Exploring research fronts and topics of Big Data and Artificial Intelligence application for cultural heritage and museum research

Man Zhao¹, Xue Wu¹, Han-Teng Liao*, and Yu Liu¹
¹ New Media Research Center, Sun Yat-sen University Nanfang College, Guangzhou, Guangdong, 510970, China
*Corresponding author’s e-mail: h.liao@oxon.org

Abstract. Modern technologies have supported the preservation, curation and reuse of human cultural heritage assets and resources, the latest development in Artificial Intelligence (AI) and Big Data technologies are expected to have great potentials in the datafication, digitization, and reuse of such resources. With the aim to reveal the research fronts and topics of such potential development, the paper conducted a bibliometric analysis based on 206 related articles from the Web of Science (WoS) database. The paper presents the findings on publication trends, main publication outlets and disciplines, main keyword clusters and research gaps. The paper also finds that the current research scope of AI and Big Data in cultural heritage and museum applications is still in the early stage of storage, calculation and presentation of Big Data. The paper help researchers, funding agencies and practitioners to understand the current status of AI and Big Data applications in cultural heritage and museum.

1. Introduction
Modern technologies have supported the preservation, curation and reuse of human cultural heritage assets and resources. For instance, the use of Information Communication Technologies (ICT) has been incorporated into educational activities of museums, and recommendations have been made to creatively apply ICTs to engage museum visitors with digital educational resources, use various educational ICTs[1]. Recent research has also observed that the ICT methods and tools have helped museums to transform their framework of cultural communication, therefore important part of museum strategies to promote arts and cultural heritage activities[2]. The global dialogues also explore the ways in which ICT and education innovation can contribute to the sustainable development goal for education (SDG4), as confirmed in the 2017 Qingdao Statement [3,4], promoting the use of ICT to achieve education targets in sustainable development goals.

Artificial Intelligence (AI) and Big Data, as part of the latest development in the ICTs, are expected to have great potentials in the datafication, digitization, and reuse of such educational and cultural resources [5,6], including interconnecting the knowledge structures of digital heritage with the social web using Big Data[7]. Using AI and Big Data may enhance museum collections management, visitor engagement, etc., thereby curatorial decisions (which collections to display, how to display them, and where to place them) can be made based on evidence[8]. Specific literature reviews have been conducted for cultural heritage and museum (e.g. recommender systems [9]), but no systematic review has been conducted to assess the current status and research frontiers of AI and Big Data in cultural heritage and museum applications.
2. Data and methods
To fill the gap, the paper conducted a bibliometric analysis based on articles collected from the Web of Science (WoS) database. The paper used the advanced query of the WoS as below:

\[ \text{TS} = \left( \text{AI OR Artificial Intelligence OR machine learning OR Big Data} \right) \text{ AND TS} = \left( \text{museum OR heritage} \right) \]

In November 2019, the query above gave us 206 articles. To overcome the analytical challenges of different terms in author keywords, we also built a thesaurus that treat terms such as “cultural heritage (CH)”, “cultural-heritage”, etc. as the same concept of “cultural heritage”. Then the research conducted various bibliometric analysis using VOSviewer.

3. Research findings

Annual trends
Annual trends of publications and citations reveal that the related research has experienced growth, with less than 8 articles per year before 2015 and about 34 articles on average per year after 2015.

3.1. Top publication and bibliographic coupling network and clusters
Since both co-citation and bibliographic coupling networks can reveal research fronts and intellectual bases [10], we visualized such networks at the level of sources (journals and conference proceedings). Figure 1 shows the relationship and distance among different sources, and their respective disciplines. The top journal, *Remote Sensing*, mostly related to cultural heritage via the disciplines of physical geography, geology, etc., has several research articles applying machine learning and feature extraction to understand cultural heritage. The second journal, *ACM Journal on Computing and Cultural Heritage*, largely an interdisciplinary publication outlet among humanities and computer science, has articles on the use of technologies for presenting and curating cultural heritage. There are two journals at the third ranking. While *In Situ-Revue De Patrimoines* is also interdisciplinary, *Journal of Cultural Heritage* is largely multi-disciplinary, mostly archeology and art, also chemistry geology, material science and spectroscopy. Categorized as “green & sustainable science technology”, *Sustainability* and *Journal of Cultural Heritage Management and Sustainable Development*, has articles utilizing Big Data for better environment understanding.

Revealing another aspect of citations (i.e. intellectual bases), several clusters have identified as the important cited outlets (see Figure 2): tourism studies (e.g. *Tourism Management*, in blue), computer science (e.g. *Lecture Notes in Computer Science*, in green), remote sensing (in yellow), and other science journals (e.g. *Nature and PLOS One*, in red).

These above findings reveal the research fronts and intellectual bases of Big Data and AI in the fields of heritage and museum research: while they are interdisciplinary, they nonetheless are related to geography, geology, archeology, environment, sustainability, etc.
3.2. Keyword map and Clusters
Keyword data is visualize based on the interconnection of the keywords resulting in a keyword co-occurrence network. Figure 3(a) show 4 clusters: (a) of: Big Data and cultural heritage (in blue color), AI (in red color), museum (in yellow color) and machine learning (in green color).

![Overall keyword map](image)

![Big Data as the node](image)

![AI as the node](image)

![Digital heritage as the node](image)

Figure 3. A keyword map based on co-occurrence data

As shown in Figure 3(b), Big Data is more prominent than AI. Well-connected to AI, cultural heritage and museums, Big Data is at the core in the discussions. As shown in Figure 3(c) and 3(d), although connected with Big Data, cultural heritage and machine learning, AI applications appear to be at the peripheral, suggesting more research needed in filling such gaps.

Indeed, more research appears to focus on the storage, calculation, and reproduction of data to enhance digital experiences and digital services (e.g. [8], [9],[11]) than the application of AI.

3.3. Importance of Internet platforms
One emerging research area we found relates to the potentials of AI and Big Data services provided by Internet platform companies. Empirical research includes the remote sensing Big Data using the Google Earth Engine Cloud Platform [12][13], the Google Arts & Culture projects, the Shrewsbury Museum & Art Gallery project based on Microsoft Azure Search Cognitive Search.

4. Conclusion
The purpose of this literature review is to provide a comprehensive, albeit preliminary, understanding of the current status of AI and Big Data applications in cultural heritage and museum applications.
Although the articles visualized and discussed here are limited to the 206 articles we have analyzed, they nonetheless reveal specific trends, main publication outlets and disciplines, main keyword clusters and research gaps. Future research must address such gaps.

The paper help researchers, funding agencies and practitioners to understand the current status of AI and Big Data applications in cultural heritage and museum. Given the scale and importance of the Internet platforms, future research on this topic can explore how AI and Big Data enhance the digital management, digital services and digitization of cultural heritage and museum through open collaboration and AI and Big Data services on the Internet platforms.

Acknowledgments
The research is funded by a project of HCI for Education Development (2018WQNCX283), under the 2018 Major Projects of Guangdong Province Department of Education “Young Innovative Talents” Grants, and partly by the Application-oriented Curriculum Development Project of “Information Visualization Design” (NFU 02-40250), under the Guangdong Province Department of Education 2018.

Man Zhao, Xue Wu, and Han-Teng Liao made equal contributions to the paper, including the data collection, analysis, and interpretation. Man Zhao, and Han-Teng Liao have been involved in drafting and revising the manuscript and have given the approval of the current version for publication.

References
[1] UNESCO IITE (2011) ICTs in Museum Education. (UNESCO Institute for Information Technologies in Education (IITE))
[2] Sylaiou, S., Papaioannou, G. (2019) ICT in the Promotion of Arts and Cultural Heritage Education in Museums. Strategic Innovative Marketing and Tourism Springer Proceedings in Business and Economics ed A Kavoura, E Kefallonitis and A Giovanis (Cham: Springer International Publishing) pp 363–70
[3] UNESCO IITE (2018) Global Dialogue on ICT and Education Innovation – Towards Sustainable Development Goal for Education (SDG 4). Ministerial Forum Proceedings (Moscow, Russian: UNESCO Institute for Information Technologies in Education (IITE))
[4] UNESCO (2017) Leveraging information and communication technology to achieve Education 2030: report of the UNESCO 2017 International Forum on ICT and Education 2030. (Qingdao, China: UNESCO)
[5] UNESCO (2019) Artificial intelligence for Sustainable Development: synthesis report, Mobile Learning Week 2019. (Paris, France: UNESCO)
[6] UNESCO (2019) Artificial intelligence in education, compendium of promising initiatives: Mobile Learning Week 2019. (Paris, France: UNESCO)
[7] Bonacchi, C., Krzyzanska, M. (2019) Digital heritage research re-theorised: ontologies and epistemologies in a world of big data. Int. J. Herit. Stud. 25: 393–47.
[8] Tavčar, A., Csaba, A., Butila, E.V. (2016) Recommender System for Virtual Assistant Supported Museum Tours. Informatica 40:
[9] Pavlidis, G. (2019) Recommender systems, cultural heritage applications, and the way forward. J. Cult. Herit. 35: 183–96.
[10] Persson, O. (1994) The intellectual base and research fronts of JASIS 1986–1990. J. Am. Soc. Inf. Sci. 45: 31–8.
[11] Bogdanovych, A., Rodríguez-Aguilar, J.A., Simoff, S.J., Cohen, A.S. (2010) Authentic interactive reenactment of cultural heritage with 3D virtual worlds and artificial intelligence. Appl. Artif. Intell. 24: 617–47.
[12] Agapiou, A. (2017) Remote sensing heritage in a petabyte-scale: satellite data and heritage Earth Engine© applications. Int. J. Digit. Earth 10: 85–102.
[13] Luo, L., Wang, X., Guo, Q., Lasaponara, R., Shi, P., Bachagha, N., li, li, Yao, Y., Masini, N., Chen, F., Ji, W., Cao, H., Li, C., Hu, N. (2018) Google Earth as a Powerful Tool for Archaeological and Cultural Heritage Applications: A Review. Remote Sens. 10: 1558