Algorithmic Ways of Seeing: Using Object Detection to Facilitate Art Exploration

Louie Søs Meyer[*]
IT University of Copenhagen
Copenhagen, Denmark
lomm@itu.dk

Johanne Engel Aaen[*]
IT University of Copenhagen
Copenhagen, Denmark
jaae@itu.dk

Anitamalina Regitse Tranberg
IT University of Copenhagen
Copenhagen, Denmark
anitatranberg@gmail.com

Peter Kun
IT University of Copenhagen
Copenhagen, Denmark
peku@itu.dk

Matthias Freiberger
IT University of Copenhagen
Copenhagen, Denmark
m.freiberger@gmail.com

Sebastian Risi
IT University of Copenhagen
Copenhagen, Denmark
sebr@itu.dk

Anders Sundnes Løvlie
IT University of Copenhagen
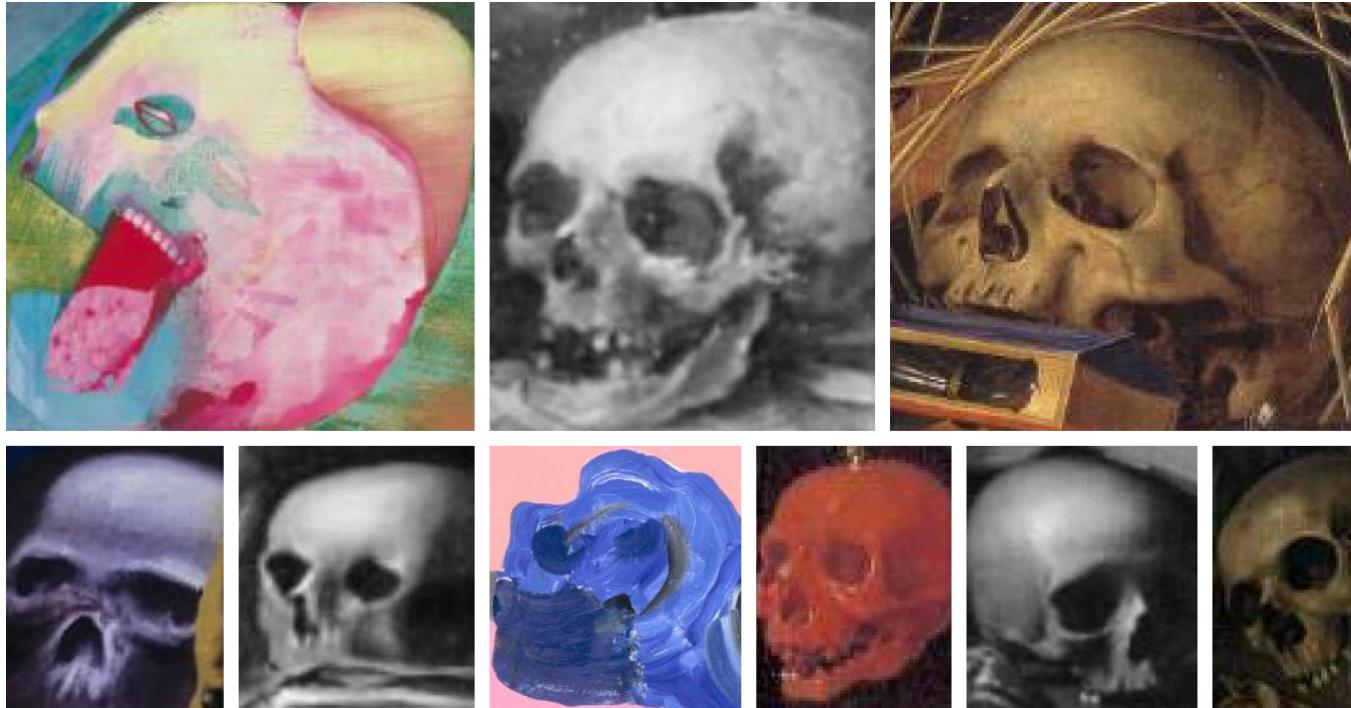Copenhagen, Denmark
asun@itu.dk

Figure 1: A selection of objects classified as "skull" in paintings from the collection of the National Gallery of Denmark.

[*]Both authors contributed equally to this research.

ABSTRACT

This Research through Design paper explores how object detection may be applied to a large digital art museum collection to facilitate new ways of encountering and experiencing art. We present the design and evaluation of an interactive application called SMKExplore, which allows users to explore a museum's digital collection

of paintings by browsing through objects detected in the images, as a novel form of open-ended exploration. We provide three contributions. First, we show how an object detection pipeline can be integrated into a design process for visual exploration. Second, we present the design and development of an app that enables exploration of an art museum's collection. Third, we offer reflections on future possibilities for museums and HCI researchers to incorporate object detection techniques into the digitalization of museums.

CCS CONCEPTS

• **Human-centered computing** → *Hypertext / hypermedia*; **Empirical studies in interaction design**; • **Applied computing** → **Fine arts**; • **Computing methodologies** → *Object detection.*

KEYWORDS

Object Detection, Art, Experience Design, Exploratory Search, Computer Vision

1 INTRODUCTION

Recent progress in computer vision has led to algorithms that are comparable to human performance on some tasks [24]. Previously, vision algorithms have been limited by what has been termed the "cross-depiction" problem [35]: While achieving impressive performance in detecting objects on photographic images, algorithms would struggle to detect the same objects in other depictions such as drawings, paintings, and other art styles. However, recent advances in machine learning algorithms trained on multimodal image-text datasets and approaches such as Contrastive Language-Image Pre-training (CLIP) [63] and Grounded Language-Image Pre-Training (GLIP) [49] offer promising performance across visual domains. In this paper, we explore how these technologies may be applied to the large art collection of the National Gallery of Denmark (Danish abbreviation: SMK) to facilitate new ways of exploring and experiencing art. While techniques such as computer vision and, more broadly, Artificial Intelligence (AI) raise both legal and ethical concerns relating to authorship, bias, trust, and more [20, 29], these technologies also offer the potential to make art collections more accessible and to offer new ways of experiencing art.

As suggested by Lev Manovich, while visual art and aesthetics are traditionally experienced and studied by looking at individual images and artworks, computational analysis of images opens up the perspective of exploring large datasets, inviting us to shift our perspective from unique exemplars to *"seeing one billion images"* and the patterns therein [56]. Within Human-Computer Interaction (HCI) such perspectives have been explored through the design of novel systems for visualization and exploratory search [26, 87, 90]. Exploratory search is of particular importance for museums and art collections, because it allows non-expert users to find pathways to explore and discover art that they don't know about and so wouldn't know how to search for in a traditional search interface -

an issue that is strongly aligned with museums' mission to inspire and educate [8, 88, 92].

Museums have long served as a productive environment for research and experiments in Human-Computer Interaction (HCI) [40], and debates around the use of computer vision in museums have become increasingly prominent [17, 84]. The recent significant developments in object detection suggest that computer vision may be applied now to museum collections to make them searchable not just through metadata about the images but through the subject matter of the artworks - i.e. the objects appearing in the images. Presently, such search has been limited by the extent to which information about the objects has been manually entered into the collection metadata by museum curators - which is often far more sparse (if at all available) than the rich visual information in the images [4, 17]. Thus, this paper presents a research through design [95] exploration of the following research question:

> **RQ:** *How can object detection be used to support exploration of an art museum's digital collection?*

As this approach represents a novel application of object detection techniques, a significant part of the effort has been to implement an object detection workflow for extracting object data about the art collection. We present the design and evaluation of an interactive application, SMKExplore, which allows users to explore a museum's digital collection of art paintings by browsing through objects detected in the images, as a novel form of exploration.

In this paper, we provide three contributions. First, we show how an object detection pipeline can be integrated into a design process for visual exploration. Second, we present the design and development of an app that enables exploration in the context of a museum collection. Third, we offer reflections on future possibilities for museums and HCI researchers to incorporate object detection techniques in digital museum collections.

2 RELATED WORK

2.1 Object Detection and Artwork

Along with rapid advances in machine learning, scholars have debated how designers may use machine learning as *design material* [7, 23, 25, 33, 45, 93]. In this paper, we focus on one particular application of machine learning: Object detection in art images.

With large numbers of home users gaining access to the world wide web in the early 2000s, large online image collections emerged as users shared their personal photographs and drawings. With the advent of these collections, research interest in automatic image annotation and image retrieval to make sense of them has strongly increased [22]. Important early work includes probabilistic approaches to automatically match text and images [5, 9], efficient scene matching techniques [82], as well as web-based tools to crowd-source image-labelling at scale [70]. Recent research directions on using large image-text datasets leverage deep learning [47] to develop new approaches for various computer vision tasks.

While deep neural networks have shown great promise when trained to recognize different objects [2], image recognition systems are brittle and may struggle if images are slightly grainy or noisy and are vulnerable to manipulation [24, 38, 78]. Neural networks struggle even more to recognize objects depicted in different styles, such as drawings or paintings. This is known as the cross-depiction

problem [35, 86]. The cross-depiction problem reveals a weakness in image recognition systems compared to human vision: While humans are versatile and can recognize even relatively minimal line drawings with ease, neural networks are highly specialized and perform poorly when confronted with an image style that is different from the data used in training [12]. One of the sources of this weakness might be that computer vision algorithms tend to be biased towards focusing on texture, unlike human vision, which tends to focus more on shape [31]. However, texture bias can be reduced by modifying the training data (while using standard architectures) [31]. Kadish and colleagues adapted the technique presented in [31] and applied it to object detection on the artworks in the People-Art dataset, achieving a 10% improvement in state of the art for this dataset [43].

Recently, great progress has been made with generative models such as Midjourney, DALL·E [64, 65], and Stable Diffusion [67, 74], which are enabled by contrastive training between images and text [49, 63]. In this approach, a given image and its description text are mapped into a shared high-dimensional vector space. This approach has brought remarkable progress on zero-shot computer vision tasks for image-text retrieval, image captioning, or visual Q&A [48]. To illustrate the progress of these approaches on non-photographic content, consider Pablo Picasso's series of lithographs titled "The Bull", which demonstrates a range of depiction styles from a lifelike drawing to ever more abstract styles. Fig. 2a shows the result of a test the authors of this paper ran in 2021, exploring how well a state-of-the-art image recognition algorithm (Fast R-CNN [34], pre-trained on Common Objects in Context (COCO)) [50] could classify these drawings. At this time, the algorithm could only identify three drawings showing a "cow" (as the COCO labelset had no separate category for bull). In 2023, the authors tested the Grounded Language-Image Pre-Training (GLIP) [49] algorithm on the same image, correctly classifying even the most abstract drawings as bulls (see Fig. 2b). This illustrates that these new algorithms have enabled a great leap in cross-depiction object detection, now making it feasible to use these technologies on art collections with a hitherto unmatched degree of precision.

2.2 AI and Museums

Although recently a surge of artists have been working with generative AI systems to create art, sometimes called "AI Art" [3, 10, 11, 14, 27, 61, 96], in this paper we focus on the use of computer vision technology for *Experience Design* in an art museum context, rather than creating new forms of art.

There is a long history of HCI research conducted in collaboration with museums, both using the museum context as a testbed for new technologies, as well as drawing insights from the interplay of technology with art and heritage collections [40, 85]. Museums around the world have spent much efforts in digitizing their collections, both as a way to secure the contents of the collections for the future and also to increase public access – particularly to the vast number of artifacts in archives that cannot be exhibited physically in the museums. It is commonly suggested, that large European museums usually have a very small share of their collections on display in their physical exhibition spaces, both due to limitations in space and other capacity limitations, as well as many artworks

being too brittle to exhibit. In the case of the National Gallery, only 0.7% of the collection is on display [1]. Making digital reproductions of the collections available online allows for the public to access the entirety of the collection at any point in time. This digitalization agenda in the museum and cultural heritage sector has also fostered the implementation of AI techniques for collection management, audience predictions, art authentication, and more [16, 17, 84].
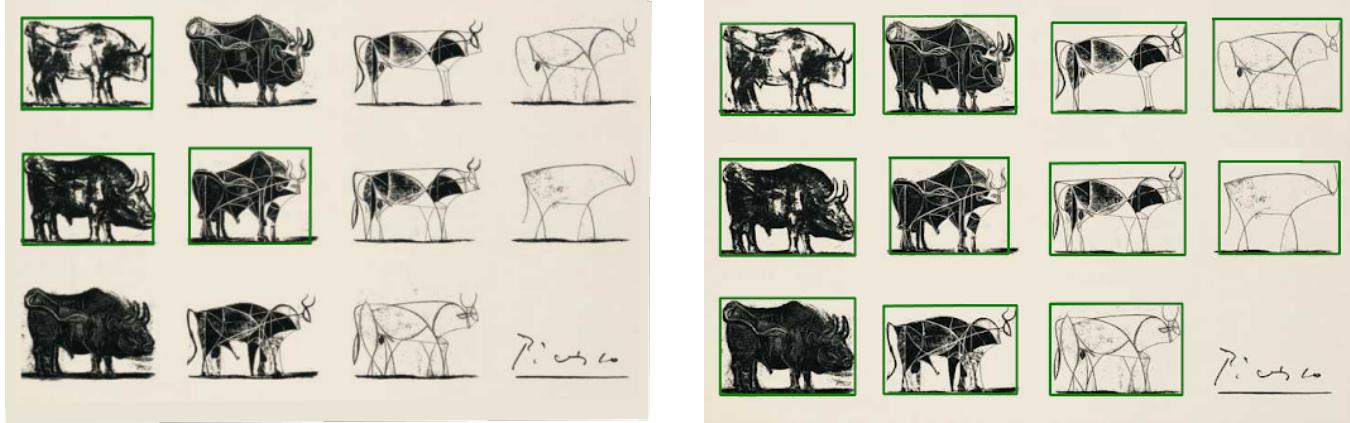
Museum collections often contain vast amounts of data that can be used for many purposes, including categorizing and collection management, as well as a means to investigate patterns of (dis-)similarities between artists, cultures, time periods, or even across an entire collection [32, 59]. Describing and allocating metadata to museum and cultural heritage items is an essential yet labor-intensive task, which often requires highly specialized domain expertise from museum curators and researchers [59].

There is a growing interest in applying computer vision in museums [17, 83, 84]. Museums commonly have experimented with computer vision techniques to enrich their metadata [73, 77, 79]. Such algorithmic enrichment may be helpful both for museum researchers who frequently need to search their collection, as well as for non-expert audiences. For example, the Harvard Art Museum invites online users to investigate how computers see and process images of artwork, comparing metadata generated by humans to metadata generated by AI technologies developed by Amazon, Clarifai, Imagga, Google, and Microsoft [37]. Another example is the Princeton Art Museum, which has experimented with computer vision for various purposes, including detecting visual similarities in Chinese paintings from different eras [84]. Museums have also explored using computer vision techniques in experiences for visitors to the physical museum exhibition, for instance through mobile apps which use image recognition to allow visitors to point their camera at an artwork and receive information about it [52, 53, 80].

Although object detection shows great promise for museums, there are potential pitfalls concerning bias. Museums hold widespread artworks in their collections, among them pieces representing controversial, challenging, and painful parts of history and contemporary society [84]. Research has found that computer vision is limited by bias, specifically in terms of cultural and gender biases [17, 84]. Ciecko and colleagues [17] question whether museums should withhold some artworks from being classified by computer vision algorithms, highlighting colonization, slavery, and genocide as particularly challenging topics. The art collection used in the study at hand is a broadly themed collection of historical and modern art and does not focus specifically on such challenging topics. However, it is hard to predict where bias may appear when analysing a large and varied collection of artworks. In this study, the results of the computer vision analysis were only presented to a small group of test participants, thus reducing the risk of unintended offense or controversy. If the design presented here should be made available to the public at large, the museum would need to carefully consider the risks associated with bias and possible mitigation strategies.

2.3 Exploratory Search in Digital Collections

The metadata of an art museum's digital collection is a complex information space, as these collections are constructed and used

(a) *Le Taureau (The Bull)* by Pablo Picasso, analyzed in 2021 using the Fast R-CNN [34] algorithm pre-trained on COCO [50]. The boxes drawn around three of the drawings indicate that these drawings are classified as "cow" by the algorithm.

(b) *Le Taureau (The Bull)* by Pablo Picasso, analyzed in 2023 using GLIP [49]. The boxes drawn around each drawing indicate that they are all correctly classified as "bull".

Figure 2: Comparison of image analysis results.

by different professionals performing various complex tasks that go beyond "variants on a search box" [68]. As an alternative to restricted and result-oriented keyword-based search, *exploratory search* emerged in HCI to support open-ended, interactive, and evolving processes as a strategy for information seeking [44, 57, 94]. While it lacks a rigid definition, its essence revolves around the journey of searching rather than the user arriving at precisely defined outcomes [62, 75]. Unlike the linear trajectory of keyword searches, exploratory search is iterative; it fosters a dynamic evolution of user needs as they garner new insights [60, 71, 87]. Exploratory search encourages leisurely browsing, inviting users on unexpected voyages through data [26]. Ultimately, its goal leans towards an engaging user experience rather than just efficiently returning search results [89].

The visual nature of cultural heritage collections has afforded visualization strategies to enable experts and non-experts to interact with such datasets, also motivated by going beyond keyword-based searches [90]. A specific type of interactive visualization of cultural heritage collections is described as *generous interfaces* which are rich, browsable interfaces that reveal the scale and complexity of digital heritage collections [88]. Such interfaces have shown that multiple entry points to a collection and navigating via multiple paths allow rich opportunities for exploration and discovery [21, 42, 81].

Many museum visitors might be unaware of the collection's contents, potential attractions, or even their own interests and objectives during a visit [28, 55]. For this reason scholars working with information visualization and design relating to museums and libraries have often been interested in facilitating serendipitous discovery [21, 81, 90], meaning chance encounters with items of interest [8, 26]. Cole and colleagues [18] present several approaches to facilitating exploratory search and serendipitous discovery using techniques such as similarity search and formal concept analysis (see also [91]). Thus, the concepts of exploratory search and generous interfaces offer a helpful perspective in designing for exploration of art collections. This paper contributes by investigating how to apply object detection to create a novel interface for exploration of the collection.

3 APPROACH AND METHODOLOGY

The current project is a research through design [95] exploration based on an interdisciplinary collaboration between scholars in machine learning and human-computer interaction (HCI). The overarching goals for the project are to use the interdisciplinary collaboration to break new ground both for computer vision - applied to art images of different depiction types and styles - as well as for human-computer interaction, in particular regarding the use of machine learning as a design material applied to experiences with visual art. One of the authors of this paper has extensive experience with HCI research in art museums, but none of the authors have specific domain expertise in art analysis.

Thus, the project's starting point was a technical exploration of the state of the art in object detection on art images, carried out by two of the authors during the second half of 2022. This led to the setup described in section 4 below, using GLIP to annotate all 6,750 paintings in the National Gallery's digital collection with object labels.

Once this material had been established we began a design exploration aiming to create an interactive application allowing a general museum audience to browse and experience the collection in novel ways. This was done through an iterative process involving stakeholders at the National Gallery and their audiences from February to August 2023. This process also revealed a need to revise and iterate on the object annotations, as described below in section 4.1.

The resulting interactive application was evaluated with test users recruited on-site at the National Gallery 11-12 Aug 2023, as described in section 6 below.

4 OBJECT DETECTION

Since no ground truth is available on the National Gallery's collection, our object detection approach has relied on pre-trained models. With COCO [50] being the most prominent object detection dataset available, many approaches applied to artwork are trained or pre-trained on it; for instance, see [43, 76]. Designed to mirror modern-day visual environments, the COCO dataset emphasizes contemporary objects, rendering its classes heavily limited for the nuanced motifs or historical themes of traditional art paintings. For this reason, we decided to utilize a contrastively pre-trained model with the possibility to customize the labels of detected objects in a way suitable in the context of artwork.

Our approach involves defining a set of labels (up to 120), which is then presented as a single string of words separated by full stops to the GLIP model [49]. The pre-trained GLIP model then represents each class label as a vector. This vector representation of an object label can then be compared to similar representations extracted from image patches. The model matches the image patch and object label vectors most similar to each other and generates labeled bounding boxes in the image based on these similarities. To evaluate the principal suitability of our approach for art images, we have applied it to the People-Art [86] test set, where we achieved an average precision (AP) of 0.56 (label = "person", confidence cutoff = 0.25, intersection over union $0.5 - 0.95$). In comparison, recent work [43, 76] reported APs of 0.36 and 0.44, respectively. We therefore concluded that our approach is suited for object detection in the National Gallery's collection.

4.1 Defining Custom Labels

In order to minimize errors, we built a dataset consisting only of digitized artworks from the museum collection that were labeled as "painting" in the collection metadata, resulting in a set of 6,750 artworks. Initially, we applied the approach described above, using the 80 object categories in COCO. This gave impressive results: The system could recognize a wide range of objects even in crowded scenes (Fig. 3a) and with unclear depiction styles (Fig. 3b). However, the system appeared to have a bias towards modern object categories: for instance, mislabelling some old books as *suitcases* (Fig. 3c) or the shield of a female warrior figure as a *handbag* - in the latter case perhaps also revealing a gender bias.

In order to mitigate this problem we pivoted to IconClass, a comprehensive index for classification of objects depicted in art images [19]. We needed to construct a custom set of labels, as IconClass contains over 28,000 unique concepts, whereas our system could only accept a maximum of 120 labels. We iteratively explored categories and labels from IconClass and observed the frequency of such objects in a random subset of the National Gallery's paintings. As the first iteration of object detection with the COCO labels had shown a strong dominance of the label *Person* (44% of all the detected objects), we prioritized including various labels relating to people and clothing. However, we omitted labels for small details such as mouth and eyes, as we expected such objects would most often be too small to crop in high-resolution and, therefore, difficult to apply in our user interface. Furthermore, we also included several labels relating to themes we observed occurring often in the artworks, such as religious themes, architecture, food items,

musical instruments, furniture, weapons, vehicles, and nature. This process resulted in a list of 120 labels, as seen in Table 1.

4.2 Selecting a Subset

Based on 6,750 paintings from the museum collection and the aforementioned 120 labels, a total of 109,145 objects were detected on 6,477 of the paintings. Analyzing the data revealed a skewed distribution of objects with 4 categories (*Human*, *Nature*, *Architecture*, and *Clothing*) representing more than 70% of the total objects detected. Similarly a high variance existed in the number of objects detected per label: the object *Man* had an instance count of 5,975 whereas *Bird Cage* only had a total count of 5. Due to technical constraints, we needed to reduce the dataset to one-tenth of the total data. In order to get a more uniform representation of objects, we defined the subset by retrieving up to 100 object instances per label, selected based on highest confidence level. This resulted in a dataset consisting of 10,775 objects detected on 3,906 of the paintings from the collection.

Finally we developed a script that cropped the detected objects from images of the original paintings, leaving us with an image collection of the singular objects as illustrated in Figure 4. These individual object images were used as a key component throughout the design and development of the interactive application.
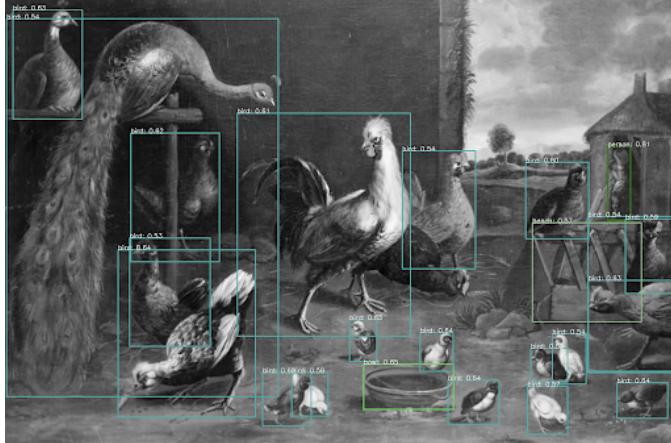
5 SMKEXPLORE

Working with the data as described above gave us valuable insights that helped inform our design process. In combination with the results from the object detection, we drew inspiration from existing literature on designing for exploration (as summarized in Section 2) to create SMKExplore: A web application that allows users to browse and explore a digital art collection through the objects detected in the paintings, as well as to use the objects to create new images using a generative image algorithm. In the following subsection we first present the design process and insights leading up to the final design, which is presented in the subsequent subsection.

5.1 Design Process

The design process leading to the design of SMKExplore ran from January-August 2023. The design and development was carried out by the three first authors of this paper, and was structured as a combination of UX Design and Agile Software Development using SCRUM, thus combining a total of five design sprints and software development sprints based on the model presented in [36].

From the outset, our aim was to explore how object detection data could be used to facilitate exploration of a digital museum collection, using the object data to create new entry points and alternative ways of browsing. The data processing described in Section 4 was conducted prior to the design process. Exploring the data helped us frame the design space and guide the process.

During the initial phases we investigated opportunities and qualities inspired by techniques from interaction-driven design as presented by [54]. This led us to design a preliminary concept that fostered immersive interactions with the objects in a 3D gallery which we envisioned implemented in WebVR. However, we had
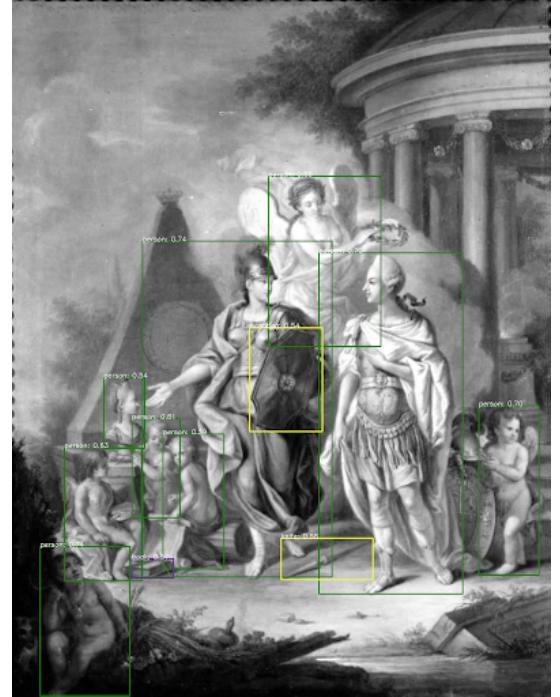
(a) The painting "Between hens, roosters and chickens, a few peacocks can be seen" (1749 – 1848) by Unknown Artist from the collection of the National Gallery of Denmark, analyzed by GLIP. Even in a crowded composition with many overlapping objects the algorithm is able to correctly label a large number of objects.

(b) The painting "Lillehammer. February. Timber Sledges" (1937) by Harald Leth from the collection of the National Gallery of Denmark, analyzed by GLIP. The algorithm can detect objects even in a relatively non-realistic painting style.

(c) The painting "Saint Jerome" (16th century) by Unknown Dutch Artist from the collection of the National Gallery of Denmark, analyzed by GLIP. The books are mislabelled as "suitcase".
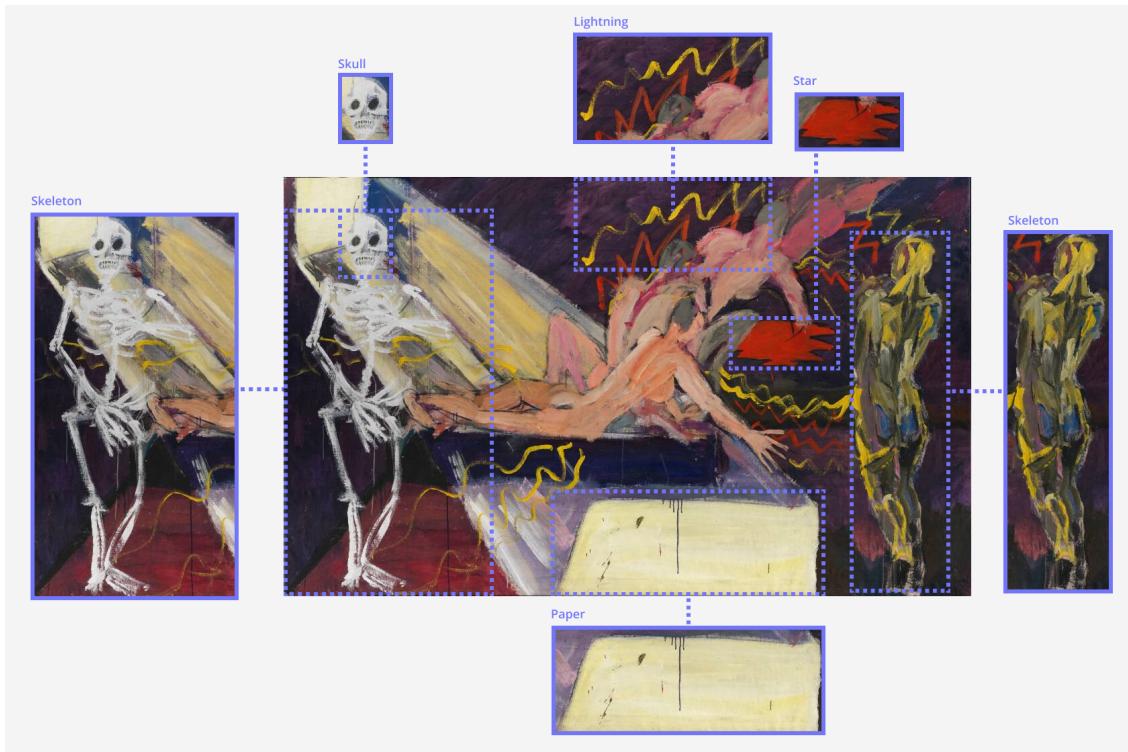

(d) The painting "Christian VII as the protector of art and science. Allegory" (1770) by Georg Mathias Fuchs from the collection of the National Gallery of Denmark, analyzed by GLIP. The shield of the warrior figure is labelled as "handbag".

Figure 3: Comparison of image analysis results.

Table 1: Custom set of labels that were used for object detection with the GLIP model.

Category	Labels
Animal	Bird, Butterfly, Cat, Chicken, Cow, Dog, Donkey, Fish, Horse, Insect, Mouse, Rabbit, Reptile, Sheep
Architecture	Bridge, Castle, Church, Door, House, Mill, Pillar, Staircase, Window
Christianity	Angel, Cross, Devil, God, Jesus Christ, Saint, Virgin Mary
Clothing	Bag, Belt, Cane, Crown, Dress, Gloves, Hat, Jewellery, Mask, Shoes, Tie, Umbrella
Food	Apple, Banana, Bread, Cheese, Grapes, Lobster, Orange, Pineapple, Vegetable, Watermelon, Wine
Furniture	Bathtub, Bed, Chair, Easel, Sofa, Table
Human	Baby, Child, Face, Hand, Man, Woman
Instrument	Drum, Flute, Guitar, Harp, Piano, Violin
Interior	Bird Cage, Book, Bottle, Bow, Cup, Drapery, Flag, Globe, Lamp, Mirror, Paper, Vase
Nature	Bush, Cloud, Fire, Flower, Lake, Lightning, Moon, Mountain, Plant, Rock, Sea, Sky, Sun, Tree
Occultism	Demon, Ghost, Skeleton, Skull, Star
Vehicle	Airplane, Bicycle, Boat, Car, Carriage, Ship, Train, Wheel
Weaponry	Armor, Arrow, Bow, Firearm, Hammer, Helmet, Rope, Shield, Spear, Sword,

Figure 4: Objects detected in *"The King and Queen Surrounded by Swift Nudes"* by Inge Ellegaard (1982).

concerns about complications regarding usability and motion sickness (cf. [15]) as well as technical feasibility, and chose instead to develop a simpler 2D concept.

Based on insights from the research literature, we established a set of design principles for our system. First, as suggested by [90] and [26] we emphasized finding a balance between overview of the data as well as opportunities to explore information in detail. This led us to establish a clear information hierarchy that allows the user to gain an overview of the collection, while we also designed pathways to detailed information on each artwork. Dörk and colleagues [26] furthermore inspired us to use *visual cues* to design navigational paths and enhance the possibility of serendipitous discoveries. Inspired by [21, 90] we additionally decided to enable users to save objects that caught their interest, as a means to revisit parts of the collection they enjoyed.

Furthermore, inspired by [57, 89] we decided to cluster the objects based on similarity and to provide possibilities to filter the data

as a means to create overview and support various information-seeking strategies, such as comparing, combining and evaluating. Insights from [57, 81, 88] led us to design for accessing the collection through multiple entry points and navigating via multiple paths (cf. [26]), in order to enhance the sense of free exploration and varied forms of interaction with the items of the collection.

In addition to this, past research [39, 51] has highlighted the benefits of using playful elements to advance and encourage non-expert user engagement and maintain user attention. Thus, we decided to include a playful element in the application in the form of an interactive canvas where users, with the help of generative AI, can create their own art with objects of personal interest from the collection.

With the aim to establish a clear connection between the application, SMKExplore, and the National Gallery, we defined the visual identity with inspiration from the museum's website. In particular, we drew upon the color scheme, font types, square frames and the layout of the Painting Screen (Fig. 5).

The design was revised and implemented through five iterations (sprints), informed in part by usability testing and in part by technical aspects, until an initial prototype was tested on users in May 2023. Based on findings and feedback from this test a last revision of the design was conducted in the beginning of August 2023.

5.2 Final Design

In the following, we present the version of SMKExplore (Fig. 5) that was used during the evaluation as reported in section 6.

SMKExplore allows *"bottom-up search"*, where the user encounters the digital collection moving from details (i.e., objects) to the full-sized painting they appear in originally. The design concept focuses on navigating the collection based on thematic interest, shunning more traditional goal-oriented search. The aim is to allow users to compare depictions of similar objects in different paintings across time periods, styles, etc., and to help users discover artworks and details they otherwise might not have noticed.

When the user enters the application, they are met by the Home Screen, which showcases a slider with three examples of objects that have been detected in the collection. By clicking the "Start Exploring" button, they are led to a screen displaying the 13 categories defined for the objects in the data processing (see Table 1). The Category Screen constitutes the first level of an information hierarchy, in which the objects are presented only as a high-order category.

Once the user chooses a category they are directed to the Object Screen, where all objects within that particular category are displayed. Users may choose to filter the category further by selecting a label, thus being presented only with images of one object label, for instance skulls in the category Occultism.

Clicking on an object leads the user to the Painting Screen, which presents the entire painting on which the object appears. Alongside the painting more detailed metadata are provided, such as title, artist, technique, production year, and color palette. Other detected objects on the painting are also displayed, making it possible to navigate to other types of objects. Detected objects can also be discovered by hovering over the painting itself.

Users can save objects that catch their interest to a list of favorites. The list of favorites provides users with an opportunity to revisit these parts of the collection later on. The saved objects can also be used to create new imagery using the interactive canvas. The canvas utilizes the outpainting function of OpenAI's DALL·E 2 API, which generates an image based on visual input(s) and a text prompt. The user may place objects on the canvas, resizing them as needed and leaving a generous amount of white space. Once they are satisfied with the composition they type in a text prompt describing, for instance, the desired image's style or theme. When the image is generated, the user has the possibility of creating a new image using the same or other objects, searching the collection further, or comparing their image to the original paintings by navigating through the list displaying the objects that were used on the canvas.

6 EVALUATION

SMKExplore was evaluated on-site at the National Gallery 11-12 Aug 2023 during museum opening hours. Two authors were placed in the museum's foyer, inviting visitors to participate in a short user test. All test participants were visitors we encountered at the museum and were unknown to us beforehand. In total 22 participants interacted with the application and were interviewed, aged 18 to 59 (median: 26), 9 males and 13 females. The participants represented 14 nationalities across Europe, North America, Asia and Oceania.

6.1 Procedure

Before interacting with SMKExplore, the content and functionality were briefly explained to the participants. Subsequently, they were instructed to explore the application freely, without time constraints. Towards the end of the session the participants were asked to generate an image on the Canvas Screen. The interactions with the application were documented using system logs as well as screen recordings.

A semi-structured interview was conducted immediately following the participants' interaction with the application. The interviews focused on the participants' experience of utilizing detected objects as the primary visual entry point to the collection. The participants were also asked about their general thoughts on using AI in art contexts and whether they noticed or reflected upon mislabelled objects. The interviews were recorded using an audio recorder. All participants were informed about the data collected and signed an information statement in accordance with the university's policies and the General Data Protection Regulation (GDPR).

The following analysis is based on findings from the system logs, interviews, and observations during the tests. Screen recordings have been used to supplement and clarify some details. The system logs were analyzed through descriptive statistics. The interviews were analyzed through thematic analysis, following the phases and guidelines presented by Braun and Clarke [13]. Additionally, following the guidelines by McDonald et al. [58], consistency and validity of qualitative results was ensured through agreement among the authors by collaboratively developing the coding schemes through iterative discussions.

The initial phase of the thematic analysis was conducted by the first and second authors, who both had been part of the design team. They initially familiarized themselves with the data by transcribing

The Home Screen showcasing one of three sliders

The Category Screen

The Object Screen displaying the category 'Human'

The Object Screen with saved objects in the Favorite List

The Painting Screen displaying the full painting, with objects hovered

The Canvas Screen with objects placed on canvas

The Image being generated with help from GenAI

Generated Image with reference to the objects used

Figure 5: Screens from SMKExplore.

and iteratively reading the interviews. From the transcripts, one author generated the initial codes for all interviews utilising the software ATLAS.ti. This amounted to a total of 158 distinctive codes related to our research question. These codes were assessed and revised by the second author. Subsequently, 16 groups of codes were established combining patterns such as "questioning own interpretation", "noticing details through objects" and "surprised by personal interest". In the following phases two additional authors, who did not take part in the design process nor the interviews, took part in the analysis and discussion to broaden the perspective. Through this process we found several overlaps in the 16 groups, which led us to narrow down to six potential themes touching upon object representation, attention to detail, interest driven search and discoveries, mislabelling and interpretation, and lastly contextualising the objects. Through an additional, conclusive, phase of analysis the themes and underlying patterns were reevaluated and the final six themes were established. These will be unfolded in the following.

6.2 Overall Experience

In general the participants became immersed rather quickly in browsing through the objects, looking concentrated throughout the process. When interacting with the canvas, they became more talkative towards the end of their session and seemed both entertained and surprised by the resulting image they generated. They spent between 4 and 15 minutes with the application (median: 8 minutes). The majority wanted to continue their exploration or said they could imagine themselves trying it again.

Generally, the participants interacted intuitively with the different features. They initially navigated from the Home Screen to the Category Screen and onward to the Object Screen. On their first visit to the Category Screen, they quickly (on average 10 seconds) found a category of interest to investigate further. Only 3 participants needed guidance on how to save an object to their favorite list. In addition, 4 participants said the functionality of the canvas could have been more apparent to them, while 3 participants said they could have used more tips on how it works. These issues primarily concerned confusion about how to resize objects on the canvas.

20 of the 22 test participants said they enjoyed using the application and described the overall experience with words such as "fun", "interesting", "intuitive", "enjoyable", and "ludic". 2 participants had somewhat more mixed feelings.

6.3 Representation of Objects

The participants generally spent the most time on the Object Screen (see Table 2), which shows all the detected objects for a specific label or category. During the interviews, several participants shared that exploring the collection through objects made them reflect on the different depictions of these objects and the wide range of motifs represented in the collection. *"It was nice to see bikes from different paintings [...] I have never thought about looking at paintings and being like, oh, this is a bike here, and there is also a bike there"* (P1). Several participants mentioned the wide variety of object types as an element of surprise to them: *"Wow, there are many of these objects I have never noticed in many of the artworks before"* (P7).

Table 2: The average time spent on each screen of the application.

Screen	Average time spent
Object Screen	3 Minutes & 22 Seconds
Canvas Screen	3 Minutes & 4 Seconds
Painting Screen	1 Minute & 27 Seconds
Category Screen	27 Seconds
Home Screen	14 Seconds

In addition to the rich variation of objects, the participants commented on the effect of seeing the different depictions of these objects side by side on the Object Screen: *"Many motifs are reappearing. It makes sense, but when you see it like this, it is wild"* (P5). The participants shared how distinct representations of the same objects made them reflect on different styles of painting through time. *"It shows how different artists from different parts of the world, during different times have treated that object. Say, an apple would be very different in the Renaissance than today"* (P16).

6.4 Focusing on Details

When asked to describe their experience of accessing the collection primarily from the objects as opposed to the entire painting, several participants emphasized that it offered them a perspective on the artwork that made them notice things they would usually disregard. Removing the objects from their original context also made many aware of the complexity that goes into a painting, which they might not have discovered otherwise. In addition to this, several of the participants also expressed that experiencing the collection in this manner made them pay attention to what they were seeing and inspired them to look at the details more: *"I think you just become more thoughtful of what actually is happening in a painting like this and what is present"* (P1).

Some also suggested that focusing on details can serve as an interesting new way to discover the entire paintings, as browsing the objects made them aware of artworks that they had not noticed when going through the exhibition: *"[...] by going through details that maybe struck me, I also had the chance to pay attention to paintings that maybe I disregarded in the exhibition"* (P3).

While most participants enjoyed or found it interesting to experience the collection through the objects, some also expressed lacking the context of the objects as problematic or something they did not enjoy. Particularly, worries about losing the entire vision of the artist, were mentioned by those who would rather see the entire painting up front: *"I like the whole vision that the artist had rather than just a small piece of it that somebody else had decided I would look at"* (P4).

6.5 Interests and Discoveries

A recurring pattern in the participants' interview answers was the ability to explore the museum collection based on their personal interests. They shared reflections on how this influenced their navigation in the application and that they discovered patterns in what caught their attention: *"I learned what I am interested in when I look at art"* (P19). Several participants stated they had found new

and unexpected artworks by pursuing their interest in particular objects: *"I didn't think I would be interested in a painting of cows, but that was very surprising and interesting"* (P16).

8 of the 22 participants mentioned the categories as an element that helped them follow their interests while exploring the collection. On average, each participant visited the Category Screen 6 times during their session. Throughout the 22 sessions all categories were visited, however the popularity of the categories varied, as illustrated in figure 6.
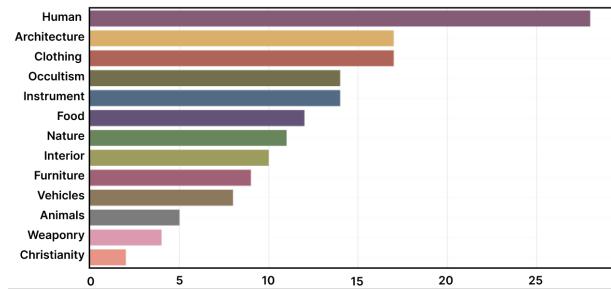

Figure 6: Total number of visits per category throughout the 22 sessions.

The log data revealed that 21 out of 22 participants explored the same category more than once during their session. In the interviews, multiple participants said they were drawn to unfold the content of the categories further, either going back and forth to it or by selecting filters in the category to narrow their search.

> *"I chose human, I think, which is a bit broad, and then I went into it and I was like, I am interested to see women in the collection, so I started to select those to go deeper into a more narrow category."* (P19)

Similarly, the log data, showcasing which objects the participants saved during their session, supports the notion that participants wanted to investigate categories of interest more in-depth. The participants on average saved 6 objects during their session, and most participants (17) saved more than one object from the same category. One particularly eager participant saved 25 objects, all from the category "Weaponry".

6.6 Mislabelling and Interpretation

Out of the 22 test participants, 12 said that they noticed objects that were not correctly tagged. However, when asked if this influenced their experience none of the 12 participants said that it bothered them. Interestingly, participants seemed to express a large degree of understanding and perhaps even sympathy for the algorithm's mislabelling. Some suggested that it couldn't necessarily be determined what the correct label should be, pointing out that strict interpretation is not always possible. Others suggested that it was understandable why the object detection model would classify a given object as something other than it actually is because of its visual similarities, for instance a spear being labelled as a flute because of similar shape and colour:

> *"Especially with the flute, he showed a lot of pictures of long, thin objects, which I do understand why he would think is a flute. And I always find this very interesting, because this is quite difficult, especially analyzing photos. It's quite difficult for artificial intelligence to do it. And as a human, you take a single look at it and you instantly know."* (P12)

Several of the participants that encountered incorrect labels found it interesting and said that being confronted with the AI's "interpretation" made them question their own interpretation:

> *"For example, for a mirror, there was one that was a full painting. That's why I clicked on it, I think, at some point, because I was like, that's not really a mirror. But I thought it was interesting because it kind of made you question whether it was you or the AI that was making a mistake, or it made you explore that."* (P6)

Several said that the incorrect labeling made them reflect or think differently about the potential visual interpretations of a particular object when taken out of context, thus challenging their own interpretation:

> *"I guess it made it a bit more exciting, because you didn't know if it was going to be the actual thing. One of them said it was a guitar, but it was open-heart surgery. It looked like a guitar. It was quite interesting."* (P14)

Interview participants often speculated why the AI had labeled an object the way it had. Particularly, participants noticed discrepancies in how the AI labelled the objects in contrast to how a human might interpret them. Similarly, people also speculated on what shared visual characteristics objects might have and how these shared characteristics would lead the AI to recognize a particular object incorrectly, but consistently: *"I began to think about what the AI saw to think it was that object and what similarities it would have to the other objects"* (P10).

One participant stated that they thought the flaws were "charming" and that it made the AI seem more human. The same participant, however, also reflected on being misled and becoming suspicious of whether they could trust the AI at all when noticing a wrongly labeled object:

> *"[...] all of a sudden, I became very aware that I suddenly couldn't trust it, that something I had clicked on and that it almost had me convinced was a skull, and I was like maybe it isn't that at all. That I am looking at it all of a sudden as an abstract, kind of distorted skull, but maybe it isn't."* (P7)

6.7 (Re)contextualizing the Objects

Second to the Object Screen, the participants spent the most time on the Canvas Screen (Table 2), on which they could generate a new image using objects they had saved. When asked if playing with the objects on the canvas contributed to their interest in exploring the art or their experience of the artwork, most participants shared reflections concerning (re)contextualization, composition in paintings, and piecing together different styles and details.

"I think it's just interesting to maybe take some details or take things in general, change the context and see what happens by reframing this relation." (P3)

The opportunity to play with positioning objects on the canvas and creating new imagery made several participants contemplate how the objects were represented in the collection.

"Putting these objects together, you could give it your own context, and that also changed the way the objects were in the collection [...] it is an interesting way of combining images, not just generating completely new images, but combining specific objects from images to a completely new image was an interesting experience." (P10)

By placing the objects in a new and different context, the participants expressed that they became aware of compositional aspects of the art. This awareness concerned the composition of existing paintings and the composition they were creating on the canvas: *"For instance, in the Baroque exhibition, there were a few areas in the paintings where there weren't a lot going on, and that didn't make sense to me. It made me think you could have added something"* (P8).

Furthermore, several participants also reflected on how combining objects from various parts of art history could reveal differences and similarities between styles across time:

"I thought it was cool how you could compose different pieces together. And I think if you did a bunch of different art movements together, you could learn a lot about how they evolved and how they could be intertwined." (P6)

During testing, we noticed that the canvas increased the inclination of the participants to explore the collection further. All but one, spent a large enough amount of time exploring the various objects that they were asked to stop exploring and generate an image. When asked to do this, some asked for more time to find other objects. Others reflected on this aspect during the interviews, in which they expressed that had they been given more time, they would have gone back and collected more or different objects to use in their image, indicating a desire to investigate the collection even further.

When asked if anything unexpected happened while interacting with the application, the most frequent answer was that they were surprised by the resulting image they had generated on the canvas:

"I was really surprised to see an actual painting that could hang in a museum [...] The painting that AI generated reminded me really of one of my favorite artists. But none of the pictures were from him. So that's quite interesting. I really like that." (P22)

Participants were in particular surprised by the way the outpainting functionality worked, saying that they did not expect the canvas to take the size and position of the object into account in the final result (although in fact most participants had opened an instruction screen which explained and visualized how the Canvas Screen worked). While several participants were familiar with other generative image systems such as Midjourney, participants were generally not familiar with outpainting.

7 DISCUSSION

In the following we will reflect on four themes coming out of our design process as well as the testing and evaluation presented above: How the system affected the participants' view on the artworks, how the labelset influenced the design, the participants' experience of errors in the labelling, and the participants' experience with the Canvas Screen. Finally, we reflect on some implications for design.

7.1 Experiencing Art Through the Lens of AI

As proposed by Lev Manovich, machine learning offers new ways to experience art and visual culture by enabling the exploration of large collections and patterns, contrary to the traditional approach of inspecting artworks individually [56]. Through the evaluation we found that participants reflected on patterns in the museum collection, specifically objects recurring over multiple paintings. Seeing the recurrence of these objects side by side made them reflect on the motifs repeatedly depicted by artists and their various styles. In addition, the evaluation highlights that participants were inspired to focus more on details by exploring the collection through objects instead of full-sized paintings.

It is particularly interesting to consider the participants' experience in light of the fact that they encountered our prototype after having visited the physical exhibitions at the museum. Many users expressed that they noticed new details and recurring objects in the art when exploring the application, which they had not noticed in the museum exhibition beforehand. Thus, the experience of exploring the art collection based on the objects detected by the machine learning model offered participants a new perspective on the art collection compared to the physical visit to the museum exhibition.

7.2 Labelling

The labels applied in an object detection model determine what objects can be detected - what the computer vision system can "see" in the image. The process of constructing the list of labels described above in section 4 demonstrates the importance of building a set of labels that enable the model to detect the most relevant objects. Our setup was limited by the amount of labels that could be fitted as an input string to the GLIP model - 120. This means that the model would not be able to label objects that were not included on the list in table 1 - thus the objects labelled by the model represents only a partial view of all the objects in the collection, limited by the selection we had made. This may help explain the mislabelling of some details, such as that mentioned by participant 4 in section 6.6, mistaking surgery for a guitar (see Figure 9): Since the model did not include any labels relating to surgery, it could not label it correctly - instead settling on a label with some visual similarity but very different meaning. Indeed, in our first iteration of the object detection analysis using labels from COCO (see section 4) this same detail was labeled 'bowl'; COCO does not have a label called 'guitar', nor any other labels that seem appropriate for this detail.

Given more time and technical resources, it might have been possible for us to increase the amount of different labels, by re-running GLIP over the artwork collection with different sets of labels each time. This might result in a dataset with several different labels for the same object, which would either have to be disambiguated through a separate process - or we could simply adjust the design

APPLIED OBJECTS

Table from *Les brioches* by Emilie Othon Friesz

Sun from *OH MY GOD!! That flower is a fucking vacuumcleaner!* by Christian Schmidt-Rasmussen

Bed from E*n Families genvordigheder* by Wilhelm Freddie

Bed from *Le moulin brûlé* by Maurice Estéve

PARTICIPANT PROMPT

"The most amazing paintings are the ones that are lively and full of color"

Canvas Composition

Generated Image

Figure 7: Canvas composition, applied objects, prompt and generated image by Participant 22.

to allow for multiple (possibly contradictory) labels for the same object, inviting users to reflect on the resulting ambiguity. We can only speculate on how such a larger set of labels would affect the user experience: One might hope that it would allow for an even richer experience with more nuance and more opportunities for surprising discoveries. However, it is also possible that adding more labels would increase the proportion of mislabelling, as the system would have to contend with a larger amount of categories overall while applying a limited ontology for each run of the object detection algorithm.

Future developments of GLIP and similar algorithms may lead to an increase in the number of labels that can be applied at a time. However, it is unlikely that this will remove all limitations on the ability of vision algorithms to detect objects in artwork. First, it may take some time before models can include a sufficiently large number of labels without forgoing precision: A comprehensive classification like Iconclass contains over 28,000 unique concepts, whereas the Getty Art & Architecture Thesaurus contains 73,831 concept records, over 600 times larger than the number of labels used in our setup. Furthermore, even if a future system would allow a very long list of labels, there would remain some fundamental challenges with mapping between concepts and images precisely and comprehensively. Debates about the large image classification dataset ImageNet [69] have demonstrated that classifying images of humans with labels from a lexical database can lead to unintended consequences and controversy [20]. Similar complications may occur in object detection, as some concepts may mean different things at different times and cultural contexts. For instance, gender labels have acquired new meanings in recent times, adding nuance to what was formerly mostly considered a binary concept. Many other concepts relating to technology, societal roles, norms, institutions, or culture have changed meaning over time and in different societal and cultural contexts. Ciecko and colleagues [17] provide a striking example of how the use of image classification might inadvertently

trigger controversy: An image of iron ankle manacles from Australia's convict history that is labeled as "Fashion Accessory" and "Jewelry" by a commonly used image classification algorithm. One could easily imagine that if a similar mislabelling were to occur in the collection of a museum relating to the history of slavery or the Holocaust, this could be offensive and hurtful for visitors and highly problematic for the museum.

In the first version of our labelset we included the label 'non-binary person', in order to accommodate a broader variation of gender identities and supplement the labels 'man' and 'woman'. However, the results made us question the classification. GLIP returned 210 bounding boxes with this label, of which the majority were depictions of children and/or nude people with displeased or uncomfortable facial expressions. We judged this to be potentially both inaccurate and harmful, and for these reasons we omitted this label from the final version of the labelset with the consequence that our application only provides two labels reflecting gender. This is unfortunate. While we do not have ground truth data available that could help us verify whether there are (few or many) images of people in the collection that should be tagged as non-binary person, the absence of this label might render invisible to the model a broader variety in gender identities. However, it seems that capturing nuances in gender presentation is difficult to do with the technology used in this study. It is worth reflecting on whether it is possible at all to classify gender with computer vision techniques that rely solely on visual appearance. For future work, it could be interesting to explore other ways to classify motifs of people in art instead of (or in addition to) gender, e.g. by clothing, hair, age etc.

7.3 Mislabelling and Trust

Seen in light of the challenges with labelling objects correctly outlined above, it is striking that the test participants generally trust our system's algorithmic labeling. Only 12 of the 22 participants noticed objects that were incorrectly labeled, even though mislabeled
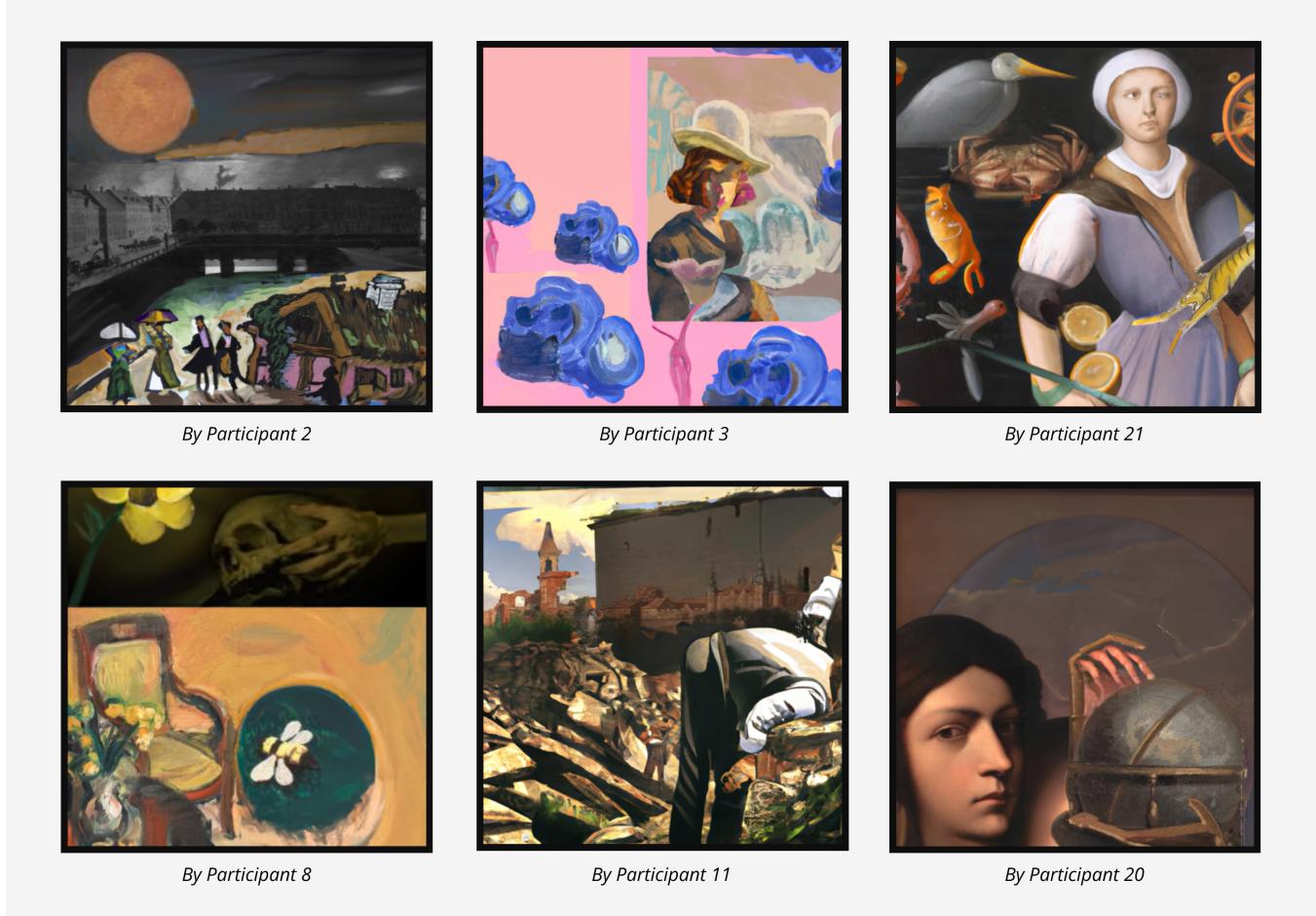
By Participant 2   *By Participant 3*   *By Participant 21*

By Participant 8   *By Participant 11*   *By Participant 20*

Figure 8: Example of images generated on the canvas.

objects – or at least questionably labeled – can easily be found in most categories. (Consider, for instance, some of the objects shown in Fig. 1 and 4.) Those who did notice questionable labels often seemed willing to offer explanations on behalf of the algorithm, one participant even personifying the algorithm: *"...he showed a lot of pictures of long, thin objects, which I do understand why he would think is a flute"* (P12). Others suggested the mislabelling made them question their own interpretations - which is well aligned with the typical ideals of art education in museums, which often emphasize questioning one's preconceptions and interpretation and opening oneself up to seeing artworks in different ways.

While these observations align with other research pointing towards a tendency to overtrust in AI systems [6, 41, 66], we do not have data to assess clearly why the participants were so willing to trust the system or make excuses for its errors. However, there is a striking similarity with the observations done by Benford and colleagues when exploring the use of emotion detection AI in an art museum: "...visitors tended to construct post hoc rationalizations of their emotional experience that agreed with, or at least accommodated, the 'results' reported by the system, even when this differed from their initial reflections" [6, p.12].

We can only speculate about why visitors appear so willing to trust in the output of these computer vision systems – object detection in our case, emotion detection in [6]. First, several factors in the presentation at the museum may inspire trust among visitors: The system is presented to them by university researchers, which may influence the participants to see the system as trustworthy and authoritative; and the context of the museum as a highly trusted institution may add to this impression. Second, visitors may be extra understanding towards the system's errors due to the application domain, as interpreting art is both a difficult task and often seen not to have a single correct answer and one to which computer systems are not commonly applied. Third, given the large amount of visual information in the interface and the focus on exploration, it is possible that some mislabellings - like those showing unclear images and shapes - were overlooked as "noise" as participants focused on the higher resolution, and thus, more clear and recognizable images.

To some degree, the discussion above has presupposed that there is a correct and an incorrect label for each object in an art image. That assumption might be challenged in several ways. First, art images frequently appear ambiguous and resist interpretation from

Figure 9: *Surgery* by Jørgen Thomsen (1943-44). The detail in the middle lower part of the image was mislabelled as a "guitar" by GLIP.

Figure 10: Classification of objects is sometimes affected by the context. In this image, five sea-faring vessels of varying sizes are correctly classified as 'boat'. However the small, left-most blue bounding box surrounds the outline of the castle Kronborg which is wrongly classified as 'boat' - presumably due to the vicinity to the sea and other boats. Artwork: *The Russian Ship of the Line "Asow" and a Frigate at Anchor near Elsinore* by C.W. Eckersberg (1828).

audiences, art critics and scholars alike. For instance, should Salvador Dalí's "Lincoln in Dalivision" be classified as depicting the face of a bearded man, or a naked woman standing by a window? Much art is even more abstract and difficult to interpret unambiguously; and as ambiguity is a central quality of art, removing ambiguity from art is not a desirable goal. Second, one might argue that computer vision may encode ways of seeing objects that subtly differ from our assumptions about how objects should be seen - and which may offer interesting perspectives. For instance, Leahu demonstrates that a vision algorithm may end up encoding not just objects as discrete entities, but also aspects of the *relations* that constitute them - such as when a neural network trained to recognize dumbbells also encodes the images of arms holding the dumbbells [46]. For Leahu, this raises the possibility of "ontological surprises" - that computer vision algorithms may reveal unexpected relations between objects. In our analysis with GLIP we could sometimes see that the context surrounding an object might affect the algorithm's classification of objects, as in the example in Figure 10. It would be an interesting challenge for future research to explore whether this sensitivity to context - or other particular aspects of the way computer vision encodes objects - could be used to help art viewers or even art scholars discover new ways of seeing art.

7.4 Creating New Images

As highlighted in Section 5, we included the canvas feature to support user engagement in exploration of the collection. Through the test, we found that the canvas encouraged participants to continue their search for objects: When given the task of using the Canvas

Screen to make an image, many users were eager to go back and look for more objects they could use to make images. Several also said they would have liked to spend more time going back and forth between the Canvas Screen and the collection. One particularly eager user (P8) spent a long while creating multiple images and would only stop when we insisted that we needed to end the testing session. These observations confirm that the generative feature helped support engagement.

In addition, we found that generating an image through positioning and combining objects on the canvas made the participants reflect on the artworks' context, time periods, styles, details, and composition. With outpainting, the participants were able to visually experience how styles and details can be merged into something new that goes beyond the original context of the object(s) (Fig. 7). This suggests that outpainting may have a promising potential as a device to facilitate practice-based learning about these dimensions of visual art in a manner that would be much more rapid and less dependent on practical skills than traditional exercises in drawing and painting.

7.5 Implications for Design

Based on the observations outlined above, we suggest a few topics that might be relevant to consider for designers working with object detection in digital art collections.

First, designers should pay close attention to the labelset used for object detection. As long as object metadata for the collection is unavailable or incomplete, it will be difficult to assess - other than by trial and error - which types of objects are prevalent in the collection and can be detected reliably. However, working with subject experts like museum curators or art historians might help identify appropriate labels, particularly when working with collections dominated by older art.

Second, designers might be interested in deliberately introducing flaws or errors in labeling as a way to provoke reflection and nudge users to question their own interpretation. However, our observations suggest that such errors might need to stand out strongly to ensure that users will notice them and identify them as errors. If users place trust and even some sympathy with the algorithm, then designers who wish to inspire critical reflection about the algorithm will need to work carefully on communicating to users that the algorithm is not necessarily to be trusted fully. Designers might explore ways to include confidence measures or other ways of visualizing uncertainty in labelling, however this would need to be balanced against the need to avoid disrupting the aesthetic of the art presentation, which is a strong design norm in art museums. Alternatively, designers might create deliberately ambiguous presentations of the algorithm's outputs in order to provoke critical reflection, following tactics similar to those presented in [30, 72].

Third, future designers and art educators might use generative systems (such as in our Canvas Screen) to facilitate learning about visual art and composition. For instance, one might use a more narrowly curated set of paintings based on time period or style to provide insight into frequently depicted motifs and typical composition. This could be supported by predefined text prompts that exemplify styles, details, and compositions recurring within the particular collection of artworks, allowing the user to explore objects and visually experiment with image generation while working towards more focused learning outcomes.

8 CONCLUSION

We have presented an approach to using object detection to facilitate exploration of a large digital art collection. First, we have demonstrated that recent leaps in computer vision, in particular the emergence of multimodal models like CLIP and GLIP, has made it feasible to use object detection on digitised art images with sufficient precision to support a meaningful and satisfying user experience for a general art-interested audience such as the visitors to the National Gallery of Denmark.

Second, we have presented the design of a web application that use the object detection data as basis for an interface that allow users to explore the collection in a novel way, using objects of interest as an entry point, and using a generative system with outpainting to facilitate creative and playful exploration. The evaluation has demonstrated that this interface has inspired test participants to see the art in a new light and discover new things about the art. We have highlighted the importance of constructing an appropriate labelset

for the object detection, and drawn awareness to the participants' tendency to trust the system's output and perhaps overlook errors in the object labelling. Finally, we have suggested some design implications that might inform future work with object detection in artwork.

Our study has been limited to only artworks classified as paintings in the museum collection. Further research would be needed to explore whether the technology can be applied across diverse media types such as drawings and sketches, sculptures, photos and video, engravings, and so on. Furthermore, there is a need for cross-disciplinary research collaboration with art experts (for instance in art history or the digital humanities) to explore the aesthetic and pedagogical implications of extracting details from their original context in the artworks and presenting them to users as lists of objects from a variety of different artworks, styles, periods and artistic agendas. While such an approach may seem problematic for some curators as it means that image fragments are presented detached from their original context in the artwork, our study has demonstrated that it has the potential to inspire and engage museum visitors to discover and learn more about art. Tapping into this potential would be beneficial for both museums and their visitors - and would break new ground for the use of computer vision in art education and dissemination.

REFERENCES

[1] 2018. SMK Open: Setting art free. https://www.smk.dk/en/article/smk-open/
[2] Yali Amit, Pedro Felzenszwalb, and Ross Girshick. 2020. Object detection. *Computer Vision: A Reference Guide* (2020), 1–9. https://doi.org/10.1007/978-3-030-03243-2_660-1
[3] Sofian Audry. 2021. *Art in the age of machine learning*. MIT Press. https://doi.org/10.7551/mitpress/12832.001.0001
[4] Kevin Bacon. 2019. AI as provocation rather than solution. https://gifting.digital/brighton-museum/
[5] Kobus Barnard, Pinar Duygulu, David Forsyth, Nando De Freitas, David M Blei, and Michael I Jordan. 2003. Matching words and pictures. *The Journal of Machine Learning Research* 3 (2003), 1107–1135.
[6] Steve Benford, Anders Sundnes Løvlie, Karin Ryding, Paulina Rajkowska, Edgar Bodiaj, Dimitrios Paris Darzentas, Harriet Cameron, Jocelyn Spence, Joy Egede, and Bogdan Spanjevic. 2022. Sensitive Pictures: Emotional Interpretation in the Museum. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems (CHI '22)*. Association for Computing Machinery, New York, NY, USA. https://doi.org/10.1145/3491102.3502080 event-place: New Orleans, LA, USA.
[7] Jesse Josua Benjamin, Arne Berger, Nick Merrill, and James Pierce. 2021. Machine Learning Uncertainty as a Design Material: A Post-Phenomenological Inquiry. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*. 1–14. https://doi.org/10.1145/3411764.3445481
[8] Lennart Björneborn. 2017. Three key affordances for serendipity: Toward a framework connecting environmental and personal factors in serendipitous encounters. *Journal of Documentation* 73, 5 (13 Oct. 2017), 1053–1081. https://doi.org/10.1108/JD-07-2016-0097
[9] David M Blei and Michael I Jordan. 2003. Modeling annotated data. In *Proceedings of the 26th annual international ACM SIGIR conference on Research and development in informaion retrieval*. 127–134.
[10] Margaret A. Boden and Ernest A. Edmonds. 2019. *A Taxonomy of Computer Art*. Chapter 2, 23–59. https://doi.org/10.7551/mitpress/8817.003.0005
[11] Ian Bogost. 2019. The AI-Art Gold Rush Is Here. https://www.theatlantic.com/technology/archive/2019/03/ai-created-art-invades-chelsea-gallery-scene/584134/ Section: Technology.
[12] Padraig Boulton and Peter Hall. 2019. Artistic Domain Generalisation Methods are Limited by their Deep Representations. *arXiv:1907.12622 [cs]* (July 2019).

http://arxiv.org/abs/1907.12622 arXiv: 1907.12622.

[13] Virginia Braun and Victoria Clarke. 2006. Using thematic analysis in psychology. *Qualitative Research in Psychology* 3, 2 (2006), 77–101. https://doi.org/10.1191/1478088706qp063oa

[14] Eva Cetinic and James She. 2022. Understanding and Creating Art with AI: Review and Outlook. *ACM Transactions on Multimedia Computing, Communications, and Applications (TOMM)* 18, 2, Article 66 (feb 2022), 22 pages. https://doi.org/10.1145/3475799

[15] Hwei Teeng Chong, Chen Kim Lim, Minhaz Farid Ahmed, Kian Lam Tan, and Mazlin Bin Mokhtar. 2021. Virtual reality usability and accessibility for cultural heritage practices: Challenges mapping and recommendations. *Electronics* 10, 12 (2021), 1430.

[16] Brendan Ciecko. 2017. Examining the impact of artificial intelligence in museums. *Museums and the Web* (2017). See, https://mw17.mwconf.org/paper/exploring-artificial-intelligence-in-museums.

[17] Brendan Ciecko. 2020. AI Sees What? The Good, the Bad, and the Ugly of Machine Vision for Museum Collections. In *Museums and the Web 2020*. Museums and the Web, Online. https://mw20.museweb.net/paper/ai-sees-what-the-good-the-bad-and-the-ugly-of-machine-vision-for-museum-collections/

[18] Richard J. Cole, Frithjof Dau, Jon Ducrou, Peter W. Eklund, and Tim Wray. 2019. Navigating Context, Pathways and Relationships in Museum Collections using Formal Concept Analysis. *International Journal for Digital Art History* 4 (Dec. 2019), 5.13–5.27. https://doi.org/10.11588/dah.2019.4.72070

[19] Leendert D. Couprie. 1983. Iconclass: an iconographic classification system. *Art Libraries Journal* 8, 2 (1983), 32–49. https://doi.org/10.1017/S0307472200003436

[20] Kate Crawford and Trevor Paglen. 2019. Excavating AI: The Politics of Training Sets for Machine Learning. https://excavating.ai/

[21] Rossana Damiano. 2019. Investigating the Effectiveness of Narrative Relations for the Exploration of Cultural Heritage Archives: A Case Study on the Labyrinth system. In *Adjunct Publication of the 27th Conference on User Modeling, Adaptation and Personalization*. 417–423. https://doi.org/10.1145/3314183.3323870

[22] Ritendra Datta, Dhiraj Joshi, Jia Li, and James Z Wang. 2008. Image retrieval: Ideas, influences, and trends of the new age. *ACM Computing Surveys (Csur)* 40, 2 (2008), 1–60.

[23] Nicholas Davis, Chih-PIn Hsiao, Kunwar Yashraj Singh, Lisa Li, Sanat Moningi, and Brian Magerko. 2015. Drawing Apprentice: An Enactive Co-Creative Agent for Artistic Collaboration. In *Proceedings of the 2015 ACM SIGCHI Conference on Creativity and Cognition* (Glasgow, United Kingdom) *(C&C '15)*. Association for Computing Machinery, New York, NY, USA, 185–186. https://doi.org/10.1145/2757226.2764555

[24] Samuel Dodge and Lina Karam. 2017. A study and comparison of human and deep learning recognition performance under visual distortions. In *2017 26th international conference on computer communication and networks (ICCCN)*. IEEE, 1–7. https://doi.org/10.1109/ICCCN.2017.8038465

[25] Graham Dove, Kim Halskov, Jodi Forlizzi, and John Zimmerman. 2017. UX design innovation: Challenges for working with machine learning as a design material. *Conference on Human Factors in Computing Systems - Proceedings* 2017-May (2017), 278–288. https://doi.org/10.1145/3025453.3025739

[26] Marian Dörk, Sheelagh Carpendale, and Carey Williamson. 2011. The Information Flaneur: A Fresh Look at Information Seeking. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '11)*. ACM, New York, NY, USA, 1215–1224. https://doi.org/10.1145/1978942.1979124

[27] Ziv Epstein, Aaron Hertzmann, the Investigators of Human Creativity, Memo Akten, Hany Farid, Jessica Fjeld, Morgan R. Frank, Matthew Groh, Laura Herman, Neil Leach, Robert Mahari, Alex "Sandy" Pentland, Olga Russakovsky, Hope Schroeder, and Amy Smith. 2023. Art and the science of generative AI. *Science* 380, 6650 (2023), 1110–1111. https://doi.org/10.1126/science.adh4451 arXiv:https://www.science.org/doi/pdf/10.1126/science.adh4451

[28] John H. Falk and Lynn Diane Dierking. 2012. *The Museum Experience Revisited*. Routledge, London, England. https://doi.org/10.4324/9781315417851

[29] Anna Foka, Lina Eklund, Anders Sundnes Løvlie, and Gabriele Griffin. 2023. Critically Assessing AI/ML for Cultural Heritage: Potentials and Challenges. In *Handbook of Critical Studies of Artificial Intelligence*, Simon Lindgren (Ed.). Edward Elgar, Cheltenham.

[30] William W. Gaver, Jacob Beaver, and Steve Benford. 2003. Ambiguity as a Resource for Design. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Ft. Lauderdale, Florida, USA) *(CHI '03)*. Association for Computing Machinery, New York, NY, USA, 233–240. https://doi.org/10.1145/642611.642653

[31] Robert Geirhos, Patricia Rubisch, Claudio Michaelis, Matthias Bethge, Felix A Wichmann, and Wieland Brendel. 2019. ImageNet-trained CNNs are biased towards texture; increasing shape bias improves accuracy and robustness. In *International Conference on Learning Representations (ICLR 2019)*. New Orleans, Louisiana, United States. https://doi.org/10.48550/arXiv.1811.12231

[32] T. Giannini and J. P. Bowen. 2019. *Museums and Digital Culture*. Springer, 3–48. https://doi.org/10.1007/978-3-319-97457-6

[33] Marco Gillies, Rebecca Fiebrink, Atau Tanaka, Jérémie Garcia, Frédéric Bevilacqua, Alexis Heloir, Fabrizio Nunnari, Wendy Mackay, Saleema Amershi, Bongshin Lee, et al. 2016. Human-centred machine learning. In *Proceedings of the 2016 CHI conference extended abstracts on human factors in computing systems*. 3558–3565. https://doi.org/10.1145/2851581.2856492

[34] Ross Girshick. 2015. Fast r-cnn. In *Proceedings of the IEEE international conference on computer vision*. 1440–1448. https://doi.org/10.1109/ICCV.2015.169

[35] Peter Hall, Hongping Cai, Qi Wu, and Tadeo Corradi. 2015. Cross-depiction problem: Recognition and synthesis of photographs and artwork. *Computational Visual Media* 1, 2 (June 2015), 91–103. https://doi.org/10.1007/s41095-015-0017-1

[36] Rex Hartson and Pardha S. Pyla. 2019. *Chapter 4 & Chapter 29*. Morgan Kaufmann.

[37] HarvardArtMuseums. [n. d.]. *About the AI Explorer*. https://ai.harvardartmuseums.org/about Acessed March 8th 2023.

[38] Douglas Heaven. 2019. Why deep-learning AIs are so easy to fool. *Nature* 574 (Oct. 2019), 163–166. https://doi.org/10.1038/d41586-019-03013-5

[39] Uta Hinrichs, Holly Schmidt, and Sheelagh Carpendale. 2008. EMDialog: Bringing information visualization into the museum. *IEEE transactions on visualization and computer graphics* 14, 6 (2008), 1181–1188. https://doi.org/10.1109/TVCG.2008.127

[40] Eva Hornecker and Luigina Ciolfi. 2019. Human-Computer Interactions in Museums. *Synthesis Lectures on Human-Centered Informatics* 12, 2 (2019), i–171. https://doi.org/10.2200/S00901ED1V01Y201902HCI042 _eprint: https://doi.org/10.2200/S00901ED1V01Y201902HCI042.

[41] Ayanna Howard. 2020. Are we trusting AI too much? Examining human-robot interactions in the real world. In *Proceedings of the 2020 ACM/IEEE International Conference on Human-Robot Interaction*. 1–1. https://doi.org/10.1145/3319502.3374842

[42] Pauline Junginger, Dennis Ostendorf, Barbara Avila Vissirini, Anastasia Voloshina, Timo Hausmann, Sarah Kreiseler, and Marian Dörk. 2020. The Close-up Cloud: Visualizing Details of Image Collections in Dynamic Overviews. *International Journal for Digital Art History* 5 (Dec. 2020). https://doi.org/10.11588/dah.2020.5.72039

[43] David Kadish, Sebastian Risi, and Anders Sundnes Løvlie. 2021. Improving Object Detection in Art Images Using Only Style Transfer. In *The International Joint Conference on Neural Networks (IJCNN)*. IEEE, Virtual Event. https://doi.org/10.1109/IJCNN52387.2021.9534264 _eprint: 2102.06529.

[44] Bill Kules and Ben Shneiderman. 2008. Users can change their web search tactics: Design guidelines for categorized overviews. *Information Processing & Management* 44, 2 (2008), 463–484. https://doi.org/10.1016/j.ipm.2007.07.014

[45] Michael Kuniavsky, Elizabeth Churchill, and Molly Wright Steenson. 2017. Designing the user experience of machine learning systems. In *AAAI Spring Symposium Proceedings (Technical Report SS-17-04)*. 27–29.

[46] Lucian Leahu. 2016. Ontological Surprises: A Relational Perspective on Machine Learning. In *Proceedings of the 2016 ACM Conference on Designing Interactive Systems* (Brisbane, QLD, Australia) *(DIS '16)*. Association for Computing Machinery, New York, NY, USA, 182–186. https://doi.org/10.1145/2901790.2901840

[47] Yann LeCun, Yoshua Bengio, and Geoffrey Hinton. 2015. Deep learning. *nature* 521, 7553 (2015), 436–444.

[48] Junnan Li, Dongxu Li, Caiming Xiong, and Steven Hoi. 2022. BLIP: Bootstrapping Language-Image Pre-training for Unified Vision-Language Understanding and Generation. In *Proceedings of the 39th International Conference on Machine Learning (Proceedings of Machine Learning Research, Vol. 162)*, Kamalika Chaudhuri, Stefanie Jegelka, Le Song, Csaba Szepesvari, Gang Niu, and Sivan Sabato (Eds.). PMLR, 12888–12900. https://proceedings.mlr.press/v162/li22n.html

[49] Liunian Harold Li, Pengchuan Zhang, Haotian Zhang, Jianwei Yang, Chunyuan Li, Yiwu Zhong, Lijuan Wang, Lu Yuan, Lei Zhang, Jenq-Neng Hwang, Kai-Wei Chang, and Jianfeng Gao. 2022. Grounded Language-Image Pre-Training. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. 10965–10975.

[50] Tsung-Yi Lin, Michael Maire, Serge Belongie, James Hays, Pietro Perona, Deva Ramanan, Piotr Dollár, and C Lawrence Zitnick. 2014. Microsoft coco: Common objects in context. In *Computer Vision–ECCV 2014: 13th European Conference, Zurich, Switzerland, September 6-12, 2014, Proceedings, Part V 13*. Springer, 740–755. https://doi.org/10.1007/978-3-319-10602-1_48

[51] Irene Lopatovska, Iris Bierlein, Heather Lember, and Eleanor Meyer. 2013. Exploring requirements for online art collections. *Proceedings of the American Society for Information Science and Technology* 50, 1 (2013), 1–4. https://doi.org/10.1002/meet.14505001109

[52] Anders Sundnes Løvlie, Lina Eklund, Annika Waern, Karin Ryding, and Paulina Rajkowska. 2021. Designing for Interpersonal Museum Experiences. In *Museums and the Challenge of Change: Old Institutions in a New World*, Graham Black (Ed.). Routledge, London & New York, 223–238.

[53] Kristin MacDonough. 2018. Smartify. *Multimedia & Technology Reviews* (April 2018). https://doi.org/10.17613/95t4-2t63

[54] Seungwoo Maeng, Youn-kyung Lim, and KunPyo Lee. 2012. Interaction-driven design: A new approach for interactive product development. In *Proceedings of the Designing Interactive Systems Conference*. 448–457.

[55] Matei Mancas, Donald Glowinski, P Brunet, F Cavallero, C Machy, Pieter-Jan Maes, Stella Paschalidou, M Rajagopal, S Schibeci, and Laura Vincze. 2009. Hypersocial Museum: addressing the social interaction challenge with museum scenarios and attention-based approaches. *QPSR of the numediart research program* 2 (01 2009), 91–96.

[56] Lev Manovich. 2020. *Cultural Analytics*. The MIT Press, Cambridge, MA. https://mitpress.mit.edu/books/cultural-analytics Publisher: The MIT Press.

[57] Gary Marchionini. 2006. Exploratory search: from finding to understanding. *Commun. ACM* 49, 4 (2006), 41–46. https://doi.org/10.1145/1121949.1121979

[58] Nora McDonald, Sarita Schoenebeck, and Andrea Forte. 2019. Reliability and Inter-Rater Reliability in Qualitative Research: Norms and Guidelines for CSCW and HCI Practice. *Proc. ACM Hum.-Comput. Interact.* 3, CSCW, Article 72 (nov 2019), 23 pages. https://doi.org/10.1145/3359174

[59] Christofer Meinecke. 2022. Labeling of Cultural Heritage Collections on the Intersection of Visual Analytics and Digital Humanities. In *2022 IEEE 7th Workshop on Visualization for the Digital Humanities (VIS4DH)*. 19–24. https://doi.org/10.1109/VIS4DH57440.2022.00009

[60] Pavol Navrat. 2012. Cognitive traveling in digital space: from keyword search through exploratory information seeking. *Central European Journal of Computer Science* 2 (2012), 170–182. https://doi.org/10.2478/s13537-012-0024-6

[61] Jonas Oppenlaender. 2022. The Creativity of Text-to-Image Generation. In *Proceedings of the 25th International Academic Mindtrek Conference* (Tampere, Finland) *(Academic Mindtrek '22)*. Association for Computing Machinery, New York, NY, USA, 192–202. https://doi.org/10.1145/3569219.3569352

[62] Emilie Palagi, Fabien Gandon, Alain Giboin, and Raphaël Troncy. 2017. A Survey of Definitions and Models of Exploratory Search. In *Proceedings of the 2017 ACM Workshop on Exploratory Search and Interactive Data Analytics* (Limassol, Cyprus) *(ESIDA '17)*. Association for Computing Machinery, New York, NY, USA, 3–8. https://doi.org/10.1145/3038462.3038465

[63] Alec Radford, Jong Wook Kim, Chris Hallacy, Aditya Ramesh, Gabriel Goh, Sandhini Agarwal, Girish Sastry, Amanda Askell, Pamela Mishkin, Jack Clark, et al. 2021. Learning transferable visual models from natural language supervision. In *International conference on machine learning*. PMLR, 8748–8763.

[64] Aditya Ramesh, Prafulla Dhariwal, Alex Nichol, Casey Chu, and Mark Chen. 2022. Hierarchical Text-Conditional Image Generation with CLIP Latents. arXiv:2204.06125 [cs.CV]

[65] Aditya Ramesh, Mikhail Pavlov, Gabriel Goh, Scott Gray, Chelsea Voss, Alec Radford, Mark Chen, and Ilya Sutskever. 2021. Zero-shot text-to-image generation. In *International Conference on Machine Learning*. PMLR, 8821–8831.

[66] Paul Robinette, Wenchen Li, Robert Allen, Ayanna M Howard, and Alan R Wagner. 2016. Overtrust of robots in emergency evacuation scenarios. In *2016 11th ACM/IEEE international conference on human-robot interaction (HRI)*. IEEE, 101–108. https://doi.org/10.1109/HRI.2016.7451740

[67] Robin Rombach, Andreas Blattmann, Dominik Lorenz, Patrick Esser, and Björn Ommer. 2022. High-Resolution Image Synthesis With Latent Diffusion Models. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. 10684–10695.

[68] Stan Ruecker, Milena Radzikowska, and Stéfan Sinclair. 2016. *Visual interface design for digital cultural heritage: A guide to rich-prospect browsing*. Routledge.

[69] Olga Russakovsky, Jia Deng, Hao Su, Jonathan Krause, Sanjeev Satheesh, Sean Ma, Zhiheng Huang, Andrej Karpathy, Aditya Khosla, Michael Bernstein, Alexander C. Berg, and Li Fei-Fei. 2015. ImageNet Large Scale Visual Recognition Challenge. *International Journal of Computer Vision (IJCV)* 115, 3 (2015), 211–252. https://doi.org/10.1007/s11263-015-0816-y

[70] Bryan C Russell, Antonio Torralba, Kevin P Murphy, and William T Freeman. 2008. LabelMe: a database and web-based tool for image annotation. *International journal of computer vision* 77 (2008), 157–173.

[71] Tony Russell-Rose and Tyler Tate. 2012. *Designing the search experience: The information architecture of discovery*. Newnes.

[72] Phoebe Sengers and Bill Gaver. 2006. Staying Open to Interpretation: Engaging Multiple Meanings in Design and Evaluation. In *Proceedings of the 6th Conference on Designing Interactive Systems* (University Park, PA, USA) *(DIS '06)*. Association for Computing Machinery, New York, NY, USA, 99–108. https://doi.org/10.1145/1142405.1142422

[73] Jonas Heide Smith. 2019. SMK's collection search levels up. https://medium.com/smk-open/smks-collection-search-levels-up-cf8e967e9346

[74] Jascha Sohl-Dickstein, Eric Weiss, Niru Maheswaranathan, and Surya Ganguli. 2015. Deep unsupervised learning using nonequilibrium thermodynamics. In *International conference on machine learning*. PMLR, 2256–2265.

[75] Ayah Soufan, Ian Ruthven, and Leif Azzopardi. 2022. Searching the Literature: An Analysis of an Exploratory Search Task. In *ACM SIGIR Conference on Human Information Interaction and Retrieval (CHIIR '22)*. ACM, Regensburg, Germany, 12 pages. https://doi.org/10.1145/3498366.3505818

[76] Matthias Springstein, Stefanie Schneider, Christian Althaus, and Ralph Ewerth. 2022. Semi-supervised Human Pose Estimation in Art-historical Images. In *Proceedings of the 30th ACM International Conference on Multimedia*. 1107–1116. https://doi.org/10.1145/3503161.3548371

[77] John Stack. 2020. *Computer Vision and the Science Museum Group Collection*. Science Museum Group Digital Lab. https://lab.sciencemuseum.org.uk/computer-vision-and-the-science-museum-group-collection-a6c20efb0ac9

[78] Jiawei Su, Danilo Vasconcellos Vargas, and Kouichi Sakurai. 2019. One pixel attack for fooling deep neural networks. *IEEE Transactions on Evolutionary Computation* 23, 5 (2019), 828–841. https://doi.org/10.1109/TEVC.2019.2890858

[79] Loic Tallon. 2018. Creating Access beyond metmuseum. org: The Met Collection on Wikipedia. *The Met Museum Blog, February* 7 (2018). https://www.metmuseum.org/blogs/now-at-the-met/2018/open-access-at-the-met-year-one

[80] Zenonas Theodosiou, Marios Thoma, Harris Partaourides, and Andreas Lanitis. 2022. A Systematic Approach for Developing a Robust Artwork Recognition Framework Using Smartphone Cameras. *Algorithms* 15, 9 (2022). https://doi.org/10.3390/a15090305

[81] Alice Thudt, Uta Hinrichs, and Sheelagh Carpendale. 2012. The bohemian bookshelf: supporting serendipitous book discoveries through information visualization. In *Proceedings of the SIGCHI conference on human factors in computing systems*. 1461–1470. https://doi.org/10.1145/2207676.2208607

[82] Antonio Torralba, Rob Fergus, and Yair Weiss. 2008. Small codes and large image databases for recognition. In *2008 IEEE Conference on Computer Vision and Pattern Recognition*. IEEE, 1–8.

[83] Elena Villaespesa. 2019. Museum collections and online users: development of a segmentation model for the metropolitan museum of art. *Visitor Studies* 22, 2 (2019), 233–252. https://doi.org/10.1080/10645578.2019.1668679

[84] Elena Villaespesa and Oonagh Murphy. 2021. This is not an apple! Benefits and challenges of applying computer vision to museum collections. *Museum Management and Curatorship* 36, 4 (2021), 362–383. https://doi.org/10.1080/09647775.2021.1873827

[85] Annika Waern and Anders Sundnes Løvlie (Eds.). 2022. *Hybrid Museum Experiences*. Amsterdam University Press, Amsterdam. https://www.aup.nl/en/book/9789048552849/hybrid-museum-experiences

[86] Nicholas Westlake, Hongping Cai, and Peter Hall. 2016. Detecting People in Artwork with CNNs. In *Computer Vision – ECCV 2016 Workshops (Lecture Notes in Computer Science)*, Gang Hua and Hervé Jégou (Eds.). Springer International Publishing, 825–841. https://doi.org/10.1007/978-3-319-46604-0_57

[87] Ryen W. White and Resa A. Roth. 2009. *Exploratory Search: Beyond the Query—Response Paradigm*. Springer International Publishing. https://doi.org/10.1007/978-3-031-02260-9

[88] Mitchell Whitelaw. 2015. Generous interfaces for digital cultural collections. *Digital Humanities Quarterly* 9, 1 (2015). https://www.digitalhumanities.org/dhq/vol/9/1/000205/000205.html

[89] Max L Wilson, Bill Kules, Ben Shneiderman, et al. 2010. From keyword search to exploration: Designing future search interfaces for the web. *Foundations and Trends® in Web Science* 2, 1 (2010), 1–97. https://doi.org/10.1561/1800000003

[90] Florian Windhager, Paolo Federico, Günther Schreder, Katrin Glinka, Marian Dörk, Silvia Miksch, and Eva Mayr. 2019. Visualization of Cultural Heritage Collection Data: State of the Art and Future Challenges. *IEEE Transactions on Visualization and Computer Graphics* 25, 6 (2019), 2311–2330. https://doi.org/10.1109/TVCG.2018.2830759

[91] Tim Wray, Peter Eklund, and Karlheinz Kautz. 2013. Pathways through information landscapes: Alternative design criteria for digital art collections. *International Conference on Information Systems (ICIS 2013): Reshaping Society Through Information Systems Design* 3.

[92] Tim Wray, Peter W. Eklund, and Karlheinz Kautz. 2013. Pathways through information landscapes: Alternative design criteria for digital art collections. In *International Conference on Information Systems*. Milan, Italy. https://www.researchgate.net/publication/259010843_Pathways_through_information_landscapes_Alternative_design_criteria_for_digital_art_collections

[93] Qian Yang, Aaron Steinfeld, Carolyn Rosé, and John Zimmerman. 2020. Re-Examining Whether, Why, and How Human-AI Interaction Is Uniquely Difficult to Design. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) *(CHI '20)*. Association for Computing Machinery, New York, NY, USA, 1–13. https://doi.org/10.1145/3313831.3376301

[94] Ka-Ping Yee, Kirsten Swearingen, Kevin Li, and Marti Hearst. 2003. Faceted Metadata for Image Search and Browsing. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Ft. Lauderdale, Florida, USA) *(CHI '03)*. Association for Computing Machinery, New York, NY, USA, 401–408. https://doi.org/10.1145/642611.642681

[95] John Zimmerman, Jodi Forlizzi, and Shelley Evenson. 2007. Research Through Design As a Method for Interaction Design Research in HCI. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '07)*. ACM, New York, NY, USA, 493–502. https://doi.org/10.1145/1240624.1240704

[96] Joanna Zylinska. 2020. *AI Art: Machine Visions and Warped Dreams*. Open Humanites Press. http://www.openhumanitiespress.org/books/titles/ai-art/