

## Wiki Loves Monuments: Crowdsourcing the Collective Image of the Worldwide Built Heritage

Azizifard, Narges; Gelauff, Lodewijk; Gransard-Desmond, Jean-Olivier; Redi, Miriam; Schifanella, Rossano

Veröffentlichungsversion / Published Version  
Zeitschriftenartikel / journal article

### Empfohlene Zitierung / Suggested Citation:

Azizifard, N., Gelauff, L., Gransard-Desmond, J.-O., Redi, M., & Schifanella, R. (2022). Wiki Loves Monuments: Crowdsourcing the Collective Image of the Worldwide Built Heritage. *Journal on Computing and Cultural Heritage*, 16(1), 1-27. <https://doi.org/10.1145/3569092>

### Nutzungsbedingungen:

Dieser Text wird unter einer CC BY-NC-ND Lizenz (Namensnennung-Nicht-kommerziell-Keine Bearbeitung) zur Verfügung gestellt. Nähere Auskünfte zu den CC-Lizenzen finden Sie hier: <https://creativecommons.org/licenses/by-nc-nd/4.0/deed.de>

### Terms of use:

This document is made available under a CC BY-NC-ND Licence (Attribution-Non Commercial-NoDerivatives). For more information see: <https://creativecommons.org/licenses/by-nc-nd/4.0>



# Wiki Loves Monuments: Crowdsourcing the Collective Image of the Worldwide Built Heritage

NARGES AZIZIFARD, University of Turin, Italy

LODEWIJK GELAUFF, Stanford University, USA

JEAN-OLIVIER GRANSARD-DESMOND, ArkéoTopia, France

MIRIAM REDI, Wikimedia Foundation, USA

ROSSANO SCHIFANELLA, University of Turin, Italy, and ISI Foundation, Italy

The wide adoption of digital technologies in the cultural heritage sector has promoted the emergence of new, distributed ways of working, communicating, and investigating cultural products and services. In particular, collaborative online platforms and crowdsourcing mechanisms have been widely adopted in the effort to solicit input from the community and promote engagement. In this work, we provide an extensive analysis of the *Wiki Loves Monuments* initiative, an annual, international photography contest in which volunteers are invited to take pictures of the built cultural heritage and upload them to Wikimedia Commons. We explore the geographical, temporal, and topical dimensions across the 2010–2021 editions. We first adopt a set of CNN-based artificial systems that allow the learning of deep scene features for various scene recognition tasks, exploring cross-country (dis)similarities. To overcome the rigidity of the framework based on scene descriptors, we train a deep convolutional neural network model to label a photo with its country of origin. The resulting model captures the best representation of a heritage site uploaded in a country, and it allows the domain experts to explore the complexity of cross-national architectural styles. Finally, as a validation step, we explore the link between architectural heritage and intangible cultural values, operationalized using the framework developed within the World Value Survey research program. We observe that cross-country cultural similarities match to a fair extent the interrelations emerging in the architectural domain. We think this study contributes to highlighting the richness and the potential of the Wikimedia data and tools ecosystem to act as a scientific object for art historians, iconologists, and archaeologists.

CCS Concepts: • **Applied computing** → **Arts and humanities**; • **Computing methodologies** → **Computer vision**; • **Information systems** → **Wikis**;

Additional Key Words and Phrases: Cultural heritage, cross-cultural study, Wiki Loves Monuments

## ACM Reference format:

Narges Azizifard, Lodewijk Gelauff, Jean-Olivier Gransard-Desmond, Miriam Redi, and Rossano Schifanella. 2022. Wiki Loves Monuments: Crowdsourcing the Collective Image of the Worldwide Built Heritage. *ACM J. Comput. Cult. Herit.* 16, 1, Article 20 (December 2022), 27 pages.  
<https://doi.org/10.1145/3569092>

Lodewijk Gelauff also with Wiki Loves Monuments in a volunteer capacity as a co-founder and former member of the international coordinating team.

Authors' addresses: N. Azizifard, University of Turin, Corso Svizzera 185, Turin, Italy, 10149; email: [narges.azizifard@unito.it](mailto:narges.azizifard@unito.it); L. Gelauff, Stanford University, Stanford, CA; email: [lodewijk@stanford.edu](mailto:lodewijk@stanford.edu); J.-O. Gransard-Desmond, ArkéoTopia, Paris, France; email: [jogd@arkeotopia.org](mailto:jogd@arkeotopia.org); M. Redi, Wikimedia Foundation, San Francisco, CA; email: [mredi@wikimedia.org](mailto:mredi@wikimedia.org); R. Schifanella, University of Turin, Corso Svizzera 185, Turin, Italy, 10149, and ISI Foundation, Via Chisola 5, Turin, 10126, Italy; email: [rossano.schifanella@unito.it](mailto:rossano.schifanella@unito.it).



This work is licensed under a Creative Commons Attribution-NonCommercial-NoDeriv International 4.0 License.

© 2022 Copyright held by the owner/author(s).

1556-4673/2022/12-ART20

<https://doi.org/10.1145/3569092>

ACM Journal on Computing and Cultural Heritage, Vol. 16, No. 1, Article 20. Publication date: December 2022.

## 1 INTRODUCTION

Cultural heritage is a complex, multifaceted mosaic that holds up a mirror to who we were, who we are, and who we aspire to be [45]. Broadly defined, cultural heritage encompasses the extraordinarily rich set of *tangible* objects and materials in the collections of cultural institutions (movable); along with the heritage represented in the built environment (immovable) and in landscapes (natural). At the same time, cultural heritage includes an *intangible* dimension that captures elements such as folklore, traditions, customs, rituals, and, in general, the knowledge base of a society [7]. As pointed out in various contexts, the promotion and preservation of cultural heritage have an enormous potential to shape vibrant, innovative, and prosperous societies; improve quality of life; drive economic growth; and open up employment opportunities [7]. Recently, the wide adoption of digital technologies and applications, such as online social networks, multimedia sharing platforms, and digital libraries, is profoundly influencing and shaping how culture is experienced in contemporary society. The cultural heritage sector, as are other sectors, is following this transformation with the emergence of new, distributed ways of working, communicating, and investigating cultural products and services. In particular, we are interested in the role of collaborative online platforms, like Wikipedia, as a digital gateway to cultural heritage and a platform to promote engagement. Crowdsourcing mechanisms have been widely adopted to enrich the information about cultural heritage by soliciting input from the community. Even though some ethical and practical concerns have been raised [29], crowdsourcing has been proven effective in encouraging audiences to curate and modify the content of museums, libraries, and archives, which needs indexing or digitization [53]. This paradigm shift was able to place the individual at the center of content production and enrichment, pushing engagement with cultural heritage outside the boundaries of traditional institutions. In this work, we focus on the **Wiki Loves Monuments**<sup>1</sup> (WLM) initiative, an annual, international photography contest in which volunteers are invited to take pictures of built cultural heritage and upload them to Wikimedia Commons,<sup>2</sup> a sister project and the shared media repository of Wikipedia. WLM represents a way of capturing a snapshot of a nation's architectural heritage for future generations and documenting a country's most important historical sites. While Wikimedia Commons hosts visual representations of a wide variety of heritage items, e.g., digitized art objects and natural heritage, it is worth mentioning that the contribution of WLM focuses on one aspect of a broader picture, that is, the built heritage component. The collection of images gathered over more than 10 years through the competition is growing into an incredibly useful historical resource. Since the first edition in 2010, the project helped to collect information on 1.6M monuments from 93 national competitions, with more than 2.8M pictures submitted by over 86K participants. To characterize the impact of the competition, we provide a quantitative analysis of the WLM photographic collection along the dimensions of space, time, and topical coverage. Different from previous work that focused mostly on specific countries or a particular edition [49], we explore the entire corpus across the time frame 2010–2021. We think this effort could shed light on the potential of a crowdsourced initiative and a collaborative platform like Wikipedia in the promotion and conservation of the worldwide cultural heritage alongside standard cultural institutions.

As part of human activity, cultural heritage has long been linked to the notion of culture as the manifestation of a shared history and a common collective identity [24]. As a validation step, our work explores this connection with quantitative instruments, focusing on the relation between architectural heritage and cultural values. In other words, as observed in a variety of empirical studies [3, 42], we plan to test if the cross-country distances in the cultural sphere are correlated with the corresponding cross-country distances in the built heritage sphere. Several efforts have been devoted to the quantitative evaluation and characterization of culture [21, 22, 26, 37, 38, 56, 59], with researchers from different fields investigating a wide variety of different facets [42]. This heterogeneity leads often to the realization that cultural dimensions are hard to be interpreted as absolute scores, and that cross-country cultural models can only be meaningfully interpreted

<sup>1</sup><http://wikilovesmonuments.org>.

<sup>2</sup><https://commons.wikimedia.org/>.

in a comparative framework instead [23]. To operationalize culture, we adopt a subset of the data from the **World Value Survey** [27] (WVS), which is a large-scale, cross-national, and cross-sectional longitudinal survey research program devoted to the scientific and academic study of social, political, economic, religious, and cultural values of people in the world.<sup>3</sup> In particular, we refer to the **Inglehart-Welzel Index (IW Index)** [27] as a multi-dimensional construct to characterize the emerging cultural traits of a country.

To summarize, the main contributions of this work can be listed as:

- For the first time, we provide a comprehensive characterization of the WLM initiative across the geographical, temporal, and topical dimensions.
- Leveraging the crowdsourced WLM photographic collection, we train a per-country visual descriptor to capture a digital representation of the architectural heritage of a nation. This visual descriptive model enables the researcher to explore the distinctive traits of a country heritage and, mostly, it provides a comparative framework to highlight (dis)similarities across countries.
- As a validation step, we provide quantitative evidence of a link between the visual representation of a nation's architectural heritage and the set of social, political, religious, and cultural values operationalized by the Inglehart-Welzel Index. This is consistent with the hypothesis that cultural heritage and societal beliefs are intertwined.

## 2 RELATED WORK

**Digital Cultural Heritage.** Since the mid-2000s, the adoption of digital tools and **information and communication technologies (ICTs)** has radically changed the digital acquisition, storage, conservation, representation, and promotion of worldwide cultural heritage, enabling a variety of potential novel applications and services. In particular, online social networks, multimedia sharing platforms, open digital libraries, and collaborative knowledge bases have provided a heterogeneous ecosystem of data sources and tools that reshaped how we experience an exhibition, a museum, an archaeological site, or monuments in historical city centers. Semantic Web technologies have been extensively used to enrich cultural heritage data [11, 12, 25, 58], to provide multilingual access [10], to implement semiautomatic pipelines for content annotation, and to facilitate the curation, collaborative management, interoperability, and engagement with digital collections [13]. A wide body of literature in the Cultural Heritage domain explored the use of IoT systems [47] to bridge the digital and real dimensions into new smart environments and innovative forms of browsing of cultural spaces. For instance, [2] **Smart Context-aware Browsing assistant for cultural Environments (SCRABS)** proposes a scalable prototype for the management and context-driven browsing of cultural environments. The system is characterized by the integration of distributed and heterogeneous pervasive data sources, context awareness, and advanced smart services such as retrieval, analytics, and recommendations based on users' preferences and the surrounding environment. The application of mobile phones in the framework of museums and exhibition spaces is steadily increasing due to their improved computational capabilities, which provide embedding interactive applications and rich graphics; their portability and widespread adoption; or their ability to shape new museum narratives [54]. Recently, the adoption of **Augmented Reality (AR)** and **Virtual Reality (VR)** technologies has acknowledged a great boost in terms of adoption and research perspectives. One common application of VR/AR involves the relocation of museum objects and cultural properties in the archaeological sites they originate from. Re-collocating allows the users to materialize their original location, scale, and function, creating a connection with the cultural identities to which they refer [15, 39, 48].

**Cultural Heritage and Wikipedia.** Wikipedia has been widely adopted as a digital gateway to tangible and intangible cultural heritage items and a platform to promote engagement. Starting from the observation that often digitized CH items have short descriptions and lack rich contextual information, Agirre et al. [1] explored

<sup>3</sup><https://www.worldvaluessurvey.org/>.



the feasibility of matching metadata annotations and high-quality descriptors from Wikipedia articles to objects in the Europeana<sup>4</sup> initiative. Hall et al. [19] implemented a methodology to promote and make visible the unpopular resources of a CH collection by augmenting the navigational interface of matching Wikipedia articles with thumbnail images that are linked to external collections. The open source analysis architecture Contropedia [46], focusing on the notion of memory work, provided a platform to study the discourse around cultural heritage focusing on the identification and visualization of heritage-related disputes within an article and their comparison across language versions. The patterns of production, consumption, and geographical distribution of Wikipedia pages describing UNESCO World Heritage sites [40] have been recently investigated underlying the distinctive role of Wikipedia in characterizing how people experience and relate to the past. Controversies in articles were analyzed showing the prevalence of hyper-local and process issues over broader concerns, e.g., violence, destruction, and preservation. Even though the authors stressed how Wikipedia is a valuable and important source of social information on heritage, they underlined some concerns about spatial coverage, the skewness of editors' cultural and sociodemographic provenance, and the prevalence of edits from the Anglosphere. In [50], the authors investigated the gaps in Wikipedia's coverage of the visual arts by comparing the extent and quality of the coverage of 100 artists and 100 artworks from Western and non-Western canon. They observed how the platform strongly favors the Western canon, with European language versions leaning toward Western artists and Asian being more global, with notable exceptions. Callahan and Herring [8] described a similar work comparing Polish and English Wikipedia articles about celebrities to check the differences in content and viewpoints across cultures. Laufer et al. [31] studied the differences in cultural explanations of culinary operations in several European language editions of Wikipedia pages and suggested a way to explore cultural relevance among food descriptions and concerns in different languages. This work explores a complementary angle, focusing on the role of the photographic contest *Wiki Loves Monuments* in characterizing the tangible cultural heritage of a country and its connections with the backbone of its social and cultural values.

**Image Analysis in the Cultural Heritage Domain.** Recently, image analysis has been successfully applied to a wide variety of tasks and application domains, e.g., object detection [51, 52, 63] or image captioning [20], with a surge in contributions in the context of cultural heritage projects [6, 34, 41]. Examples of relevant applications cover the automatic enrichment of digital visual collections, improved search capabilities [62], automated dating [43], photogrammetric analysis of historical images for virtual reconstruction [35], and historical buildings classification [32]. To avoid the cost of training from scratch complex deep learning architectures, a popular method, adopted also in our work, is transfer learning, which refers to utilizing “what has been learned in one setting [. . .] to improve generalization in another setting” [17]. Wevers [62] showed the role of transfer learning to improve scene detection applied to a historical press photo collection. Kulkarni et al. [30] applied transfer learning to predict which cultural heritage site an image refers to, allowing improved searches through specific terms, and helping in the studying and understanding of heritage assets. Within the context of the INCEPTION European project, authors in [33] applied transfer learning to the classification of cultural heritage images, showing a substantial improvement in accuracy over SVM and random forest baselines. A generally accepted concept in photography underlines how a shot reverberates the photographer's inner sense, which translates to the observation that often the content of a photo can contribute to characterizing people's attitudes and values on a subject [28]. In this direction, exploiting a continuously re-training deep learning model, authors in [28] analyzed the visual content of a corpus of Flickr photos made by tourists to capture the tourists' urban image of a place as a reflection of the unique landscapes, cultural characteristics, and traditional elements of the visited region. Similarly, Pan et al. [44] analyzed 145 travel photos submitted to the *The New York Times* to explore the interplay between motivations, image dimensions, and affective qualities of places as a reflection of the inner feelings of the photographers. In this work, we embrace this vision within the cultural heritage domain; we aim

---

<sup>4</sup><https://www.europeana.eu>.

at exploring the interplay between the choice of subjects of the WLM submissions and the inner representation of the importance of the cultural heritage assets of the photographers.

### 3 WIKI LOVES MONUMENTS PHOTOGRAPHY CONTEST

WLM is an international photography contest organized by volunteers around the world in which participants are invited to take pictures of local built cultural heritage sites, e.g., a building, ruin, or complex of historic significance, and upload them to the Wikimedia Commons platform. The competition is organized under a federated model, where national teams organize the competition in their specific country over 1 month (traditionally September), with its own monuments, organization, partners, jury process, and awards. After the national juries have concluded their work, each national jury can submit up to 10 photographs to an international finale.

The national competitions typically establish some kind of list or definition of what buildings qualify as a “monument,” usually following official lists and definitions (e.g., “listed building” or “scheduled monument” in the United Kingdom, “*monument historique*” in France, and “*tombamento*” in Brazil) and may apply additional rules to the competition based on the national situation, such as copyright restrictions and antiquity laws. The photos are then uploaded to Wikimedia Commons under a free copyright license, so that the images can be used to illustrate relevant articles, and the aforementioned lists are generally accessible and reusable, contributing to expanding the quality of the knowledge base and helping in promoting the history and national heritage of all participating countries. WLM started as a pilot project in the Netherlands in 2010 and, due to its increasing success, organizers in many countries joined the competition; the competition has been organized in over 90 countries and was in 2021 running in more than 35 countries around the world.<sup>5</sup> It has not only managed to collect more than 2.8 million images of built cultural heritage worldwide but also inspired the creation of a database<sup>6</sup> of heritage sites so that photographers would know what to photograph and where. This database currently holds more than 1.6 million data points collected from various heritage organizations, likely the largest in existence. Arguably, WLM is a notable example of a successful crowdsourcing campaign and it is acknowledged by the Guinness Book of World Records as the largest photography competition in the world.<sup>7</sup>

**Temporal and Geographical Characterization.** In this work, we focus on the WLM editions in the period 2010–2021 for which we collect all the submitted photos with the corresponding metadata. Photos are characterized by an identifier representing the depicted site, a reference to the WLM national organizer, and the filename that might contain a textual reference to the monument photographed. Figure 1 shows the spatial coverage of the resulting dataset; each photo has been assigned to a country according to the WLM contest of submission. We observe a highly heterogeneous spatial coverage with a strong presence of European countries; in fact, the top 6 countries for volume of submissions are in Europe, as shown in Figure 2. Within the top 30 most represented countries in this dataset, we find representatives of Asia and North/South America, while Africa is present only with Egypt. To characterize the temporal variability of contributors’ participation, Figure 2 shows the volume of submissions (dashed bars) and the number of countries (solid bars) that are involved in each edition. After the first 2 years of growth, WLM reached an average engagement in the range of 200K to 250K photographs per year. The color-coded split by continent in Figure 2 confirms the predominance of Europe followed by Asia as the most prolific sources of information. Despite the European focus and the difficulty of reaching and engaging with a wider community, WLM has reached more than 90 different countries in 7 continents, and it arguably provides an extensive and diverse dataset to characterize the built cultural heritage across national boundaries.

**Characterization of the Contributor Base.** To shed light on the characteristics of the typical WLM contestant and contextualize the nature of the dataset, we explore several activity metrics at the contributor level. Even though in the rest of the manuscript the unit of interest will be a country and the data will be analyzed only

<sup>5</sup><https://wikiloves.toolforge.org/monuments>.

<sup>6</sup>[https://commons.wikimedia.org/wiki/Commons:Monuments\\_database](https://commons.wikimedia.org/wiki/Commons:Monuments_database).

<sup>7</sup><https://www.guinnessworldrecords.com/world-records/largest-photography-competition>.

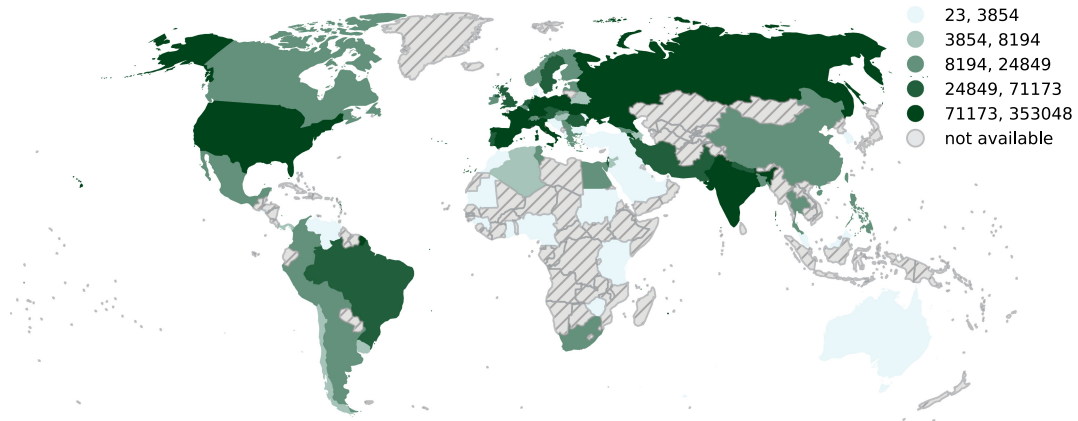


Fig. 1. Geographical coverage of the WLM dataset (Antarctica not displayed).

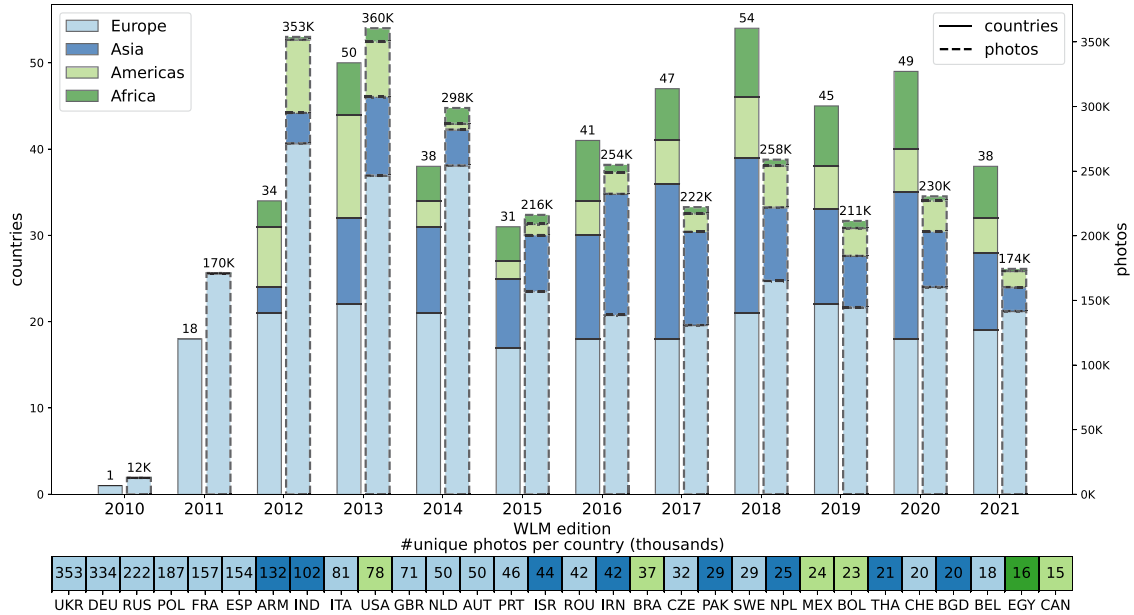


Fig. 2. Temporal evolution of the volume of uploaded photos in different WLM editions (top). Top 30 countries by volume of uploaded photos grouped by country and continent (bottom). Country names are abbreviated with the ISO 3166-1 standard codes.

in an aggregated form, we aim at characterizing to what extent the profile of a geographical region depends on the activity of a limited group of contributors. In this direction, Figure 3 shows the empirical cumulative distribution function of (a) the number of photos uploaded per contributor, (b) the distribution of the number of countries a contributor uploaded pictures for, and (c) the number of years of participation in the contest. We observe that 95% of the participants uploaded during the period of study up to 66 photos in a single country, entering a maximum of two editions of the contest. However, even though the vast majority of the contributors do not exhibit extreme behaviors, we observe the presence of a limited group of very active contributors. The top

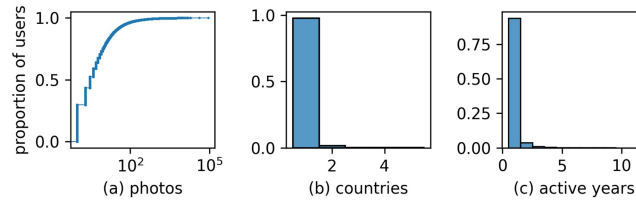


Fig. 3. Per-contributor statistics in terms of volume of photos uploaded, number of unique countries photographed, and years of activity. The 95th percentile of the participants shot 66 or less photos, participated in a single country, and was active for up to two editions.

10 most prolific contestants are responsible for about 265K submissions (around 9% of the collection). A handful of contributors (25) took part in all the editions, and even though usually participants upload photos within their country of origin, 25 contributors submitted photos across 10 or more countries.

To limit the impact of over-representation of a small set of active contributors, we apply the following filters to the original data:

- (1) **Pre-processing:** Images were downloaded at 300px format on the longest side and resized to a 224px or 150px square format, according to the input requirements of the vision pipelines in Sections 3.1 and 4.
- (2) **Near-duplicates removal:** To reduce the effect of contributors uploading nearly identical photos, we apply a similarity search routine at the country level. Each photo in the corpus is represented by a vector of visual features extracted from VGG16 [57], a state-of-the-art deep learning model for object detection initialized with pre-trained weights on ImageNet [55]. For each photo within a country, we compute the set of  $k$ -nearest neighbors and their similarity score, with  $k = 100$ . We define as duplicates the neighbors that show a similarity score greater than a threshold  $\delta$ , computed as the inverse of the cosine distance of the visual vectors extracted in the previous step normalized in the range  $[0,1]$ . After exploring different configurations and evaluating them through visual inspection, we set the threshold  $\delta = 0.98$ . We finally keep a randomly selected representative for each duplicate set, discarding the remaining photos. To enable a fast retrieval of the neighbors set, we use the Python library Faiss<sup>8</sup> that provides efficient similarity search and clustering of dense vectors. This filter removes 2% of the initial dataset.
- (3) **Submissions per contributor:** To limit the contribution of very active contributors, we randomly sample up to 1,000 photos as the representatives of the activity of a participant. This reduces the impact of individual preferences and styles across the geographical and temporal dimensions.
- (4) **Support:** To ensure sufficient support, only countries with at least 5K images remaining are considered.

This process results in a dataset of around 2.66M images by 82,716 contributors from 56 countries.

### 3.1 Toward a Topical Characterization of the Wiki Loves Monuments Corpus

After the geographical and temporal dimensions, in this section, we shift the focus to the content of a photo to capture which are the visual features that characterize the submissions in each national competition. Since the metadata do not always provide topical annotations nor an easy way to connect a submission to a monument in the corresponding national monuments registry, we refer to computer vision techniques to capture this aspect. In particular, we adopt the Places365 [65] scene recognition framework, a set of state-of-the-art **convolutional neural network (CNN)**-based artificial systems that allow the learning of deep scene features for various scene recognition tasks. Places are categorized by function making use of the spatial associations among the objects in an image. The framework is built upon a large-scale database<sup>9</sup> containing 10M images, which are labeled with

<sup>8</sup><https://github.com/facebookresearch/faiss>.

<sup>9</sup><http://places2.csail.mit.edu/>.

Table 1. Set of Scenes within the Places365 Taxonomy (Indoor/Outdoor Cultural Places) We Consider in Our Analysis

Scene Type	Categories
<b>outdoor</b>	amphitheater, aqueduct, arch, archeological excavation, army base, bazaar, campus, castle, cemetery, church, courthouse, courtyard, fountain, kasbah, library, lighthouse, mausoleum, medina, moat water, mosque, museum, oast house, pagoda, palace, plaza, ruin, synagogue, temple asia, tower, viaduct, windmill
<b>indoor</b>	aquarium, arcade, archive, art gallery, art school, art studio, artists loft, auditorium, beauty salon, booth, burial chamber, catacomb, church, classroom, conference center, jail cell, kindergarten classroom, lecture room, legislative chamber, library, movie theater indoor, museum, music studio, natural history museum, orchestra pit, science museum, stage indoor, television studio, throne room, ticket booth

365 scene semantic categories organized in a three-level taxonomy [64]. At the root of the hierarchy, there is the dichotomy indoor/outdoor places; the latter are grouped into natural and man-made outdoor scenes. Since we are interested in capturing the presence of elements related to the cultural heritage sphere, in this work, we focus on the *indoor/cultural* and *outdoor/man-made/cultural* scenes that count for 61 categories such as amphitheater, church, catacomb, museum, and others. The complete set of scene descriptors is listed in Table 1.

We adopt the pre-trained *VGG16-places365* model provided in [64] to extract for each photo the scene categories with a prediction confidence score  $\rho > 0.15$  that measures the reliability of class assignments. The threshold has been selected to filter out the annotations that have a confidence score less than the 30th percentile of the overall distribution. We explored different quality thresholds obtaining comparable results.

Figure 4 (top-left) shows the top 5 indoor and outdoor scenes in our aggregated dataset, being, respectively, church, burial chamber, museum, catacomb, throne room, and church/outdoor, synagogue/outdoor, cemetery, tower and ruin. These associations are not surprising due to the geographical coverage dominated by European countries. However, shifting to a country-specific analysis, differences between cultural areas emerge: for instance, the most frequent outdoor cultural scenes in Thailand (Figure 4, top-center) are temple, pagoda, and palace, while in the Netherlands (Figure 4, top-right) oast house, synagogue, and church are predominant. To give a hint to the cross-country heterogeneity, we provide a representative for each geographical region. We list in the Additional Material the most frequent categories for the remaining national competitions (Figures A1–A4).

To explore in more detail the complexity of the category assignments, we refer to a comparative framework in which each country is represented by a vector  $v_c^{places365} = \langle v_1, \dots, v_n \rangle$ , where  $v_i$  models the frequency of the category  $i$  within the corpus of the photos submitted in that country and  $n$  is the number of categories with a not null frequency ( $n = 60$ ). This representation enables to explore the similarity among countries according to the most frequently detected scenes. To have a first descriptive image of cross-national similarities, we refer to the **T-distributed Stochastic Neighbor Embedding (t-SNE)** [61] approach to embed the high-dimensional  $v_c^{places365}$  vectors in a two-dimensional space. Figure 5(a) shows the map limited to the countries with at least 5K photos tagged with at least one cultural category (56 countries, 1.5M photos in total). To ease the navigation of the map, we color the points according to the corresponding country’s geographical region. Interpreting closeness as a similarity measure, we observe the emergence of interesting patterns: in several cases, geographical proximity matches the proximity in the embedding space, e.g., the case of Bangladesh, India, and Pakistan, or the South American cluster of Colombia, Argentina, Chile, Uruguay, and Bolivia, or in Southeast Asia the cases of Thailand, China, and Nepal, suggesting the presence of homogeneous cultural areas with a similar representation of categories among the photos. However, there are some notable examples when the weak geographic association fails: e.g., Ireland exhibits close proximity to Georgia and Serbia in the embedding space while being geographically distant. A similar observation holds for the Philippines and Brazil, or South Africa and Ukraine. These patterns might be linked to limitations of the adopted scenes’ detection framework or to specific characteristics of the



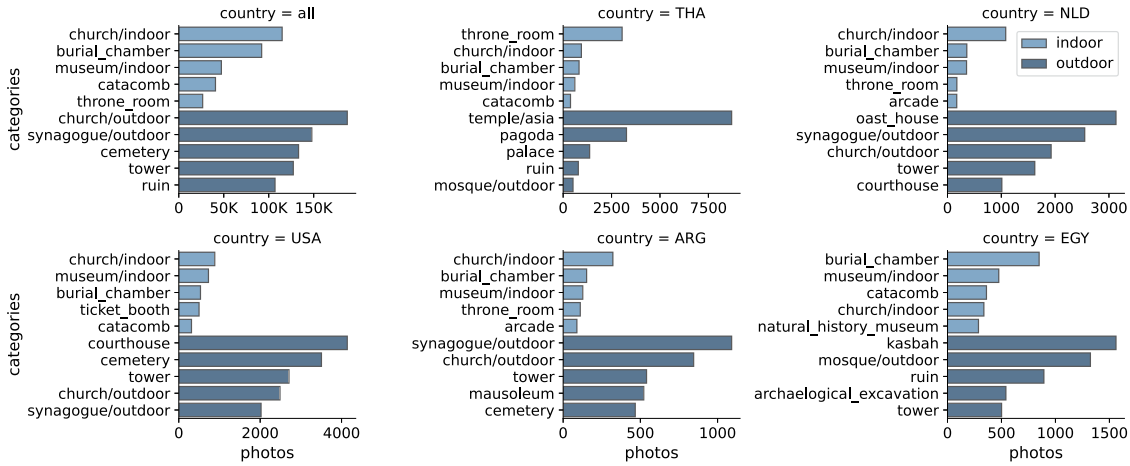


Fig. 4. Top 5 indoor and outdoor Places365 scenes in the entire corpus (upper-left) and in Thailand (Asia), the Netherlands (Europe), the United States (North America), Argentina (South America), and Egypt (Africa) as representatives of the main geographical regions.

national competition, e.g., local suggestions on what to photograph. Figure 5(b) provides an alternative view to explore these geographical relations: the dendrogram resulting from an agglomerative hierarchical clustering of the  $v_c^{places365}$  vectors (distance function = cosine, linking criterion = average) is shown in a circular layout. Countries are colored with the corresponding geographical area and they are placed close by according to the similarity in their scenes' frequency profiles. The hierarchy shows the existence of semantically and geographically related groups as well as counterexamples. As previously pointed out, despite the ability to detail high-level similarities, this approach presents a few drawbacks: First, it is based on a limited set of visual descriptors that are not able to capture the complexity and diversity of the worldwide cultural heritage. Second, certain categories are prone to misclassifications: e.g., jail cell is likely to capture anything that has a metallic grid pattern (Figure 6(a)), or throne room often refers to indoor rooms in historical palaces with a chair or even an altar in a church (Figure 6(b)). Similarly, the natural history museum category is often triggered by the presence of statues of animals in facades (Figure 6(c)). Third, the nature and the guidelines of the photographic contest tend to bias the choice of the subjects of a photo since most competitions limit themselves to officially recognized heritage sites. This means that the government policy on what to acknowledge as heritage, but also which of those sites are easily accessible to photographers, will under perfect labeling determine the country's vector. For example, the fourth most frequently detected scene globally is cemetery, which often depicts headstones of famous individuals or commemorative plaques of historical events. Similarly, two of the three most frequent scenes in Italy appear to be cemetery and fountain, highlighting the choice of Italian contestants to focus on those architectural elements independently of their relevance to the national heritage. Finally, artistic styles are hardly captured by generic categories like palace (Figures 6(d) and 6(e)) or arch (Figures 6(f) and 6(g)), which hide the diversity of cultural movements and historical epochs; i.e., a palace in different geographical regions might show significant visual and artistic differences.

#### 4 MODELING THE COLLECTIVE IMAGE OF A COUNTRY CULTURAL HERITAGE

To overcome the drawbacks and limitations of the previous approach based on scene detectors, we aim at training a model able to capture the best visual representation of a heritage site uploaded in a country. We frame the problem as a multi-class classification task, in which the model learns to label a photo with its country of origin.



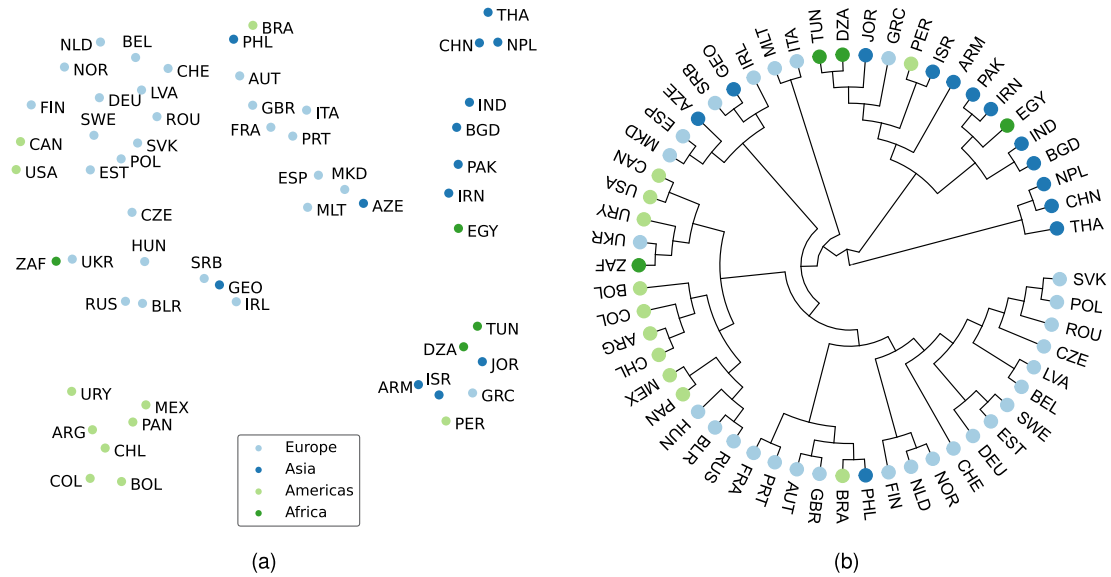


Fig. 5. Panel (a) shows a two-dimensional map of the Places-365 category vectors  $v_c^{places365}$  using the t-SNE method (metric = cosine, learning rate = 0.2, early exaggeration = 10, perplexity = 5). Colors are assigned to geographical regions. Panel (b) shows the dendrogram resulting from an agglomerative hierarchical clustering of the  $v_c^{places365}$  vectors (distance function = cosine, linking criterion = average). These figures provide an overview of the structure of the cross-country similarities induced by the Places365 analysis.



Fig. 6. Examples of possible classification issues emerging from the use of the Places365 scene detectors.

To this extent, we adopt a deep CNN architecture, called Xception [9], in its implementation available in Keras.<sup>10</sup> Instead of training the model from scratch, which might take a conspicuous amount of time on large datasets, we apply the idea of transfer learning; starting from an Xception model pre-trained on ImageNet [55] for object detection, we freeze the model weights except for the top layer, which is customized to solve our multi-class problem. At last, we run a fine-tuning routine, which consists of unfreezing the entire model from the previous step and re-training it on the WLM data with a low learning rate ( $1e^{-4}$ ). In this framework, we limit our analysis to the countries with at least 5K photos to ensure a smooth learning phase, resulting in 56 countries. To keep a relative balance between classes, we consider a maximum number of 30K photos per country sampled randomly. We adopt an 80–20 split for training and test sets; 20% of the training data is held as a validation set for hyperparameter tuning.

<sup>10</sup><https://keras.io/api/applications/xception/>.

Table 2. Classification Summary

Precision	Recall	F1-score	MCC	Top 1 Acc.	Top 3 Acc.	Top 5 Acc.
0.47	0.45	0.45	0.44	0.45	0.63	0.71

Precision, recall, and F1-score averages are weighted by the per-class support to account for label imbalance. The Matthews Correlation Coefficient (MCC) is a performance measure that combines the contribution of the true and false positives and negatives. A coefficient of +1 represents a perfect prediction, 0 no better than random prediction, and -1 total disagreement between prediction and observation. Top k accuracy computes the number of times the correct country is among the top k predicted countries.

Table 3. Per-class Classification Report

	NPL	THA	ARM	NLD	BOL	IND	BRA	PRT	BGD	IRN	ISR	CZE	GBR	MEX
f1-score	0.76	0.7	0.69	0.69	0.67	0.6	0.57	0.57	0.56	0.56	0.56	0.55	0.52	0.52
support	5,191	4,382	5,864	5,587	4,644	5,826	5,503	5,638	4,017	5,600	5,650	5,471	5,644	4,967
	ROU	AUT	PAK	EGY	PAN	DEU	ESP	MLT	CHN	FRA	ITA	USA	PHL	RUS
f1-score	0.5	0.49	0.49	0.47	0.46	0.44	0.41	0.41	0.4	0.4	0.4	0.4	0.39	0.39
support	5,598	5,673	5,941	3,280	1,090	5,952	5,879	1,821	2,929	5,906	5,765	5,778	1,954	5,924
	SWE	MKD	NOR	BEL	IRL	UKR	POL	TUN	CHE	GEO	JOR	GRC	PER	LVA
f1-score	0.38	0.37	0.37	0.36	0.36	0.36	0.33	0.32	0.31	0.31	0.29	0.28	0.28	0.26
support	5,821	2,654	2,396	3,717	1,859	5,950	5,902	1,955	4,102	2,404	1,274	2,476	1,684	1,732
	FIN	CAN	COL	SVK	ARG	AZE	BLR	ZAF	SRB	EST	URY	DZA	CHL	HUN
f1-score	0.25	0.24	0.24	0.22	0.21	0.21	0.21	0.21	0.2	0.19	0.18	0.17	0.16	0.12
support	1,655	3,076	2,076	2,241	2,014	1,151	1,490	2,536	2,260	1,943	1,517	1,056	1,622	1,639

Countries are sorted in decreasing order per F1-score.

In Table 2 we present the performance of the model at an aggregate level. We observe an F1-score of 0.45 computed as a weighted average of the per-class F1-score to account for label imbalance. Additionally, to address the case of a multi-class imbalanced scenario, we compute the **Matthews Correlation Coefficient (MCC)**, which is a performance measure that takes into account all the dimensions of a confusion matrix combining true/false positives and negatives. In essence, *MCC* is a correlation coefficient value between  $-1$  and  $+1$ ; a coefficient of  $+1$  represents a perfect prediction,  $0$  no better than random prediction, and  $-1$  total disagreement between prediction and observation. We obtain an  $MCC = 0.44$ , which indicates a solid gain with respect to a random classifier. Using the class probability vectors, we compute the top  $k$  accuracy with  $k \in \{1, 3, 5\}$ , which indicates the number of times the correct country is among the top  $k$  predicted countries. We observe that a photo's true country is 63% in the top 3 and 71% in the top 5 predicted countries. Note that in a scenario with high visual heterogeneity and cross-country spillovers due to geographical or cultural proximity, we believe the model shows valid performance, especially in the information retrieval setting. To explore the diversity of performance across countries, in Table 3 we provide the per-class F1-score sorted in decreasing order. We observe a high variability across countries, with Nepal, Thailand, Armenia, the Netherlands, and Bolivia at the top of the ranking and Estonia, Uruguay, Algeria, Chile, and Hungary at the bottom. Qualitatively, countries with a low support tend to perform worse due to the lack of enough data to train a high-quality discriminator. However, this effect is not due only to underfitting in the case of a limited training set, but it is important to note how difficult the task might be to discriminate between countries with close artistic styles and monument types. Classification errors are not random: the model tends to misclassify photos between countries that have similar visual profiles. To quantify this

Table 4. Top 3 Countries for Number of Misclassifications Weighted by Per-class Support

UKR	SWE	PAK	RUS	GBR	CAN	CZE
RUS-BLR-ROU	FIN-NOR-RUS	IRN-IND-PER	UKR-FIN-BLR	IRL-NLD-BEL	USA-FIN-NLD	SVK-AUT-POL

We interpret the likelihood of misclassifications as a form of similarity measure, indicating the inability to easily discriminate between countries. For example, Russia is misclassified with high probability with Ukraine, Romania, or Finland, and Spain with Portugal, Italy, or France. For simplicity, in this table we present the seven cases where the concentrations of the errors are highest.



Fig. 7. For three pairs of countries (Czechia/Austria, Belgium/the Netherlands, and Ukraine/Russia) we show three pairs of photos with a high visual similarity score. We compute the similarity between two photos as the cosine similarity between their feature vectors extracted by our model. A high similarity is a proxy of the difficulty to discriminate between classes: in fact, some of the examples are hardly distinguishable to the human eye, which underlines how class boundaries are sometimes fuzzy and, consequently, how difficult is the multi-class task we are trying to solve.

observation, for each label we compute the top 3 countries for number of misclassifications weighted by per-class support (Table 4). We interpret the likelihood of misclassifications as a form of similarity measure, indicating the inability to easily discriminate between labels. For example, Russia is misclassified with high probability with Ukraine, Finland, or Belarus, and Czechia with Slovakia, Austria, or Poland. For simplicity, in Table 4 we present the seven countries where the errors are more concentrated in a few classes. To explore this concept visually, in Figure 7 we show for a pair of countries and three sets of photos with the highest visual similarity score. We compute the similarity between two pictures as the cosine similarity between their feature vectors extracted with our model. It is immediate to note how some of these pictures are strikingly similar even if they belong to two different classes, indicating once again the inherent difficulty of the task and the fuzziness of the definition of the class boundaries.

In Figure 8 we visually explore the top 10 countries by F1-score showing a sample of 10 photos with the highest class assignment probability as a measure of confidence in the prediction. We refer to them as visual representatives for their high discriminative power. Not surprisingly, the diversity of subjects and artistic styles across countries emerges especially in the areas with different cultural backgrounds.

Finally, we refer to the same comparative framework in Section 3 where each country is represented by a vector  $v_c^{WLM} = \langle v_1, \dots, v_m \rangle$ , which is computed as the average of all the feature vectors extracted from each photo belonging to a specific country. In the case of our fine-tuned model, a photo is represented by an  $m$ -dimensional vector, where  $m = 1,000$  is the number of object classes the Xception model was originally trained for. Figure 9(a) shows the map using t-SNE for the 56 countries covered by our analysis. Consistently with the case of the Places365 characterization, the geographical proximity and the cultural similarity among countries positively correlate with the distance in the embedding space. However, differently than the previous case, we observe (1) a higher degree of separability in the embedding space, with denser and more homogeneous clusters, and (2) a more



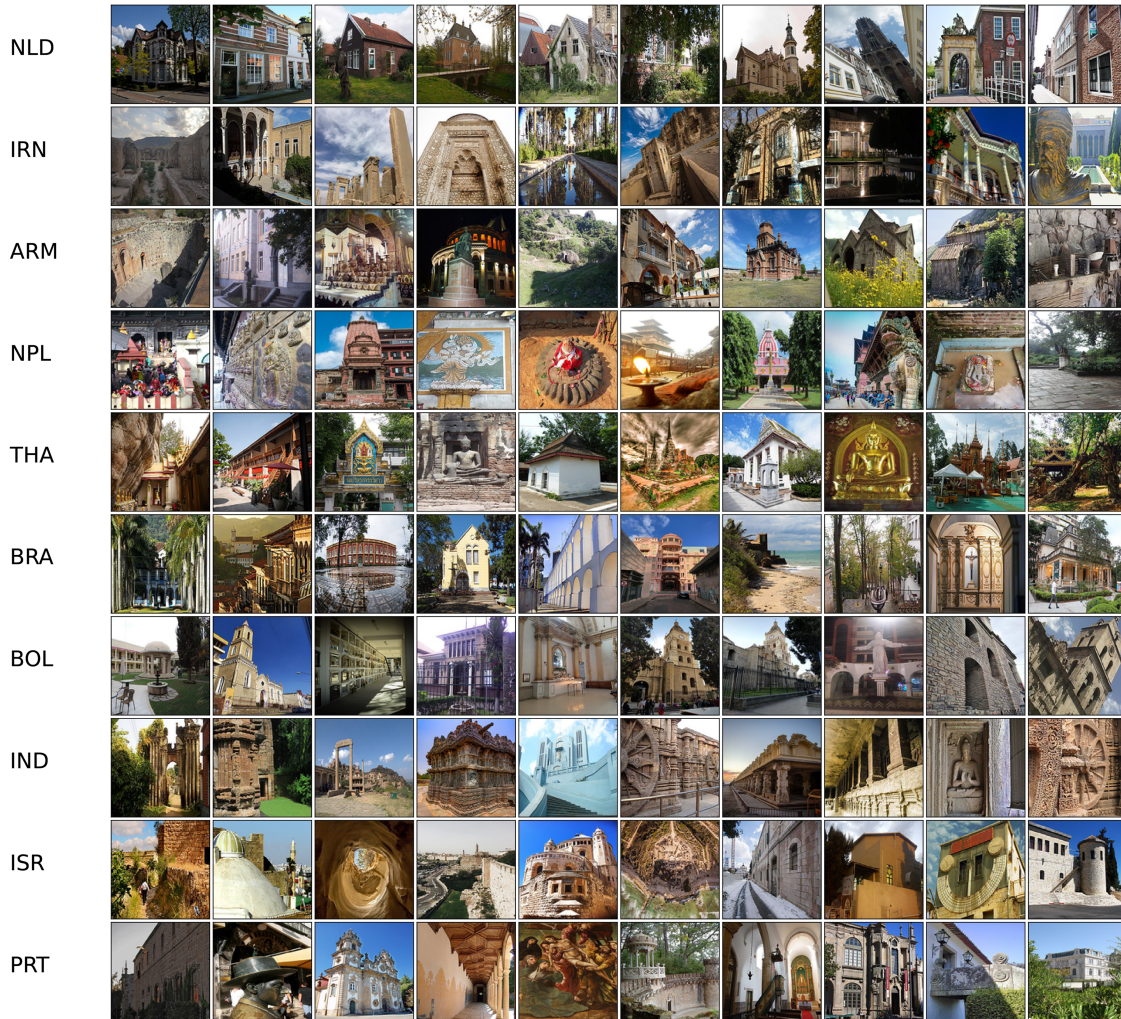


Fig. 8. A sample of photos with the highest model classification probability for the top 10 countries according to per-class F1-score. Photos have been resized to fit in the grid.

coherent positioning in terms of geographical regions that emerges from the consistent coloring within groups and the absence of some idiosyncrasies present in the scenes framework—e.g., Ireland is currently close to Great Britain instead of Georgia and Peru is no longer part of the Mediterranean cluster. We think the improved cross-country placement highlights the advantages of the framework based on the multi-class classification pipeline. However, it is interesting to underline how geography does not explain entirely the observed relations: in fact, our approach captures features that go beyond the obvious geographic clustering; e.g., countries like Brazil and Portugal that are culturally strongly linked appear close even if geographically very distant.

## 5 ON THE INTERPLAY BETWEEN HERITAGE AND CULTURAL VALUES

Cultural heritage has long been linked to a notion of culture, which bundles together ethnicity, collective identity, territory, and the idea of a common origin and shared past [24]. In this framework, which maintains a dominant

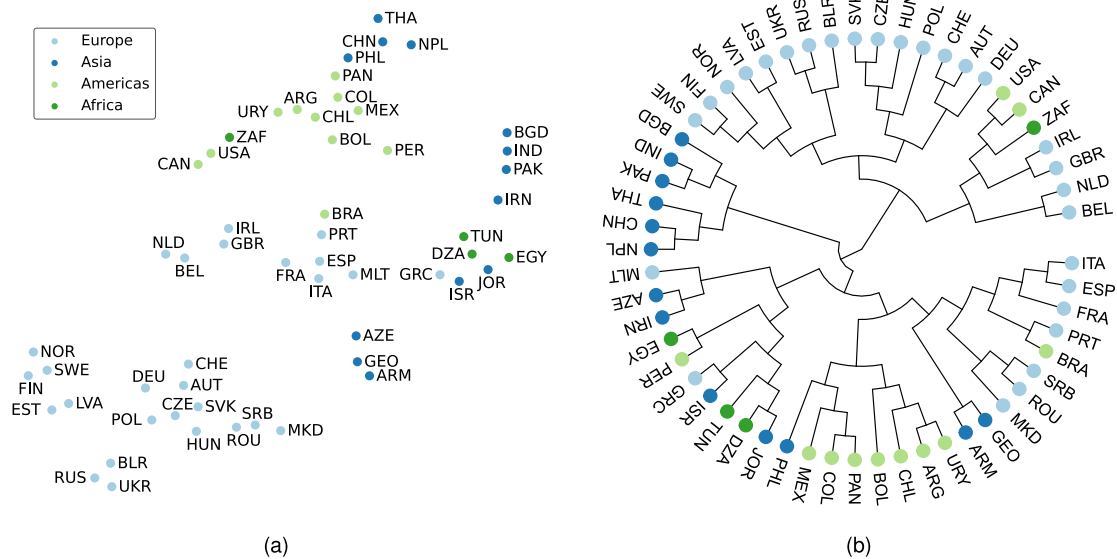


Fig. 9. Panel (a) shows a two-dimensional map of the  $v_c^{WLM}$  using the t-SNE method (metric = cosine, learning rate = 0.2, early exaggeration = 10, perplexity = 5). Colors are assigned to geographical regions. Panel (b) shows the dendrogram resulting from an agglomerative hierarchical clustering of the  $v_c^{WLM}$  vectors (distance function = cosine, linking criterion = average).

position in the mainstream heritage discourses since the 19th century, cultural heritage is the manifestation of a shared cultural history and a common collective identity. In this section, we aim at exploring with quantitative instruments this connection, focusing on the relation between built heritage and culture as a way to validate the cross-national differences we observed in Sections 3 and 4. Among the wide spectrum of models proposed in the literature, we adopt the mainstream etic approach in cross-cultural studies that assumes the existence of a set of universal cultural dimensions that are equally relevant to all cultures and that allow positioning a society relative to other societies along each of the dimension continuums. Even though several attempts to use large-scale online data, e.g., Facebook topical preferences [42] or Twitter news-oriented posts [3], have been made, social surveys are the de facto standard instrument within the cross-cultural studies community to quantitatively model cultural dimensions. The idea is to exploit the aggregation of individual opinions and the prevalence of some cultural traits to define national cultures, whereas cross-country comparisons are used to measure the cultural distance between countries or clusters of countries. In our work, we consider the WVS [27] that is a large-scale, cross-national, and cross-sectional longitudinal survey research program devoted to the scientific and academic study of social, political, economic, religious, and cultural values of people in the world.<sup>11</sup> The project grew out of the European Values Study<sup>12</sup> in 1981, and the corresponding social survey has been running every 5 years operating in more than 120 world societies. Extensive geographical and thematic scope, free availability of survey data, and project findings for the broad public turned the WVS into one of the most authoritative and widely used cross-national surveys in the social sciences. To maximize geographical coverage, we adopt as a reference dataset the **Integrated Value Survey (IVS)** [4, 18] that represents an effort to merge the WVS and EVS trend files in a single resource covering the time frame 1981–2021. Following the same methodology used in the construction of their Cultural Map [27], Inglehart and Welzel propose the *IW Index* [27] as a multi-dimensional construct to

<sup>11</sup><https://www.worldvaluessurvey.org/>.

<sup>12</sup><https://europeanvaluesstudy.eu/>.

characterize the cultural traits of a country. Operationally, they select 10 variables from the IVS that explore the dimensions of happiness, trust, respect for authority, voice (availability of people to express personal opinions by signing petitions), the importance of God, justification of homosexuality, abortion, national pride, individual rankings of social values, and obedience/independence [5].

In this scenario, each country is identified by a vector  $v_c^{IWI} = \langle v_1, \dots, v_{10} \rangle$ , where each component  $v_i$  represents the average value of the  $i^{th}$  question across IVS participants in country  $c$ . Since cultural dimensions are hard to interpret as absolute scores independently of the quality of the underlining model, we embrace the belief that cross-country cultural models can only be meaningfully interpreted and adopted in a comparative framework [23]. To enable comparisons among countries, we compute the cross-country distance matrix  $M_{IWI}$ , where the generic element  $m_{ij} = \text{cosine}(v_i^{IWI}, v_j^{IWI})$  contains the cosine distance between the two countries  $i$  and  $j$ , in other words, how different they are according to the 10 selected survey dimensions. We aim at exploring two questions: (1) Are the cross-national cultural relations linked to the built heritage ecosystem emerging from the WLM dataset? (2) Which approach (Places365 scene detectors or fine-tuned visual model) describes this relation better? To this purpose, we compute the matrices  $M_{places365}$  and  $M_{WLM}$  that encode, respectively, relative distances in the embedding space built from the Places365 and WLM fine-tuned visual models. We adopt the Mantel [36] test, which is a statistical test widely used in ecology to quantify the extent to which two distance matrices are correlated. We use the implementation in the scikit-bio Python package<sup>13</sup> and the Spearman's rank correlation coefficient as correlation metric. Our analysis is limited to the set of countries that are shared between the IVS and the WLM datasets, resulting in 49 countries. We observe a correlation coefficient  $\rho = 0.34$  and  $\rho = 0.45$  for, respectively, the  $M_{places365}$  and  $M_{WLM}$  cases. Both correlations are statistically significant, with  $p\text{-value} = 0.001$ . This means that countries that share similar cultural values tend to have a similarly built heritage profile in this dataset. Moreover, the approach based on the multi-class classification task and the fine-tuned visual model seems to capture these relations to a finer extent. However, not all countries behave the same: with this approach, we are able to quantify which pairs of countries are the closest/farthest, or which pairs have significantly different similarity estimates in the cultural values and WLM spaces (refer to Table A1 for an overview of the cross-country distance matrix for WLM). For example, United States/Canada, Chile/Argentina, and Serbia/Romania appear to be the pairs with the closest relation. The countries that are the most consistent with the cross-country cultural map are China, Belarus, Austria, Hungary, and Uruguay, while the ones that differ the most are Iran, Pakistan, Bangladesh, Armenia, and Thailand. The consistency score for a country is calculated as the sum of the pairwise differences in the distances estimation across the two representations. A more in-depth analysis of these relations along with their theoretical investigation is out of the scope of this work, which mainly aims at providing an example of how the crowdsourced collective knowledge emerging from WLM can be adopted to tackle relevant dimensions of the cross-cultural studies domain.

## 6 DISCUSSION

Beyond an interest in the WLM communities, this study highlights the richness that emerges from the data of the Wikimedia ecosystem as a scientific object. Indeed, part of our results ultimately bear on the question of stylistics for an art historian/iconologist or chrono-typology for an archaeologist. This approach, which dates back to the 19th century, has always been governed by the amount of data that can be processed to be as reliable as possible, even if it is only relative dating. The WLM contest based on Wikimedia Commons tools offers the possibility of working on perfectly documented Big Data. What this study confirms is that today's data analysis and processing tools offer new avenues of innovation in research for art historians, iconologists, archaeologists, and more broadly digital humanities. The automation of our research raises the question of the diffusion of architectural styles throughout the world that should be worked with by art historians interested in a global approach. One of our results is therefore to explicitly account for what WLM in particular and the wider pool of

<sup>13</sup><http://scikit-bio.org>.



media in Wikimedia Commons represent, namely a pool of data ready to be explored at different geographical, chronological, and sociological levels, allowing international as well as local approaches. When treated with care, the media database of Wikimedia Commons could prove to provide a treasure of insights. Moreover, from a platform management perspective, the longitudinal and geographical analysis summarized in Figure 2 gives a consolidated visualization of the countries that have participated in the contest. With some effort, this analysis could be repeated and deepened for an expanded dataset, to include other built heritage photos on Wikimedia Commons and other free and less free repositories. It would be particularly valuable to better understand which components of the world’s cultural heritage are poorly described with free photos, especially when combined with an analysis of identified and unidentified heritage sites.

The classification of images by using CNNs is a methodology that could be applied more broadly by assisting in the maintenance and suggestion of labels in the Wikimedia universe. While this would require adapted models, it may be helpful to surface input labels for microtasks. The method could also be adapted in combination with the monuments database to recognize characteristics along different axes than country, such as style and building type, that could be used to detect possible identification errors and labels.

**Limitations.** Our approach carries a few inherent limitations that are worth discussing, in particular:

- Due to the original objective of the competition, which is limited to built cultural heritage, this study is not able to fully characterize the different dimensions within the cultural heritage concept, focusing only on a fraction of it. There are several initiatives and different perspectives on how to fill this gap. For example, **Wiki Loves Heritage (WLH)**<sup>14</sup> is a photo contest organized for the first time in 2018 by Wikimedia Belgium that aims at broadening the set of items that are allowed to enter the competition to movable, immaterial, maritime heritage, landscape, and museum collections. However, WLH was run only in Belgium, limiting participation and adoption; moreover, there is no agreement that a new combined competition merging WLM-WLH would be the most effective way to fill the current gaps. Another approach would be to combine available datasets with a different focus, such as natural heritage (Wiki Loves Earth<sup>15</sup>) or intangible heritage (Wiki Loves Folklore<sup>16</sup> and Wiki Loves Africa<sup>17</sup>), that have been modeled after WLM. However, creating a broader data repository merging images from heterogeneous sources would create some challenges that we were able to sidestep in this exercise.
- Focusing on the community of WLM organizers, our methodological pipeline reinforces the importance of collecting and surfacing metadata on submissions. For example, roughly half of the photos uploaded to the contest have geographical coordinates embedded in the EXIF data, and therefore we chose to use the readily available and complete country data instead. It might be an interesting starting point for the study of regional and local trends. While the Wiki Loves Monuments dataset is relatively well documented with monument identifiers that connect a subset of the photos to a database, and Wikimedia Commons has a rich category structure, it is far from trivial to utilize this for the whole set. This may be the topic of further research, and the construction of a richer combined dataset may be worthwhile.
- The data used in this study suffers from a variety of sample biases that are intrinsic to the nature of the contest and the platform used for their collection. There is a number of factors that one could argue to both be introducing bias and be an effective influence on what constitutes a country’s cultural heritage:
  - **Participating countries:** Even though the competition spans 90 countries in 7 continents over its 11 years, the coverage is highly heterogeneous, and European countries are over-represented both in number and coverage. Likely, the presence of an active community, resources, and a culture of

<sup>14</sup>[https://commons.wikimedia.org/wiki/Commons:Wiki\\_Loves\\_contests](https://commons.wikimedia.org/wiki/Commons:Wiki_Loves_contests).

<sup>15</sup><https://wikilovesearth.org/>.

<sup>16</sup>[https://commons.wikimedia.org/wiki/Commons:Wiki\\_Loves\\_Folklore](https://commons.wikimedia.org/wiki/Commons:Wiki_Loves_Folklore).

<sup>17</sup><https://www.wikilovesafrica.net/>.

Table 5. Attributions for Images Included in Figures 6–8

Hilversum_-_Oude_Enghweg_8.jpg	Johan Bakker	CC BY-SA 4.0
Woudrichem_-_Hoogstraat_4_In_de_vergulden_helm.jpg	JHGS	CC BY-SA 3.0 NL
Huize_Vrede_(Luchtenveld,_Drachtstercompagnie).JPG	Baykedeuries	CC BY-SA 4.0
Ophemert.jpg	Peter van der Wielen	CC BY-SA 3.0 NL
Bossesteeg_-_vakwerkhuisje_achterzijde.jpg	Ciell	CC BY-SA 3.0 NL
Herenhuis_in_Overgangsstijl_met_electische_Chalet_en_Art_Nouveaustijl-elementen_1904_-_4.jpg	Hilda Morassi	CC BY-SA 3.0 NL
Oud_Groevenbeek-0117_-_Landhuis_Manor_-_Back_view.jpg	Michele Ahin	CC BY-SA 3.0 NL
Utrecht_Domtoren_from_Servetstraat,_2017.jpg	Samuel Mudrik	CC BY-SA 4.0
Sebastiaansdoelen_3_t.o.JPG	Elisa.rolle	CC BY-SA 3.0 NL
37768_Weteringstraat_19a.jpg	M. Stoffer	CC BY-SA 3.0 NL
Persepolis,1.jpg	Sima Heshmati	CC BY-SA 4.0
3_Gonbad_photo3.jpg	Ehsan Danandeh	CC BY-SA 4.0
Ka'ba-ye_Zartosht_(MG_5510-2).jpg	Herbert karim masihi	CC BY-SA 4.0
Alborz_Highschool_2016-4.jpg	Sepideh Shahbazi	CC BY-SA 4.0
Wonders_of_Shapouri_Garden.png	Amin Dehdarian	CC BY-SA 4.0
Stepan_Shahumyan_school9152164_10.JPG	Lilitik22	CC BY-SA 3.0
Alexander_Spendiaryan.jpg	Antimion	CC BY-SA 3.0
Church_in_Kazachi_post_56.JPG	Arman musikyan	CC BY-SA 3.0
Kobayr_monastery_1111_07.jpg	Dav Sargsyan	CC BY-SA 4.0
Kathmandu-Pashupatinath-Tempel-10-Aussichtsplattform-gje.jpg	Gerd Eichmann	CC BY-SA 4.0
Seto_Machindranath_Temple_(47).jpg	Nabin K. Sapkota	CC BY-SA 4.0
Jagannaath_Temple.jpg	Prasanna Bajracharya	CC BY-SA 4.0
Jal_Binayak_19.jpg	Anup Raj Rai	CC BY-SA 4.0
Carved_Stone_in_shape_of_Sun.jpg	Girisumit1990	CC BY-SA 4.0
Patan_Durbar_Square_image-7_2016.jpg	Biraj Mishra	CC BY-SA 4.0
Bishnu_Temple,Saptari_(4).JPG	NirajkKarn	CC BY-SA 3.0
Line_Sculpture_at_Patan_Durbar_Square.jpg	Kumud.parajuli	CC BY-SA 4.0
Bhudhanilkhantha_dhaneshwor8.jpg	Lusifor, manoj shrestha	CC BY-SA 4.0
Guheswori_Woripari_16.jpg	Anup Raj Rai	CC BY-SA 4.0
Wat_Phathanaram_funeral_pyre_1.jpg	Ahoerstemeyer	CC BY-SA 4.0
Dhara_Devi_spirit_house.jpg	Kate Branch	CC BY-SA 4.0
Palácio_dos_Príncipes_de_Joinville_075.JPG	Ehardt	CC BY-SA 4.0
Pôr_do_sol_em_Ouro_Preto.jpg	Plínio Daniel Lins Brandão Veas	CC BY-SA 4.0
Palacete_Provincial.jpg	Alexinaldo Granela Borja	CC BY-SA 4.0
Vista_da_fachada_do_Fórum.jpg	Leonardo F.B.	CC BY-SA 4.0
Rio_de_Janeiro_Arcos_da_Lapa_por_Carlos_Alkmin_-6714.jpg	Carlos Alkmin	CC BY-SA 4.0
A_colorida_Casa_de_Cultura.jpg	LopesRobson	CC BY-SA 4.0
Capela_de_Santo_Antônio_dos_Velasques_Vera_Cruz_Bahia_2018-0812.jpg	Paul R. Burley	CC BY-SA 4.0
Escaliers_de_la_rue_Foyatier_D161109.jpg	Yuri Granata Delalibera	CC BY-SA 4.0
Igreja_de_Nossa_Senhora_da_Palma_Salvador_Sacristy_Altar_2018-0653.jpg	Paul R. Burley	CC BY-SA 4.0
Casa_das_Rosas_Avenida_Paulista.jpg	PedroBDoprattsen	CC BY-SA 4.0
Jardines_Principales_del_Gran_Hotel_Cochabamba.JPG	Nicolas Elchakieh	CC BY-SA 3.0
2templo_santa_clara.jpg	Casilla Carreño Mireya	CC BY-SA 4.0
Cementerio_cbba_21.jpg	Valverde Angelica	CC BY-SA 4.0
Facultad_de_Medicina_4.jpg	victor manuel lozano	CC BY-SA 3.0
02_CATEDRAL_METROPOLITANA.jpg	Villegas Luján Santiago	CC BY-SA 4.0
20180922_154519_catedral_metropolitana.jpg	PaulCayoTiella	CC BY-SA 4.0
Fachada_Catedral_Metropolitana.jpg	Elu Rodriguez	CC BY-SA 4.0
IGLESIA_SAN_ILDEFONSO-33.jpg	Daniel tribeno	CC BY-SA 4.0
Catedral_metropolitana_de_Cochabamba_Bolivia_-_septiembre_2019_(45).jpg	Jhonn William Guillen Prado	CC BY-SA 4.0
CATEDRAL_NUESTRA_SEÑORA_DE_LA_PAZ.jpg	Arlett katherine cervantes zeballos	CC BY-SA 4.0
Debottar_Estate-Ajodhya-West_Bengal_722157-DSC_0028_01.jpg	Souvik pal	CC BY-SA 4.0
Architect_at_Chitrakarini_Temple.jpg	Pattnaik.chandan	CC BY-SA 4.0
Stone_Pillars_Chitradurga.jpg	Pavithrah	CC BY-SA 3.0
South_Side_Trikuta_Structure_-_Lakshmi_Narasimha_Temple,_Nuggehalli_03.jpg	Bikashrd	CC BY-SA 4.0
Cathedral_of_Mary_Help_of_Christians_side_view.jpg	Chinnu24	CC BY-SA 4.0
Many_wheels.JPG	Kinkiniroy2012	CC BY-SA 3.0
Rear_View_of_Cheluva_Narayanawamy_Temple.jpg	Anand Radhakrishnan	CC BY-SA 4.0
Ellora_Caves-Maharashtra_t-205.jpg	Akshatha Inamdar	CC BY-SA 4.0
Ajanta_Caves_22.jpg	Anupamg	CC BY-SA 4.0
Konark_sun_temple_Art-konark-Odisha-001.jpg	Phadke09	CC BY-SA 4.0

Table 6. Attributions for Images Included in Figures 6–8

Saraya_Building_-_Tzfai_-_Palmach_2273.jpg	Herman1975	CC BY-SA 3.0
Pousada_Sta_Maria_Bouro_032-1.jpg	Nittygritty	CC BY-SA 3.0
A_brasileira01.jpg	Taniagrafino	CC BY-SA 3.0
Lamego_-_Santuário_de_N._Sra_dos_Remédios.jpg	António Amen	CC BY-SA 3.0
Mosteiro_de_Tibães_(2).jpg	Sqjaques	CC BY-SA 4.0
Descida_da_Cruz_lateral_altar_mor.jpg	Ruipa	CC BY-SA 4.0
Jardins_de_la_Quinta_da_regaleira_-_Sintra_Portugal.jpg	Elaeudania278	CC BY-SA 3.0
O_Altar_Mor_e_o_púlpito.jpg	Senhormario	CC BY-SA 4.0
Igreja_da_Misericórdia_de_Penafiel_-_3.jpg	Reis Quarteu	CC BY-SA 4.0
Quinta_do_Morgado_1.jpg	Alegna13	CC BY-SA 3.0
Palácio_Sotto_Mayor_7287.jpg	João Carvalho	CC BY-SA 3.0
Codalet_(66)_Abbaye_Saint-Michel_de_Cuxa_Cloître_Chapiteau_Détail_07.JPG	GO69	CC BY-SA 3.0
Exterior_of_the_Castle_of_Valencay_24.jpg	Krzysztof Golik	CC BY-SA 4.0
Arch_within_arch_(Zoom_in).jpg	Shivaligupta	CC BY-SA 4.0
Dormition_Church123.JPG	I Bohdan	CC BY-SA 3.0
شهر تاریخی ماداکو .jpg	سیده پرگل شمسی	CC BY-SA 4.0
8_خانه قدکی .jpg	Mohandessahrsaz	CC BY-SA 4.0
12_باغ دولت آبادنمای داخلی .jpg	سیده زینب صفوی	CC BY-SA 4.0
6_مسجد دانشگاه تهران .jpg	Mehrshidroudian	CC BY-SA 4.0
Գաղիվանք վանական համալիր S-M_76 .jpg	Soghomon Matevosyan	CC BY-SA 4.0
Էջմիածնի եկեղեցական համալիր_34 .jpg	Voskanyan	CC BY-SA 4.0
Գյումրիի շին նշաններ_1 .jpg	Արարատ թրվանդ	CC BY-SA 4.0
Home_Gyumri_(Տուն_Գյումրի)_05 .jpg	Poghosyan97	CC BY-SA 4.0
Ախթալայի վանք_ Akhtala_monastery_14.jpg	Voskanyan	CC BY-SA 4.0
շին նշաններ_Գյումրի_24 .jpg	Soghomon Matevosyan	CC BY-SA 4.0
วัดพระพุทธฉาย อ.เมือง จ.สระบุรี (11) .jpg	กสิณธร ราชโอรส	CC BY-SA 4.0
บ้านหลังหลี่ เขตคลองสาน กรุงเทพมหานคร (15) .jpg	กสิณธร ราชโอรส	CC BY-SA 4.0
วัดปรีชนิก274 .jpg	CKasa	CC BY-SA 3.0
พระปรางค์สามยอด อ.เมือง จ.ลพบุรี (3) .JPG	กสิณธร ราชโอรส	CC BY-SA 4.0
วัดบางกะพ้อม อ.อัมพวา จ.สมุทรสงคราม(26) .jpg	กสิณธร ราชโอรส	CC BY-SA 4.0
วัดพระศรีสรรเพชญ์ HDR_1.jpg	Suttipong Kehawan	CC BY-SA 3.0
วัดราชโอรสราชวรวิหาร เขตจอมทอง กรุงเทพมหานคร (87) .jpg	กสิณธร ราชโอรส	CC BY-SA 4.0
วัดศรีเทพวรวิหาร เขตพระนคร กรุงเทพมหานคร (2) .jpg	กสิณธร ราชโอรส	CC BY-SA 4.0
212_בִּירְכֵי בִּירְכֵי לְרֵמֶת לְאֶחָד הַתְּנַנִּים בְּמִקְוֵה בְּבֵרֵי סֵת הַכּוֹפֵר .jpg	רויב וועבמ	CC BY-SA 4.0
בית הפילה מוסלמי עם שעון יפו מאחורי .jpg	ניק	CC BY-S 3.0
הפעמון במערה .jpg	Liorca	CC BY-SA 4.0
אפולוניה ממבט הצפון .jpg	Gilad Topaz	CC BY-SA 4.0
הצגה סרטי הכנסת הרוסית .jpg	Yehuda asulin	CC BY-SA 3.0
בית כנסת הב"ד .jpg	Litaln7	CC BY-SA 3.0
שעון השממ .jpg	Shosh sharet	CC BY-SA 3.0
วัดราชินีตาดารมKAI_4579 .jpg	Cat2you	CC BY-SA 4.0

participating as a volunteer or in a competition like this affect both the likelihood of organizing the competition in the first place and the number of participants it draws.

- **Heritage lists:** The competition uses heritage lists as the starting point of what photos can be submitted. The sites that are selected to be included in these lists are often selected by conservative mechanisms that may favor colonial-era heritage and limited viewpoints of what constitutes heritage. In addition, some countries are more inclusive. For example, Germany is estimated to have several hundred thousand sites on its lists, the Netherlands has 60K on its national list alone, and India only includes a few thousand sites. The fact that the competition uses primarily a pre-approved definition rather than the photographer’s interpretation of what constitutes “built heritage,” and that this mechanism differs somewhat in each country, can introduce a bias of what gets photographed in which country.
- **Access:** The accessibility of such sites to participating photographers and the public at large determines which sites in a country are more likely to be photographed.
- **Interest:** The willingness and interest of photographers to go out and capture the representation of any site may be affected by many factors, including whether the photo is exciting enough to result in a potential prize-winning photo.

- **Power users:** There is a small number of very active participants that photograph a lot of sites and upload them to the competition. While their drive is admirable, it may result in an over-representation of the country or the region that they happen to live in or visit. We attempted to limit the effect of this factor by applying filters.
- **Laws:** Many countries effectively prohibit more recent (post-colonial) or ancient heritage sites' photos to be published through copyright (lack of Freedom of Panorama [14]) or antiquity laws [16, 60]. This may either make it harder for organizers to collect and publish visual documentation of heritage or highly restrict and bias what can be documented.
- The image processing pipeline adopted in our work is based on deep convolutional neural network architectures that do not easily provide human-readable explanations for their choices. The current analysis lacks an interpretation layer that would provide domain experts with an actionable definition of the characteristic traits of the built cultural heritage of a country as emerging from the WLM corpus. Future work in this direction will reduce the gap between quantitative and qualitative interpretations, potentially enriching the toolbox available to quantitative art historians, iconologists, and archaeologists.

## 7 CONCLUSIONS

In this work, we presented an extensive characterization of the Wiki Loves Monuments photographic contest along the geographical, temporal, and topical dimensions. We highlighted the potential of the Wikimedia ecosystem to act as a scientific object for art historians, iconologists, and archaeologists, as well as the limitations and the biases that such a dataset introduces. Several avenues for future work are open: First, the analysis of the Wikimedia Commons Monuments Database<sup>18</sup> and the systematic integration with the WLM dataset will help shed light on intra-national or regional patterns that in many countries show high heterogeneity and complexity (e.g., cultural differences and artistic styles between regions in North and South Italy). Moreover, the integration with the geographic coordinates, when available, would provide a refined understanding of which heritage sites are most attractive for a photographer and, possibly, their relation with the official visitation statistics. Second, while in this work we rely on a snapshot that aggregates the 11 years of the competition, it would be interesting to explore the within-country temporal evolution of photographed sites to spot trending venues or the emergence of particular styles in the portrayed subjects or in the photographic techniques. Third, a more in-depth analysis of the visual content of the uploaded photos could provide a more nuanced framework to disentangle the effects of photographic styles and the characteristics of different artistic movements, providing the researcher with an explainable, multi-dimensional characterization of a country built heritage. In this direction, several deep learning frameworks have been recently applied to the art domain proposing aesthetic measures, artistic style detectors, and salient object detection tools.

In addition to the results presented, we hope to contribute to attracting new scientists to the richness and the potential of the projects within the Wikimedia ecosystem. This data mining exploration just scratched the surface of what is possible and it underlines the importance to work under precise hypotheses. This means the need for a closer dialogue with the digital humanities that hold those hypotheses and the adoption of interdisciplinary methodologies and teams.

<sup>18</sup>[https://commons.wikimedia.org/wiki/Commons:Monuments\\_database](https://commons.wikimedia.org/wiki/Commons:Monuments_database).

APPENDIX

A ADDITIONAL MATERIAL

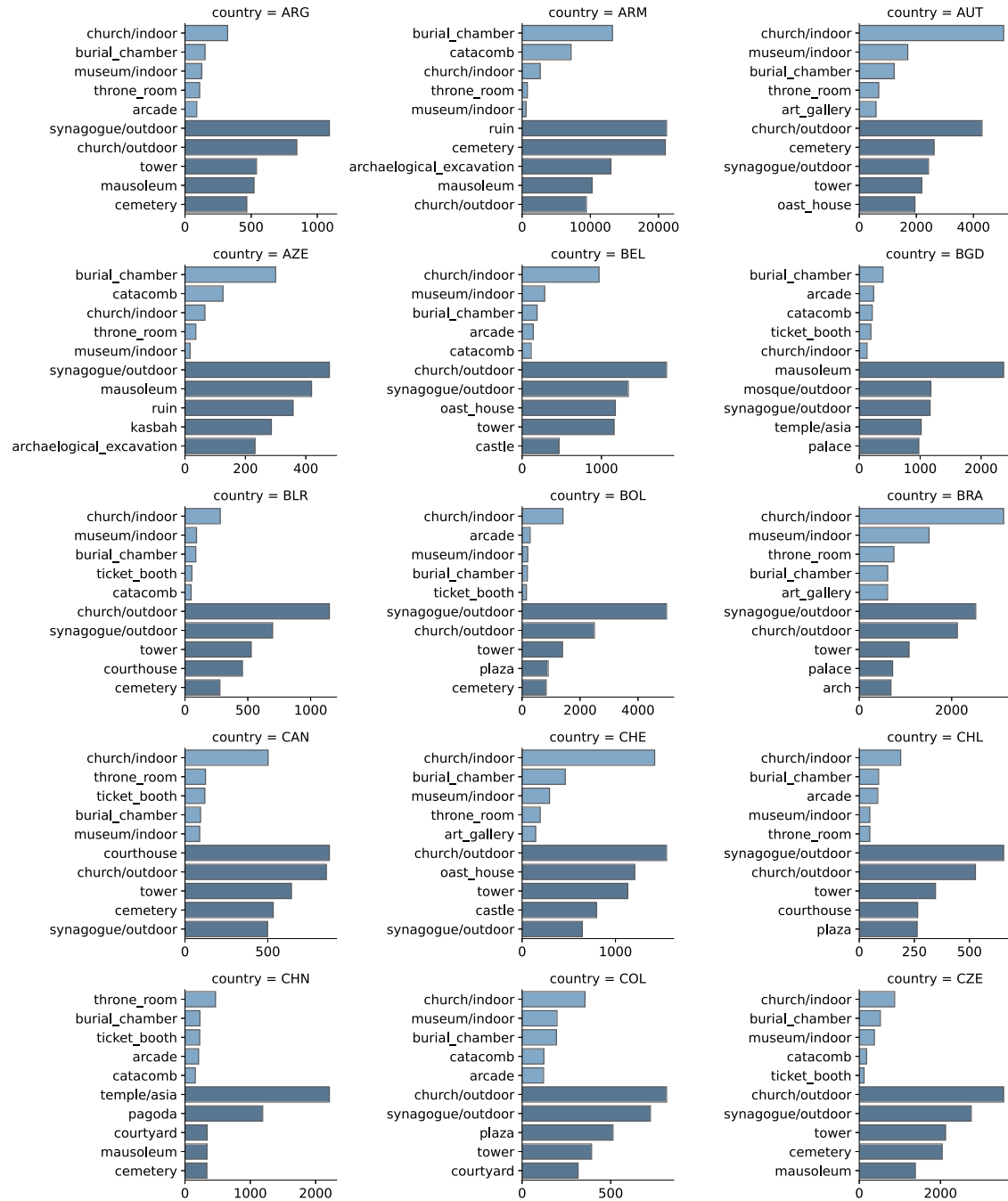


Fig. A1. Top 5 indoor and outdoor Places365 scenes per country.

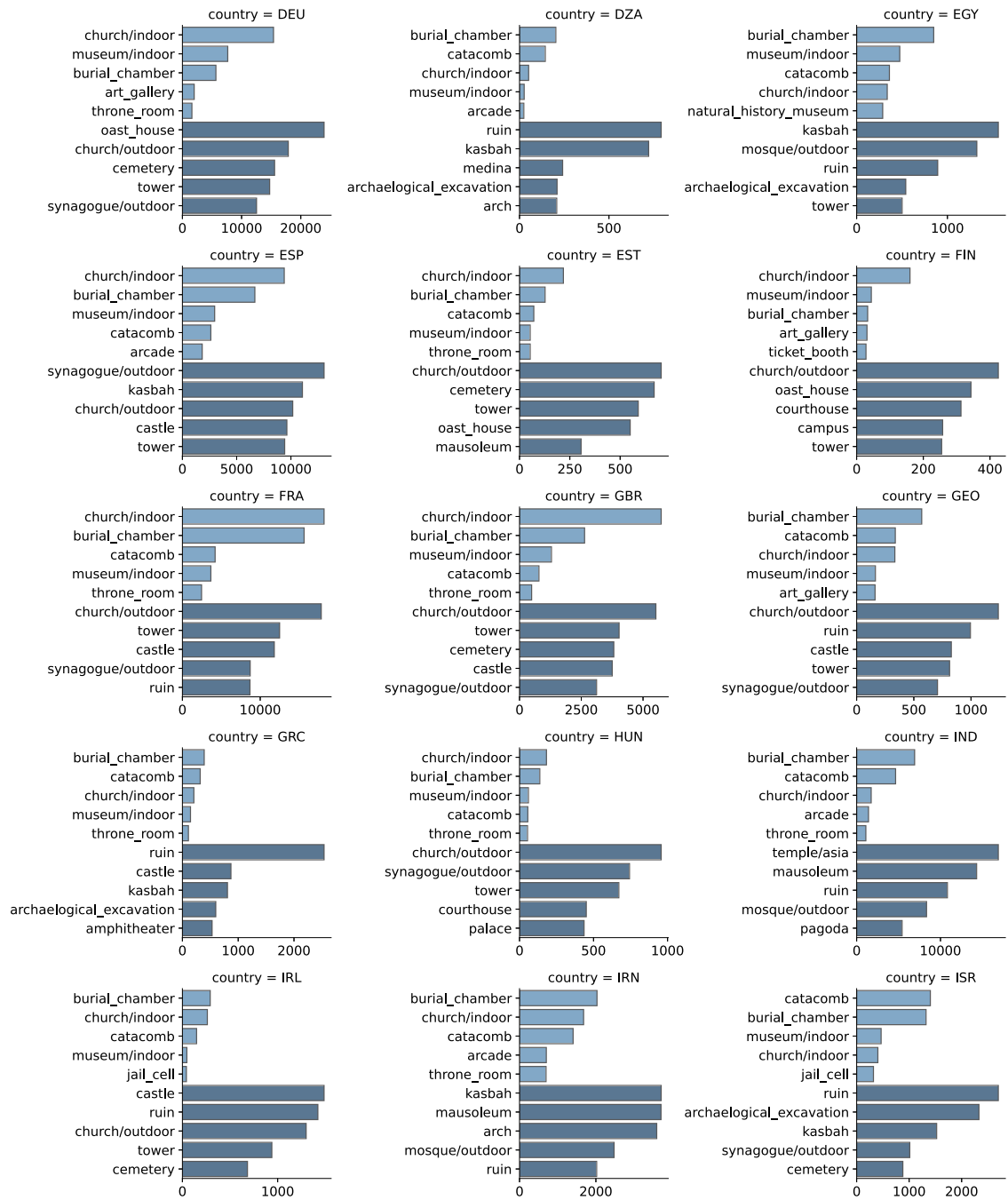


Fig. A2. Top 5 indoor and outdoor Places365 scenes per country.



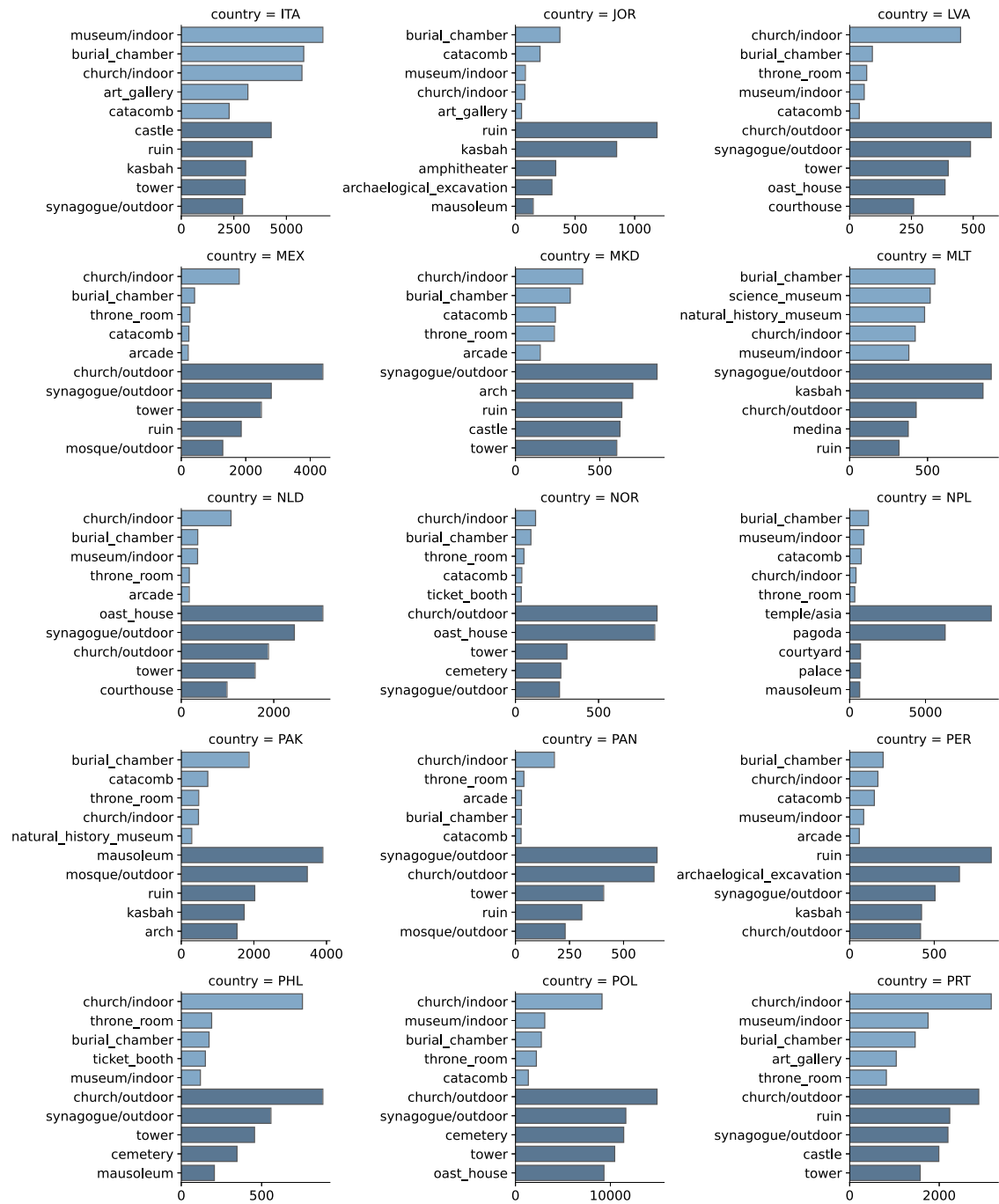


Fig. A3. Top 5 indoor and outdoor Places365 scenes per country.

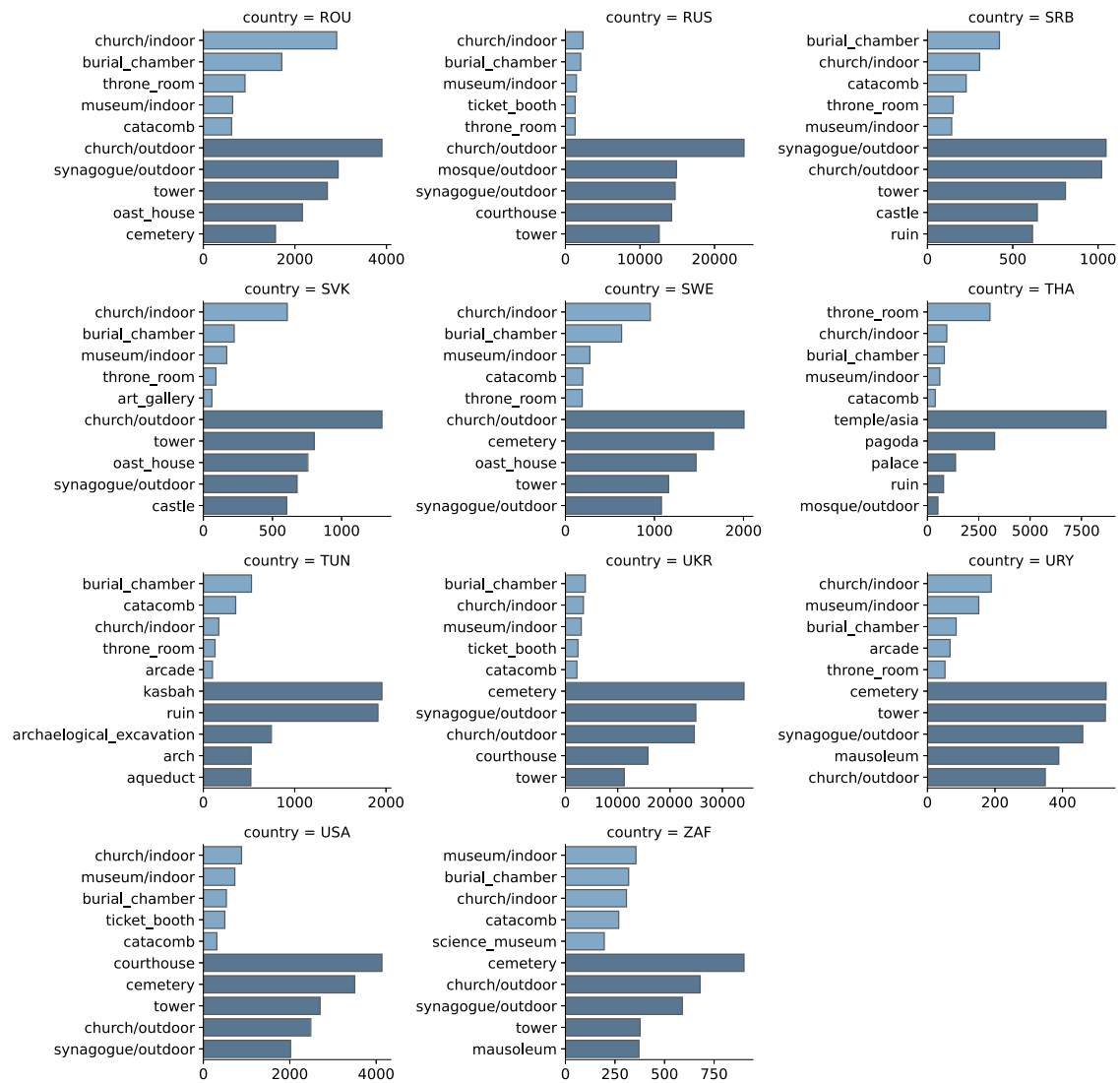


Fig. A4. Top 5 indoor and outdoor Places365 scenes per country.

Table A1. Summary of the Top 5 Most Similar Countries for Each National Contest

ARG	URY	0.08	CHL	0.08	COL	0.12	ZAF	0.15	USA	0.15
ARM	GEO	0.14	AZE	0.23	DZA	0.26	SRB	0.27	GRC	0.28
AUT	CZE	0.09	SVK	0.09	HUN	0.10	CHE	0.10	POL	0.11
AZE	ESP	0.15	GEO	0.15	SRB	0.16	GRC	0.17	DZA	0.18
BGD	IND	0.16	PAK	0.23	PHL	0.27	CHN	0.28	COL	0.28
BLR	RUS	0.07	UKR	0.09	POL	0.10	SRB	0.15	EST	0.16
BOL	CHL	0.17	COL	0.17	ARG	0.19	MEX	0.19	URY	0.22
BRA	PRT	0.12	COL	0.14	ITA	0.16	CHL	0.17	ARG	0.18
CAN	USA	0.05	FIN	0.14	NOR	0.16	GBR	0.16	ZAF	0.16
CHE	AUT	0.10	FRA	0.16	HUN	0.17	DEU	0.17	SVK	0.17
CHL	ARG	0.08	COL	0.11	URY	0.11	ZAF	0.15	MEX	0.15
CHN	COL	0.21	PHL	0.21	USA	0.22	CHL	0.23	CAN	0.24
COL	CHL	0.11	ARG	0.12	MEX	0.13	BRA	0.14	URY	0.14
CZE	SVK	0.07	POL	0.07	HUN	0.07	AUT	0.09	DEU	0.13
DEU	POL	0.13	AUT	0.13	CZE	0.13	CHE	0.17	SVK	0.18
DZA	JOR	0.12	TUN	0.12	GRC	0.13	EGY	0.15	PER	0.17
EGY	JOR	0.14	DZA	0.15	PER	0.19	TUN	0.20	IRN	0.22
ESP	ITA	0.11	PRT	0.11	FRA	0.12	GRC	0.15	AZE	0.15
EST	SWE	0.12	POL	0.12	FIN	0.13	RUS	0.15	BLR	0.16
FIN	SWE	0.08	NOR	0.11	EST	0.13	CAN	0.14	POL	0.15
FRA	ESP	0.12	GBR	0.13	ITA	0.14	PRT	0.15	AUT	0.16
GBR	FRA	0.13	CAN	0.16	USA	0.16	POL	0.19	ZAF	0.19
GEO	ARM	0.14	SRB	0.15	AZE	0.15	ROU	0.19	GRC	0.19
GRC	DZA	0.13	ITA	0.14	ESP	0.15	JOR	0.16	TUN	0.17
HUN	CZE	0.07	SVK	0.09	SRB	0.10	AUT	0.10	POL	0.10
IND	PAK	0.15	BGD	0.16	DZA	0.25	EGY	0.25	COL	0.27
IRN	PAK	0.19	DZA	0.20	AZE	0.20	TUN	0.20	EGY	0.22
ITA	ESP	0.11	FRA	0.14	GRC	0.14	PRT	0.14	SRB	0.16
JOR	DZA	0.12	EGY	0.14	GRC	0.16	TUN	0.17	PER	0.22
MEX	COL	0.13	ARG	0.15	CHL	0.15	ESP	0.16	PRT	0.17
MKD	SRB	0.13	ROU	0.16	GRC	0.19	ESP	0.19	PRT	0.20
NLD	GBR	0.27	CAN	0.27	USA	0.30	DEU	0.32	POL	0.33
NOR	SWE	0.10	FIN	0.11	CAN	0.16	USA	0.17	EST	0.17
PAK	IND	0.15	IRN	0.19	DZA	0.23	EGY	0.23	BGD	0.23
PER	CHL	0.17	DZA	0.17	COL	0.18	EGY	0.19	MEX	0.20
PHL	COL	0.16	CHL	0.20	BRA	0.21	CHN	0.21	MEX	0.21
POL	CZE	0.07	SVK	0.10	HUN	0.10	BLR	0.10	AUT	0.11
PRT	ESP	0.11	BRA	0.12	ITA	0.14	FRA	0.15	COL	0.15
ROU	SRB	0.08	SVK	0.10	HUN	0.10	UKR	0.13	POL	0.13
RUS	UKR	0.05	BLR	0.07	POL	0.13	EST	0.15	FIN	0.15
SRB	ROU	0.08	HUN	0.10	POL	0.13	MKD	0.13	SVK	0.13
SVK	CZE	0.07	AUT	0.09	HUN	0.09	POL	0.10	ROU	0.10
SWE	FIN	0.08	NOR	0.10	EST	0.12	POL	0.14	CAN	0.16
THA	CHN	0.28	PHL	0.34	COL	0.34	BRA	0.36	IND	0.37
TUN	DZA	0.12	GRC	0.17	JOR	0.17	AZE	0.19	ESP	0.19
UKR	RUS	0.05	BLR	0.09	POL	0.11	ROU	0.13	SRB	0.13
URY	ARG	0.08	CHL	0.11	COL	0.14	ZAF	0.15	USA	0.16
USA	CAN	0.05	ZAF	0.12	ARG	0.15	CHL	0.16	FIN	0.16
ZAF	USA	0.12	CHL	0.15	ARG	0.15	URY	0.15	CAN	0.16

Distances are computed using the cosine similarity between the per-country  $v_c^{WLM}$  vectors.

## ACKNOWLEDGMENTS

We thank the many volunteers who worked on Wiki Loves Monuments both in national and international teams, whose work resulted in this collection of photos. The authors acknowledge the creators of the images included in Figures 6–8. Attributions are provided in Tables 5 and 6.

## REFERENCES

- [1] Eneko Agirre, Ander Barrena, Oier Lopez de Lacalle, Aitor Soroa, Samuel Fernando, and Mark Stevenson. 2012. Matching cultural heritage items to Wikipedia. In *Proceedings of the 8th International Conference on Language Resources and Evaluation (LREC'12)*. European Language Resources Association (ELRA), Istanbul, Turkey, 1729–1735. [http://www.lrec-conf.org/proceedings/lrec2012/pdf/1021\\_Paper.pdf](http://www.lrec-conf.org/proceedings/lrec2012/pdf/1021_Paper.pdf).
- [2] Flora Amato, Vincenzo Moscato, Antonio Picariello, Francesco Colace, Massimo De Santo, Fabio A. Schreiber, and Letizia Tanca. 2017. Big data meets digital cultural heritage: Design and implementation of SCRABS, a smart context-AwaRe browsing assistant for cultural environments. *J. Comput. Cult. Herit.* 10, 1, Article 6 (April 2017), 23 pages. <https://doi.org/10.1145/3012286>
- [3] Issa Annamoradnejad, MohammadAmin Fazli, Jafar Habibi, and Sadjad Tavakoli. 2019. Cross-cultural studies using social networks data. *IEEE Transactions on Computational Social Systems* 6, 4 (2019), 627–636. <https://doi.org/10.1109/TCSS.2019.2919666>
- [4] GESIS Data Archive. 2021. *EVS (2021): EVS Trend File 1981-2017*. Technical Report ZA7503 Data file Version 2.0.0. GESIS. DOI : <https://doi.org/10.4232/1.13736>
- [5] Luca De Benedictis, Roberto Rondinelli, and Veronica Vinciotti. 2021. The network structure of cultural distances. arXiv:2007.02359 [stat.AP]
- [6] Samarth Bhargav, Nanne van Noord, and Jaap Kamps. 2019. Deep learning as a tool for early cinema analysis. In *Proceedings of the 1st Workshop on Structuring and Understanding of Multimedia HeritAge Contents (SUMAC'19)*. Association for Computing Machinery, New York, NY, 61–68. <https://doi.org/10.1145/3347317.3357240>
- [7] Karol Jan Borowiecki, Neil Forbes, and Antonella Fresa. 2016. *Cultural Heritage in a Changing World*. Springer Nature, Cham. <https://doi.org/10.1007/978-3-319-29544-2>
- [8] Ewa S. Callahan and Susan C. Herring. 2011. Cultural bias in Wikipedia content on famous persons. *Journal of the American Society for Information Science and Technology* 62, 10 (2011), 1899–1915. <https://doi.org/10.1002/asi.21577> arXiv:<https://onlinelibrary.wiley.com/doi/pdf/10.1002/asi.21577>
- [9] François Chollet. 2016. Xception: Deep learning with depthwise separable convolutions. *CoRR* abs/1610.02357 (2016), 1800–1807. arXiv:1610.02357 <http://arxiv.org/abs/1610.02357>.
- [10] Dana Dannélls, Aarne Ranta, Ramona Enache, Mariana Damova, and Maria Mateva. 2013. Multilingual access to cultural heritage content on the Semantic Web. In *Proceedings of the 7th Workshop on Language Technology for Cultural Heritage, Social Sciences, and Humanities*. Association for Computational Linguistics, 107–115. <https://aclanthology.org/W13-2715>.
- [11] Victor de Boer, Jan Wielemaker, Judith van Gent, Michiel Hildebrand, Antoine Isaac, Jacco van Ossenbruggen, and Guus Schreiber. 2012. Supporting linked data production for cultural heritage institutes: The Amsterdam museum case study. In *The Semantic Web: Research and Applications*, Elena Simperl, Philipp Cimiano, Axel Polleres, Oscar Corcho, and Valentina Presutti (Eds.). Springer, Berlin, 733–747.
- [12] Victor de Boer, Jan Wielemaker, Judith van Gent, Marijke Oosterbroek, Michiel Hildebrand, Antoine Isaac, Jacco van Ossenbruggen, and Guus Schreiber. 2013. Amsterdam museum linked open data. *Semant. Web* 4, 3 (July 2013), 237–243.
- [13] Mauro Dragoni, Sara Tonelli, and Giovanni Moretti. 2017. A knowledge management architecture for digital cultural heritage. *J. Comput. Cult. Herit.* 10, 3, Article 15 (July 2017), 18 pages. <https://doi.org/10.1145/3012289>
- [14] Mélanie Dulong de Rosnay and Pierre-Carl Langlais. 2017. Public artworks and the freedom of panorama controversy: A case of wikimedia influence. *Internet Policy Review* 6, 1 (2017), 1–27. <https://doi.org/10.14763/2017.1.447>
- [15] Bruno Fanini, Alfonsina Pagano, and Daniele Ferdani. 2018. A novel immersive VR game model for recontextualization in virtual environments: The  $\rho$ VRModel. *Multimodal Technologies and Interaction* 2 (2018), 20. <https://doi.org/10.3390/mti2020020>
- [16] Greek Ministerial Order GDAPK/DMEEP/C2/F51-5. 2005. Determination of Fees for: a) Photography - Filming - Video Recording in Museums, Monuments and Archaeological Sites, b) Publishing Photographs of Archeological Content for Commercial Exploitation, c) Electronic Publishing, d) Use of Images over the Internet, e) Events in Ancient Theaters and Other Archeological Sites, f) Procedure for Paying Fees for Impressions, Copies and Illustrations of Immovable and Movable Monuments that Belong to the State, g) Procedure for Paying Fees for the Use of Immovable and Movable Monuments on a Logo. WIPO Lex Database. <https://wipolex.wipo.int/en/legislation/details/9073>.
- [17] Ian J. Goodfellow, Yoshua Bengio, and Aaron Courville. 2016. *Deep Learning*. MIT Press, Cambridge, MA.
- [18] C. Haerpfer, R. Inglehart, A. Moreno, C. Welzel, K. Kizilova, J. Diez-Medrano, M. Lagos, P. Norris, E. Ponarin, B. Puranen, et al. 2021. *World Values Survey Time-Series (1981–2020) Cross-National Data-Set*. Technical Report Data File Version 2.0.0. JD Systems Institute and WVSA Secretariat. DOI : <https://doi.org/10.14281/18241.15>

- [19] Mark Michael Hall, Oier Lopez de Lacalle, Aitor Soroa Etxabe, Paul Clough, and Eneko Agirre. 2012. Enabling the discovery of digital cultural heritage objects through Wikipedia. In *Proceedings of the 6th Workshop on Language Technology for Cultural Heritage, Social Sciences, and Humanities*. Association for Computational Linguistics, 94–100. <https://aclanthology.org/W12-1013>.
- [20] Simao Herdade, Armin Kappeler, Kofi Boakye, and Joao Soares. 2019. Image captioning: Transforming objects into words. *CoRR abs/1906.05963* (2019), 10. arXiv:1906.05963 <http://arxiv.org/abs/1906.05963>.
- [21] Geert Hofstede. 1980. *Culture's Consequences: International Differences in Work Related Values*. Sage Publications, Beverly Hills, CA.
- [22] Geert Hofstede. 2001. *Culture's Consequences: Comparing Values, Behaviors, Institutions, and Organizations Across Nations* (2nd ed.). Sage Publications, London, UK.
- [23] Geert Hofstede. 2011. Dimensionalizing cultures: The Hofstede model in context. *Online Readings in Psychology and Culture* 2, 1 (2011), 26. <https://doi.org/10.9707/2307-0919.1014>
- [24] C. Holtorf, A. Pantazatos, and G. Scarre. 2018. *Cultural Heritage, Ethics and Contemporary Migrations* (1st ed.). Routledge, United Kingdom. <https://doi.org/10.4324/9780429464300>
- [25] Eero Hyvönen. 2012. *Publishing and Using Cultural Heritage Linked Data on the Semantic Web*. Morgan and Claypool Publishers. <http://dx.doi.org/10.2200/S00452ED1V01Y201210WBE003>
- [26] Ronald Inglehart. 1997. *Modernization and Postmodernization: Cultural, Economic, and Political Change in 43 Societies*. Princeton University Press, Princeton, NJ.
- [27] Ronald Inglehart and Christian Welzel. 2005. *Modernization, Cultural Change, and Democracy: The Human Development Sequence*. Cambridge University Press, United Kingdom. <https://doi.org/10.1017/CBO9780511790881>
- [28] Youngok Kang, Nahye Cho, Jiyoun Yoon, Soyeon Park, and Jiyeon Kim. 2021. Transfer learning of a deep learning model for exploring tourists' urban image using geotagged photos. *ISPRS International Journal of Geo-Information* 10, 3 (2021), 20. <https://doi.org/10.3390/ijgi10030137>
- [29] Jenny Kidd. 2018. *Public Heritage and the Promise of the Digital*. Oxford University Press, United Kingdom. <https://doi.org/10.1093/oxfordhb/9780190676315.013.9>
- [30] Uday Kulkarni, S. M. Meena, Sunil V. Gurlahosur, and Uma Mudengudi. 2019. Classification of cultural heritage sites using transfer learning. In *2019 IEEE 5th International Conference on Multimedia Big Data (BigMM'19)*. IEEE, 391–397. <https://doi.org/10.1109/BigMM.2019.00020>
- [31] Paul Laufer, Claudia Wagner, Fabian Flöck, and Markus Strohmaier. 2015. Mining cross-cultural relations from Wikipedia: A study of 31 European food cultures. In *Proceedings of the ACM Web Science Conference*. Association for Computing Machinery, New York, NY, 10. <https://doi.org/10.1145/2786451.2786452>
- [32] Jose Llamas, Pedro Leronés, Roberto Medina, Eduardo Zalama, and Jaime Gómez-García-Bermejo. 2017. Classification of architectural heritage images using deep learning techniques. *Applied Sciences* 7 (2017), 992. <https://doi.org/10.3390/app7100992>
- [33] Jose Llamas, Pedro Leronés, Eduardo Zalama, and Jaime Gómez-García-Bermejo. 2016. Applying deep learning techniques to cultural heritage images within the INCEPTION project. In *Digital Heritage. Progress in Cultural Heritage: Documentation, Preservation, and Protection*, Vol. 10059. Springer International Publishing, Cham, 25–32. [https://doi.org/10.1007/978-3-319-48974-2\\_4](https://doi.org/10.1007/978-3-319-48974-2_4)
- [34] Tino Mager, Seyran Khademi, Ronald Siebes, Carola Hein, Victor Boer, and Jan Gemert. 2020. *Visual Content Analysis and Linked Data for Automatic Enrichment of Architecture-Related Images*. Springer, Cham, 279–293. [https://doi.org/10.1007/978-3-030-15200-0\\_18](https://doi.org/10.1007/978-3-030-15200-0_18)
- [35] Ferdinand Maiwald, T. Vietze, D. Schneider, Frank Henze, Sander Münster, and Florian Niebling. 2017. Photogrammetric analysis of historical image repositories for virtual reconstruction in the field of digital humanities. *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences XLII-2/W3* (2017), 447–452. <https://doi.org/10.5194/isprs-archives-XLII-2-W3-447-2017>
- [36] Nathan Mantel. 1967. The detection of disease clustering and a generalized regression approach. *Cancer Research* 27 2 (1967), 209–20.
- [37] Joseph J. DiStefano and Martha L. Maznevski. 1995. *Measuring Culture in International Management: The Cultural Perspectives Questionnaire*. University of Western Ontario Working Paper Series, Canada.
- [38] Joseph J. DiStefano and Martha L. Maznevski. 2004. *Culture, Leadership, and Organizations: The Globe Study of 62 Societies*. Sage Publications, Thousand Oaks, CA.
- [39] Bibiana Martínez, Sergio Casas-Yrurzum, María Vidal-González, Lucia Vera, and Inma García-Pereira. 2018. TinajAR: An edutainment augmented reality mirror for the dissemination and reinterpretation of cultural heritage. *Multimodal Technologies and Interaction* 2 (2018), 33. <https://doi.org/10.3390/mti2020033>
- [40] Ben Marwick and Prema Smith. 2021. World heritage sites on Wikipedia: Cultural heritage activism in a context of constrained agency. *Big Data and Society* 8, 1 (2021), 19. <https://doi.org/10.1177/20539517211017304>
- [41] Florian Niebling, Jonas Bruschke, Heike Messemer, Markus Wacker, and Sebastian von Mammen. 2020. *Analyzing Spatial Distribution of Photographs in Cultural Heritage Applications*. Springer International Publishing, Cham, 391–408. [https://doi.org/10.1007/978-3-030-37191-3\\_20](https://doi.org/10.1007/978-3-030-37191-3_20)
- [42] Nick Obradovich, Ömer Özak, Ignacio Martín, Ignacio Ortuño-Ortín, Edmond Awad, Manuel Cebrián, Rubén Cuevas, Klaus Desmet, Iyad Rahwan, and Ángel Cuevas. 2020. *Expanding the Measurement of Culture with a Sample of Two Billion Humans*. NBER Working Papers 27827. National Bureau of Economic Research, Inc. <https://ideas.repec.org/p/nbr/nberwo/27827.html>.

- [43] Frank Palermo, James Hays, and Alexei Efros. 2012. Dating historical color images. In *Computer Vision (ECCV'12)*, Vol. 7577. Springer, Berlin, 499–512. [https://doi.org/10.1007/978-3-642-33783-3\\_36](https://doi.org/10.1007/978-3-642-33783-3_36)
- [44] Steve Pan, Jinsoo Lee, and Henry Tsai. 2014. Travel photos: Motivations, image dimensions, and affective qualities of places. *Tourism Management* 40 (2014), 59–69. <https://doi.org/10.1016/j.tourman.2013.05.007>
- [45] Hermann Parzinger. 2020. Togetherness: A new heritage deal for Europe. In *Big Ideas*. European Investment Bank, European Union, 44.
- [46] Christian Pentzold, Esther Weltevrede, Michele Mauri, David Laniado, Andreas Kaltenbrunner, and Erik Borra. 2017. Digging Wikipedia: The online encyclopedia as a digital cultural heritage gateway and site. *Journal on Computing and Cultural Heritage* 10, 1, Article 5 (2017), 19 pages. <https://doi.org/10.1145/3012285>
- [47] Charith Perera, Arkady Zaslavsky, Peter Christen, and Dimitrios Georgakopoulos. 2014. Context aware computing for the internet of things: A survey. *IEEE Communications Surveys Tutorials* 16, 1 (2014), 414–454. <https://doi.org/10.1109/SURV.2013.042313.00197>
- [48] Cristina Portalés, Joao Rodrigues, Alexandra Gonçalves, Ester Alba, and Jorge Sebastián Lozano. 2018. Digital cultural heritage. *Multimodal Technologies and Interaction* 2 (2018), 58. <https://doi.org/10.3390/mti2030058>
- [49] Emilio José Rodríguez Posada, Ángel Gonzalez Berdasco, Jorge A. Sierra Canduela, Santiago Navarro Sanz, and Tomás Saorín. 2012. Wiki loves monuments 2011: The experience in Spain and reflections regarding the diffusion of cultural heritage. *Digithum* 14 (2012), 94–104. <https://raco.cat/index.php/Digithum/article/view/254241>.
- [50] Martin Poulter and Waqās Ahmed. 2021. Representation of Non-Western Cultural Knowledge on Wikipedia: The Case of the Visual Arts. Preprints 2021, 2021040770. DOI : 10.20944/preprints202104.0770.v1
- [51] Joseph Redmon and Ali Farhadi. 2018. YOLOv3: An incremental improvement. *ArXiv abs/1804.02767* (04 2018), 6.
- [52] Shaoqing Ren, Kaiming He, Ross Girshick, and Jian Sun. 2015. Faster R-CNN: Towards real-time object detection with region proposal networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 39 (2015), 1137–1149. <https://doi.org/10.1109/TPAMI.2016.2577031>
- [53] Mia Ridge. 2013. From tagging to theorizing: Deepening engagement with cultural heritage through crowdsourcing. *Curator: The Museum Journal* 56, 4 (October 2013), 435–450.
- [54] Maria Roussou and Akrivi Katifori. 2018. Flow, staging, wayfinding, personalization: Evaluating user experience with mobile museum narratives. *Multimodal Technologies and Interaction* 2 (2018), 32. <https://doi.org/10.3390/mti2020032>
- [55] Olga Russakovsky, Jia Deng, Hao Su, Jonathan Krause, Sanjeev Satheesh, Sean Ma, Zhiheng Huang, Andrej Karpathy, Aditya Khosla, Michael Bernstein, Alexander Berg, and Li Fei-Fei. 2014. ImageNet large scale visual recognition challenge. *International Journal of Computer Vision* 115 (2014), 211–252. <https://doi.org/10.1007/s11263-015-0816-y>
- [56] Shalom H. Schwartz. 1994. *Beyond Individualism/Collectivism: New Cultural Dimensions of Values*. Sage Publications, London, United Kingdom.
- [57] Karen Simonyan and Andrew Zisserman. 2015. Very deep convolutional networks for large-scale image recognition, In *3rd International Conference on Learning Representations (ICLR'15)*, Conference Track Proceedings, Yoshua Bengio and Yann LeCun (Eds.). *ArXiv abs/1409.1556*, 14. <http://arxiv.org/abs/1409.1556>.
- [58] Pedro Szekely, Craig A. Knoblock, Fengyu Yang, Xuming Zhu, Eleanor E. Fink, Rachel Allen, and Georgina Goodlander. 2013. Connecting the Smithsonian American art museum to the linked data cloud. In *The Semantic Web: Semantics and Big Data*, Philipp Cimiano, Oscar Corcho, Valentina Presutti, Laura Hollink, and Sebastian Rudolph (Eds.). Springer, Berlin, 593–607.
- [59] Fons Trompenaars. 1993. *Riding the Waves of Culture: Understanding Diversity in Global Business*. Irwin Professional Publishing, Chicago, IL.
- [60] Gazzetta Ufficiale. 2004. Codice dei beni culturali e del paesaggio, ai sensi dell'articolo 10 della legge 6 luglio 2002, n.137, decreto legislativo 22 gennaio 2004 n.42. Gazzetta Ufficiale della Repubblica Italiana. <https://www.gazzettaufficiale.it/eli/gu/2004/02/24/45/so/28/sg/pdf>.
- [61] Laurens van der Maaten and Geoffrey Hinton. 2008. Visualizing data using t-SNE. *Journal of Machine Learning Research* 9, 86 (2008), 2579–2605. <http://jmlr.org/papers/v9/vandermaaten08a.html>.
- [62] Melvin Wevers. 2021. Scene detection in De Boer historical photo collection. In *Proceedings of the 13th International Conference on Agents and Artificial Intelligence (ICAART'21), Volume 1, Online Streaming*, Ana Paula Rocha 0001, Luc Steels, and H. Jaap van den Herik (Eds.). SCITEPRESS, Portugal, 601–610. <https://doi.org/10.5220/0010288206010610>
- [63] Baoyuan Wu, Weidong Chen, Yanbo Fan, Yong Zhang, Jinlong Hou, Jie Liu, and Tong Zhang. 2019. Tencent ML-images: A large-scale multi-label image database for visual representation learning. *IEEE Access* PP (2019), 1–1. <https://doi.org/10.1109/ACCESS.2019.2956775>
- [64] Bolei Zhou, Aditya Khosla, Ágata Lapedriza, Antonio Torralba, and Aude Oliva. 2016. Places: An image database for deep scene understanding. *Journal of Vision* 17 (2016). <https://doi.org/10.1167/17.10.296>
- [65] Bolei Zhou, Ágata Lapedriza, Aditya Khosla, Aude Oliva, and Antonio Torralba. 2017. Places: A 10 million image database for scene recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence* PP (2017), 1–1. <https://doi.org/10.1109/TPAMI.2017.2723009>

Received 26 February 2022; revised 12 June 2022; accepted 15 June 2022