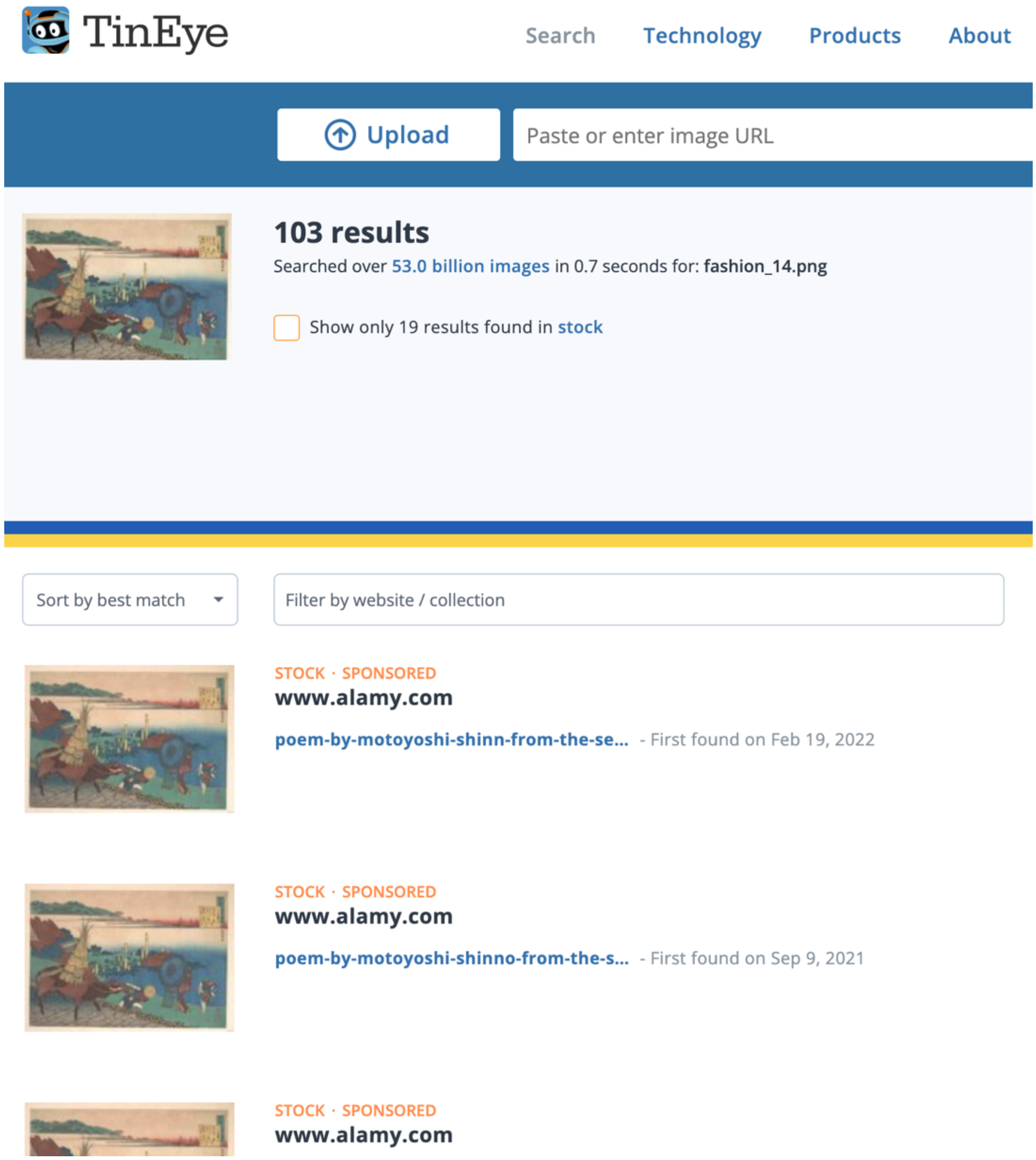


Figure 13: Content-based image retrieval on TinEye.[12]

[12] "TinEye Reverse Image Search," n.d., online, Internet, 13 Apr. 2022. , Available: https://tineye.com/.

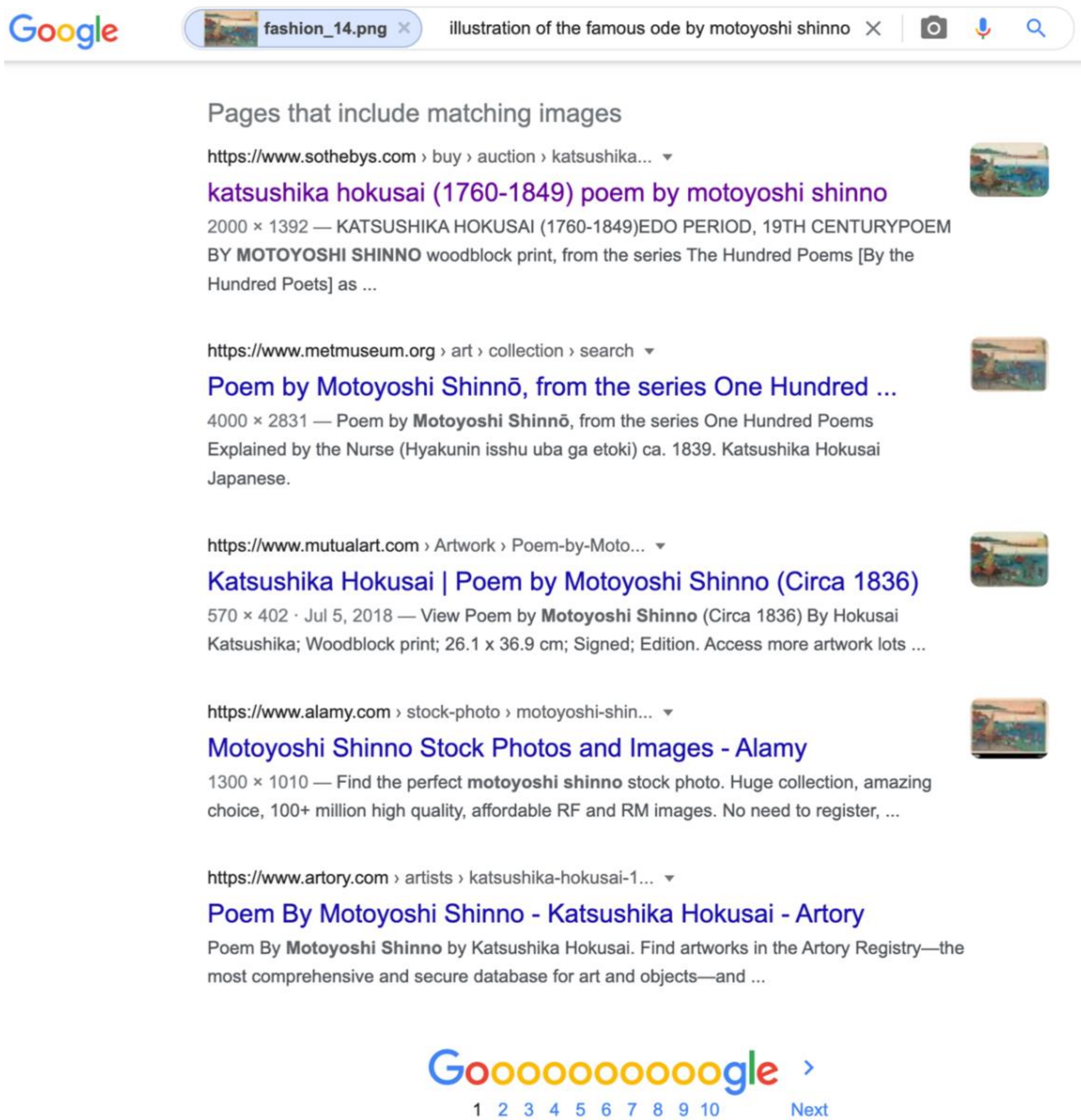


Figure 14: Content-based image retrieval on Google Image.[13]

[13] "Google Images," n.d., online, Internet, 13 Apr. 2022. , Available: https://images.google.com/.

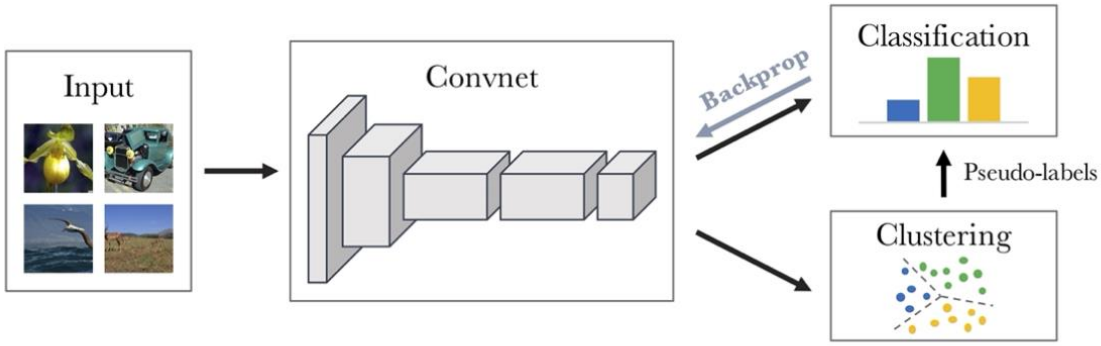


Figure 15: The structure of DeepCluster algorithm.[14]

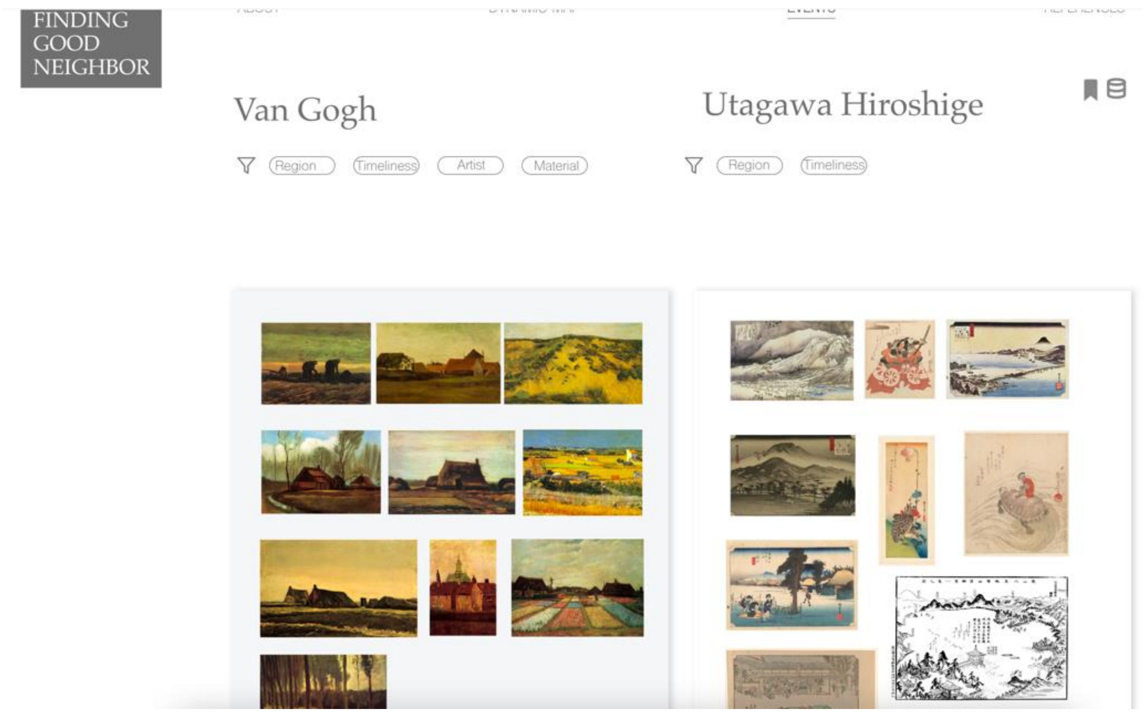


Figure 16: The user interface designed by Adobe XD. This is a demo of multi-panel comparative search between Van Gogh and Utagawa Hiroshige.[15]

---

[14] Ibid.
[15] *The Metropolitan Museum of Art Open Access CSV*, vols. (The Metropolitan Museum of Art, 2022), online, Internet, 1 Mar. 2022. , Available: https://github.com/metmuseum/openaccess.

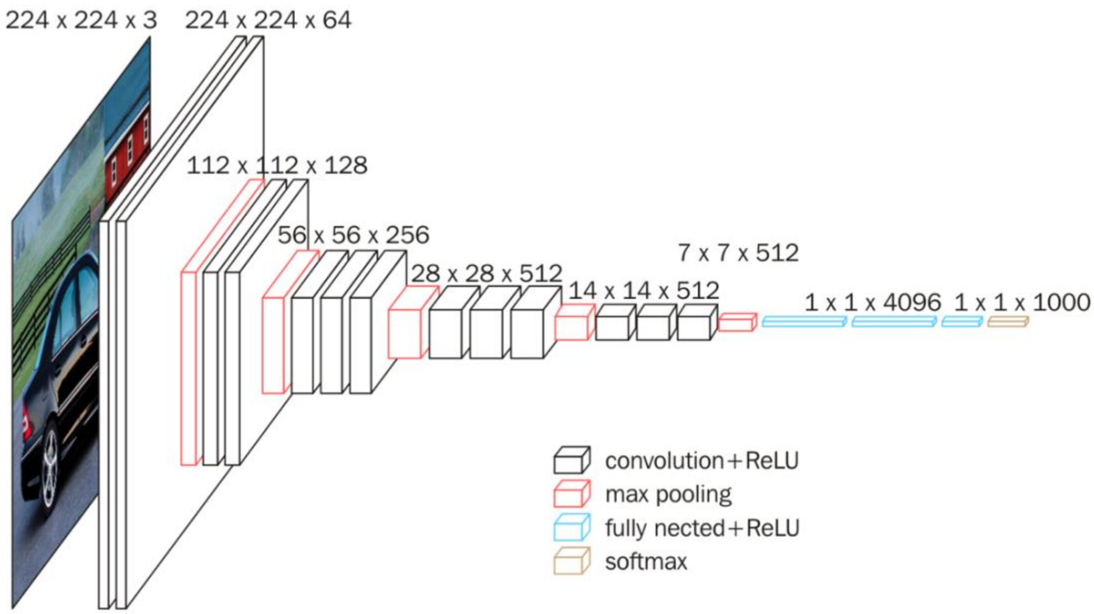


Figure 17: VGG-16's layer-by-layer network structure.[16]

[16] Simonyan and Zisserman, "Very Deep Convolutional Networks for Large-Scale Image Recognition."

Figure 18: The visual features extracted by VGG-16 deep neural network. The features are stored in the format of array with a size of 4096.

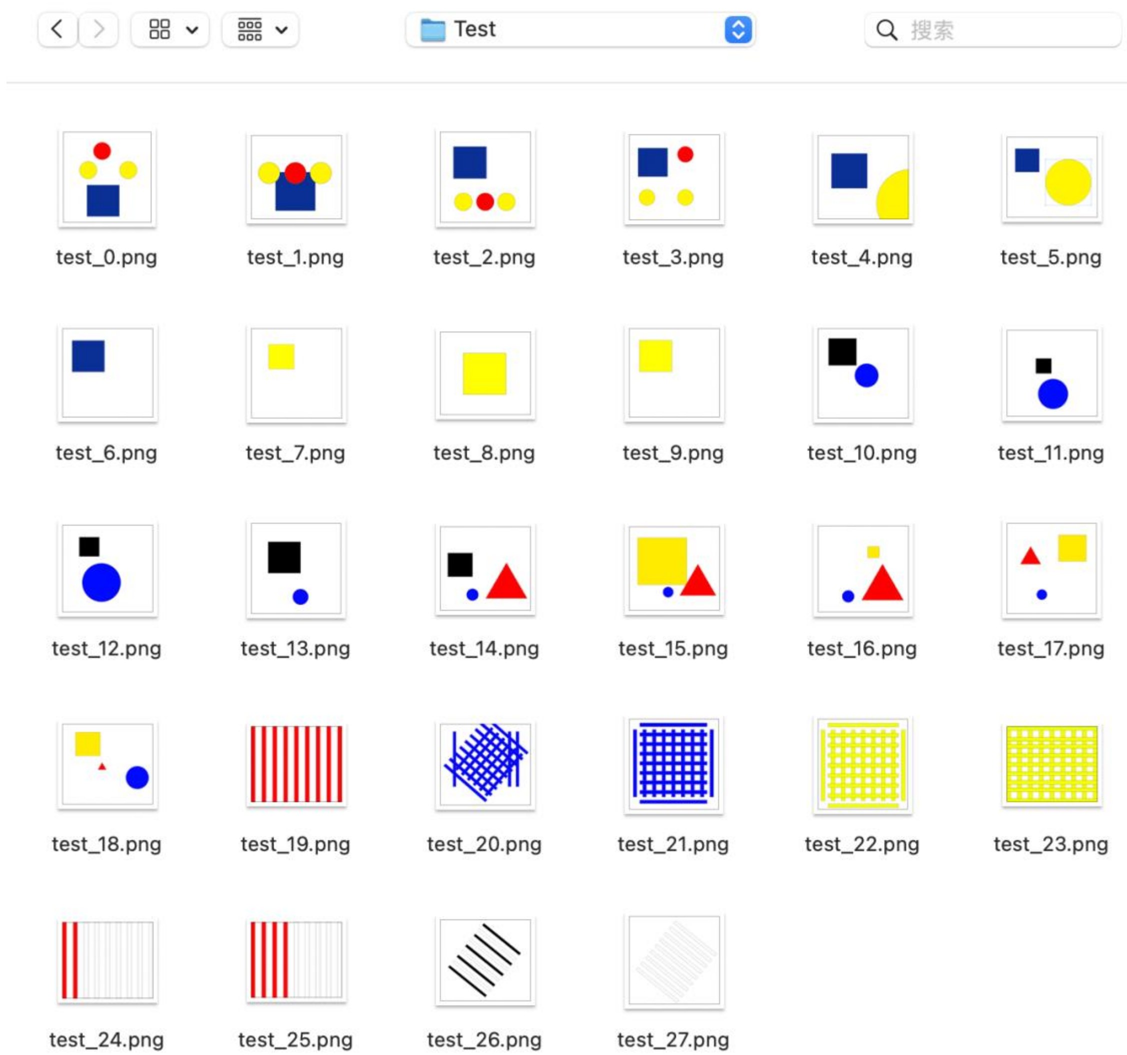


Figure 19: Demo dataset to test the capability of the model.

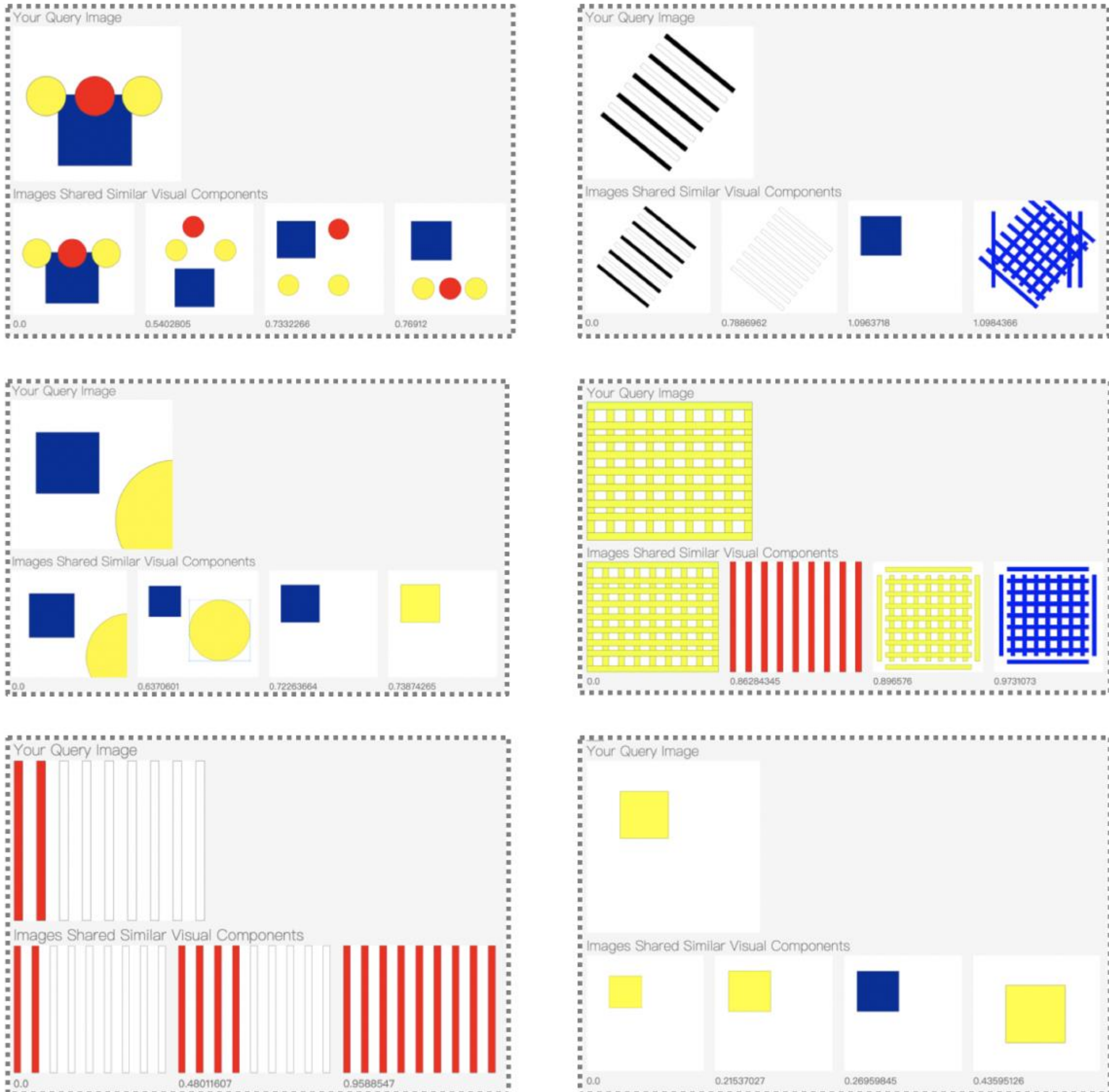


Figure 20: Six test cases of the content-based retrieval system based on features extracted by VGG-16 model.

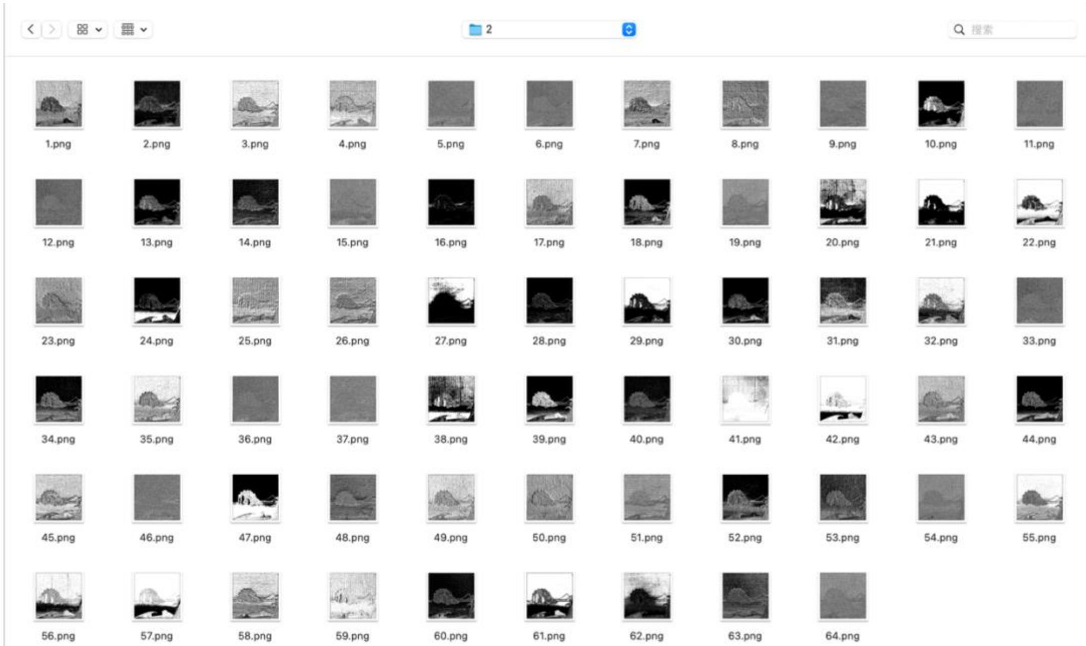


Figure 21: Visualization of features extracted by channels in VGG-16's second layer.

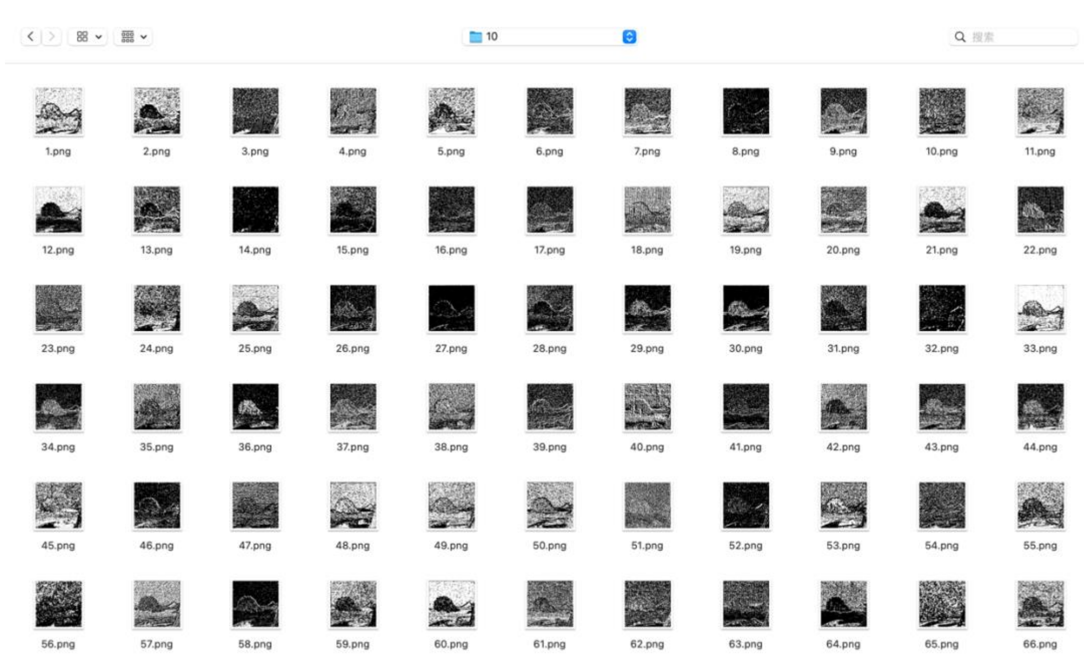


Figure 22: Visualization of features extracted by channels in VGG-16's tenth layer.

Figure 23: Visualization of features extracted by channels in VGG-16's sixteenth layer.

Marble head of a Ptolemaic queen (ca. 270–250 B.C.). Greek

Google Cloud Vision	Amazon Rekognition	IBM Watson
sculpture, classical sculpture, stone carving, forehead, art, head, chin, statue, artifact, museum	head, art, **sculpture**, statue, human, person, figurine, archaeology	ancient person, person, **sculpture**, alabaster color

Figure 24: Example of art image tags generated by three object detection systems.

Great Indian Fruit Bat (ca. 1777–82). Painting attributed to Bhawani Das

Google Cloud Vision	Amazon Rekognition	IBM Watson
bat, illustration, batman, art	bull, animal, mammal, wildlife, **bat**, art, silhouette	brown bat, **bat,** mammal, animal, European brown bat, fruit bat, big brown bat, little brown bat, coal black color

Figure 25: Example of art image tags generated by three object detection systems.

Art_CBIR — -zsh — 96×28

```
silviawsw@MacBook-Pro ~ % cd Art_CBIR
silviawsw@MacBook-Pro Art_CBIR % cd bin
silviawsw@MacBook-Pro bin % ls
activate                django-admin            pip3                    sqlformat
activate.csh            f2py                    pip3.8                  ttx
activate.fish           f2py3                   pyftmerge               wheel
activate.nu             f2py3.8                 pyftsubset              wheel-3.8
activate.ps1            flask                   python                  wheel3
activate_this.py        fonttools               python3                 wheel3.8
deactivate.nu           pip                     python3.8
silviawsw@MacBook-Pro bin % source activate
(frontend) silviawsw@MacBook-Pro bin % cd ..
(frontend) silviawsw@MacBook-Pro Art_CBIR %
```

Figure 26: Terminal command to set up a python virtual environment using *virtualenv*.

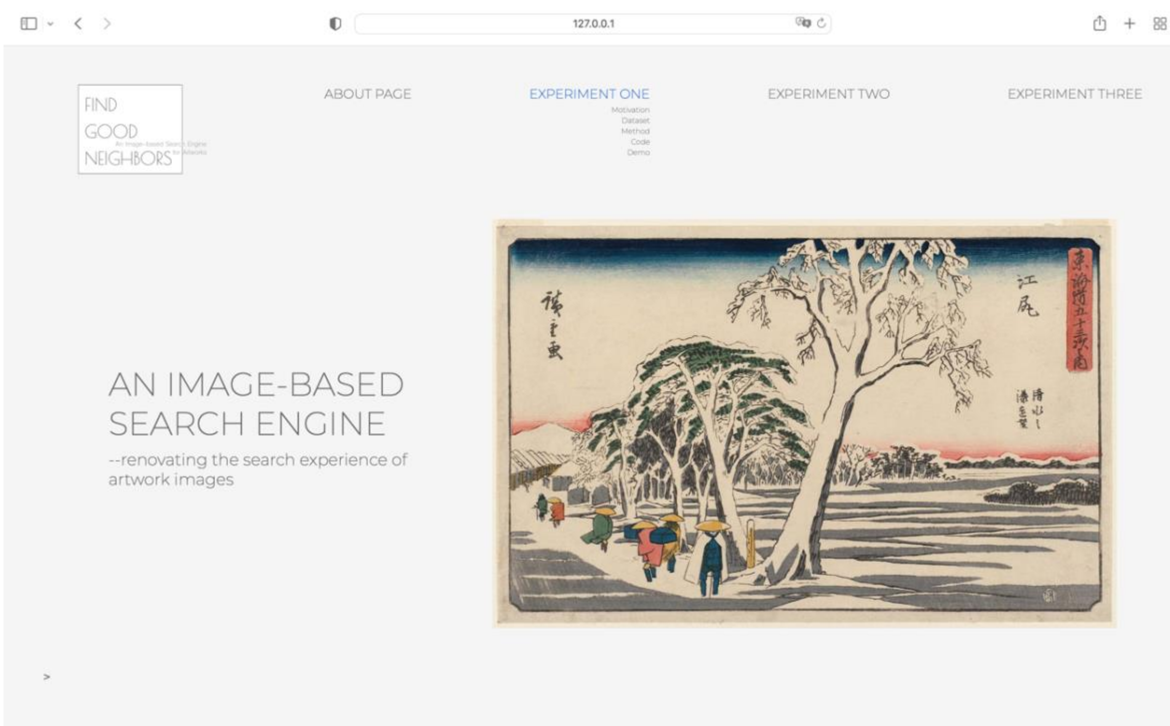


Figure 27: The home page of the website.

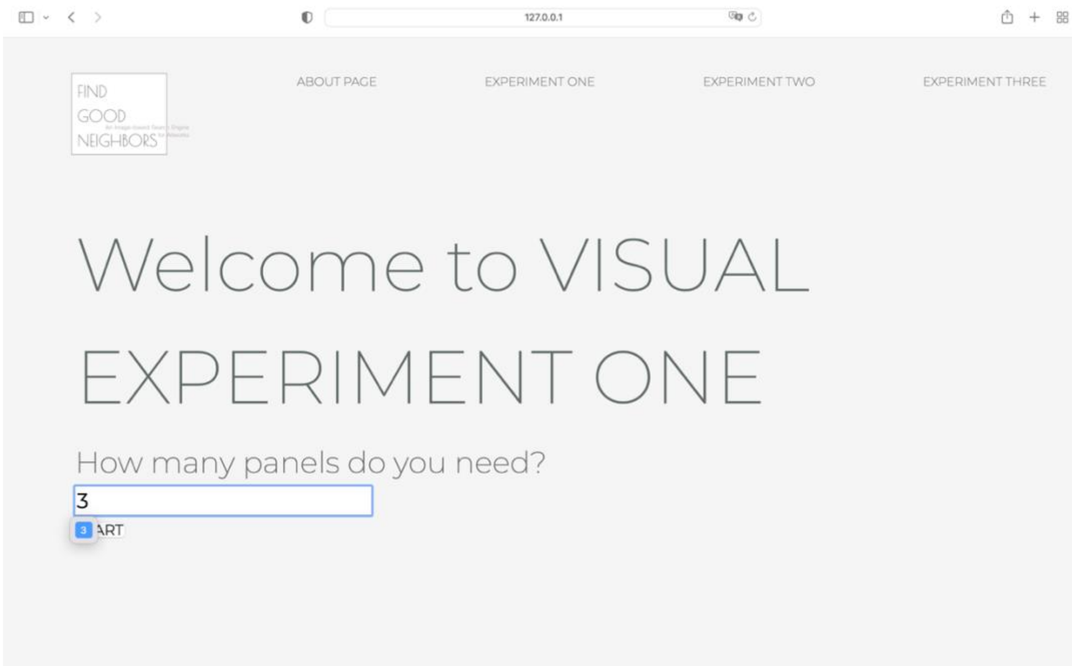


Figure 28: Customize the panels.

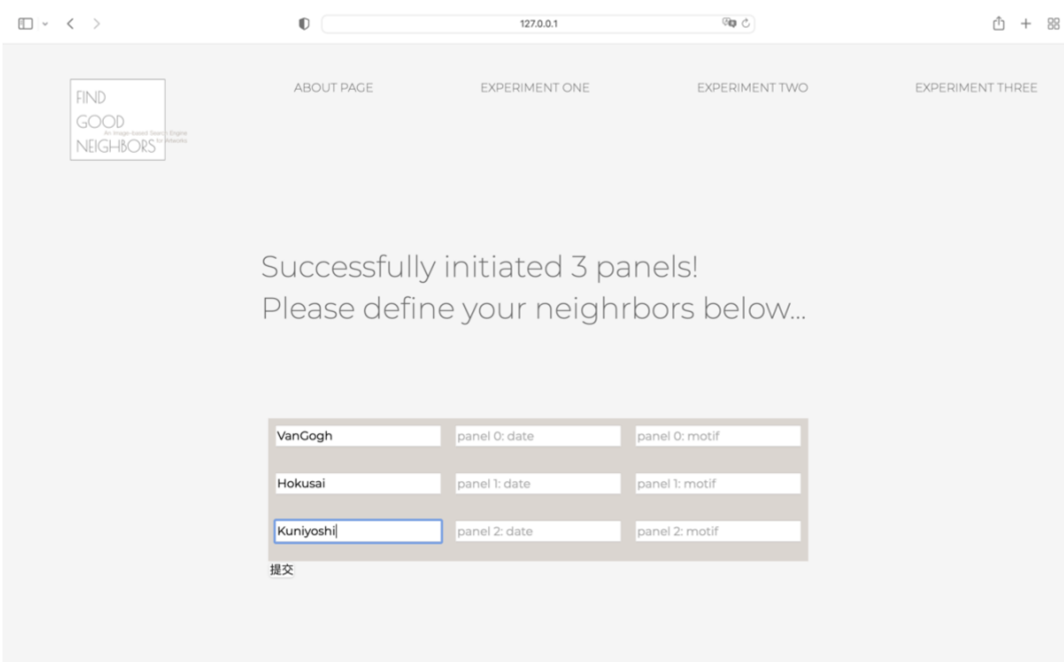


Figure 29: Customize the scope of artworks for each panel based on users' research interests.

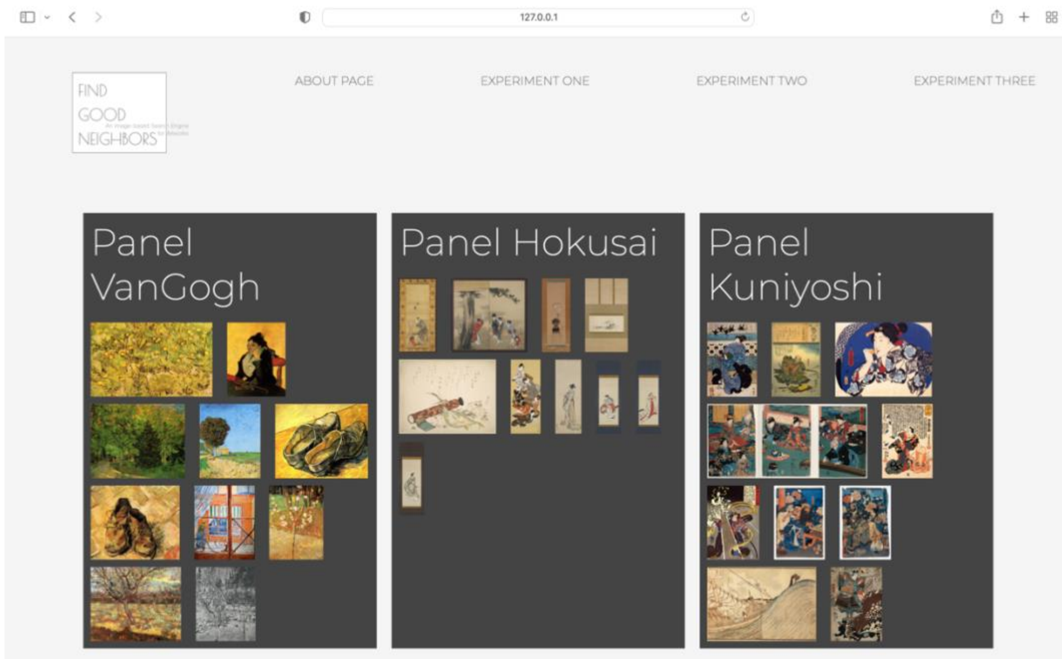
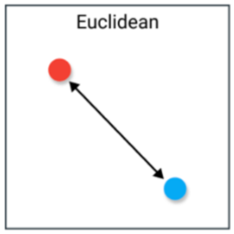
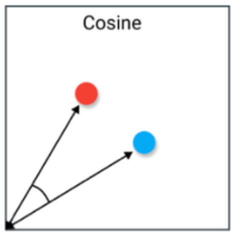


Figure 30: Returned image panels based on the user-defined parameters.



$$D(x, y) = \sqrt{\sum_{i=1}^{n}(x_i - y_i)^2)}$$



$$D(x, y) = cos(\theta) = \frac{x \cdot y}{\|x\| \, \|y\|}$$

```
query = features_Hiroshige[int(query_img)]
query_img = img_paths_Hiroshige[int(query_img)]
dists = np.linalg.norm(features_Hokusai-query, axis=1)
dists_cosine = 1 - np.dot(features_Hokusai,query) / (np.linalg.norm(features_Hokusai)*np.linalg.norm(query))
```

Figure 31: Functions of Euclidean distance and Cosine distance.[17]

[17] Maarten Grootendorst, "9 Distance Measures in Data Science" *Medium.* , 7 Dec. 2021, online, Internet, 13 Apr. 2022. , Available: https://towardsdatascience.com/9-distance-measures-in-data-science-918109d069fa.

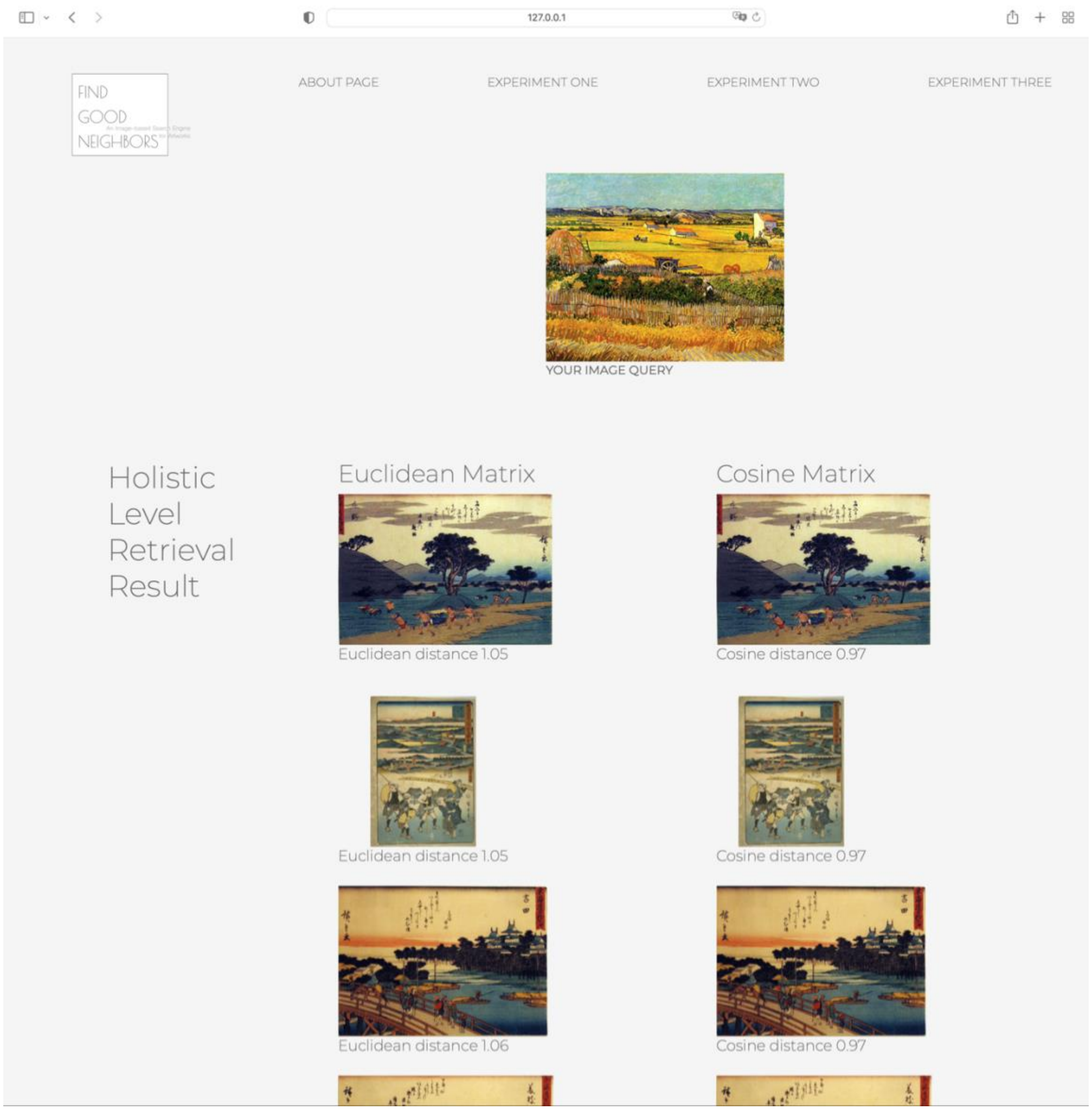


Figure 32: Outputs of the content-based image retrieval.

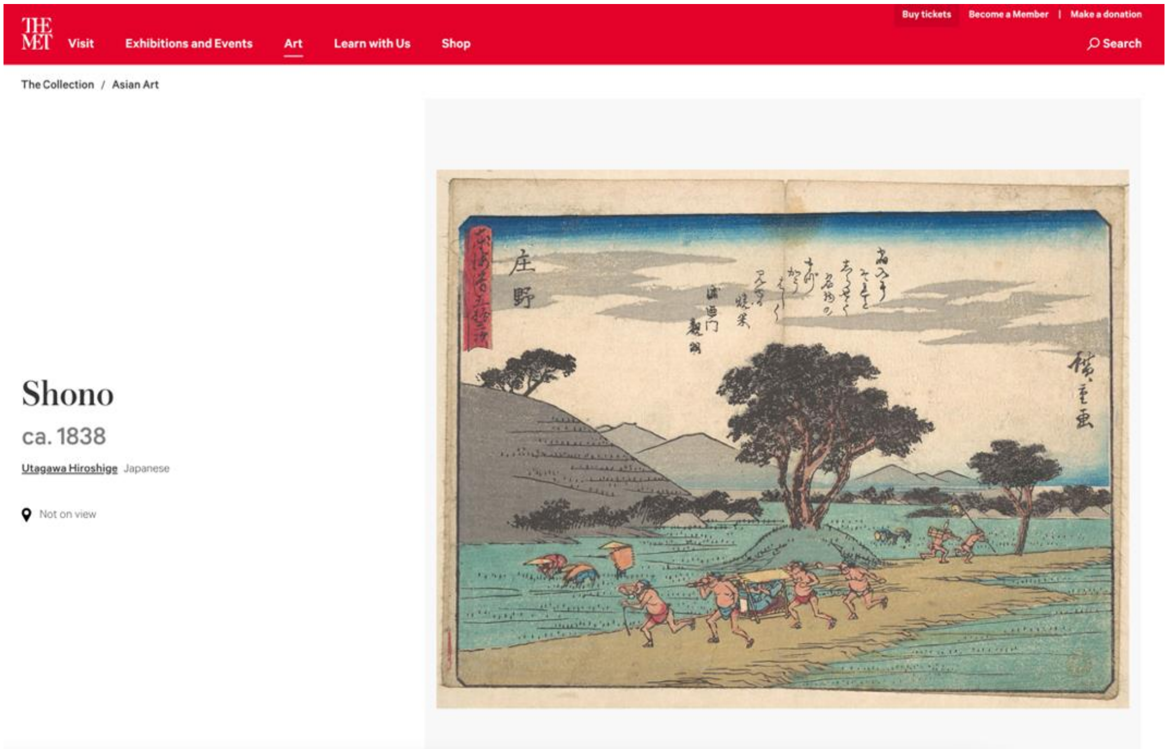


Figure 33: Utagawa Hiroshige's *Shono* in MET's Public Domain Collection.[18]

[18] "Utagawa Hiroshige | Shono | Japan | Edo period (1615–1868)" *The Metropolitan Museum of Art*. , n.d., online, Internet, 10 Mar. 2022. , Available: https://www.metmuseum.org/art/collection/search/36580.

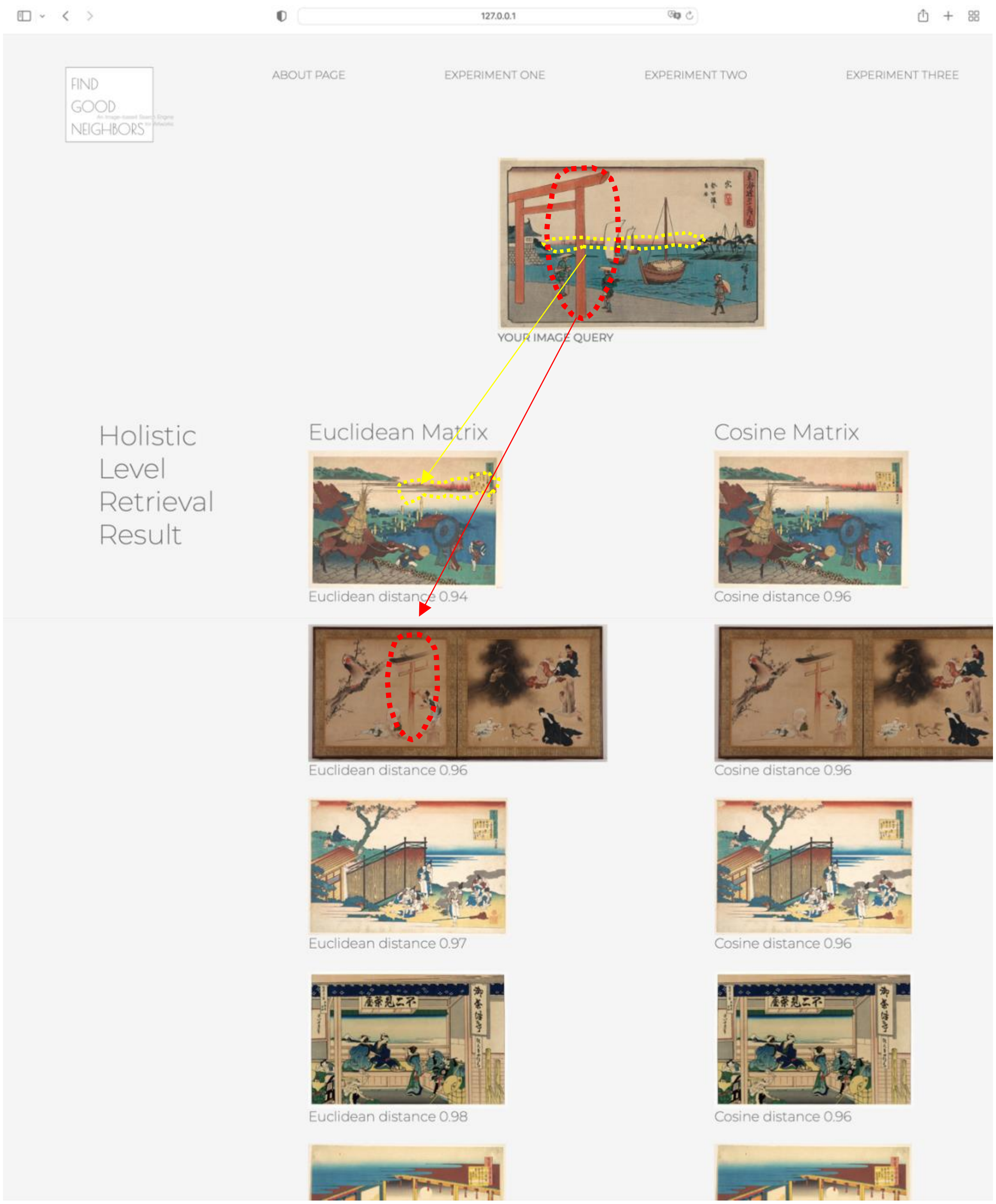


Figure 34: Outputs of content-based image retrieval.

Works Cited

Artstor. “The Plum Garden at Kameido Shrine.” Artwork, n.d. Online. Internet. 10 Mar. 2022. . Available: https://library.artstor.org/public/SS35559_35559_34101454.

Bishop, Claire. “Against Digital Art History.” *International Journal for Digital Art History*3 (2018). Online. Internet. 13 Apr. 2022. . Available: https://journals.ub.uni-heidelberg.de/index.php/dah/article/view/49915.

Bloch, Vitale. “Roberto Longhi.” *The Burlington Magazine* 113.823 (1971): 609–612.

Brown, Kathryn, ed. *The Routledge Companion to Digital Humanities and Art History*. New York: Routledge, 2020.

Caron, Mathilde et al. “Deep Clustering for Unsupervised Learning of Visual Features.” *arXiv:1807.05520 [cs]* (2019). Online. Internet. 3 Mar. 2022. . Available: http://arxiv.org/abs/1807.05520.

Cecchini, Laura Moure. “Baroque Futurism: Roberto Longhi, the Seventeenth -Century, and the Avant-Garde.” *The Art Bulletin* 101.2 (2019): 29–53.

Drucker, Johanna. “Is There a ‘Digital’ Art History?” *Visual Resources* 29.1–2 (2013): 5–13.

Geman, Stuart. “Ulf Grenander, 1923–2016.” *Journal of the Royal Statistical Society. Series A (Statistics in Society)* 179.4 (2016): 1122–1124.

Gilbert, Allan H., and Horst W. Janson. “Erwin Panofsky, Studies in Iconology: Humanistic Themes in the Art of the Renaissance.” *The Art Bulletin* 22.3 (1940): 172–175.

Grootendorst, Maarten. “9 Distance Measures in Data Science.” *Medium*, 7 Dec. 2021. Online. Internet. 13 Apr. 2022. . Available: https://towardsdatascience.com/9-distance-measures-in-data-science-918109d069fa.

Heinrich, Wölfflin. *Principles of Art History: The Problem of the Development of Style in Early Modern Art*. 2015th ed. Getty Trust Publications, 1915.

Hristova, Stefka. "Images as Data: Cultural Analytics and Aby Warburg's Mnemosyne." *International Journal for Digital Art History*2 (2016). Online. Internet. 3 Mar. 2022. . Available: https://journals.ub.uni-heidelberg.de/index.php/dah/article/view/23489.

Jones, Steven E. *Roberto Busa, S. J. , and the Emergence of Humanities Computing: The Priest and the Punched Cards*. London, UNITED KINGDOM: Taylor & Francis Group, 2016. Online. Internet. 9 Mar. 2022. . Available: http://ebookcentral.proquest.com/lib/duke/detail.action?docID=4470572.

Kienle, Miriam. "Digital Art History 'Beyond the Digitized Slide Library': An Interview with Johanna Drucker and Miriam Posner" 6.3 (2017): 6.

Klinke, Harald. "Big Image Data within the Big Picture of Art History." *International Journal for Digital Art History*2 (2016). Online. Internet. 25 Feb. 2022. . Available: https://journals.ub.uni-heidelberg.de/index.php/dah/article/view/33527.

Larsen, Logan. "On the Library as Foundation: Thinking On and Beyond Aby Warburg's Systematic Approach to Images and Culture," 2020. Online. Internet. 3 Mar. 2022. . Available: https://repositories.lib.utexas.edu/handle/2152/81441.

Li, Fei-Fei. *How we're teaching computers to understand pictures*, 1427123194. Online. Internet. 25 Feb. 2022. . Available: https://www.ted.com/talks/fei_fei_li_how_we_re_teaching_computers_to_understand_pictures.

Longhi, Roberto. *Breve ma veridica storia della pittura italiana*. Abscondita, 2013.

Manovich, Lev. "Data Science and Digital Art History." *International Journal for Digital Art History*1 (2015). Online. Internet. 10 Mar. 2022. . Available: https://journals.ub.uni-heidelberg.de/index.php/dah/article/view/21631.

Marr, D., and T. Poggio. "From Understanding Computation to Understanding Neural Circuitry" (1976). Online. Internet. 1 Mar. 2022. . Available: https://dspace.mit.edu/handle/1721.1/5782.

Marr, David. *Vision: A Computational Investigation into the Human Representation and Processing of Visual Information*. Cambridge, MA, USA: MIT Press, 2010.

Moretti, Franco. "Conjectures on World Literature." *New Left Review*1 (2000): 54–68.

Mukhopadhyay, Nitis. "A Conversation with Ulf Grenander." *Statistical Science* 21.3 (2006). Online. Internet. 28 Feb. 2022. . Available: http://arxiv.org/abs/math/0701092.

Murray, Janet H. *Inventing the Medium: Principles of Interaction Design As a Cultural Practice*. Cambridge, UNITED STATES: MIT Press, 2011. Online. Internet. 28 Feb. 2022. . Available: http://ebookcentral.proquest.com/lib/duke/detail.action?docID=3339350.

Panofsky, Erwin. *Studies in Iconology: Humanistic Themes In the Art of the Renaissance*. New York: Routledge, 2019.

Preez, Amanda Du. "Approaching Aby Warburg and Digital Art History: Thinking Through Images." In *The Routledge Companion to Digital Humanities and Art History*. Routledge, 2020.

Rauwel Alain. "Roberto Longhi: the eye and the verb." *Art Critique*, 12 Dec. 2018. Online. Internet. 3 Mar. 2022. . Available: https://www.art-critique.com/en/2018/12/roberto-longhi-the-eye-and-the-verb/.

Riordan, Michael, Lillian Hoddeson, and Conyers Herring. "The Invention of the Transistor." In *More Things in Heaven and Earth: A Celebration of Physics at the Millennium*. Ed. Benjamin Bederson, 563–578. New York, NY: Springer, 1999. Online. Internet. 10 Mar. 2022. . Available: https://doi.org/10.1007/978-1-4612-1512-7_37.

Simonyan, Karen, and Andrew Zisserman. "Very Deep Convolutional Networks for Large-Scale Image Recognition." *arXiv:1409.1556 [cs]* (2015). Online. Internet. 9 Mar. 2022. . Available: http://arxiv.org/abs/1409.1556.

Steinberg, Michael P. "THE LAW OF THE GOOD NEIGHBOR." *Common Knowledge* 18.1 (2012): 128–133.

Team, Keras. "Keras documentation: VGG16 and VGG19," n.d. Online. Internet. 10 Mar. 2022. . Available: https://keras.io/api/applications/vgg/.

Villaespesa, Elena, and Seth Crider. "A critical comparison analysis between human and machine-generated tags for the Metropolitan Museum of Art's collection." *Journal of Documentation* 77.4 (2021): 946–964.

Warburg, Aby, Christopher Johnson, and Claudia Wedepohl. "From the Arsenal to the Laboratory." *West 86th: A Journal of Decorative Arts, Design History, and Material Culture* 19 (2012): 106–124.

Weaver, Warren. "Translation." In *Proceedings of the Conference on Mechanical Translation*. Massachusetts Institute of Technology, 1952. Online. Internet. 10 Mar. 2022. . Available: https://aclanthology.org/1952.earlymt-1.1.

Zhu, Song-Chun. "A Discussion on the Origin of Computer Vision and Artificial Intelligence." The Vision Seeker, Oct. 2016.

"A View from Above: Exploratory Visualizations of MoMA Photography Collection," n.d. Online. Internet. 13 Apr. 2022. . Available: http://lab.culturalanalytics.info/2016/04/exploratory-visualizations-of-thomas.html.

"About us | Haltadefinizione," n.d. Online. Internet. 10 Mar. 2022. . Available: https://www.haltadefinizione.com/en/about-us/.

"An Introduction to the Use of Images in Art Historical Research: Context and Prospects." *Art History Library* (2020).

"Bilderatlas Mnemosyne." *The Warburg Institute*, n.d. Online. Internet. 8 Mar. 2022. . Available: https://warburg.sas.ac.uk/archive/bilderatlas-mnemosyne.

"Bilderatlas Mnemosyne | Final version." *The Warburg Institute*, n.d. Online. Internet. 8 Mar. 2022. . Available: https://warburg.sas.ac.uk/archive/bilderatlas-mnemosyne/final-version.

"Chinese Fine Art," n.d. Online. Internet. 10 Mar. 2022. . Available: https://kaggle.com/rickyjli/chinese-fine-art.

"Cultural Analytics | Books Gateway | MIT Press," n.d. Online. Internet. 22 Mar. 2022. . Available: https://direct-mit-edu.proxy.lib.duke.edu/books/book/4966/Cultural-Analytics.

"CVPR 2021 Conference Report | IEEE Computer Society," n.d. Online. Internet. 27 Feb. 2022. . Available: https://www.computer.org/publications/tech-news/events/cvpr-2021-recap/.

"Digital image." *Wikipedia*, 15 Feb. 2022. Online. Internet. 10 Mar. 2022. . Available: https://en.wikipedia.org/w/index.php?title=Digital_image&oldid=1072072337.

"Google Images," n.d. Online. Internet. 13 Apr. 2022. . Available: https://images.google.com/.

"IBM Research," REPLACE. Online. Internet. 1 Mar. 2022. . Available: https://dominoweb.draco.res.ibm.com/dominoweb.draco.res.ibm.com/a3ea9019a4ed364985256593006fd727.html.

"ImageNet," n.d. Online. Internet. 25 Feb. 2022. . Available: https://www.image-net.org/index.php.

"Library – Fondazione Roberto Longhi," n.d. Online. Internet. 3 Mar. 2022. . Available: https://fondazionelonghi.it/wordpress/en/eng_library/.

“NVIDIA LAUNCHES THE WORLD’S FIRST GRAPHICS PROCESSING UNIT: GEFORCE 256.” *ResponseSource Press Release Wire*, n.d. Online. Internet. 4 Mar. 2022. . Available: https://pressreleases.responsesource.com/news/3992/nvidia-launches-the-world-s-first-graphics-processing-unit-geforce-256/.

“Opening art with ‘art’: Roberto Longhi’s approach to art historical research and writing Chinese art history with ‘language’ (Volume one).” *Art History Library*, n.d.

“The Harvest Vincent van Gogh, 1888.” *Van Gogh Museum*, n.d. Online. Internet. 10 Mar. 2022. . Available: https://www.vangoghmuseum.nl/en/collection/s0030V1962.

The Metropolitan Museum of Art Open Access CSV. The Metropolitan Museum of Art, 2022. Online. Internet. 1 Mar. 2022. . Available: https://github.com/metmuseum/openaccess.

“The State Hermitage Museum: Digital Collection -- Powered by IBM,” n.d. Online. Internet. 1 Mar. 2022. . Available: https://web.archive.org/web/20030808170654if_/http://www.hermitagemuseum.org:80/fcgi-bin/db2www/qbicSearch.mac/qbic?selLang=English.

“TinEye Reverse Image Search,” n.d. Online. Internet. 13 Apr. 2022. . Available: https://tineye.com/.

“Utagawa Hiroshige | Shono | Japan | Edo period (1615–1868).” *The Metropolitan Museum of Art*, n.d. Online. Internet. 10 Mar. 2022. . Available: https://www.metmuseum.org/art/collection/search/36580.

“Utagawa Hiroshige | Sudden Shower over Shin-Ōhashi Bridge and Atake (Ōhashi Atake no yūdachi), from the series One Hundred Famous Views of Edo (Meisho Edo hyakkei) | Japan | Edo period (1615–1868).” *The Metropolitan Museum of Art*, n.d. Online. Internet. 10 Mar. 2022. . Available: https://www.metmuseum.org/art/collection/search/55433.

“Virtual Tour - Aby Warburg: Bilderatlas Mnemosyne exhibition at Haus der Kulturen der Welt | The Warburg Institute,” n.d. Online. Internet. 13 Apr. 2022. . Available:

https://warburg.sas.ac.uk/virtual-tour-aby-warburg-bilderatlas-mnemosyne-exhibition-haus-der-kulturen-der-welt.