Correlating Eye Gaze with Object to Enrich Cultural Heritage Knowledge Graph

Shenghui Wang[1,*], Daria Kulyk[1], Delaram Javdani Rikhtehgar[1], Dirk Heylen[1] and Carolien Rieffe[1,2]

[1]University of Twente, Drienerlolaan 5, 7522 NB Enschede, The Netherlands
[2]Leiden University, Wassenaarseweg 52, 2333 AK Leiden, The Netherlands

Abstract

Virtual Reality (VR) technology has the potential to enhance cultural heritage experiences by providing immersive environments. This study proposes a novel approach that combines VR environments with eye-tracking data to identify users' points of interest in exhibition paintings. By leveraging gaze patterns, valuable insights into user preferences, behavior, and attention can be extracted and integrated into the cultural heritage knowledge graph. To achieve this, an object detection model is fine-tuned on historical artwork datasets, and statistical tests are conducted to analyze gaze-object correlations. The results demonstrate the feasibility of using an object detection algorithm to detect points of interest and reveal correlations between eye gaze patterns and meaningful objects in paintings. This approach has the potential to enrich the knowledge graph, enabling more personalized and immersive experiences for art enthusiasts and visitors.

Keywords

Image object detection, Eye gaze, Virtual reality, Knowledge Graph, Cultural Heritage

1. Introduction

Virtual Reality (VR) has revolutionized the preservation and exploration of cultural heritage [1], yet its potential to enrich the cultural heritage knowledge graph remains largely untapped. To bridge this gap, we propose an innovative approach that combines immersive VR environments with eye-tracking data to identify users' points of interest in exhibition paintings. By leveraging users' gaze patterns, we can extract valuable insights into their content preferences, behavior, and attention [2, 3]. This information can be seamlessly integrated into the cultural heritage knowledge graph, contributing to a more comprehensive representation of the artworks.

The stored information in the knowledge graph can then be utilized to customize the information provided to exhibition visitors about the artworks. Studies have shown that providing

ISWC 2023 Posters and Demos: 22nd International Semantic Web Conference, November 6–10, 2023, Athens, Greece
*Corresponding author.
shenghui.wang@utwente.nl (S. Wang); d.k.j.heylen@utwente.nl (D. Heylen); crieffe@fsw.leidenuniv.nl (C. Rieffe)
https://people.utwente.nl/shenghui.wang (S. Wang); https://people.utwente.nl/d.k.j.heylen (D. Heylen); https://www.universiteitleiden.nl/en/staffmembers/carolien-rieffe (C. Rieffe)
0000-0003-0583-6969 (S. Wang); 0009-0007-8266-8395 (D. J. Rikhtehgar); 0000-0003-4288-3334 (D. Heylen); 0000-0002-7584-6698 (C. Rieffe)

elaborate and content-specific information about artworks enhances understanding and aesthetic appreciation [4, 5]. By leveraging the insights gained from the enriched knowledge graph, museums and cultural institutions can tailor the descriptions and information provided to visitors, matching their preferences and increasing engagement and satisfaction. This approach aligns with the transition from collection-oriented museums to visitor-oriented ones [6], recognizing the importance of catering to diverse visitor preferences and interests.

However, eye-tracking data alone does not provide semantic meaning to the areas of gaze. While it indicates where someone is looking, it lacks details about the specific objects within the paintings that capture users' attention. This limitation makes it challenging to infer the type of content users are focused on and whether their fixations primarily concentrate on background elements or *meaningful* areas of the painting.

To address this challenge, manual annotation of objects by human observers is a possible solution. However, it becomes impractical for larger datasets due to the time-consuming nature of the task. To overcome this, we propose combining gaze data with an automated object detection model to add meaning to users' eye gaze. Our research aims to achieve two objectives: assessing the feasibility of using an object detection algorithm to detect points of interest and investigating the correlations between eye gaze patterns and meaningful objects in paintings.

To achieve these objectives, we first fine-tune an object detection model on a historical artwork dataset and evaluate its predictions on the 19 paintings presented in a special VR exhibition [7]. We then utilise participants' eye-tracking data collected during a user study using this exhibition to conduct statistical tests on *meaningful* areas of the paintings and determine the influence of object types on gaze duration.

2. Object detection

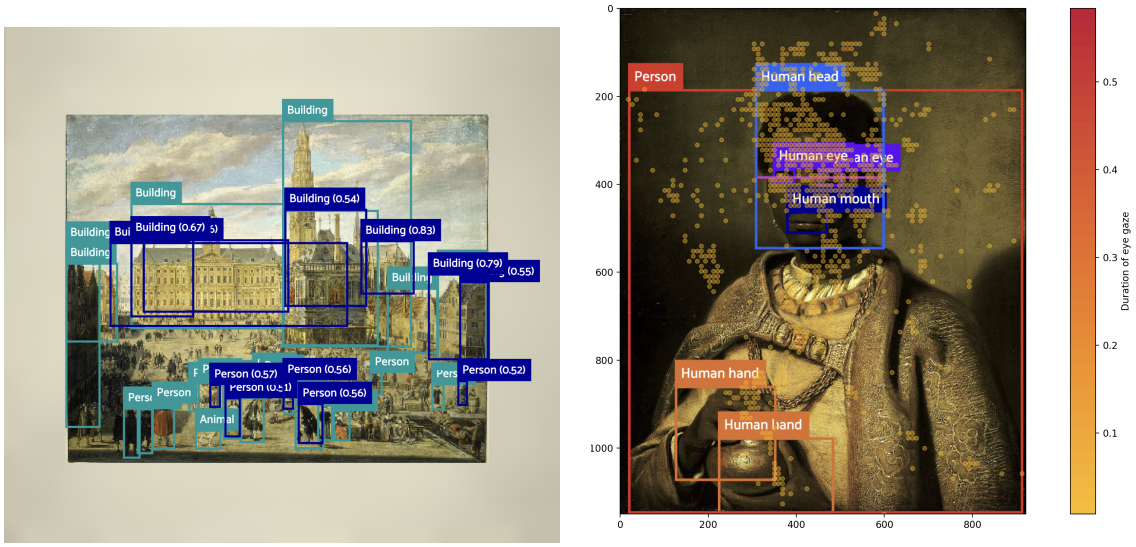
In this study, we fine-tuned the Faster R-CNN model,[1] pre-trained on MS COCO dataset,[2] using a subset of the Open Image V7 dataset[3] that contains images relevant to our study. Specifically, we focused on European fine-art paintings from the 17th century, categorized as portrait, genre, and landscape, as they closely resemble the ones presented in the VR exhibition. We sampled images of 11 Open Image categories (see Table 1 for the size of each category) that frequently appeared in the artistic descriptions of these paintings. These categories were chosen based on their relevance and occurrence in the SemArt dataset.[4] In our adapted model, we adjusted the output layer to accommodate 12 categories, including 11 categories of interest and one category representing the background.

For the VR exhibition, we manually annotated the 19 paintings using the same 11 categories. Figure 1 (a) provides an example of detected objects, marked in dark blue with their corresponding confidence scores, compared to the manual annotation marked in light blue. The performance of the fine-tuned object detection model on these paintings is summarized

[1]The Fast R-CNN model is a Convolutional Neural Networks (CNN)-based object detection framework that relies on a Region Proposal Network for efficient region detection within images [8].
[2]https://cocodataset.org/
[3]https://storage.googleapis.com/openimages/web/factsfigures_v7.html
[4]https://github.com/noagarcia/SemArt

Table 1
The size of the training set and the Average Precision scores (%) per image category

Metric	Animal	Building	Dress	Hat	Eye	Hair	Hand	Head	Mouth	Person	Tree
Train	248	2258	862	171	4684	4175	4104	5868	2373	11516	1496
AP	0	14.9	23.8	13.4	25.8	32.2	24.8	58.1	39.7	27.5	20.0
AP^{50}	0	30.4	38.7	46.1	59.4	74.1	50.3	91.2	67.0	59.8	27.1



(a) Object detection (b) Eye-gaze heat map

Figure 1: Example of object detection and eye-gaze heat map over bounding boxes

in Table 1.[5] The model demonstrated satisfactory performance in detecting various object classes, including small human-related objects like hair, eyes, and mouths. This indicates the potentials of transfer learning and fine-tuning for such classes. However, the model exhibited lower accuracy in detecting non-human classes, particularly animals, due to the limited training examples available and the small size of animal objects in the VR paintings. The performance in non-human classes such as trees, buildings, hats, and dresses also suffered, potentially attributed to the unresolved cross-depiction problem and the scarcity of training examples for garments.

3. Gaze-object correlation

As reported in [7], a total of 31 participants visited 19 paintings in the VR exhibition. For each painting, the eye-gaze data of each participant was recorded and represented as a heat map with a 100 x 100 grid overlaying the painting. The heat map captured the duration of the participant's eye gaze within each cell of the grid. Figure 1 (b) provides an example of an eye-gaze heat map from one participant, overlaid with manually annotated bounding boxes.

[5]Detailed overview of these average precision metrics can be found at https://cocodataset.org/#detection-eval.

For each gaze point, we determined whether it fell within any of the manually annotated object bounding boxes ("on-object") or outside of them ("out-object") for the painting. For gaze points marked as "on-object," we collected information on the corresponding object categories and the duration of that gaze point. By analyzing the collected data, we calculated the average duration of gaze on and off objects for each participant across all paintings, as well as the specific gaze duration on individual objects.

To ensure the reliability of our analysis, we confirmed the normality of the data using the Shapiro-Wilk test.Subsequently, a paired t-test revealed a significant positive average difference in gaze duration between objects and areas without objects ($t_{31} = 6.33, p < 0.001$). This finding indicates that, on average, participants exhibited a selective focus on meaningful areas of the paintings, demonstrating their interest and engagement with the artwork. Additionally, a one-way ANOVA test demonstrated a significant difference in the average time spent on different categories of objects ($F(2) = 6.607, p < 0.001$). Notably, participants, on average, allocated significantly more time to viewing buildings compared to human heads and figures.

4. Conclusion

This study presents a novel approach that combines VR environments, eye-tracking data, and object detection to enhance the cultural heritage knowledge graph. By fine-tuning an object detection model on historical artwork datasets, it is possible to identify and assign semantics to potential areas of interest within exhibition paintings. The statistical tests conducted reveal correlations between eye gaze patterns and meaningful objects depicted in the paintings. This approach holds great potential for enriching the knowledge graph, thereby paving the way of providing a more immersive and personalized experience for art enthusiasts and visitors.

Next, we will further explore the development of more advanced models and gather more suitable training data to improve the efficiency for object detection in paintings. Additionally, we will utilise semantic technologies to explicitly integrate detected objects and their associated knowledge via manual annotation or entity linking, further enriching the culture heritage knowledge graph. In the future, we aspire to develop eye-gaze-based interactions that will usher in personalised, immersive experiences within the realm of VR.

Acknowledgment

The authors would like to express their gratitude to Museum Rembrandthuis for their support and for providing the essential exhibition information. They would also like to thank Claudia Alessandra Libbi and Rens van der Werff for their development of the VR exhibition utilized in this study.

References

[1] M. Shehade, T. Stylianou-Lambert, Virtual Reality in Museums: Exploring the Experiences of Museum Professionals, Applied Sciences 2020, Vol. 10, Page 4031 10 (2020) 4031. doi:10.3390/APP10114031.

[2] M. Mu, M. Dohan, A. Goodyear, G. Hill, C. Johns, A. Mauthe, User attention and behaviour in virtual reality art encounter, Multimedia Tools and Applications (2022) 1–30. doi:10.1007/S11042-022-13365-2/FIGURES/22.

[3] X. Li, Y. Shan, W. Chen, Y. Wu, P. Hansen, S. Perrault, Predicting user visual attention in virtual reality with a deep learning model, Virtual Reality 25 (2021) 1123–1136. doi:10.1007/s10055-021-00512-7.

[4] V. Swami, Context matters: Investigating the impact of contextual information on aesthetic appreciation of paintings by Max Ernst and Pablo Picasso, Psychology of Aesthetics, Creativity, and the Arts 7 (2013) 285–295. doi:10.1037/a0030965.

[5] H. Leder, C. C. Carbon, A. L. Ripsas, Entitling art: Influence of title information on understanding and appreciation of paintings, Acta Psychologica 121 (2006) 176–198. doi:10.1016/j.actpsy.2005.08.005.

[6] D. Pantile, R. Frasca, A. Mazzeo, M. Ventrella, G. Verreschi, New Technologies and Tools for Immersive and Engaging Visitor Experiences in Museums: The Evolution of the Visit-Actor in Next-Generation Storytelling, through Augmented and Virtual Reality, and Immersive 3D Projections, in: Proceedings of the 12th International Conference on Signal Image Technology and Internet-Based Systems, 2017, pp. 463–467. doi:10.1109/SITIS.2016.78.

[7] D. Javdani Rikhtehgar, S. Wang, H. Huitema, J. Alvares, S. Schlobach, C. Rieffe, D. Heylen, Personalizing Cultural Heritage Access in a Virtual Reality Exhibition: A User Study on Viewing Behavior and Content Preferences, Adjunct Proceedings of the 31st ACM Conference on User Modeling, Adaptation and Personalization (2023) 379–387. doi:10.1145/3563359.3596666.

[8] S. Ren, K. He, R. Girshick, J. Sun, Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks, IEEE Transactions on Pattern Analysis and Machine Intelligence 39 (2015) 1137–1149. doi:10.1109/TPAMI.2016.2577031.