ArtEureka: A Digital Art Exploration Tool

Recommending Objects Within The Rijkmuseum's Online Environment

Sinem Ertem, Jan Janiszewski, Luuk Kolstee, Bruno Sotic
University of Amsterdam
Amsterdam, The Netherlands
sinem.ertem@student.uva.nl,jan.j.janiszewski@gmail.com,luuk.kolstee@student.uva.nl,soticbruno@gmail.com

ABSTRACT

The Rijksmuseum in Amsterdam, like many museums around the world, is developing an environment for users to enjoy the museum's collection. This digitized format presents the museum with new challenges. Normally relying upon expert annotated information, the online collection counts more than 700000 objects and is far too large for complete manual annotation. Furthermore the museum wishes to attract a diverse and large public who lack the domain knowledge required to interact with metadata driven search. Our solution is ArtEureka. It employs a recommendation system that allow users to explore the collection using both metadata filters and visual similarity cues. Cosine similarity is calculated based on classification scores from 36 different art movements, assigned by the resnet50 neural network. In addition functionality was added to prevent filter bubbles and allow for serendipitous search. The recommendation system was incorporated into a prototype and validated. Validation showed neutral scores for usability, and neutral to positive scores for object relevance. We expect ArtEureka to be a highly robust and usable exploration system after a slight refocus on prototype usability.

KEYWORDS

Digital Art Collections, Vision-language, Cosine Similarity, Recommendations Systems, Serendipity, Machine Learning, Data Science

1 INTRODUCTION

With 2.7 million visitors per year, the Rijksmuseum is the largest museum in the Netherlands [13]. Like most musea of its size, the Rijksmuseum has an online presence (i.e., Rijksstudio), which presents users with 770048 unique art objects. These objects can also be searched using a meta-data driven search engine employed on the platform. Unfortunately, the metadata-driven process poses two problems. First, many of the objects within the collection have not been extensively documented since the documentation of object metadata needs to be validated by experts, a rigorous and resource-intensive process. This creates gaps within the collection of objects which are difficult if not impossible to find. Second, it requires users to have a certain level of domain-knowledge required to navigate the system. Though the Rijksstudio platform enables keyword-driven search, it does not enable users to find objects based on visual features. This poses a significant problem to the website, since visual-based search helps website visitor with image search [20]. The platform also does not employ a recommendation system, making it difficult to use for users who do not exactly know what object they are looking for. This paper documents the creation of ArtEureka, a system designed to allow users to explore the online Rijksmuseum collection in a more intuitive manner by using both visual and metadata cues.

1.1 The Rijksmuseum

The Rijksmuseum is confusingly one amongst many 'rijksmuseums' in the Netherlands. These museums receive a subsidy from the Dutch government. Within this paper the capitalized 'The Rijksmuseum' will be used to refer to the specific museum in Amsterdam. The funds provided by the subsidy ensures that the museums have the funds to exhibit and keep the 'rijkscollecties', Dutch national collections often owned by the state. Dutch national policy on culture is subject to change, in 2016 the 'Council for Culture' advised the Dutch government not to increase The Rijksmuseum's subsidy to 7.6 million. The council pointed to the fact that the museum was reporting record profits [11]. The exact figures are irrelevant to this project, but it is important to note that the Rijksmuseum has an obligation to the Dutch state. The museum gains support from the Dutch state, and in turn has an exemplary and educative function for the entirety of the Dutch public. Even those without domain knowledge. Furthermore, the Dutch ministry for education, culture and science has set apart money for the digitization of cultural collections such as the Rijksmuseum's [10]. Q42, developer of the current Rijksstudio website provides the following statements on the intended function of the website: *"How do you make a museum website on which the user can lose themselves? And how do you make the online experience just as good as a visit to the cultural institution itself?"* [12]. Clearly the website is intended to complement the Rijksmuseum's accessible and grand stature. The goal is to engross users, enable exploration and to be easy to use for any possible user. A 'Netflix' for art, as Q42 puts it.

1.2 Problem Statement and Solution

The Rijksstudio platform's method of indexing is highly accurate and fitting of the Museum's international level of prestige, but is lacking in usability and makes object retrieval difficult. As it requires expert annotation it is also difficult to scale to the sizes required for the online collection. For users without much domain knowledge, the metadata-driven system is difficult to navigate. Furthermore the system does not provide users with similar objects through a recommendation system. Internal statistics provided by the Rijksmuseum and q42 show that the platform receives very little traffic outside of people looking for 'Rembrandt', or 'The Night's Watch'. The collection needs to be more easily explorable for the general public. This paper documents the creation of a system that will make it possible for users to explore the collection. It does so by answering the following research question. *Can a recommendation system combine existing metadata with visual object characteristics to provide users with satisfyingly similar objects?*. To answer this question, a recommending system will be created using AI modelling. The resultant recommendation system will be implemented within a prototype exploration system. Finally, the implementation will be tested for both user experience and relevance of results.

2 RELATED WORK

This chapter of the report serves as a literature review of the domains important to the topic and proposed solution.

2.1 Online Art Platforms

The importance of online access to museum collections has been apparent for a long time. A 2006 study found that augmented reality (AR) and virtual reality (VR) technologies can provide those without physical access to the museum with a fulfilling online 'visit' [17]. Such an online presence also provides a solution in situations were resources don't permit a museum to physically exhibit their entire collection. The study shows the importance of the museum 'feel' by successfully implementing online exhibitions. Such a robust implementation would however be constrained by a resource-intensive process of creating quality online exhibitions. This would leave us with problems similar to those currently faced by the scalability of annotated metadata-driven indexing. Furthermore, the usage of VR and AR does not facilitate casual browsing. A robust index of the needs of online museum visitors is necessary. Luckily, such studies exist

A 2018 study sought to create a segmentation model of the various user-archetypes identified among visitors of the online platforms of the Metropolitan Museum of Art [16]. The study identifies six segments: professional researcher, personal interest information-seeker, student researcher, inspiration-seeker, casual browser, and visit planner. The needs and wants of these various users differ. A casual browser lacks the domain knowledge professional researchers may call upon. As such the article proposes that online art platforms need to provide multiple ways to view and interact with the online collection. This supports the creation of an exploration system apart from the current Rijksstudio search engine. The diverging needs of users also necessitates the selection of a user group. An exploration system corresponds best to the casual browser, visit planner, personal interest information-seeker and inspiration seeker user groups. Facilitating the two research groups would be difficult without additional layers of information that may be overwhelming to casual browsers. A 2018 study on the online presence of the 'Museum für Kunst und Gewerbe Hamburg' (MKG) confirms that academic browsers have highly specific needs that are not always shared by other users [14]. A focus on these 'especially informed' users can be identified. This is not without merit, the MKG study identified that this group formed over half the platform's userbase, but it does leave a gap regarding more casual browsers. It is seemingly exactly this casual browser, interested in a 'Netflix for art', that platforms intend to engross. Perhaps it is indicative of the current limitations of platforms, and resulting difficulties in attracting casual browsers and visitors. As such there are opportunities to bridge the gap during this study into recommendation systems.

A study into the general behavior of users on digital museum platforms found four main characteristics [15]. It found that searching behavior has a strong visual aspect, where selection was based upon the object's visual representation rather than appended information. This supports the usage of a visually intuitive recommendation system. Secondly, it found that users often start their search from known starting points. This also supports the adoption of a recommendation system to allow users to continue from known objects. It thirdly identified that users use broad and general search terms before searching onward from a selected object. This supports a synthesis between search and key terms on the one hand, and recommendations based on visual cues on the other. Finally the concept of 'meaning making' was deemed important. Visitors attributed meaning to objects based upon the context in which they were found. A possible recommendation system has to be perceived as presenting objects within a contextually sound collection, or users will find it difficult to create meaning between objects.

2.2 Recommendation Systems

Recommendation systems can be employed when volumes of content are deemed to be overwhelming, and when there's a need for the concise and efficient delivery of information to prevent information overload[6]. A recommendation system allows the user to browse through dynamically generated content. It allows a user to move past catalogue searches when these are deemed insufficient or too complicated. The recommendation system dynamically sends a user in a certain direction based upon their past selection. This can cause a user to get 'stuck' in said direction, not allowing for any way out. This is called a filter bubble [9]. Combating filter bubbles is a balancing act. When the objective is to recommend the closest object to a user, it is not intuitive to provide the user with any but the best matches. However, one has to remember that neither a recommending agent, nor the user, are infallible. Both may cause the exploration to be pushed into an incorrect path, and in such a situation it is necessary to allow for ways out of that path. Providing users with paths divided from their current direction may also enable serendipitous findings. Objects the user was not looking for, but would be happy to find. Studies however find that this is a difficult concept to implement. If the serendipitous pathing results in a recommendation engine that is too inaccurate the user will stop putting stock in the recommendations and ignore them [9]. Serendipity is difficult to implement. It includes an emotional dimension or judgement from the user, and serendipitous discovery is rare. A 2016 study recommends the usage of context-dependent metrics to better enable serendipity within recommendation systems [8]. An example given is the usage of item popularity as an ancillary metric. Other systems forego the concept of serendipity and inform the user of their presence within a filter bubble. These solutions then provide users with additional filters that would enable them to tune their selection to their needs [18]. Whatever approach is taken, it is clear that one must plan for filter bubbles when designing a recommendation system.

3 INTERACTION DESIGN

This section of the report details any design choices made, as well as the exploratory studies into the needs of the various stakeholders. It also provides images of the prototype used for validation.

3.1 Methodology

To address the challenge of enhancing the user experience and functionality of retrieving art through the museums digital collection, this project employed a mixed-method user study that is based on the Design Thinking methodology [1]. Design thinking is a human-centered approach to problem solving that has gained widespread popularity in technology, research and design. The method is characterised by empathizing and understanding the end-user; defining the problem from the said users perspective; prototyping a solution; and validating it.

In the 'ideation' phase, the research team conducted qualitative user studies to gain a deeper understanding of the needs and motivations of the museums digital audience. The goal of the unstructured interviews was to uncover the pain points they experienced when browsing through the museums collection.

With this information, the team then brainstormed ideas to develop a range of potential solutions. The solutions are then narrowed down and the most promising one is worked out into a low-fidelity prototype.

Finally, in the usability testing stage, the prototype is assessed and qualitative as well as quantitative data is gathered fro the potential end-users in order to improve and refine the solution.

3.2 Stakeholders

Before any implementation came to be, the project's main stakeholders were identified.

3.2.1 Stakeholder: The Rijksmuseum & Q42. The Rijksmuseum, as a public institution, has been introduced earlier in this paper, as have the limitations of their current digitization efforts. Is has been noted that the Rijksmuseum wants its online presence to be attractive to a broad public of casual browsers. Within the museum's policy plans for 2020-2024, the Rijksmuseum's specific goals can be noted [5]. Within this document the museum formalizes its plans for the coming period, and how these politically allign with its position as receiver of a national subsidy. The noted core values are 'Authenticity', 'Quality', 'Personal', 'Personal', 'Innovative' and 'simplicity'. The report also notes that the museum makes its collection accessible, both online and offline, for as broad a public as possible. This is also specifically one of the core focus points from the current minister. To broaden the group reached by the Rijksmuseum. Quality is still one of the museum's core values, across all of its modes of interaction. This implies that any information or methodology on the website needs to be qualitatively sound. There is no tolerance for faulty information. The report states that the development of the museum's online presence will not be halted. This aligns with the goals set by Q42, the intermediary stakeholder between this project and the Rijksmuseum. In order to make the museum's digital collection more accessible, the organization aims to improve its search and retrieval system through vision-language models. Q42 is currently exploring methods to enhance the accessibility, retrieval, engagement, and user experience while browsing the digital collection.

3.2.2 Stakeholder: The User. There are multiple types of users for the Rijksmuseum's digital collection, but two main groups can be identified: 1) those that know exactly what they are looking for (i.e. a specific item); 2) those that are interested in exploring the collection without a specific search goal in mind. The former might relate to academics, researchers, students, hobbyists and journalists.Those who have a specific item in mind. This search task would usually be approached by indexing for the items name and/or artist and could be retrieved via metadata.

On the other hand, the second type of user might be curious about exploring the collection and does not have a specific retrieval goal in mind. Their approach to exploring is likely to be informed by their mental model of how art is categorised and/or previous experience with database search. Literature points towards this being an as of yet unreached userbase, and also the one the Dutch state and with it the museum would be most interested in reaching. The prototype proposed in this report primarily focuses on solving the issues faced by the second type of user. Based on unstructured interviews with potential end users that fit this profile, we have created the following user persona:

Carl is a tourist visiting Amsterdam who has an interest in culture. While he is considering visiting the Rijksmuseum, he wants to first explore the collection online. Carl is not familiar with advanced search and retrieval techniques, and finds indexing through search engines in museums to be time-consuming. Furthermore he would not know which terms to use to get to specific works of art that appeal to him. Carl is interested in finding artwork that appeals to his personal preferences without having to spend a lot of time and/or mental effort during the process.

Hence, the idea behind the proposed prototype is to make the content exploration process easier and more intuitive for end-users like Carl.

3.3 Requirements

With the project's main stakeholders identified, the implementation's requirements can be explicated.

3.3.1 Rijksmuseum Requirements. The Rijksmuseum wants to reach a very broad and varied userbase. With this fits a solution that is not necessarily tailor-fit to fit specific preconceptions, but which is instead intuitive to a broad base of users. As the Rijksmuseum values quality above all, the proposed solution should be qualitatively sound. It should not provide users with erroneous information at any point. Furthermore, the Rijksmuseum and Q42 wish for the site to be engaging to users. Any changes would preferably increase user engagement and enjoyment. Finally the aforementioned report states the end goal of the web presence to be to get people to visit the physical museum. A good implementation would support users who may be planning for a visit.

3.3.2 User Requirements. As noted, the Rijksmuseum seeks to reach a broad and varied group of users with its online presence. Literature on the users of online museum collections notes a substantial group of people with very little domain knowledge, who would be helped greatly by the implementation of a recommending system. These users require an implementation to be intuitive, requiring very little domain knowledge, and allowing them to work from objects they know. Visual connections between paintings should be clear and intuitive to them. A fence-sitting potential visitor should be persuaded to visit the museum, but a sure visitor should also find value in exploring the collection before their visit.

3.4 Prototypes

A lo-fi prototype tool was developed to help potential users filter through the museum collection. The design process involved user research and testing. We wanted to create a user interface that was intuitive and easy to use. To ensure that the user experience was optimized, the interface was informally tested with a sample group of potential users. Based on the user feedback, we made several adjustments to the interface. We simplified the layout and added visual cues to guide users through the filtering process. We also incorporated intuitive gestures, such as swiping and tapping. The interaction model of our prototype can be seen in figure 1.

The landing page of the prototype, shown in figure 2, is designed to align with the aesthetic of the Rijksmuseum and serves as the main filtering interface for users. Our initial user study showed that users prefer to avoid search engines when finding art. Therefore, we designed this page to help users easily filter through the museum's collection based on Century or Artist.

These options were selected based on the metadata available and their relevance to the users' mental model of how art can be categorized and stored (as indicated by our initial user study). The accompanying text on the page provides clear explanations of the functionality of each filtering option. Additionally, there is an information icon next to the title that offers additional background information about the tool and how to use it. This helps ensure that users have all the information they need to effectively navigate and utilize the filtering options.

Figure 3 shows the simple filtering system. The user can use once he clicks on one of the two filtering options and select either century or artist, then scroll through the available options and select one of the two. The system then loads a corresponding page featuring a collage of relevant paintings.

When filters are applied, the user is presented with a scrollable collage of paintings related to the chosen filters. Accompanying the colage is text that informs the user of the number of paintings that fall under that category. For example, as shown in figure 4, if the user filters by the 17th century, the text states that there are 995 paintings related to that time period. Similarly, if the user filters by Artist, as shown in figure 5, the collage displays relevant paintings and provides information on the number of paintings related to that artist, as well as their art movement affiliation. The user is then able to click on one of the paintings that appeals to them to retrieve more information and relevant paintings.

Upon selecting a painting, the user is directed to a page that provides more detailed information about the selected painting, as shown in figure 5. The page features information about the painting's quality and accompanying text, as well as scores indicating the art movement to which the painting belongs. By scrolling down, the user can access the recommendation system and view other related paintings. This additional information and recommendation system allows users to delve deeper into the museum's collection. This page is visualized in figure 6

As shown in figure 6 the user is presented with two windows showcasing recommended similar and dissimilar images. They can navigate through the collection of related or unrelated paintings using either the tap or vertical scroll option. Each painting is interactive

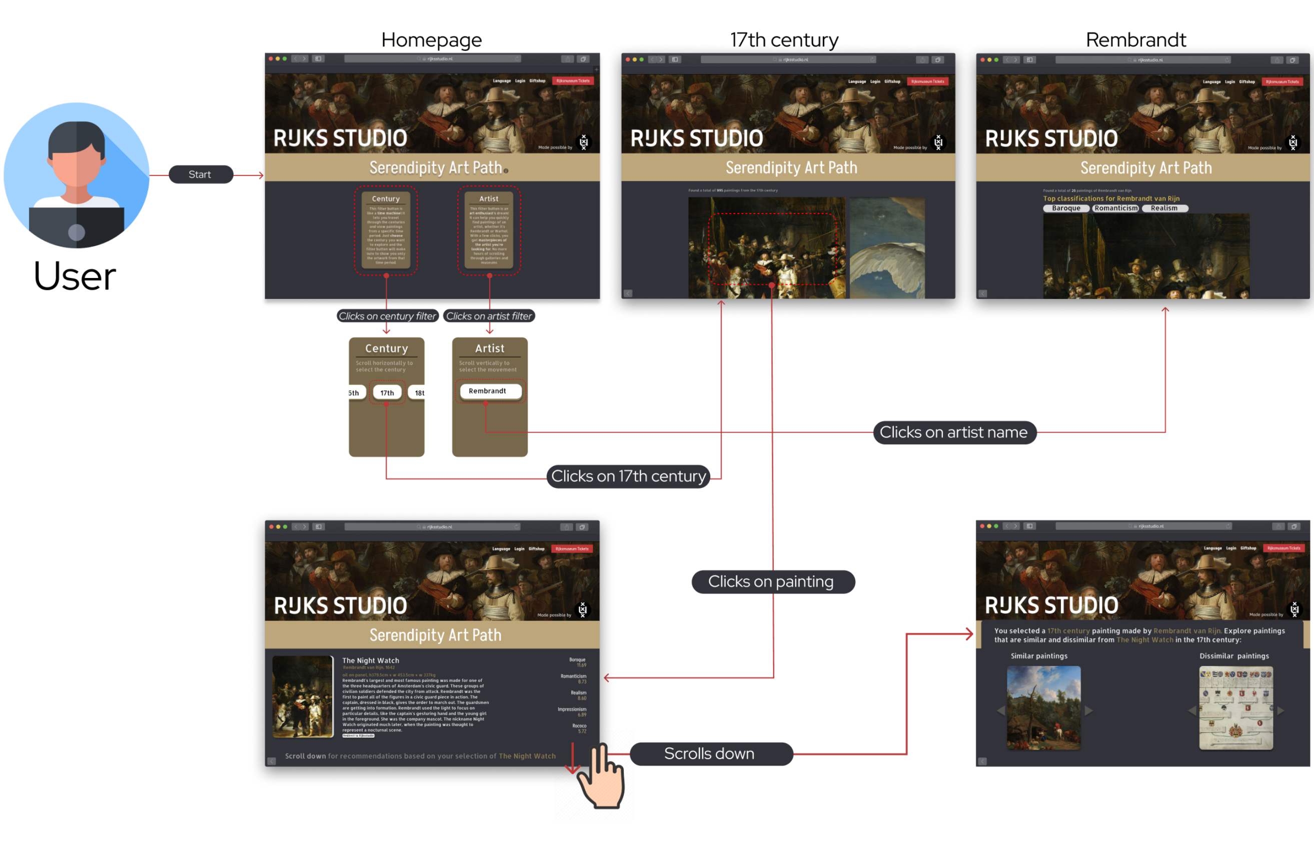

Figure 1: User Interaction Model

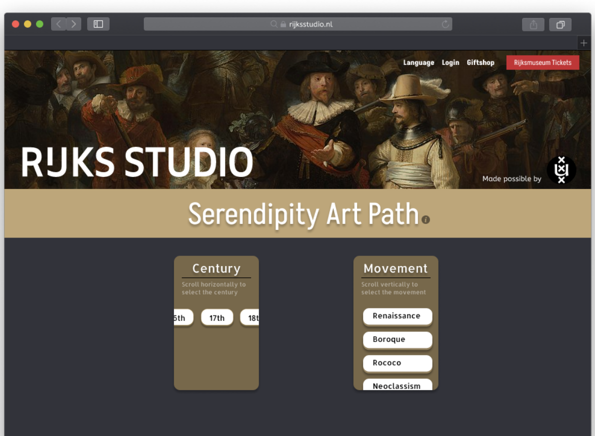

Figure 2: Homepage with Filtering Options

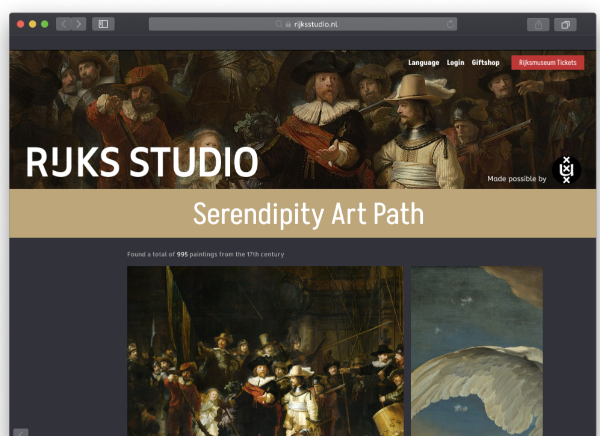

Figure 3: Page After Using Filters

and can be clicked on to access a detailed page with information about the selected piece.

Finally, figure 7 presents a functional recommendation system that operates separately from the interface.

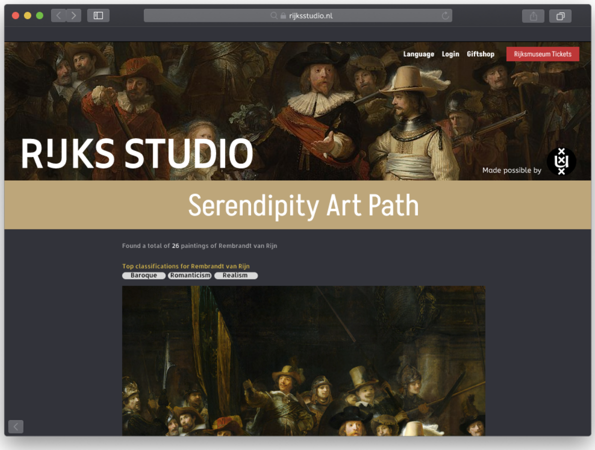

Figure 4: Scrolled Down Page After Using Filters

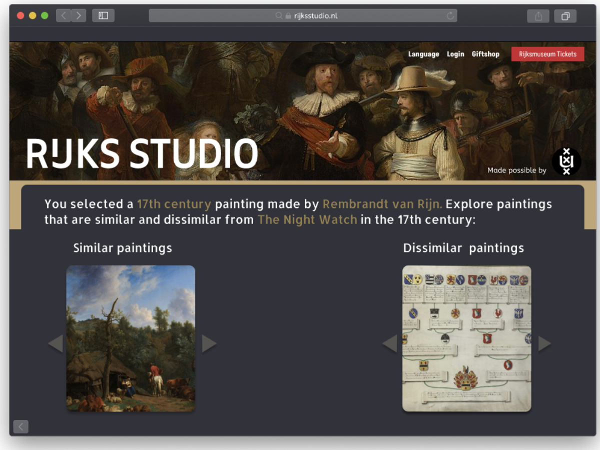

Figure 6: Recommended Paintings Presented After Scrolling Down From Individual Painting

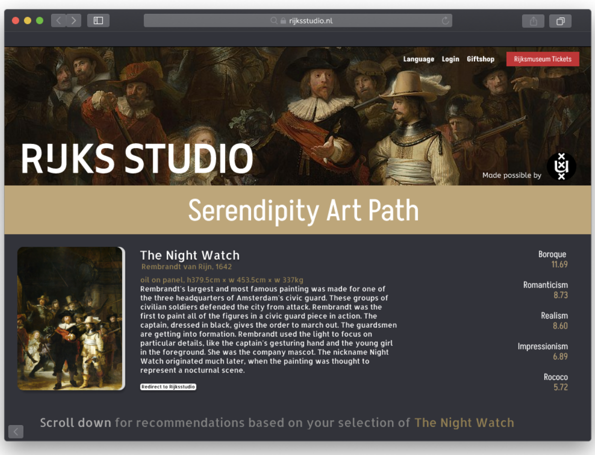

Figure 5: Page After Selecting a Painting

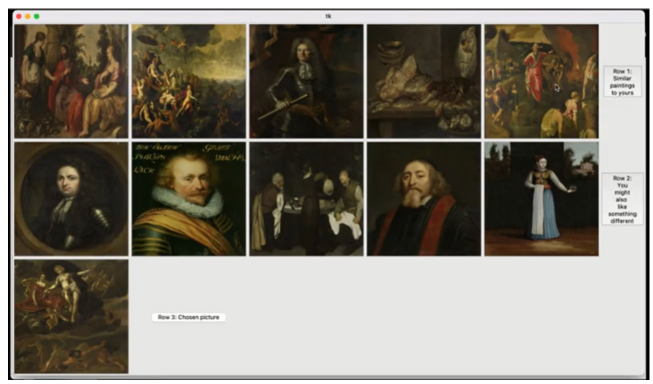

Figure 7: Recommendation System Prototype

Due to a shortage of research, the research team was unable to integrate the back-end with the frontend UI. Nevertheless, the UI was designed to mimic the backed using a limited selection of pre-defined paintings. The recommendation system displays the most similar images at the top of the screen, with the paintings in the middle becoming increasingly dissimilar to the selected one, providing the user with a broader range of options. The painting in the bottom corner represents the original selection, which serves as the basis for generating the recommendations. This prototype was also used to test and evaluate the similarity of the images in a later stage (testing).

4 SYSTEM DESIGN

This section of the report details the system design. All technical choices on data, prediction and implementation will be detailed here.

4.1 Dataset

The dataset used to train the model employed by this study was collected from Wikiart, a database of more than 250.000 artworks from all around the world, categorized into 36 different art movements (all available art movements were used to capture different painting aspects) [19]. It consisted of 19730 images of art, which were split into a training set (80%) and a test set (20%) for evaluation. Since this art collection was the biggest collection with available metadata curated by volunteer editors, we found it to be particularly suitable for the training of our model. The choice to use paintings was made because these contain the most popular items within the collection, such as 'The Night Watch', and as such are expected to attract the largest userbase. Furthermore the quality of the metadata attached to the paintings is very high, which will enable the usage of both metadata filters and visual similarity based upon the scores assigned by the model.

4.2 Preprocessing

As the model expects an image input size of 224x224, the paintings were first transformed into a format of 224x224. Further, three different preprocessing pipelines were employed for the training. The first pipeline only converted the reformatted paintings to tensors. The second pipeline added auto-augmentation to the process, which involved resizing the images and converting them to tensors. The third pipeline added random augmentation to the process in addition to resizing and converting the images to tensors.

4.3 Model Architecture

ResNet50 is a deep convolution neural network architecture that was introduced by Microsoft Research [4]. The architecture consists of a series of residual blocks, where each block contains several convolution layers, batch normalization layers, and activation functions.

At its core, the ResNet50 architecture uses 50 convolution layers to extract features from an input image. The residual blocks allow information to be passed through the network more easily, which helps to reduce the vanishing gradient problem (i.e., the problem that due to the size of the model, the algorithm does not correctly propagate the gradient information from one end of the model back to the other end of the model) and improve the network's ability to learn. The network ends with a fully connected layer that produces a set of class scores, which are used to predict the class of an input image.

To prevent catastrophic forgetting, we limited the training of the model to only the last 2 layers.

4.4 Training Procedure

The model was optimized using Stochastic Gradient Descent (SGD) with a learning rate of 0.001 and a momentum of 0.9. A Cyclic Learning Rate (CLR) scheduler was employed to vary the learning rate, with a base

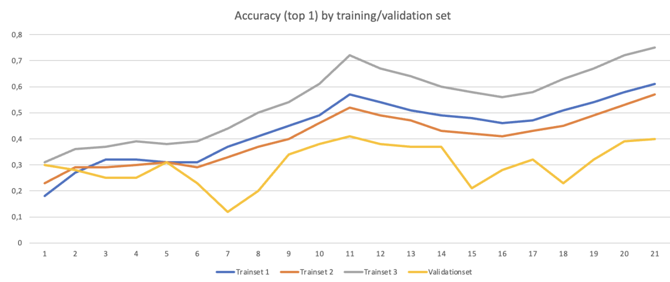

Figure 8: Accuracy for training/validationsets

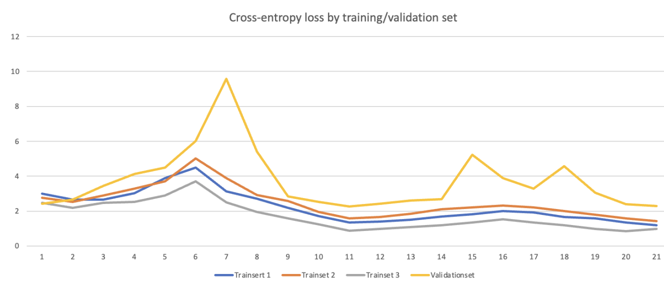

Figure 9: Accuracy for training/validationsets

learning rate of 0.001, a maximum learning rate of 0.1, and a step size of 5 using the triangular2 mode. The Cross-Entropy Loss function was used as the loss metric and the model was trained for 21 epochs with a batch size of 64, with each epoch consisting of training and validating three different datasets, provided by the three different preprocessing functions named above (resulting in 63 sub-epochs). The number of 21 epochs (63 sub-epochs) simply resulted for trying to run the model training for as long as possible and lacking time to train for longer.

4.5 Evaluation Metrics

The performance of the model was evaluated using accuracy and loss on the three training datasets resulting from the three diferrent preprocessing steps and the validation dataset per epoch (for performance plot, see Figures 8 and 9

4.6 Similarity Scores

After being scored for by the ResNet50 model, the cosine similarity between objects was calculated. Ultimately this cosine similarity score was used to match objects. The choice for cosine similarity over other methods such as euclidean similarity was made because the cosine similarity score is more measures the angle between vectors, rather than the distance. This allows for a robust method more easily scalable to sparse matrices.

5 VALIDATION METHODOLOGY

This chapter details the methodology by which the created system was validated. It details the experiment as well as participant demographics and used measuring methodology

5.1 Participants

A total of 14 participants were recruited to participate in the use study in the form a focus group, with a demographic distribution representative of the target user group of the museum. They were recruited through a convenience sample. The limitations of this approach are acknowledged (in that it might not generalize to a wider population) but this method allowed us to recruit relevant users for the study that fit the persona. Participants were provided with a brief description of the museum's digital collection and the purpose of the study before their participation. Demographic information about the participants can be seen in table 1.

The participants in this study consisted of 14 individuals with diverse backgrounds and nationalities. The average age of the participants was 25.07 years, with the youngest being 18 years old and the oldest being 31 years old. The participants in the sample consist of diverse nationalities (mainly European) and are evenly split in terms of gender.

5.2 Prototype

The prototype of the exploration platform was developed by the research team using Figma. Figma is a popular, collaborative web application for interface design that allows for highly elaborate interactive prototyping. Therefore, the program was chosen for the design of the user interface (UI). The prototype consisted of said UI for browsing the collection through filters, and also employed recommendations made by the model. As this is a prototype, these objects were added manually. This means that users could not explore beyond the created prototype's path and were instead confined to the created closed path. This was done to enable the participant to get a feel for the intended functionality. After working through this prototype, participants were shown another lo-fi prototype that allowed for endless exploration of the collection. This second prototype was used to test the participant's reaction to the linked objects and the way the recommendation system presents them.

5.3 Usability Testing

Participants were asked to use the UI to find art that appeals to them. They were asked to verbalize their thoughts and actions while doing so. This is a common approach within the UCD method since think-aloud allows the research to document and collect qualitative data regarding the user experience [2]. This open-ended task allowed us to understand the user's experience with the UI and identify potential usability issues.

5.4 Interviews & System Usability

Following the open-ended task, participants were interviewed using a semi-structured interview protocol. The interviews aimed to elicit feedback on the UI, the prototype functionality, as well as the participant's overall experience with the digital collection and the recommended items. Participants were asked to complete the 'System Usability Scale' (SUS) - a widely used tool for measuring the perceived usability of a system [7]. Eight items from the survey instrument were appropriated that measure the user experience, namely: complexity, ease of use, support needed, integration, inconsistency, ability to master the functions, confidence when using the system, and background needed to use the system. One survey item was added that measured the relevancy of the recommendation system [3]. All items were measured on a 5-point Likert scale.

5.5 Data Analysis

The data collected from the open-ended tasks, interviews, SUS questionnaire was analyzed qualitatively and quantitatively to identify patterns and trends. The data was analyzed to identify the strengths and weaknesses of the UI and functionality, as well as the participant's overall experience with the prototype. First, the mean scores from the SUS questionnaire were calculated and are discussed. Then, the transcripts from he usability-test (interviews) were grouped together based on similarity of the topic. The data was seperated on the following bases: feedback relating to UI; feedback relating to the functionality and recommendations.

Participant nr.	Gender	Age	Nationality	Background
1	Female	26	French	Medicine
2	Female	21	Italian	Marketing
3	Male	31	Latvian	Business
4	Male	25	Lithuanian	Hospitality
5	Male	28	Croatian	Computer Science
6	Female	19	German	Literature
7	Female	23	Mexican	Engineering
8	Male	29	Canadian	Biology
9	Male	24	Australian	Mathematics
10	Female	31	Swedish	Political Science
11	Female	18	Russian	Philosophy
12	Male	26	Norwegian	Finance
13	Female	20	Chinese	Psychology
14	Male	31	South African	Journalism

Table 1: Participant Demographic Table

No pre-defined codebook was used to analyze the data. Hence, the analysis in inductive in approach.

6 VALIDATION RESULTS

The described validation method was executed and analyzed. This chapter details the results of the user study.

6.1 System Usability

After conducting the focus group usability test, the participant scores for the 'System Usability Score' were aggregated and the mean was calculated. The results of the analysis are presented in Table 2.

On average, the fourteen participants evaluated the prototype slightly more positive than neutral, with an average score of 3.06 out of 5. This means that the responses and experiences were mixed in terms of user engagement, complexity, integration and functionality. The average score for the eight items relating to the UI amount to 2.99 out of 5, whereas the item measuring the relevance of the recommendations load onto an average score of 3.06 out of 5. This means that both the UI and relevancy of recommendations were perceived as neutral by the fourteen participants, instead of positive or negative.

6.1.1 UI. The results of the system usability test were mixed, with participants encountering several issues during their interaction with the system. Out of the 14 participants, a 6 reported confusion about the function of the first page, stating that it was not immediately clear that the page served as a filtering page for the collection of art. The positioning of the two buttons in the middle of the page, rather than on the left as is typical, was also seen as a source of confusion. Additionally, 11 participants noted that the text explaining the filters was too lengthy.

A common issue raised by the participants was the difficulty in changing the filters once they were applied. The participants felt that it would have been more efficient if the paintings were displayed next to the filters on the same page, rather than having to go back to the previous page to make changes. This was seen as raising the interaction cost.

Another problem encountered by the participants was the scrolling through the retrieved images. The images were often cropped and did not fit the window size, resulting in a 'wacky' (Participant nr. 5) scrolling experience. Some participants did not understand that they could click on an image to view details and see other similar works of art. The fact that the page could not be scrolled, but rather a window within the page, was also seen as negative.

On the positive side, participants found the page that displayed information about the images to be well-designed. However, some of the participants did not realize that they could scroll down to view recommended works of art, as they did not read the accompanying text.

Q	Interview Item	Mean Score
1	I found the system unnecessarily complex.	2.92
2	I thought the system was easy to use.	2.85
3	I think that I would need the support of a technical person to be able to use this system.	3.07
4	I found the various functions in this system were well integrated.	3.42
5	I thought there was too much inconsistency in this system.	2.42
6	I would imagine that most people would learn to use this system very quickly.	3.35
7	I felt very confident using the system.	2.92
8	I needed to learn a lot of things before I could get going with this system.	2.92
9	The related/recommended artwork that the system gave me was relevant for me.	3.64
Total	**UI related items**	**2.99**
Total	**Back End related items**	**3.64**
Total		**3.06**

Table 2: Validation Result Scores

6.1.2 Recommendation Relevance. While the recommended artwork was mostly perceived as relevant by the participants, the 'dissimilar' recommendations were seen as confusing in both the UI and functional backend prototype. Most of the participants commented on the explicit labeling of similar and dissimilar paintings in the UI as it is not intuitive to why this is present or explicit. The idea behind offering dissimilar items was to avoid over fitting, i.e. to give the user an option to 'escape' a potential 'bubble' of only relevant artwork, and to allow for serendipitous findings. When this was explained, the users reacted positively but emphasized that due to the initial unclarity it makes them perceive the system as inconsistent and confusing. Some of the participants suggested that the recommended art-work (both similar and dissimilar) should be presented as a scroll-able collage of paintings without explicit labeling.

7 CONCLUSION

Through a literature review and user studies, ArtEureka was designed to present the Rijksmuseum's digital collection in a more engaging manner. Initial user studies and literature studies pointed towards the necessity of a recommendation system. By using a convolutional neural network to classify the objects along 26 styles, and then matching them through cosine similarty, this project attempted to provide users with visually intuitive recommendations. User validation revealed many usability limitations within the current prototype, although the core explorations functionality was liked after an initial exploration. Filter functionality and navigation were spread across too many pages and did not allow for intuitive use. This resulted in a neutral score for the prototype implementation. The recommended images were liked by many users, but most users disliked the addition of dissimilar images. This was found to be unnecessary by many. A positive note lies in the perception of the relevancy of the similar objects, which were determined to be intuitively visually similar. The model performed at a first order accuracy of around 40%, which is not accurate enough to label the objects, but provides enough information to visually match similar objects. ArtEureka provides a core functionality that allows for the combination of metadata filters and visual object matching, which was found to work effectively.

8 DISCUSSION & FUTURE WORK

The ArtEureka project resulted in a strong recommendation system combining visual and textual cues. As a relatively small university project it was however constrained by time and resources. This means that there's much that can still be tested and implemented. The model's performance on the Rijksmuseum collection was not validated or measured beyond the judgement of users. Involving experts to validate classifications would be resource and time intensive, but without complications. This would enable us to know actual model performance, which is expected to be slightly below the performance on the WikiArt collection due to possible overfitting. This accuracy is largely irrelevant to the current implementation, which seeks to provide users with relevant objects. As was also found during literature research, the model's imperfections help prevent filter bubbles and enable 'fuzzy' object matching intuitive to human users. Our recommendation, and expectation, is for future work to focus on the presentation of the results. Mainly the dissimilar objects meant to enable serendipity were difficult on the users. Presenting these in a more intuitive manner could shift user perception and make their inclusion worthwhile. Further user testing is required for this. Another feature that may easily be added is to allow users to filter for and 'save' the works that are viewable within the museum. This would enable the usage of the platform to explore the collection and plan one's visit. Our belief is that the ArtEureka system provides a robust and effective exploration system, which still has many possibilities for improvement, most specifically within the system's user interface. To maximize the gain of further research, we therefore recommend to focus on a study that focuses on measuring user experience, keeping interaction costs low and providing the system's information in the most user-friendly and intuitive way possible.

REFERENCES

[1] David C. Chou. 2018. Applying design thinking method to social entrepreneurship project. *Computer Standards & Interfaces* 55 (1 2018), 73–79. https://doi.org/10.1016/J.CSI.2017.05.001

[2] Danielly F.O. De Paula, Bianca H.X.M. Menezes, and Cristiano C. Araújo. 2014. Building a quality mobile application: A user-centered study focusing on design thinking, user experience and usability. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* 8518 LNCS, PART 2 (2014), 313–322. https://doi.org/10.1007/978-3-319-07626-3{_}29/COVER

[3] Mojisola Erdt, Alejandro Fernández, and Christoph Rensing. 2015. Evaluating Recommender Systems for Technology Enhanced Learning: A Quantitative Survey. *IEEE Transactions on Learning Technologies* 8, 4 (10 2015), 326–344. https://doi.org/10.1109/TLT.2015.2438867

[4] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. 2016. Deep Residual Learning for Image Recognition. , 770–778 pages. http://image-net.org/challenges/LSVRC/2015/

[5] Het Rijksmuseum. 2020. Activiteitenplan 2021-2024. https://www.rijksmuseum.nl/en/support/donate/anbi-information

[6] F. O. Isinkaye, Y. O. Folajimi, and B. A. Ojokoh. 2015. Recommendation systems: Principles, methods and evaluation. *Egyptian Informatics Journal* 16, 3 (11 2015), 261–273. https://doi.org/10.1016/J.EIJ.2015.06.005

[7] Aycan Kaya, Reha Ozturk, and Cigdem Altin Gumussoy. 2019. Usability Measurement of Mobile Applications with System Usability Scale (SUS). (2019), 389–400. https://doi.org/10.1007/978-3-030-03317-0{_}32

[8] Denis Kotkov, Jari Veijalainen, and Shuaiqiang Wang. 2016. Challenges of Serendipity in Recommender Systems. *International conference on web information systems and technologies* 2 (2016), 251–256. https://doi.org/10.5220/0005879802510256

[9] Christian Matt, Alexander Benlian, Thomas Hess, and Christian Weiß. 2014. Escaping from the Filter Bubble? The Effects of Novelty and Serendipity on Users' Evaluations of Online Recommendations. In *International Conference on Interaction Sciences*.

[10] Cultuur en Wetenschap Ministerie van Onderwijs. 2018. Cultuur in een open samenleving. https://www.rijksoverheid.nl/documenten/rapporten/2018/03/12/cultuur-in-een-open-samenleving

[11] Nederlandse Omroep Stichting. 2016. Kamer wil minder subsidie voor Rijksmuseum. https://nos.nl/artikel/2108716-kamer-wil-minder-subsidie-voor-rijksmuseum

[12] Q42. 2022. Rijks website. https://www.q42.nl/werk/rijks-website

[13] Rijksmuseum. 2019. Rijksmuseum announces 2019 is its most successful year ever . https://www.rijksmuseum.nl/en/press/press-releases/rijksmuseum-announces-2019-is-its-most-successful-year-ever

[14] Antje Schmidt. 2018. MKG Collection Online: The potential of open museum collections. *Hamburger Journal für Kulturanthropologie* 7 (2018), 25–39. https://www.ssoar.info/ssoar/handle/document/62697

[15] Mette Skov and Peter Ingwersen. 2014. Museum Web search behavior of special interest visitors. *Library & Information Science Research* 36, 2 (4 2014), 91–98. https://doi.org/10.1016/J.LISR.2013.11.004

[16] Elena Villaespesa. 2019. Museum Collections and Online Users: Development of a Segmentation Model for the Metropolitan Museum of Art. *https://doi.org/10.1080/10645578.2019.1668679* 22, 2 (7 2019), 233–252. https://doi.org/10.1080/10645578.2019.1668679

[17] Krzysztof Walczak, Wojciech Cellary, and Martin White. 2006. Virtual museum exhibitions. *Computer* 39, 3 (3 2006), 93. https://doi.org/10.1109/MC.2006.108

[18] Wenjie Wang, Fuli Feng, and Liqiang Nie. 2022. User-controllable Recommendation Against Filter Bubbles. *Proceedings of the 45th International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR '22)* (6 2022), 1251–1261. https://doi.org/10.1145/3477495.3532075

[19] WikiArt Team. 2023. Wikiart.org.

[20] Xiang Sean Zhou and Thomas S. Huang. 2002. Unifying keywords and visual contents in image retrieval. *IEEE Multimedia* 9, 2 (4 2002), 23–33. https://doi.org/10.1109/93.998050