Paving the Way for Personalized Museums Tours in the Metaverse

Alex Falcon[1,*], Beatrice Portelli[1,2], Ali Abdari[1,2] and Giuseppe Serra[1]

[1]*University of Udine, Italy*

[2]*University of Naples Federico II, Italy*

Abstract

Museums play a central role in the preservation and communication of human history. With the advent of powerful and accessible Virtual Reality technologies, new Metaverse Museums started being developed, creating new possibilities for discovering and experiencing knowledge from all over the world. In anticipation of this technology becoming more widely adopted, we must prepare tools to aid future visitors in finding and navigating the museums and exhibitions which are more relevant to their current interests. In this light, Deep Learning methods could be of great use for modeling Metaverse museums, retrieving the most relevant ones, and creating personalized tours of the artifacts contained within them. In this paper, we present our research project, "Personalized Museum Tours in the Metaverse" led by the Artificial Intelligence Laboratory of Udine (AILAB Udine), detailing its objectives, proposed methodology, and future directions.

1. Introduction and Motivation

Museums are important institutions in our society, which are open to the public and conserve, acquire, exhibit, and communicate testimonies of man and their environment. As society changes and transforms in time, museums also have renewed themselves, adapting their objectives, roles, and the way they interact with the public. In particular, nowadays it is not unusual for them to leverage and incorporate digital technologies to offer a more immersive experience for the users, for example through downloadable guides, QR-codes next to the artworks to access detailed descriptions, and even online exhibitions.

The peak expression of this evolution would be a completely Digitized Museum, or a Museum in the Metaverse, accessible through Augmented Reality (AR) technologies or even completely built-in Virtual Reality (VR). There are several reasons why Metaverse Museums could be a beneficial evolution of this institution: wider access to history and knowledge from all over the

Figure 1: A still shot from two Metaverse museums. (**left**) VR Museum of Fine Arts. (**right**) Smithsonian American Art Museum "Beyond The Walls".
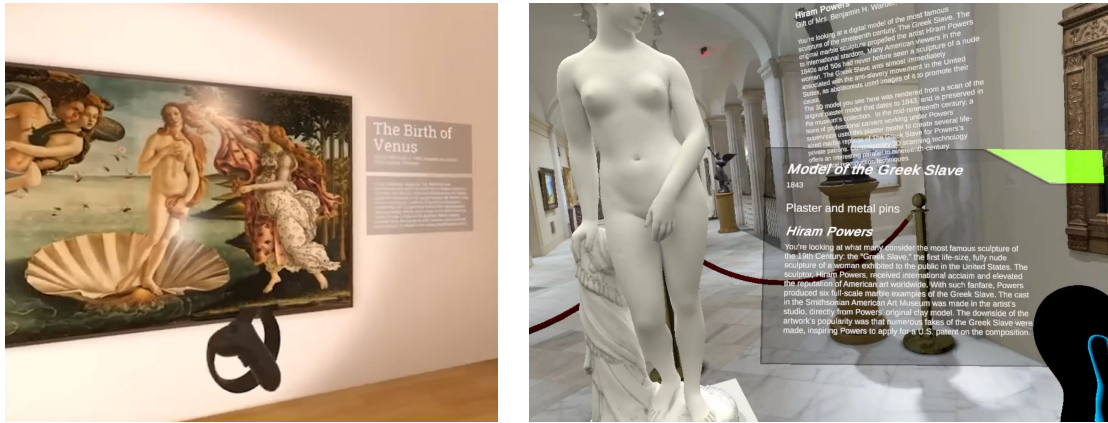
world; a reduction in carbon emissions caused by people physically traveling to the museum, especially for long-distance travels; a safer way to visit exhibitions in case of healthcare crises (e.g., the recent COVID-19 pandemic); a more accessible way to visit exhibitions for people with physical impairment or other restrictions that prevent them from safely visiting museums in person. Additionally, a museum in the Metaverse allows for more flexibility and interactivity in the contents presented to the visitors. For example, it could contain recreations of historical places and artifacts that no longer exist (e.g., project Rekrei[1]), and provide access to places too delicate for physical interaction. Consider the Lascaux Cave in France, closed due to the paintings' deterioration from light, air changes, and carbon monoxide from visitors' breath[2].

However, if Metaverse Museums were to become more common and accessible, it could be difficult for users to choose among all the possible interesting exhibitions and navigate the immense amount of artifacts and information. This is an area where Artificial Intelligence (AI) and Deep Learning (DL) techniques can be leveraged to alleviate the cognitive load on the user and enhance the discoverability of the artworks.

Our project aims to develop a multimedia system for the creation of personalized Metaverse museum tours based on the user interests, formulated through a textual query. Realizing such a system entails an in-depth analysis of heterogeneous types of data, including visual (3D scenes, paintings, statues, etc) and textual (artifact descriptions and user queries). Notably, this project deals with many innovative aspects: the analysis of Metaverse museums and their artifacts, the estimation of their relevance to the user interests, and the automatic generation of personalized tours, blending the artifacts' descriptions with the desires of the user and vast amounts of external knowledge.

---

[1]https://rekrei.org/about
[2]https://archeologie.culture.gouv.fr/lascaux/en/preservation-lascaux-hill
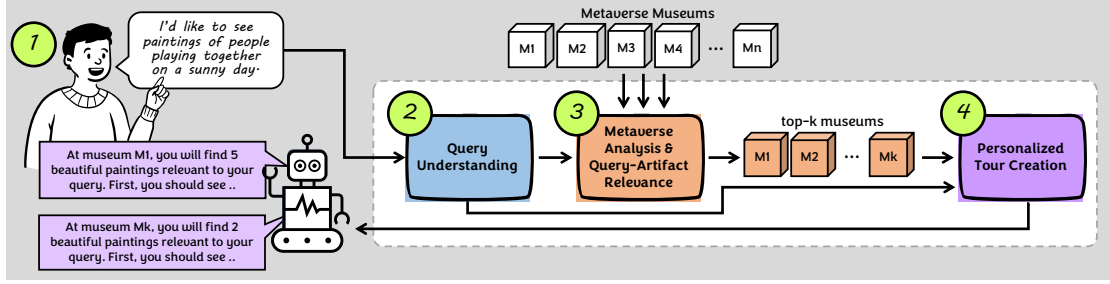
Figure 2: Overview of the proposed system. (1) The user formulates a query. (2) The Query Understanding module encodes its semantic contents with a compact representation. (3) The Metaverse Analysis & Query-Artifact Relevance module encodes and ranks the museums according to their relevance to the query. (4) The Personalized Tour Creation module generates a brief tour of each relevant museum.

2. Personalized Metaverse Museums Tour Creation: Project Description

The research project stems from the need for tools to understand, represent, and navigate complex scenes such as digitized Metaverse museums. The final objective is to develop a system to create personalized Metaverse museum tours, given a user query. Figure 2 presents an overview of the proposed system.

The user interacts with the system by formulating a query in natural language. The **Query Understanding** module analyzes the user query and represents it by means of a compact representation wherein its semantics are preserved. This representation will serve as guidance for the **Metaverse Analysis** & **Query-Artifact Relevance** module. In fact, this module processes all the available Metaverse museums and ranks them according to their relevance to the user query. After a subset of the most relevant museums is identified, the **Personalized Tour Creation** module outputs for each of them a suggested itinerary covering the relevant artifacts and explaining how they relate to the user desires.

2.1. Query Understanding: Modelling the User Query

The advancements in natural language processing made in the past few years enabled interactive applications to understand the semantics of user-generated queries. These include word-level techniques, such as word2vec [1], GloVe [2], and fastText [3], which compute a vector representation for each word, hence using the words as a unit of meaning; and sentence-level techniques, such as BERT [4], T5 [5], and GPT [6], which use sentences as a unit of meaning.

In practice, to have a better understanding of the user query and its implications, in our project we will employ a combination of both word- and sentence-level techniques.

Notably, the user queries may be **objective** and ask for specific authors or art schools, but they also may be much more **subjective** and focus on feelings or contain very general and imprecise prompts (e.g., "happy people playing on a beach" or "mysterious doctors performing an autopsy"). Therefore, it becomes crucial to integrate these two types of information into

the training procedure. To achieve this goal, we plan to do the following: for each painting, we consider both an objective description of its visual contents and one or more subjective descriptions capturing the emotions that it sparked in the observer. Then, during training, the knowledge from both types of descriptions is integrated into the representations of the visual artifacts and the user queries. Another tool to relate subjective and objective descriptions is represented by models trained for formality style transfer [7], as they paraphrase a sentence from one style to another while preserving the semantic meaning. They could be used either to generate training data or to transform the user query and use it to match more artifacts.

2.2. Metaverse Analysis & Query-Artifact Relevance: Modelling the Museum and the Cross-modal Interactions

To understand the contents of a Metaverse museum and relate them to the user query, both **local** (fine-grained, artifact-level) and **global** (coarse, museum- or room-level) information are needed. A representation is then computed for each museum by aggregating these two sources of information. Finally, they are **ranked** according to their relevance to the user query, obtaining the final output of this module.

Local Information. To provide the subsequent modules with precise knowledge of the relevant artifacts present in the room/museum, the artifacts must first be identified and analyzed. To automatize this process, several well-known and widely used object detectors in the vision community can be leveraged, such as YOLO [8] and Faster R-CNN [9]. Once the artifacts are localized, multiple methodologies can be followed to obtain the needed local information.

A first approach represents each object in the scene with its categorical and spatial attributes (e.g., size, position, and rotation in the 3D space), as done in [10]. Although these attributes can create a representation with discriminative qualities, there are two limitations: first, the obtained representation lacks visual grounding; second, it will not lie in the same embedding space as the query, hence requiring a finetuning step to align the visual and textual spaces.

Another possible approach uses a jointly trained vision-language model (e.g., CLIP [11]) to extract a representation for each artifact. Interestingly, since such an approach involves learning a joint vision-language embedding space, the artifact representation will lie in the same joint embedding space as its own description, hence making it easier to understand whether it is relevant to the query. We followed a similar approach with positive results in [12], where we developed a system to rank digital apartments according to a textual query.

Global Information. To obtain the global information, a naive solution could aggregate the pieces of local information obtained in the previous step, e.g., by using a Variational Autoencoder [10] to represent a 3D scene through its objects as done in our previous work [13] about Metaverse scene retrieval. Also, knowledge graphs can aggregate the local information by discovering relationships among the artifacts in a room (or museum) [14, 15]. An alternative method, diverging from the aggregation of local information, involves the use of 3D Convolutional Neural Networks (CNNs, e.g., [16, 17]) to capture finer-grained spatial relations between the artifacts. Spatial attention plays a key role in reducing the amount of redundancy in the learned representations due to the presence of large empty spaces such as walls, halls,

etc. However, the obtained representations would not lie in the same embedding space of the queries, hence requiring further finetuning to align the two spaces.

Query Relevance and Ranking. The final step performed by this module consists of ranking the museums according to their relevance to the user query. Two important aspects should be considered during this step. First, a relevance score should be computed for each artifact, since few of them are likely to be highly relevant to the query, and those few will impact the final relevance of the museum as well as becoming the key elements of the personalized tour. Second, the relevance value for the museum is influenced by several global aspects, including the number of relevant artifacts per room (or within the museum), and how the relevance values are distributed (e.g. if there are many low-relevance artifacts or only a few highly relevant ones). Moreover, further constraints could be added, such as ranking the museums based on how much the relevant artifacts are clustered within the museum's rooms. The output of this step is a ranking of the museums, where for each museum a few artifacts are highlighted as the most relevant ones for the query.

2.3. Personalized Tour Creation: Presenting Suggestions to the User

For each of the museums selected by the previous module, the Personalized Tour Creation module creates a textual summary of the relevant artifacts and motivation on their relevance to the user query and interests. To put together this information, we will use a Large Language Model (LLM) such as ChatGPT, or open-source alternatives such as LLaMa [18] and the recent Phi2 [19]. These models are impressive tools to automatically create coherent and well-written summaries. Moreover, since they are pretrained on large amounts of textual data from the Web, they also have external knowledge of the artifacts which may be useful in adding more insights to the personalized tour. Finally, due to the model's prompt-based nature, the tour can be personalized by creating specialized prompts. For instance, the summary obtained by prompting ChatGPT with "create a summary of why the artifacts *The Annunciation by Leonardo da Vinci, The Annunciation by Benvenuto Tisi, and The Annunciation by Sandro Botticelli* are relevant to the user interests defined by *'I want to see the graceful Holy Mary being announced the birth of Jesus by the angel'*" contains both the motivation for the relevance (e.g., "The angel Gabriel's announcement to Mary is depicted with a sense of divine tranquility, emphasizing Mary's graceful response to the news") and some key features of the artifact or the style pursued by the author (e.g., "Da Vinci's meticulous attention to detail and his ability to convey emotions through facial expressions make this artwork a captivating representation of the Annunciation").

2.4. Dataset Creation

At the time of writing, there are no public datasets containing a collection of Metaverse museums. Therefore, we decide to create a new dataset as a resource for us and future researchers. In particular, the dataset should contain a large amount of digital museums, where each room contains multiple artifacts, including paintings, 3D sculptures, and other items of interest. Moreover, to enable fine-grained cross-modal applications and to estimate the relevance of each item to the user query, detailed descriptions of the artifacts are needed. As mentioned in the

previous sections, both subjective and objective textual descriptions are crucial to include in the dataset, due to the wide range of possible user queries.

To obtain a large amount of digital museums, we resort to the creation of a synthetic dataset. This process involves several steps: first, collecting a diverse set of artifacts, including 2D and 3D art; second, obtaining meaningful descriptions for the artifacts, both objective and subjective; third, creating each 3D museum with multiple rooms, each containing several artifacts.

Metaverse Museums Creation. We are currently developing an automatic approach to create 3D museums, each containing multiple rooms with a variable number of sculptures and/or paintings. The items are either attached to the walls or placed in key positions in the room (e.g., corners or center of the room). The items will be grouped in the rooms based on topic, artist, theme, or other criteria, in order to create realistic thematic exhibitions.

2D Artwork Collection. To collect painting data and descriptions, we consider several public datasets, focusing on the type of annotations. Image captioning techniques like MPlug [20] and datasets such as ArtCap [21] provide objective descriptions. However, they both lack expert insights into interpretation and historical context, and subjective feelings expressed by observers. For the insights, SemArt [22] offers contextual details and artistic comments on over 21k paintings from the Web Gallery of Art (WGA). ArtEmis [23] includes 400k non-expert annotations about the emotions felt while observing fine art, while ArtELingo [24] adds multilingual annotations and insights into cultural variations in emotional responses to the same artwork.

3D Artwork Collection. While there are many popular datasets for 3D objects (e.g., ShapeNet [25]), large open datasets on 3D statues and sculptures are not readily available. Recently, Objaverse [26] and ObjaverseXL [27] were released, which contain over 800k and 10M annotated 3D objects respectively. Among these objects, there are also artistic statues and sculptures which are relevant for our purposes. Unfortunately, the textual annotations are sparse and do not contain an extensive description of the item's visual appearance or any subjective comments. To address this, captioning techniques could be used to obtain objective descriptions, whereas artistic comments could be scraped (e.g., from the WGA when possible).

3. Related work

3.1. Research on Applications for the Metaverse

A widespread range of applications have been designed for the Metaverse, ranging from entertainment to virtual shopping [28, 29] and apartments scouting [12], from education to healthcare [30, 31] and industrial applications [32, 33]. Although some Metaverse museums are available online (e.g., the Museum of Other Realities[3], and The Vordun Museum[4]), the research on them is still in their infancy [34, 35] and many topics on their automatic understanding are widely

---

[3]https://www.museumor.com/
[4]https://secondlife.com/destination/vordun-museum-and-gallery

unexplored. For instance, the task of retrieving multimedia-enriched scenarios by means of textual queries was only recently introduced [13]. Differently, in this project, we aim to create personalized Metaverse museum tours in an automatic way.

3.2. Retrieving and Ranking Artistic Artifacts

The task of retrieving and ranking artifacts according to their relevance to a user query is similar to the more general text-to-image retrieval task, which has seen great improvements in recent years, leading to the development of high-performing tools [11]. Conde and Turgutlu [36] developed a direct transposition of CLIP for the art domain, a pretrained model that can be used to describe images containing artworks or retrieve artworks given a description. Other interesting approaches include an unsupervised method [37] to discover the relationships among different artworks and schools only based on visual features, and different approaches employing knowledge graphs to capture the underlying relationships between other artistic attributes too, such as style, movement, subject, and period [14, 15].

3.3. Creating a Personalized Museum Tour

Suggesting a museum tour or a city itinerary to tourists is a difficult topic which sparked interest in the research community since before the advent of deep learning. Early approaches for creating personalized demo tours included three main steps: first, a user model is created by asking the users to rate artworks; second, an art recommendation step is performed based on the user model and a set of predefined topics; third, a tour creation tool selects a fixed amount of artworks from those recommended in the previous step [38, 39]. More recent literature on this topic often focuses on improving or optimizing the itinerary based on several constraints, e.g., travel costs [40] or visit duration [41]. Three main differences set our work apart from the previous ones. First, the art recommendation part is done automatically through cross-modal analysis, letting the user inform the model about their interests via text, therefore avoiding the imposition of predefined topics. This approach is suitable for modeling user interests, whether they are described using technical terms or a simpler lexicon. It also helps in learning how to contextualize the user query within the artwork visuals and textual descriptions. Moreover, our procedure does not require every artwork to be tagged based on a set of predefined topics, demanding expert knowledge to create a minimal set of comprehensive tags and extensive human efforts from experts to tag every artwork featured in a virtual museum, therefore limiting scalability. Second, the different context (Metaverse) makes some of the previously investigated constraints less relevant, such as optimizing travel costs or even visit duration, since the user can visit the museum during different sessions and across several days from the comfort of their home. Third, our aim is to not only provide a list of artifacts to the user, but to inform them, creating a brief summary explaining why the user should visit a certain museum, while specifying a selection of the most relevant artifacts. We envision LLMs as the main tool to create a coherent tour and to provide contextual information related to the artifacts. Nonetheless, external knowledge could also be integrated from knowledge bases, similarly to [42], which combines LLMs and Wikipedia to generate in-depth descriptions of artworks including background information about the history of the painting too.

4. Conclusions

In this paper, we presented a research project detailing the first steps to realize an innovative multimedia system for creating personalized museum tours in the Metaverse. This project is motivated by the increasing attention driven to the Metaverse and its vast possibilities in terms of interactive and immersive experiences, and the importance of spreading knowledge on fine arts and cultural heritage across the globe. We provided an overview of the proposed system, and in-depth description of its components and how they relate to state-of-the-art methodologies in vision-language understanding. Given the lack of suitable public datasets for experimentation, we also provided a detailed explanation of the data collection process that is being currently carried out. We hope that in the near future the dataset will be a valuable resource for researchers and practitioners working on the intersection of Metaverse and multimedia understanding, and that this project will spark interest in the computer science and cultural heritage community fostering collaborative endeavors.

Acknowledgements

References

[1] Q. Le, T. Mikolov, Distributed representations of sentences and documents, in: Proceedings of ICML, PMLR, 2014, pp. 1188–1196.

[2] J. Pennington, R. Socher, C. D. Manning, Glove: Global vectors for word representation, in: Proceedings of EMNLP, 2014, pp. 1532–1543.

[3] A. Joulin, É. Grave, P. Bojanowski, T. Mikolov, Bag of tricks for efficient text classification, in: Proceedings of EACL, 2017, pp. 427–431.

[4] J. Devlin, M.-W. Chang, K. Lee, K. Toutanova, Bert: Pre-training of deep bidirectional transformers for language understanding, in: Proceedings of NAACL-HLT, 2019, pp. 4171–4186.

[5] C. Raffel, N. Shazeer, A. Roberts, K. Lee, S. Narang, M. Matena, Y. Zhou, W. Li, P. J. Liu, Exploring the limits of transfer learning with a unified text-to-text transformer, The Journal of Machine Learning Research 21 (2020) 5485–5551.

[6] A. Radford, K. Narasimhan, T. Salimans, I. Sutskever, et al., Improving language understanding by generative pre-training (2018).

[7] S. Rao, J. Tetreault, Dear sir or madam, may I introduce the GYAFC dataset: Corpus, benchmarks and metrics for formality style transfer, in: Proceedings of NAACL-HLT, 2018, pp. 129–140.

[8] J. Redmon, S. Divvala, R. Girshick, A. Farhadi, You only look once: Unified, real-time object detection, in: Proceedings of IEEE CVPR, 2016, pp. 779–788.

[9] S. Ren, K. He, R. Girshick, J. Sun, Faster r-cnn: Towards real-time object detection with region proposal networks, Advances in neural information processing systems 28 (2015).

[10] H. Yang, Z. Zhang, S. Yan, H. Huang, C. Ma, Y. Zheng, C. Bajaj, Q. Huang, Scene synthesis via uncertainty-driven attribute synchronization, in: Proceedings of IEEE/CVF ICCV, 2021, pp. 5630–5640.

[11] A. Radford, J. W. Kim, C. Hallacy, A. Ramesh, G. Goh, S. Agarwal, G. Sastry, A. Askell, P. Mishkin, J. Clark, et al., Learning transferable visual models from natural language supervision, in: Proceedings of ICML, 2021, pp. 8748–8763.

[12] A. Abdari, A. Falcon, G. Serra, Farmare: a furniture-aware multi-task methodology for recommending apartments based on the user interests, in: Proceedings of IEEE/CVF ICCV, 2023, pp. 4293–4303.

[13] A. Abdari, A. Falcon, G. Serra, Metaverse retrieval: Finding the best metaverse environment via language, in: Proceedings of 1st International Workshop on Deep Multimodal Learning for Information Retrieval, 2023, pp. 1–9.

[14] G. Castellano, V. Digeno, G. Sansaro, G. Vessio, Leveraging knowledge graphs and deep learning for automatic art analysis, Knowledge-Based Systems 248 (2022) 108859.

[15] N. Garcia, B. Renoust, Y. Nakashima, Contextnet: representation and exploration for painting classification and retrieval in context, International Journal of Multimedia Information Retrieval 9 (2020) 17–30.

[16] J. Hou, B. Graham, M. Nießner, S. Xie, Exploring data-efficient 3d scene understanding with contrastive scene contexts, in: Proceedings of IEEE/CVF CVPR, 2021, pp. 15587–15597.

[17] S. Peng, K. Genova, C. Jiang, A. Tagliasacchi, M. Pollefeys, T. Funkhouser, et al., Openscene: 3d scene understanding with open vocabularies, in: Proceedings of IEEE/CVF CVPR, 2023, pp. 815–824.

[18] H. Touvron, L. Martin, K. Stone, P. Albert, A. Almahairi, Y. Babaei, N. Bashlykov, S. Batra, P. Bhargava, S. Bhosale, et al., Llama 2: Open foundation and fine-tuned chat models, arXiv preprint arXiv:2307.09288 (2023).

[19] M. Abdin, J. Aneja, S. Bubeck, C. César Teodoro Mendes, W. Chen, A. Del Giorno, R. Eldan, S. Gopi, S. Gunasekar, M. Javaheripi, P. Kauffmann, Y. T. Lee, Y. Li, A. Nguyen, G. de Rosa, O. Saarikivi, A. Salim, S. Shah, M. Santacroce, H. Singh Behl, A. Taumann Kalai, X. Wang, R. Ward, P. Witte, C. Zhang, Y. Zhang, Phi-2: The surprising power of small language models, https://www.microsoft.com/en-us/research/blog/phi-2-the-surprising-power-of-small-language-models/, 2023. Accessed: February 12, 2024.

[20] C. Li, H. Xu, J. Tian, W. Wang, M. Yan, B. Bi, J. Ye, H. Chen, G. Xu, Z. Cao, et al., mplug: Effective and efficient vision-language learning by cross-modal skip-connections, in: Proceedings of EMNLP, 2022, pp. 7241–7259.

[21] Y. Lu, C. Guo, X. Dai, F.-Y. Wang, Artcap: A dataset for image captioning of fine art paintings, IEEE Transactions on Computational Social Systems (2022).

[22] N. Garcia, G. Vogiatzis, How to read paintings: semantic art understanding with multimodal retrieval, in: Proceedings of ECCV Workshops, 2018.

[23] P. Achlioptas, M. Ovsjanikov, K. Haydarov, M. Elhoseiny, L. J. Guibas, Artemis: Affective language for visual art, in: Proceedings of IEEE/CVF CVPR, 2021, pp. 11569–11579.

[24] Y. Mohamed, M. Abdelfattah, S. Alhuwaider, F. Li, X. Zhang, K. Church, M. Elhoseiny,

Artelingo: A million emotion annotations of wikiart with emphasis on diversity over language and culture, in: Proceedings of EMNLP, 2022, pp. 8770–8785.

[25] A. X. Chang, T. Funkhouser, L. Guibas, P. Hanrahan, Q. Huang, Z. Li, S. Savarese, M. Savva, S. Song, H. Su, et al., Shapenet: An information-rich 3d model repository, arXiv preprint arXiv:1512.03012 (2015).

[26] M. Deitke, D. Schwenk, J. Salvador, L. Weihs, O. Michel, E. VanderBilt, L. Schmidt, K. Ehsani, A. Kembhavi, A. Farhadi, Objaverse: A universe of annotated 3d objects, in: Proceedings of IEEE/CVF CVPR, 2023, pp. 13142–13153.

[27] M. Deitke, R. Liu, M. Wallingford, H. Ngo, O. Michel, A. Kusupati, A. Fan, C. Laforte, V. Voleti, S. Y. Gadre, E. VanderBilt, A. Kembhavi, C. Vondrick, G. Gkioxari, K. Ehsani, L. Schmidt, A. Farhadi, Objaverse-xl: A universe of 10m+ 3d objects, arXiv preprint arXiv:2307.05663 (2023).

[28] A. Dawson, et al., Data-driven consumer engagement, virtual immersive shopping experiences, and blockchain-based digital assets in the retail metaverse, Journal of Self-Governance and Management Economics 10 (2022) 52–66.

[29] W. Song, Y. Gong, Y. Wang, Vtonshoes: Virtual try-on of shoes in augmented reality on a mobile device, in: IEEE ISMAR, 2022, pp. 234–242.

[30] H. Laaki, Y. Miche, K. Tammi, Prototyping a digital twin for real time remote control over mobile networks: Application of remote surgery, IEEE Access 7 (2019) 20325–20336.

[31] Y. Liu, L. Zhang, Y. Yang, L. Zhou, L. Ren, F. Wang, R. Liu, Z. Pang, M. J. Deen, A novel cloud-based framework for the elderly healthcare services using digital twin, IEEE Access 7 (2019) 49088–49101.

[32] M. H. Farhat, X. Chiementin, F. Chaari, F. Bolaers, M. Haddar, Digital twin-driven machine learning: ball bearings fault severity classification, Measurement Science and Technology 32 (2021) 044006.

[33] A. Siyaev, G.-S. Jo, Towards aircraft maintenance metaverse using speech interactions with virtual objects in mixed reality, Sensors 21 (2021) 2066.

[34] Z. Gao, T. C. Braud, et al., Vr-driven museum opportunities: digitized archives in the age of the metaverse, Artnodes (2023) 1–14.

[35] M. C. Longo, R. Faraci, Next-generation museum: A metaverse journey into the culture, Sinergie Italian Journal of Management 41 (2023) 147–176.

[36] M. V. Conde, K. Turgutlu, Clip-art: Contrastive pre-training for fine-grained art classification, in: Proceedings of IEEE/CVF CVPR, 2021, pp. 3956–3960.

[37] G. Castellano, E. Lella, G. Vessio, Visual link retrieval and knowledge discovery in painting datasets, Multimedia Tools and Applications 80 (2021) 6599–6616.

[38] L. Aroyo, N. Stash, Y. Wang, P. Gorgels, L. Rutledge, Chip demonstrator: Semantics-driven recommendations and museum tour generation, in: International Semantic Web Conference, Springer, 2007, pp. 879–886.

[39] Y. Wang, N. Stash, R. Sambeek, Y. Schuurmans, L. Aroyo, G. Schreiber, P. Gorgels, Cultivating personalized museum tours online and on-site, Interdisciplinary science reviews 34 (2009) 139–153.

[40] J. L. Sarkar, A. Majumder, A new point-of-interest approach based on multi-itinerary recommendation engine, Expert Systems with Applications 181 (2021) 115026.

[41] P. Yochum, L. Chang, T. Gu, M. Zhu, An adaptive genetic algorithm for personalized

itinerary planning, IEEE Access 8 (2020) 88147–88157.

[42] Z. Bai, Y. Nakashima, N. Garcia, Explain me the painting: Multi-topic knowledgeable art description generation, in: Proceedings of IEEE/CVF ICCV, 2021, pp. 5422–5432.