


ARTICLE

Heritage site-seeing through the visitor's lens on Instagram

Tania Loke¹, Yayoi Teramoto² , Chico Q. Camargo³ , Kathryn Eccles¹¹ Oxford Internet Institute, University of Oxford, ² Oxford Brookes University, ³ Computer Science, University of Exeter

Keywords: cultural heritage, instagram, tourism, social media analysis, computer vision, object detection

<https://doi.org/10.22148/001c.38966>

Journal of Cultural Analytics

Vol. 7, Issue 3, 2022

English Heritage is a charity that manages over 400 historic sites in the UK, from prehistoric sites to medieval castles, most of them free, non-ticketed, and unstaffed. As such, there is little information about visitor attendance and behaviour in those sites—a challenge common to other non-ticketed heritage sites. In this context, image-based social media such as Instagram appear as a possible solution, as photographs are often central to the tourist experience, and tourists present their imagined audiences with a self-narrative of their trip. Therefore, this study aims to improve our understanding of tourist behaviour in unstaffed heritage sites by analysing publicly available Instagram data. We collect posts on unstaffed English Heritage sites, finding that posting activity concentrates at a few sites. Focusing on 3,979 images each for the top five sites, we analyse image content using pre-trained object detection models. Besides off-the-shelf inference, we fine-tune a model to identify structures from particular heritage sites, and are able to describe the types of photographs taken by visitors in each site, supporting the notion of tourists as performers with the site serving as backdrop. Overall, this study demonstrates a methodology for understanding cultural behaviour at heritage sites using images from social media posts. In addition to recovering the otherwise lost connection between a heritage organisation and its visitors, our methodology can be readily extended to other tourist destinations to understand how visitors interact with and relate to these sites and the objects within them through their photographs.

1. Introduction

English Heritage is a charity managing over 400 historic buildings, monuments, and places. These include prehistoric sites, medieval castles, Roman forts, stately houses, and the ruins of numerous abbeys, priories, and villages. Access to 255 of these sites is free, representing over half of the charity's portfolio. Since these free sites are unstaffed and non-ticketed, they pose a challenge in understanding visitor behaviour. More recently, however, tourists have often registered their experience by posting reviews, photographs, and comments on social media. Such data offer an unprecedented, though particular, view of the visitor experience at these sites. Given that photography is intrinsically linked to the being, doing, and performing of tourism (Robinson and Picard 1), image-based social media data such as Instagram can serve as rich datasets revealing visitors' preferences and behaviours (Balomenou et al. 173). Besides summarising tourists' perceived destination image (Donaire et al. 27), photographs are central to the visitor experience itself, as visitors stage and enact performances to present their imagined audiences with a desired self-narrative of their trip (Belk 349; Larsen 425). This study aims to improve our understanding of cultural behaviour in unstaffed heritage sites using publicly available Instagram data. Specifically, we ask:

- (RQ1) What does the distribution of posts, across sites and across time, reveal about site popularity and temporal patterns in visitor engagement?
- (RQ2) What are visitors taking and sharing photographs of across different heritage sites?
- (RQ3) How do visitors perform and depict their interactions with heritage sites in these photographs?

This study extends existing research on tourism photography by establishing a methodology for analysing image content based on modern computer vision techniques. By profiling each site according to the proportion of its Instagram images that contain given objects, we explore what these photographs can reveal about site characteristics and visitor activities that users find attractive and hence post about whilst experiencing these sites. We demonstrate the use of object detection as a way of “cataloguing” or “indexing” a large volume of photographs based on their content; this not only provides a summary of visual attributes across sites, but also facilitates retrieving images that reveal how visitors relate to and interact with heritage sites.

Our dataset comprises public Instagram posts scraped from the hashtag and location pages relevant to 26 unstaffed heritage sites of interest, totalling 54,621 posts published in May 2014–April 2019. To examine RQ2–RQ3, we use a subset of 3,979 posts for each of the top five sites (matching the number of scraped posts for the fifth most-posted site). To analyse these images' content, we conduct off-the-shelf inference using a pre-trained deep convolutional neural network-based object detection model. As the model's pre-training dataset excludes heritage site-relevant object categories, we implement transfer learning by fine-tuning the model on a set of 520 images of a specific site, the Rollright Stones, annotated with the object labels “sculpture” and “stone”.

Through our analysis, we find that:

- Bourdieu's idea that tourist photographs typically serve to honour the unique encounter between a person and a site with high symbolic yield (36) may concern only a small fraction of visitor photographs on Instagram.
- The results presented here support the claim by Galí and Donaire that ‘tourist photographs taken in western countries tend to avoid the presence of people’ in pursuing the ‘romantic ideal of tourism consumption’ (897).
- Differently from Robinson and Picard's proposition that vernacular tourist photography ‘makes no claims towards art’ (9), the evidence presented here indicates that tourist photography may rather ‘attempt to construct idealised images which beautify the object being photographed’, as put by Urry and Larsen (169).

This paper proceeds as follows: Section 2 locates this study's research questions and contribution within existing tourism research using social media data, then reviews developments in computer vision techniques and determines their suitability for this study. Section 3 theorises the practice of travel photography on Instagram to motivate and inform the study's research questions. Section 4 details methods for collecting and analysing Instagram data on unstaffed heritage sites. Section 5 presents the study's findings, which Section 6 discusses in relation to existing research before concluding.

2. Literature review

2.1 Understanding visitor engagement using social media data

Defined by Kaplan and Haenlein (61) as 'a group of Internet-based applications (...) that allow the creation and exchange of User Generated Content' (UGC), social media constitutes a 'mega trend' with significant impact on the tourism system (Leung et al. 3). UGC on social media platforms is a primary data source for investigating tourism consumption as it is cheaper and easier to access than government- or privately-owned transactional data (Li et al. 304). While much tourism research focuses on text-based UGC such as reviews and blog posts, this study contributes to existing work on image-based UGC.

The majority of existing studies analyse the images' metadata, comprising use-related, temporal, geographical, and textual information (Li et al. 308). In demonstrating correlations between metadata-derived metrics and official visitor statistics, these studies highlight the potential of using social media data as a proxy for tourist activity. By identifying international tourists amongst Flickr users in China based on their location of origin, Su et al. show that the number and profile of these users correlate with official statistics on international tourists in 2009–2013 (30). Latorre-Martínez et al. obtain similar results for Flickr users in Zaragoza. Wood et al. demonstrate that Flickr metadata can reliably proxy empirical annual visitation rates at 836 recreational sites across the world. Likewise, Sessions et al. show that monthly Flickr activity is a statistically significant predictor of official visitor counts for 38 national parks in the United States. Analysing Instagram, Flickr, and Twitter posts geolocated to 56 national parks in Finland and South Africa, Tenkanen et al. find that social media activity is highly associated with park popularity, with social media-derived monthly visitation patterns correlating relatively well with official statistics in 2014 (1).

Analyses of geographical and temporal metadata can reveal tourists' behaviour and preferences. Popescu et al. identify tourist sites, estimate visit times, and detect panoramic spots using Flickr metadata of 723,303 photographs taken during one-day visits across 183 cities. To identify tourist attractions in cities, Kisilevich et al. and Zhou et al. perform a spatial clustering analysis on geolocated images. By clustering images with respect to Flickr users' origin, Vu et al. compare destination preferences and travel trajectories amongst Asian

and Western tourists in Hong Kong. Su et al. similarly exploit Flickr metadata on users' origin to characterise the geographical preferences of international tourists visiting China. Besides identifying popular destinations, geolocation metadata can reveal sites that lack published photographs, as Farahani et al. and Motamed and Mahmoudi Farahani demonstrate for the cities of Shiraz and Melbourne respectively.

Overall, existing studies demonstrate that metadata from image-based UGC may contain valuable insight on tourist consumption patterns, particularly concerning where and when they visit. However, demographic biases in social media usage and whether visitors perceive the site as social media-worthy may lead to discrepancies with actual visitation patterns. While RQ1 examines the distribution of posts across sites and time to investigate visitation patterns amongst Instagram users, the lack of empirical visitor data on unstaffed heritage sites precludes considering how reliably Instagram metadata might proxy for tourist activity.

Prior tourism research that does examine the images themselves commonly uses manual content analysis, whereby researchers empirically quantify visual representation in images using reliable, explicitly defined categories (Bell 13). Also focusing on heritage tourism, Farahani et al.'s content analysis of 186 photographs of Nasir-al-Molk Mosque in Shiraz revealed the site's physical and spiritual qualities that contribute to its popularity amongst Flickr, 500px, and Instagram users. McMullen similarly examines 200 popular Pinterest photographs of four heritage tourist destinations in the US to understand what users find most interesting about these destinations. Content analysis of geolocated social media photographs can indicate tourists' perceived image of a city, as Motamed and Farahani and Galí and Donaire show for Melbourne and Barcelona respectively. Donaire et al. complement their content analysis of Flickr photographs of the Boí Valley with a cluster analysis to segment tourists based on their photographs' elements and angles of perspective.

While content analysis can thus provide insight into how tourists engage with a given destination, such manual categorisation imposes significant resource demands that limit the practicality of analysing large numbers of images (Balomenou et al. 174). At the upper end, Pearce et al.'s content analysis of 10,912 photographs from blog posts by Chinese tourists to interpret their visual representations of the Great Ocean Road in Australia required four months' worth of regular work by two full-time staff (28). Li et al. thus highlight the need in social media-related tourism research for more advanced analytic techniques applied directly on images (310). Notably, Rossi et al.'s image classification framework, comprising a scale-invariant feature transform (SIFT) descriptor and support vector machine (SVM) classifier, enables them to assign 90,000 Instagram images of Venice to one of six categories, whose frequency and geographical distribution reveal tourist consumption patterns. Analysing 238,290 Flickr images geolocated to Melbourne, Miah et al. use

speeded up robust features (SURF) to represent visual content, then employ kernel density estimation to identify representative images relevant to specific tourist attractions.

Nonetheless, deep convolutional neural network (DCNN)-based feature extractors have superseded SIFT- and SURF-based models across computer vision tasks. Whereas the latter use hand-engineered filters to extract features from images, DCNNs use trainable filters to learn image features directly and automatically from data with minimal domain knowledge (Liu et al. 7). As the next section details, this study extends previous research by employing DCNN-based techniques to examine what visitors to unstaffed heritage sites post photographs of. In particular, object detection enables extracting valuable details per image across this study's sample of 19,895 Instagram images across five sites.

2.2. Object detection

Object detection is a computer vision task which aims 'to determine whether or not there are any instances of objects from the given categories (such as humans, cars, bicycles, dogs or cats) in some given image and, if present, to return the spatial location and extent of each object instance (e.g. via a bounding box)' (Liu et al. 1). Past studies using DCNNs have recorded state-of-the-art performance on object detection, as well as in other computer vision tasks such as scene recognition, fine-grained recognition, and domain adaptation (Donahue et al.) object and action classification (Oquab et al.), scene recognition, fine-grained recognition, attribute detection, image retrieval (Sharif Razavian et al.), and unsupervised image clustering (Guérin et al.). Relative to handcrafted features, DCNN-based features provide a more powerful, discriminative representation of images (Liu et al. 193).

Off-the-shelf DCNN-based inference is restricted to the target object categories a given model was trained on. Consequently, objects more specific to heritage sites (e.g. "arch", "ruin", "stone") remain undetected. As training requires large amounts of labelled data and is computationally expensive, *transfer learning* is used to transfer knowledge from models trained for a given source domain and learning task to a different target domain and learning task (Pan and Yang). Researchers have used this approach effectively for object detection on images of X-rayed baggage (Akçay et al.), traffic signs (Arcos-García et al.), and packed food products in a refrigerator (Talukdar et al.). Similarly, we fine-tune a pre-trained detector to improve its performance on images of a specific heritage site.

Therefore, this study extends existing tourism research on image-based UGC by demonstrating how DCNN-based object detection techniques can be used to understand visitor cultural behaviour at unstaffed heritage sites. In addition to revealing what visitors are taking and sharing photographs of, object detection can identify photographs depicting interactions of visitors

themselves with the site for further qualitative analysis. The next section contextualises and motivates these research questions by theorising the process of taking then sharing travel photographs on Instagram.

3. Theoretical background

To motivate and inform this study's research questions around the practice of tourist photography on Instagram, we draw on the concepts of the tourist gaze and self-presentation as central to both the 'first act' of taking photographs and the 'second act' of sharing photographs online (Nov and Ye).

3.1. The tourist gaze: How do visitors look at the site?

Urry and Larsen argue that tourism experiences are 'fundamentally visual' in nature: the 'tourist gaze' organises visitors' encounters with the 'other' to provide some sense of competence, pleasure, and structure to these experiences (14). This 'gaze' emerges from the twin birth of mass tourism and photographic techniques in the mid-19th century, from which time photography and tourism have been intrinsically linked. 'Photography is what one does on holiday, and also what makes a holiday' (Bourdieu and Whiteside 36). Photographs summarise the tourist's perceived destination image (Donaire et al. 27), and can serve as rich datasets revealing tourists' preferences and behaviours (Balomenou et al. 173). In exploiting this opportunity, this study recognises that the 'hermeneutic circle of representation' (Urry 140) and the selective framing of photographs may shape how Instagram photographs depict the tourist gaze.

Although commonly seen as providing a faithful reproduction of reality (Bourdieu and Whiteside 93), photography is an active signifying practice that attempts to construct idealised images aestheticizing its subject Urry and Larsen. Tourists' selectivity in taking and retaining photographs creates a more polished and positive set of evidences than does the experience itself Belk and Hsiu-yen Yeh. This study's particular view filtered through Instagram compounds such distortion from the camera lens as the platform's 'social currency' of likes and comments fosters 'an online culture where the primary motive is to impress rather than just inform' (Jacob 261). Even as photography may render travel into 'a strategy for accumulating photographs' (Sontag 9), Instagram adds the goal of capturing Instagram-worthy moments (Jacob 262).

Recognising that photography is thus central to the tourist gaze, RQ2 explores visitors' gaze of unstaffed heritage sites by analysing what they take and share photographs of.

3.2. Self-presentation: How do visitors look?

Tourist photography is bound up with self-presentation and 'strategic impression management' (Goffman; Larsen 424). Photographs enable tourists to create 'extensions of self in place, simultaneously capturing moments as lived and securing projected memories' (Scarles 471). In selectively taking and

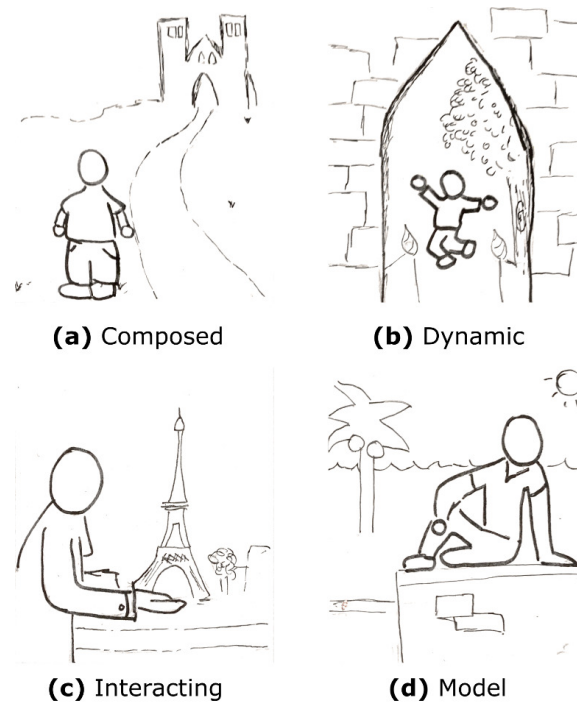


Figure 1. Four categories of posing for solo tourist photographs, adapted from Pearce and Wang (116).

The authors also include “bland”, “projecting”, “cute”, and “costume” categories.

retaining photographs, tourists collect illustrations to construct a self-narrative (Belk and Hsiu-yen Yeh 349). Photographs thus not only mark what is significant to the tourist, but also form a conscious attempt at fashioning one's self-image (Manovich).

Conceptualising the tourist as ‘playful performer and cultural producer’ Stylianou-Lambert further highlights how photography facilitates the construction of self-identities and socialities (1821). Through photography, the landscape becomes a theatrical stage in which tourists are embodied, expressive subjects enacting choreographed and experimental performances (Larsen). Pearce and Wang's ethological study of the postures of tourists at iconic tourism sites emphasises the performative nature of tourist photography. The authors derive categories of tourist poses in solo photographs, including “composed”, “dynamic”, “interacting”, and “model” poses (Fig. 1).

The concept of self-presentation thus highlights the curated, performed nature of the Instagram photographs under study. Although this study's visual analysis of these photographs casts them in a static frame, understanding their place within dynamic, social processes provides contextual knowledge for interpreting analytical results. RQ2 recognises that self-presentation motives influence what objects visitors choose to photograph and display on Instagram. By bringing both the posting of photographs, as well as the taking of them, into the analytical frame, this offers further insights into cultural behaviour around heritage sites. Moreover, besides analysing what the images contain, we consider how often visitors' photographs present themselves together with a salient site feature to determine the extent to which photographs serve as

joint signifiers of “I am here; here is the place that I am.” Subsequently, RQ3 explores the performative, embodied nature of tourist photography by applying Pearce and Wang’s typology of poses to such photographs depicting both people and site-specific features. Furthermore, we extend the notion of performativity to include visitors’ ‘extended selves’ (Belk) in the form of personal possessions and pets by considering images depicting visitors’ bicycles and dogs alongside salient site features.

4. Methodology

This section details this study’s methods of collecting and analysing Instagram data on unstaffed heritage sites.

4.1. *Instagram data*

This study’s primary data source comprises public Instagram posts on unstaffed heritage sites managed by English Heritage. We use Instagram as it is the most actively used image-based social media platform in the United Kingdom (UK): We are Flint’s (39) survey of 2,008 Internet users aged over 18 in the UK finds that 41% of respondents use Instagram, whereas 36% use Pinterest.

4.2. *Webscraping*

We select 26 unstaffed heritage sites of key interest as determined in consultation with English Heritage. For each site, we manually search Instagram to determine its relevant hashtag and location pages, which we then scrape using Instaphyte.¹ Appendix A of the Supplementary Information lists all scraped pages.

Besides downloading images from the given hashtag or location page, Instaphyte generates a CSV file containing the images’ metadata. These metadata include each post’s unique code, Unix timestamp, caption, user ID, number of likes, number of comments, and image URL. For posts containing multiple images, Instaphyte retrieves the first image only.

Since a post may include both hashtag and location information (and thus be scraped twice), we keep only unique posts as identified by their code. We narrow this sample to posts published in the five-year period from 1 May 2014 (which comes after the earliest post for 18 sites) to 30 April 2019 (which precedes the last post for 23 sites). This yields 54,621 posts across the 26 sites, ranging from three (Roman Wall of St Albans) to 12,087 (Castlerigg Stone Circle) posts per site.

This study’s sample is representative only of visitors who use public Instagram accounts and who add the site’s relevant hashtag and/or location to their posts. Gender and age biases exist amongst Instagram users: We are Flint’s (39) survey

¹ Available from <https://github.com/ScriptSmith/instaphyte> (accessed 23 February 2022).

of UK Internet users finds that more women (48%) than men (35%) report using Instagram, and that usage is dominant amongst those aged 18–34. Furthermore, both the practice of sharing photographs online and the decision to do so publicly may be linked to visitors' demographic and motivational characteristics. Lo et al.'s survey of Hong Kong residents indicates that those who post travel photographs online tend to be younger, better educated, and earn a higher income than those who do not. Surveying Flickr users, Nov and Ye find that those expressing greater commitment to the online community and having a higher number of contacts tend to share more photographs publicly. Additionally for Instagram, the decision to share posts publicly may be an unconscious one since posts are public by default, unless users make their account private so that only approved followers can view their posts ("Controlling Your Visibility"). Therefore, this study's findings concern a specific subset of visitors and cannot be generalised to all who visit the heritage sites under study.

Following the University of Oxford's Central University Research Ethics Committee best practice guidance for Internet-based research, users' specific consent is not required when collecting their publicly available Instagram posts. When displaying these data, we de-identify images to protect users.

4.3. Deriving posting activity from the metadata

To analyse these posts' distribution across sites and time (RQ1), we narrow the sample to the top 10 sites by number of posts, totalling 45,286 posts. We operationalise site popularity as the number of posts per site over the five-year period under study. As a robustness check, we derive the number of photograph user-days (PUD) per site. PUD counts each unique user once per day, i.e. a user uploading multiple posts per day is only counted once. PUD thus provides a more user-centric operationalisation of site popularity than post counts by adjusting for potential cases where a single user (or bot) generates a large fraction of posts. Past studies show that PUD derived from social media metadata can provide a reliable estimate of official visitor statistics (Wood et al.; Sessions et al.; Tenkanen et al.). To consider temporal patterns in posting activity, we convert each post's Unix timestamp to human-readable local time, assuming the London time zone.

4.4. Object detection

This study applies object detection to Instagram images of the top five sites to characterise what visitors post photographs of RQ2 and to identify relevant images for exploring visitors' presented interactions at these sites (RQ3). As the fifth most-posted site has 3,979 scraped posts, we randomly sample the same

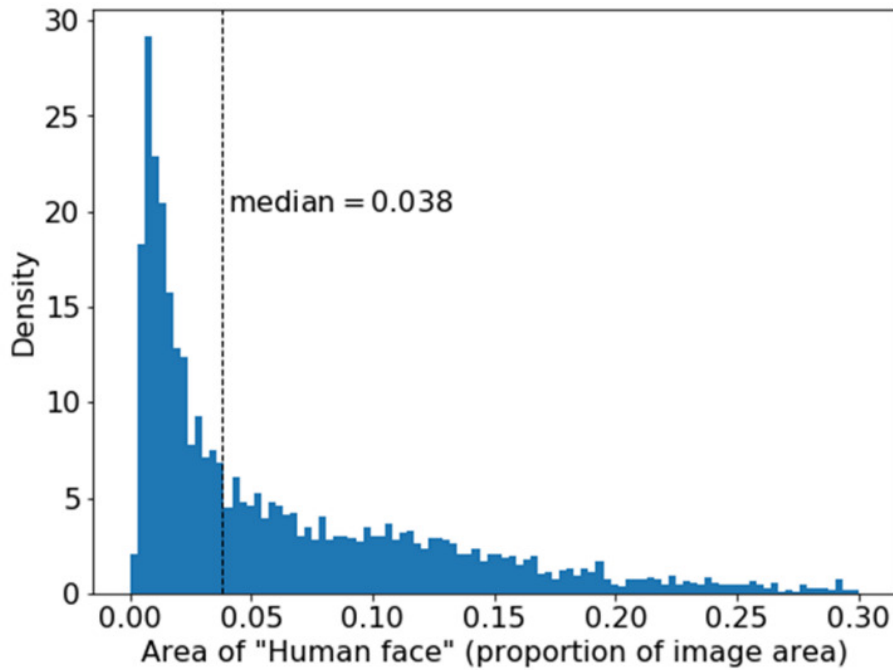


Figure 2. Distribution of areas of “human face” bounding boxes in the dataset. Horizontal axis restricted to [0.00, 0.30] for clarity (maximum value is 0.82).

amount from the top four sites, yielding 19,895 images. We use the Tensorflow Object Detection API² (Huang et al.) to implement off-the-shelf inference and transfer learning.

An object detection model predicts whether an image contains any instances of objects from predefined categories. For each object detected, the model returns its predicted location (bounding box), category, and confidence level. The most commonly used metric to evaluate model performance is Average Precision (AP), whose mean (mAP) over all object categories is used to compare model performance (Liu et al. 9).

OFF-THE-SHELF INFERENCE

For off-the-shelf inference, we use a Faster R-CNN detector with Inception ResNet V2 feature extractor pre-trained on Open Images V4, available from the TensorFlow detection model zoo.³ We select this model as its Open Images mAP of 54 is the highest amongst all available models. Since this study analyses a static dataset (rather than continuous stream) of images without requiring real-time inference, we prioritise model performance over speed.

Initial results from off-the-shelf inference include 600 possible object categories per Open Images V4, which the model was trained on. We retain detections with over 50% confidence and select 12 categories of interest,

² . Available from https://github.com/tensorflow/models/tree/master/research/object_detection (accessed 22 February 2022).

³ Available from https://github.com/tensorflow/models/blob/master/research/object_detection/g3doc/detection_model_zoo.md (accessed 22 February 2022) as faster_rcnn_inception_resnet_v2_atrous_oidv4.

namely: “person”, “human face”, “tree”, “castle”, “building”, “house”, “tower”, “sculpture”, “flower”, “bird”, “dog”, and “bicycle”. Additionally, we incorporate “boy”, “girl”, “man”, and “woman” into the pre-existing category “person”; “blue jay”, “canary”, “duck”, “eagle”, “falcon”, “goose”, “owl”, “raven”, “sparrow”, and “swan” into “bird”; “bronze sculpture” into “sculpture”; and “rose” into “flower”.

We derive the category “selfie” from “human face” detections covering at least 3.8% of the image area, then discard “human face” detections otherwise (as these would typically be superfluous to an existing “person” detection). This threshold corresponds to the median area of “human face” bounding boxes in the dataset. These bounding-box areas have a skewed distribution (Fig. 2): after a large peak of small-area detections, there is a long tail of larger areas that more likely convey users’ ‘desire to frame the self in a picture taken to be shared with an online audience’, that is, a selfie (Dinhopl and Gretzel 130).

However, as noted above, off-the-shelf inference using the pre-trained detector does produce false positives (incorrect detections) and false negatives (undetected ground truths). Given a lack of annotated ground-truth images, evaluating the detector’s mAP on this study’s dataset is beyond the current scope. Nonetheless, we conduct transfer learning to address the problem of false negatives due to heritage site-relevant object categories being absent from the pre-trained model’s training dataset.

TRANSFER LEARNING

Taking the Rollright Stones site as a case study, we fine-tune a Faster R-CNN detector pretrained on the MS COCO dataset to obtain more informative results concerning RQ2–RQ3. We select the Rollright Stones for two main reasons. Firstly, models pre-trained on Open Images or MS COCO cannot detect “stone” since both datasets exclude this category. Although Open Images includes the category “sculpture”, woven-wood sculptures at the Rollright Stones do not resemble sculptures in Open Images that typically are carved or sculpted from clay, metal, stone, or wood.⁴ Secondly, the site has the fifth-highest number of scraped images, thus providing adequate training data.

From the 3,979 images of the Rollright Stones, we randomly select 1,000 images for annotation. We evenly distribute these images between four authors for annotation with “stone” and “sculpture” as target object categories. Annotation yields 650 images containing at least one sculpture or stone, with 119 sculptures and 1,889 stones in total. We randomly split these with an 80:20 ratio into a training set of 520 images (97 sculptures, 1,532 stones) and a testing set of 130 images (22 sculptures, 357 stones).

⁴ Available from <https://storage.googleapis.com/openimages/web/visualizer/index.html> (accessed 23 February 2022).

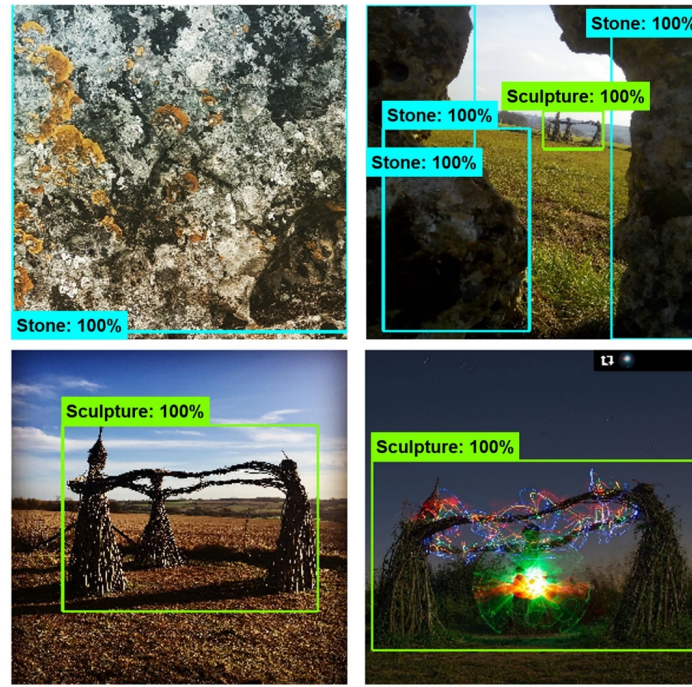


Figure 3. Detections of “sculpture” and “stone” using the fine-tuned model on Rollright Stones images.

Using the training set, we fine-tune a Faster R-CNN detector with Inception V2 feature extractor pre-trained on the MS COCO dataset. Relative to other COCO-trained models available from the TensorFlow detection model zoo, this model provides a good trade-off between performance (COCO mAP of 28 reported) and speed (58 ms per image)—Appendix C of the Supplementary Information provides more details. The fine-tuned model records a mAP of 60.85 on the testing set, which is the average of its APs of 76.95 for “sculpture” and 44.74 for “stone” (calculated at an intersection over union threshold > 0.5). Results for the fine-tuned model are shown in [Fig. 3](#).

To address RQ2–RQ3, we incorporate results from both manual annotation and transfer learning. Since manual annotation only covers 1,000 images, we use the fine-tuned model to obtain predictions on the remaining 2,979 images. This provides “sculpture” and “stone” detections on all 3,979 images of the Rollright Stones, though the model may tend to overestimate the occurrence of “stone”: proportionally, “stone” occurs in 54.7% of the manually-annotated images, but is detected in 70.1% of the remaining images by the fine-tuned model. We combine these detections with those from off-the-shelf inference for each image. This yields a more comprehensive account of what the images contain for RQ2, and enables identifying images in which people (detected via off-the-shelf inference) co-occur with “sculpture” or “stone” (detected via manual annotation or transfer learning) for RQ3.

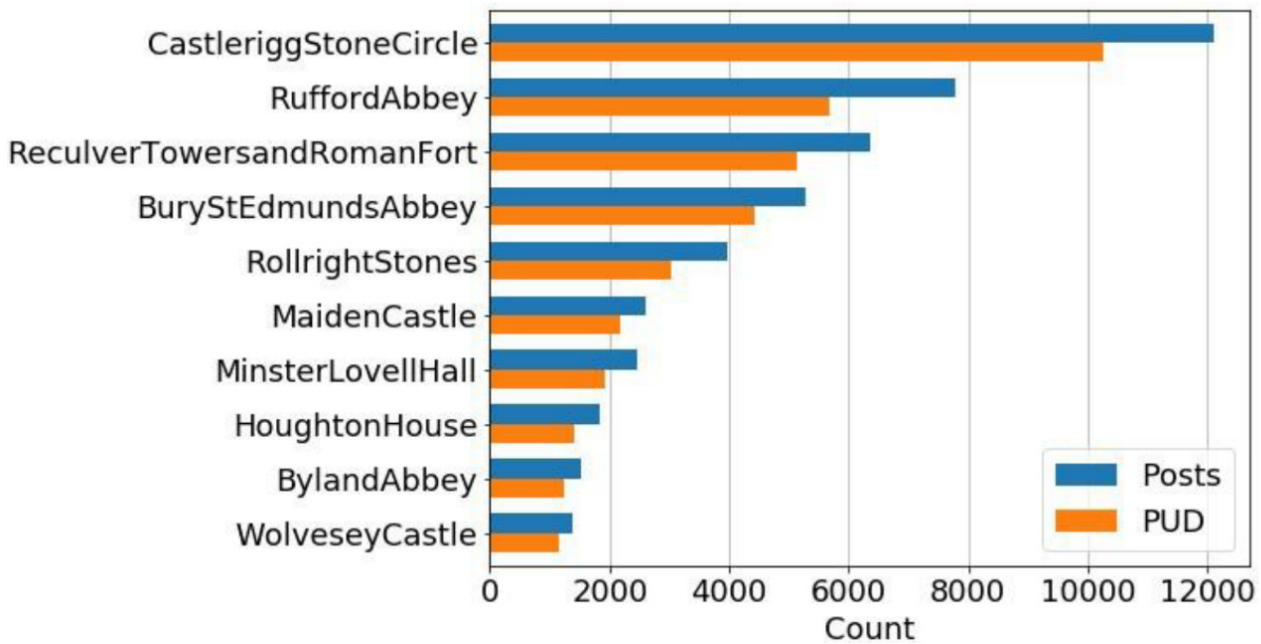


Figure 4. Top 10 most-posted sites in May 2014–April 2019, shown with total number of posts and photograph user-days (PUD).

PUD counts each unique user once per day.

5. Results

5.1. Posting patterns across sites and time

[Figure 4](#) shows the top 10 sites by number of posts and photograph user-days (PUD) in the five-year period under study (Appendix B of the Supplementary Information presents results on all sites scraped). Both metrics agree in ranking the top 10 sites amongst Instagram users. Castlerigg Stone Circle has the highest number of posts by far, at 12,087 or 26.7% of the 45,286 posts in the top 10. Rufford Abbey is a distant second at 7,786 posts (17.2% in the top 10), after which the difference between successive sites becomes less stark: Rufford Abbey is more closely followed by Reculver Towers and Roman Fort (6,353 or 14.0%), Bury St Edmunds Abbey (5,279 or 11.7%), and the Rollright Stones (3,979 or 8.8%). The last five sites together account for 9,802 or 21.7% of posts in the top 10. Overall, these results indicate that the present sample of 26 unstaffed sites comprises a small number of very popular sites, and a long tail of sites less popular amongst Instagram users.

All top 10 sites display a clear concentration of posts on Saturdays, Sundays, and Mondays than other days of the week ([Fig. 5a](#)). This supports the notion of having visitors having leisurely “days out” to these sites on weekends. Users mostly publish their posts between 4 pm and 9 pm ([Fig. 5b](#)), possibly indicating that Instagram photographs serve not only as immediate markers of “I am here, right now,” but also to share memories of “I was there.”

The distribution of Instagram posts across sites and time allows us to answer RQ1. We find that user interest concentrates on a few key sites, revealing not only expected peaks on weekends, but also deviations from this norm possibly

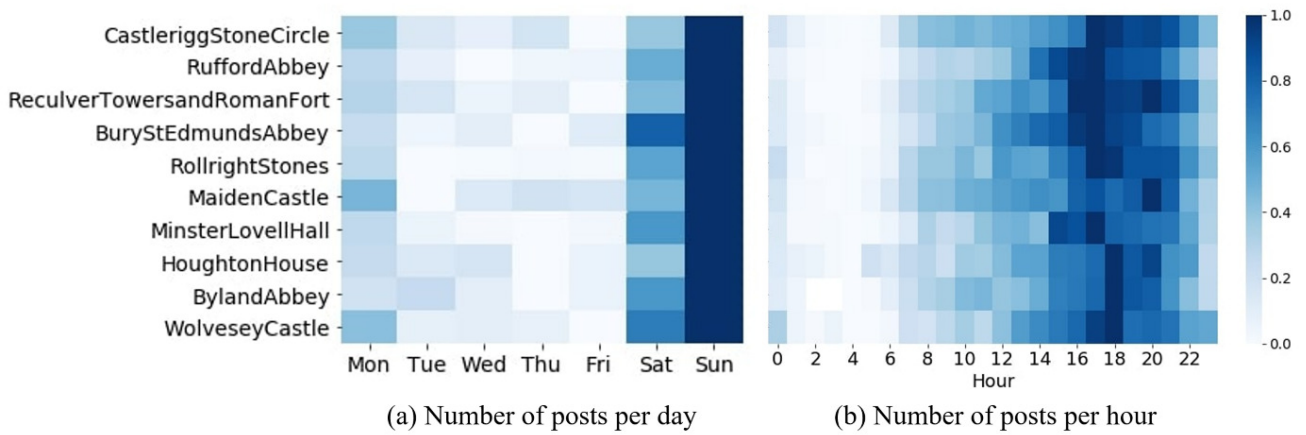


Figure 5. Daily and hourly posting patterns across the top 10 sites. Hours in 24-hour format, indicating posts with timestamps within the hour. Values indicate number of posts in May 2014–April 2019, normalised to $[0, 1]$ for each site.

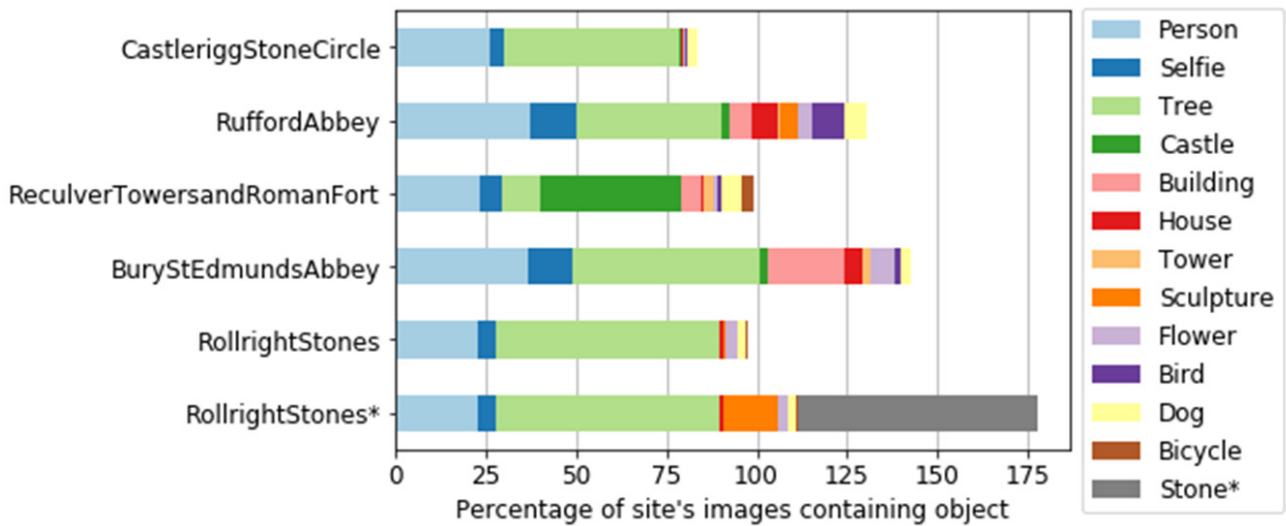


Figure 6. Percentage of images containing a given object across the top five sites, out of 3,979 images per site.

Objects detected with at least 50% confidence using a Faster R-CNN detector pre-trained on Open Images V4. Each site's cumulative percentage does not equal 100% since images can contain multiple categories or none. The bottom row describes the object found in the Rollright Stones after applying transfer learning to also detect sculptures and stones.

due to on-site events. Naturally, these results concern a particular subset of visitors, namely those uploading tagged images onto their public Instagram accounts.

5.2. What visitors take photographs of

[Figure 6](#) characterises each of the top five sites by depicting the percentage of its images containing a given object, out of a sample of 3,979 images per site.

Overall, object detection recovers the main features of every site. Generally across all sites, the object categories most commonly detected using off-the-shelf inference are trees and people, reflecting the sites' outdoor setting and the presence of visitors. Although object detection cannot infer whether the people photographed are users themselves or other visitors, it provides a means

of retrieving these images for further analysis. Nonetheless, only between 22.7% (the Rollright Stones) and 37.6% (Rufford Abbey) of images depict people, indicating that the majority of photographs exclude human presence.⁵

Besides considering the object categories that do appear in images, it is instructive to note that not all images contain instances of the target object categories (see Appendix B, Supplementary Information, for example images). Castlerigg Stone Circle and the Rollright Stones have 28.7% and 19.9% of images in which no objects were detected, partly due to the inability of off-the-shelf inference to identify the sites' stones as an object category. Transfer learning on the Rollright Stones images crucially supplements off-the-shelf inference by detecting the site's stones and woven-tree sculptures: "stone" and "sculpture" respectively feature in 66.2% and 15.0% of the site's images, decreasing the proportion of images without detected objects to 5.8%. This demonstrates that visitors not only view the site's prehistoric stones as its distinctive feature, but also find attractive the woven-tree sculptures added in 2017 ("The Three Fairies Sculpture"). This improvement, however, may include false-positive detections. Reculver Towers and Roman Fort's proportion of images lacking detected objects is relatively high at 26%—these include photographs of the site's coastal surroundings and sunset views (Fig. A5 in Appendix B, Supplementary Information). Without target categories that suitably capture these site characteristics, object detection thus overlooks information on visitor behaviour contained in these images.

Given that tourist photography is bound up with self-presentation and performativity, [Fig. 7](#) explores how often photographs serve as joint signifiers of person-and-place. For clarity, we combine "building", "castle", "house", and "tower" into the category "structure". As shown for Rufford Abbey, Reculver Towers and Roman Fort, and Bury St Edmunds Abbey, people and structures are depicted more often in isolation from each other than in the same image. Reculver Towers and Roman Fort has the highest overlap ratio (intersection over union, IOU) between "person" and "structure" images at 6.8%, i.e. of all images containing either or both of "person" and "structure", only 6.8% have both categories co-occurring. Bury St Edmunds Abbey and Rufford Abbey have smaller IOUs at 4.3% and 2.2% respectively.

Using object detection to retrieve images depicting "structure" without "person" can reveal how visitor photography aestheticizes the site. Taking Rufford Abbey and Bury St Edmunds Abbey as examples, [Fig. 8](#) shows that such images may thus focus more on portraying the site's architectural qualities rather than visitors themselves. Besides the buildings' iconic exteriors, these images highlight details including archways, ceilings, doorways, and windows that visitors find Instagram-worthy in experiencing the site.

⁵ Taking "selfie" into account only marginally increases these proportions to 23.1% and 39.0% respectively, as "selfie" tends to co-occur with "person".

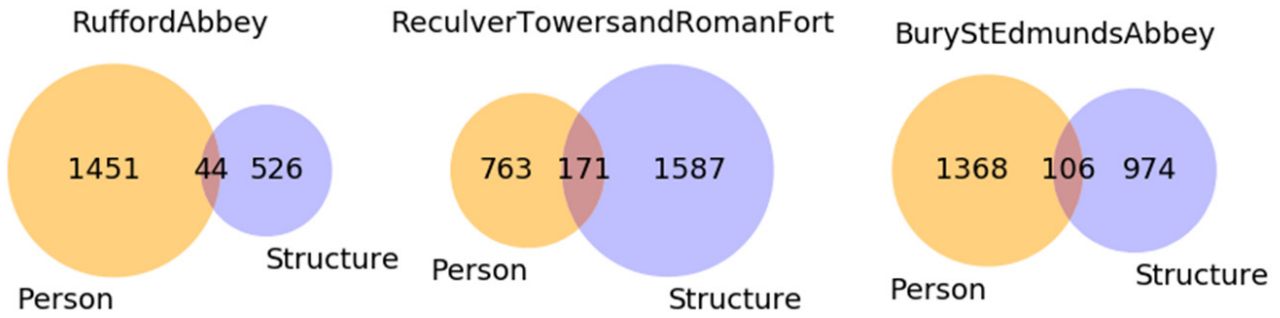


Figure 7. Venn diagrams for co-occurrences of “person” and “structure” (combining “building”, “castle”, “house”, and “tower”).

Numbers are image counts in each subset, e.g. Rufford Abbey has 1,451 images depicting “person” without “structure”.



Figure 8. Images depicting “building” without “person”.

Answering RQ2, results from object detection show not only what but also how often visitors include particular object categories in their Instagram photographs. Characterising sites by the proportion of images containing given object categories both profiles sites in terms of the visual characteristics that attract visitors, and reveals common visitor behaviour at these sites. For instance, visitors to Rufford Abbey and Bury St Edmunds Abbey interact with and photograph both the built and natural environments at these sites. Both sites also have the highest proportion of images with people and selfies. Reculver Towers and Roman Fort is most popular for depicting cycling and dog-walking amongst Instagram users. Considering co-occurrence patterns, only a small fraction of images containing either or both of people and structures depict both categories co-occurring.

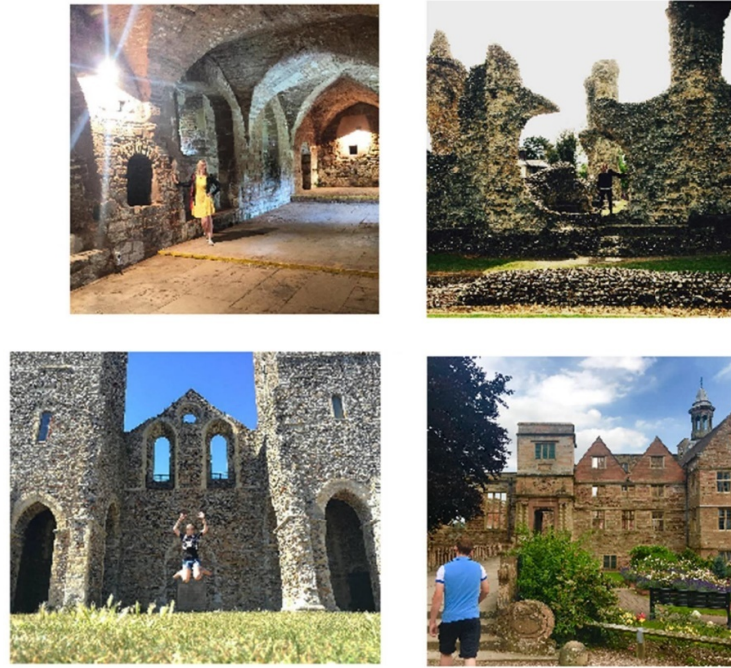


Figure 9. Images depicting visitors adopting “composed” poses (top left), “interacting” poses (top right), “dynamic poses” (bottom left), and facing away from the camera (bottom right).

5.3. *Visitors’ performativity in photographs*

To investigate how visitors perform and depict their interactions with heritage sites, we use object detection to identify images in which people and their extended selves co-occur with salient site features, such as represented by the overlapping regions in [Fig. 7](#). While a content analysis of all images retrieved is beyond this study’s scope, we present examples that explore and illustrate visitors’ performativity in their Instagram photographs.

Drawing on Pearce and Wang’s (116) typology of tourists’ poses in solo photographs, which includes “composed”, “dynamic”, “interacting”, and “model” poses ([Fig. 1](#)), from the images in which “person” co-occurs with “building” or “house”, we find visitors adopting “composed” poses in which they appear relaxed and may lean against a balustrade or wall, sometimes using the site as an artistic stage upon which individuals enact and showcase their special occasion with “model” poses. Some photographs exhibit more playful performances as visitors adopt “interacting” and “dynamic” poses by climbing, sitting, jumping and using dramatic lighting effects to draw more attention to themselves, or even facing away from the camera and walking towards the site, as shown on [Fig. 9](#). The latter images do not conform to Pearce and Wang’s proposed categories, but we suggest they connote a sense of “exploring”, as if visitors are inviting their online audience to “come along” on their journeys at the site.

[Fig. 10a](#) shows that sculptures at Rufford Abbey encourage playful behaviour amongst visitors, such as mimicking the sculpture’s bodily pose or facial expression, incorporating the sculpture as if it were another person in a group



Figure 10. Images with “person” and “sculpture”, showing visitors’ “interacting” and “composed” poses at the sites.

pose, and using the sculpture as a photographic frame encircling the person. These “interacting” poses suggest that the sculptures’ size and playful appearance encourage direct and playful interaction in visitors’ photographic performances. Similarly, images with “person” and “sculpture” at the Rollright Stones exhibit “interacting” poses where visitors “dance” with the Three Fairies Dancing Sculpture and use the woven-tree archway as photographic frame (Fig. 10b). “Composed” poses include visitors in costume, who thereby engage with the folklore and legend associated with the site.

Visitors’ performances of their extended selves also include photographing their bicycles or pet dogs alongside site features. As Reculver Towers and Roman Fort has the highest proportion of images with bicycles (Fig. 6), Fig. 11a illustrates how Instagram users perform as both cyclist and tourist by positioning their bicycle as central prop on the stage of the site’s iconic “castle”. Likewise, images where dogs and key site features co-occur exhibit users’ self-presentation as heritage site visitor and dog-owner (Fig. 11b). Visitors frame their dogs in a similar way to how people are depicted at these sites, even including portrait-style shots and deliberate playfulness in setting their pets atop the sites’ ledges or stones.

To address RQ3, object detection provides a useful means of retrieving images in which people and salient site features co-occur. Examples depicting people co-occurring with built structures show that besides “composed” poses against iconic views of the site, visitors may employ framing and posing to direct the viewer attention towards their photographic performances with the site as backdrop. Smaller-sized sculptures such as at Rufford Abbey and the Rollright Stones tend to evoke more playful, “interacting” poses amongst visitors.



(a) “Castle” and “bicycle”

(b) “Castle” and “dog”

Figure 11. Images from Reculver Towers and Roman Fort (a) with “castle” and “bicycle” or (b) with “castle” and “dog”.

Considering visitors’ presentation of their extended selves, photographs depicting bicycles or dogs alongside site features demonstrate an interplay between capturing visitors’ idealised view of the site and including their own experience and identity as cyclists or dog-owners.

6. Discussion and conclusion

This study aimed to improve our understanding of cultural behaviour in unstaffed heritage sites by analysing publicly available Instagram data. We focused on what the distribution of posts, considered across sites and across time, reveals about site popularity and temporal patterns in visitor engagement (RQ1), what visitors are taking and sharing photographs of across different sites (RQ2), and how visitors perform and depict their interactions with heritage sites through their photographs (RQ3).

We scraped the hashtags and location pages relevant to 26 unstaffed heritage sites of interest, obtaining 54,621 posts published in May 2014–April 2019. Concerning RQ1, ranking the popularity of sites by number of posts (with a robustness check based on the number of photograph user-days, i.e. counting each unique user once per day) reveals that posting activity is concentrated in a small fraction of very popular sites, followed by a long tail of sites that have received less attention from Instagram users. On a larger geographical and temporal scale, this result mirrors Farahani et al.’s finding that a small fraction of heritage sites accounts for the majority of 186 social media images geo-tagged within the historic city of Shiraz in 2015 (205).

To address RQ2–RQ3, we used pre-trained DCNN-based object detection models for off-the-shelf inference and transfer learning to analyse 19,895 images across the top five sites. Compared to previous studies using manual content analysis on tourist photographs (Donaire et al.; Galí and Donaire; Pearce et al.), this study's method is less labour-intensive and scales more easily to larger datasets.

We demonstrated three main findings regarding RQ2. Firstly, comparing the proportion of images depicting object categories of interest within and between sites can reveal site characteristics and visitor activities more associated with visitor engagement and photography. Secondly, only a minority of photographs depict people, out of which a small fraction depict people along with sites' built structures. And thirdly, images depicting built structures in the absence of people reveals architectural details that visitors find attractive and picture-worthy.

Amongst the top five sites, images of Rufford Abbey and Bury St Edmunds Abbey have the widest variety of visual content. Besides the sites' built structures, their natural surroundings within Rufford Abbey Country Park and Bury St Edmunds Abbey Gardens include flowers and birds that visitors depict in Instagram images tagged to the sites. Reculver Towers and Roman Fort especially attracts cyclists and dog-owners. Woven-tree sculptures at the Rollright Stones often feature in the site's images, alongside its prehistoric stones. In a similar vein, Rossi et al. characterise tourism consumption in Venice by classifying Instagram photographs to one of six predefined categories using traditional computer vision techniques (based on handcrafted features). While we likewise demonstrate that Instagram photographs can provide insight into how tourists engage with heritage sites, using object detection methods enables the consideration of multiple potential object categories per image rather than a single category, and facilitates a comparative perspective across different sites. This approach should also be beneficial to audience segmentation tools, as different objects might help identify audience “types”, depending on people's interest in heritage, cultural activities, or days out.

We found that the majority of Instagram images across the top five sites exclude people. This accords with previous studies that also consider the degree of human presence in tourist photographs, collected via visitor-employed photography (Garrod) or social media (Donaire et al.; Galí and Donaire). Our findings support Galí and Donaire's claim that ‘tourist photographs taken in western countries tend to avoid the presence of people’ in pursuing the ‘romantic ideal of tourism consumption’ (897).

To interrogate the notion of tourist photographs serving as joint signifiers of “I (person) am here; here (structure) is the place that I am,” we measured the cooccurrence between people and built structures in images of three sites. While the emphasis on “image as evidence” in tourist photography is well-documented (Sontag 6; Urry and Larsen 179; Jacob 262), this study

contributes new insight by showing that images where people depict themselves with heritage sites' built structures only form a small share of such evidence on Instagram: of all images containing either or both of people and structures in three sites under study, only up to 6.8% depict both object categories. Therefore, Bourdieu's suggestion that tourist photographs typically serve to consecrate the unique encounter between a person and a site with high symbolic yield (36) may concern only a small fraction of visitor photographs on Instagram. We further posit that Instagram's role as curator of users' exhibition spaces online may reduce visitors' felt need to include themselves in framing their photographs since the platform automatically associates each post with the respective user's account (Hogan).

Images depicting the sites' built structures in the absence of people include examples focusing on architectural details that visitors find attractive and hence choose to curate and display for themselves and their imagined audience. Contrary to Robinson and Picard's suggestion that vernacular tourist photography 'makes no claims towards art' (9), these examples show that tourist photography may rather 'attempt to construct idealised images which beautify the object being photographed' (Urry and Larsen 169). Despite not depicting the visitors themselves, these images affirm the notion of self-presentation in online tourist photography, whereby users seek 'to capture images that could make a place more appealing to others because of their exceptional photographic eyes and perspectives' (Lo and McKercher 111).

Answering RQ3, we found that visitors display a range of poses in enacting performances against the backdrop of salient site features. We employed Pearce and Wang's categorisation of tourist poses to characterise these performances. At Rufford Abbey, Reculver Towers and Roman Fort, and Bury St Edmunds Abbey, visitors commonly adopt "composed" poses in front of the sites' built structures. These photographs resemble Stylianou-Lambert's description of online tourist photographs that serve as a proof of 'being there', whereby the act of posing in front of a landmark follows specific conventions of 'frontality, eye-level shooting, smiling, posing, and letting the landmark show' (1830). Crang notes that in turning their backs on the site to face the camera, visitors separate themselves from their present experience as they perform for an imagined audience elsewhere (366–367). While the examples observed in this study mainly conform to this pattern and to Pearce and Wang's categorisation, a few exceptions depicted visitors facing away from the camera and towards the site. We suggest that these exceptions connote an "exploring" pose of visitors inviting their imagined audience to journey with them.

Sculptures at Rufford Abbey and the Rollright Stones evoke "interacting" poses as visitors engaged with the sculpture by mimicking or touching it. These performances convey a playful attitude and sense of ownership (Stylianou-Lambert 1830), and demonstrate the holiday snap's conscious celebration of disjuncture with normal work-related behaviours (Robinson and Picard 6).

Whereas the literature on performativity in tourist photography mainly focuses on visitors' embodied actions, we expanded this notion to include visitors' 'extended selves' (Belk) of bicycles and dogs. Examples for both cases reflect an interplay between capturing an idealised view of the site and personalising the site, as visitors simultaneously present their identities as visitor and cyclist (or dog owner). Despite excluding visitors themselves, these examples demonstrate Instagram users' selectivity in framing and sharing photographs so as to reflect their desired self-image (Lo and McKercher).

In sum, this study has shown that analysing publicly available Instagram data can improve our understanding of tourist behaviour at unstaffed heritage sites by gauging relative temporal patterns in posting activity, revealing site features and visitor activities that are commonly photographed across sites, and illustrating how visitors present themselves through photographic performances against the backdrop of salient site features. Since these findings concern Instagram users with public accounts who include the site's relevant hashtag and/or location in their posts, they are not representative of visitors in general.

This study concedes several limitations, which also suggest avenues for further research. Firstly, our analysis neglected image captions. As the online equivalent of visitors' spoken commentaries accompanying their recollections around traditional photographic albums (Robinson and Picard 14), the caption crucially states the photograph's signifying intention (Bourdieu and Whiteside 92). Further research using natural language processing alongside computer vision techniques may recover some of this lost context, thus providing a fuller understanding of how tourists engage with heritage sites in terms of Instagram users' own motives and meanings ascribed to their posts.

Secondly, results from object detection included false-positive and false-negative predictions. Although we evaluated the fine-tuned model's performance on a testing dataset, the lack of annotated data precluded evaluating both the off-the-shelf detector and the fine-tuned model on testing sets representing all sites. Moreover, the restriction of object detectors to their target categories inevitably overlooks other visual attributes of interest, such as a site's coastal or countryside surroundings. Overall, it would be useful to explore different training configurations and target categories, consider transfer learning between different target datasets to reduce the amount of annotation required (e.g. using this study's fine-tuned model to detect stones in images of Castlerigg Stone Circle), and complement these results using methods for holistic scene understanding (Xiao et al.).

Thirdly, our analysis of visitors' performativity in their photographs drew upon illustrative examples that may not be representative of all images in which people and salient site features co-occur. Nonetheless, this study's

methodology for readily identifying subsets of such images from large datasets provides a useful starting point for a more systematic content analysis of these images.

Finally, this study's analysis of Instagram data is removed from the physical and social contexts creating these posts. Consequently, it cannot infer the motivational and circumstantial reasons behind posting activity and image content, and might overlook ways in which visitors interact with the site which might not appear as photos on Instagram. Future research can employ participant observation and qualitative interviewing of visitors to better understand not only what their photographs include (and exclude), but also *why*.

Overall, this study has afforded an unprecedented view into visitor behaviour at unstaffed heritage sites through the lens of Instagram. Since this study mainly relied on off-the-shelf inference with a pre-trained object detection model, the present methodology can be readily applied in other contexts (be they different destinations, collections of images, or social media platforms) to understand how visitors engage with tourist destinations through their photographs. Secondly, object detection methods can be used by charities such as English Heritage itself to monitor what their visitors are posting about across their unstaffed sites. This would help recover the otherwise lost connection between the charity and its visitors concerning the latter's experience at heritage sites, thus informing the charity's mission 'to bring the story of England to life' for its visitors (*English Heritage: Annual Report 2017/18* 3).

Data Repository

<https://doi.org/10.7910/DVN/OGAMMY>

Acknowledgements

This work was supported by a Knowledge Exchange Fellowship from The Oxford Research Centre in the Humanities (TORCH) at the University of Oxford, Grant number 0005946.

Submitted: August 31, 2022 EST, Accepted: September 19, 2022 EST



This is an open-access article distributed under the terms of the Creative Commons Attribution 4.0 International License (CCBY-4.0). View this license's legal deed at <http://creativecommons.org/licenses/by/4.0> and legal code at <http://creativecommons.org/licenses/by/4.0/legalcode> for more information.

REFERENCES

- Akçay, Samet, et al. "Using Deep Convolutional Neural Network Architectures for Object Classification and Detection within X-Ray Baggage Security Imagery." *IEEE Transactions on Information Forensics and Security*, vol. 13, no. 9, 2018, pp. 2203–15, [doi:10.1109/tifs.2018.2812196](https://doi.org/10.1109/tifs.2018.2812196).
- Arcos-García, Álvaro, et al. "Evaluation of Deep Neural Networks for Traffic Sign Detection Systems." *Neurocomputing*, vol. 316, 2018, pp. 332–44, [doi:10.1016/j.neucom.2018.08.009](https://doi.org/10.1016/j.neucom.2018.08.009).
- Balomenou, Nika, et al. "Making Sense of Tourists' Photographs Using Canonical Variate Analysis." *Tourism Management*, vol. 61, 2017, pp. 173–79, [doi:10.1016/j.tourman.2017.02.010](https://doi.org/10.1016/j.tourman.2017.02.010).
- Belk, Russell, and Joyce Hsiu-yen Yeh. "Tourist Photographs: Signs of Self." *International Journal of Culture, Tourism and Hospitality Research*, vol. 5, no. 4, 2011, pp. 345–53, [doi:10.1108/17506181111174628](https://doi.org/10.1108/17506181111174628).
- Belk, Russell W. "Possessions and the Extended Self." *Journal of Consumer Research*, vol. 15, no. 2, 1988, p. 139, [doi:10.1086/209154](https://doi.org/10.1086/209154).
- Bell, Philip. "Content Analysis of Visual Images." *Handbook of Visual Analysis*, edited by Theo Van Leeuwen and Carey Jewitt, Sage Publications, 2001, pp. 10–34, [doi:10.4135/9780857020062.n2](https://doi.org/10.4135/9780857020062.n2).
- Bourdieu, Pierre, and Shaun Whiteside. *Photography: A Middle-Brow Art*. New ed., Stanford University Press, 1996.
- "Controlling Your Visibility." *Instagram*, <https://help.instagram.com/116024195217477>. Accessed 23 Feb. 2022.
- Crang, Mike. "Picturing Practices: Research through the Tourist Gaze." *Progress in Human Geography*, vol. 21, no. 3, 1997, pp. 359–73, [doi:10.1191/030913297669603510](https://doi.org/10.1191/030913297669603510).
- Dinhopl, Anja, and Ulrike Gretzel. "Selfie-Taking as Touristic Looking." *Annals of Tourism Research*, vol. 57, 2016, pp. 126–39, [doi:10.1016/j.annals.2015.12.015](https://doi.org/10.1016/j.annals.2015.12.015).
- Donahue, Jeff, et al. "Decaf: A Deep Convolutional Activation Feature for Generic Visual Recognition." *International Conference on Machine Learning*, PMLR, 2014.
- Donaire, José A., et al. "Tourist Clusters from Flickr Travel Photography." *Tourism Management Perspectives*, vol. 11, 2014, pp. 26–33, [doi:10.1016/j.tmp.2014.02.003](https://doi.org/10.1016/j.tmp.2014.02.003).
- English Heritage: Annual Report 2017/18*. English Heritage, https://www.english-heritage.org.uk/siteassets/home/about-us/annual-reports/eh-annual-report-2018_-web.pdf. Accessed 23 Feb. 2022.
- Farahani, Leila Mahmoudi, et al. "Investigating Heritage Sites through the Lens of Social Media." *Journal of Architecture and Urbanism*, vol. 42, no. 2, 2018, pp. 199–211, [doi:10.3846/jau.2018.7057](https://doi.org/10.3846/jau.2018.7057).
- Galí, Núria, and José A. Donaire. "Tourists Taking Photographs: The Long Tail in Tourists' Perceived Image of Barcelona." *Current Issues in Tourism*, vol. 18, no. 9, 2015, pp. 893–902, [doi:10.1080/13683500.2015.1037255](https://doi.org/10.1080/13683500.2015.1037255).
- Garrod, Brian. "Understanding the Relationship between Tourism Destination Imagery and Tourist Photography." *Journal of Travel Research*, vol. 47, no. 3, 2009, pp. 346–58, [doi:10.1177/0047287508322785](https://doi.org/10.1177/0047287508322785).
- Goffman, Erving. *The Presentation of Self in Everyday Life*. Doubleday, 1959.
- Guérin, Joris, et al. "CNN Features Are Also Great at Unsupervised Classification." *arXiv Preprint arXiv:1707.01700*, 2017.

- Hogan, Bernie. "The Presentation of Self in the Age of Social Media: Distinguishing Performances and Exhibitions Online." *Bulletin of Science, Technology & Society*, vol. 30, no. 6, 2010, pp. 377–86, [doi:10.1177/0270467610385893](https://doi.org/10.1177/0270467610385893).
- Huang, Jonathan, et al. "Speed/Accuracy Trade-Offs for Modern Convolutional Object Detectors." *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2017, [doi:10.1109/cvpr.2017.351](https://doi.org/10.1109/cvpr.2017.351).
- Jacob, Jerin. "Image as Evidence: The Traveller's Need to Instagram." *Indian Journal of Mental Health*, vol. 5, no. 2, 2018, pp. 260–64.
- Kaplan, Andreas M., and Michael Haenlein. "Users of the World, Unite! The Challenges and Opportunities of Social Media." *Business Horizons*, vol. 53, no. 1, 2010, pp. 59–68, [doi:10.1016/j.bushor.2009.09.003](https://doi.org/10.1016/j.bushor.2009.09.003).
- Kisilevich, Slava, et al. "Event-Based Analysis of People's Activities and Behavior Using Flickr and Panoramio Geotagged Photo Collections." *2010 14th International Conference Information Visualisation*, IEEE, 2010, [doi:10.1109/iv.2010.94](https://doi.org/10.1109/iv.2010.94).
- Larsen, Jonas. "Families Seen Sightseeing: Performativity of Tourist Photography." *Space and Culture*, vol. 8, no. 4, 2005, pp. 416–34, [doi:10.1177/1206331205279354](https://doi.org/10.1177/1206331205279354).
- Latorre-Martínez, María Pilar, et al. "Image-Focused Social Media for a Market Analysis of Tourism Consumption." *International Journal of Technology Management*, vol. 64, no. 1, 2014, p. 17, [doi:10.1504/ijtm.2014.059234](https://doi.org/10.1504/ijtm.2014.059234).
- Leung, Daniel, et al. "Social Media in Tourism and Hospitality: A Literature Review." *Journal of Travel & Tourism Marketing*, vol. 30, no. 1–2, 2013, pp. 3–22, [doi:10.1080/10548408.2013.750919](https://doi.org/10.1080/10548408.2013.750919).
- Li, Jingjing, et al. "Big Data in Tourism Research: A Literature Review." *Tourism Management*, vol. 68, 2018, pp. 301–23, [doi:10.1016/j.tourman.2018.03.009](https://doi.org/10.1016/j.tourman.2018.03.009).
- Liu, Shaopeng, et al. "A Novel Scene Classification Model Combining ResNet Based Transfer Learning and Data Augmentation with a Filter." *Neurocomputing*, vol. 338, 2019, pp. 191–206, [doi:10.1016/j.neucom.2019.01.090](https://doi.org/10.1016/j.neucom.2019.01.090).
- Manovich, Lev. *Instagram and Contemporary Image*. CUNY, 2017.
- McMullen, Melissa. "'Pinning' Tourist Photographs: Analyzing the Photographs Shared on Pinterest of Heritage Tourist Destinations." *Current Issues in Tourism*, vol. 23, no. 3, 2020, pp. 376–87, [doi:10.1080/13683500.2019.1568399](https://doi.org/10.1080/13683500.2019.1568399).
- Miah, Shah Jahan, et al. "A Big Data Analytics Method for Tourist Behaviour Analysis." *Information & Management*, vol. 54, no. 6, 2017, pp. 771–85, [doi:10.1016/j.im.2016.11.011](https://doi.org/10.1016/j.im.2016.11.011).
- Motamed, Bahareh, and Leila Mahmoudi Farahani. "The Evaluative Image of the City through the Lens of Social Media: Case Study of Melbourne CBD." *Journal of Architecture and Urbanism*, vol. 42, no. 1, 2018, pp. 24–33, [doi:10.3846/jau.2018.1474](https://doi.org/10.3846/jau.2018.1474).
- Nov, Oded, and Chen Ye. "Community Photo Sharing: Motivational and Structural Antecedents." *ICIS 2008 Proceedings*, 2008.
- Oquab, Maxime, et al. "Learning and Transferring Mid-Level Image Representations Using Convolutional Neural Networks." *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2014, [doi:10.1109/cvpr.2014.222](https://doi.org/10.1109/cvpr.2014.222).
- Pan, Sinno Jialin, and Qiang Yang. "A Survey on Transfer Learning." *IEEE Transactions on Knowledge and Data Engineering*, vol. 22, no. 10, 2009, pp. 1345–59, [doi:10.1109/tkde.2009.191](https://doi.org/10.1109/tkde.2009.191).
- Pearce, Philip L., et al. "The Spectacular and the Mundane: Chinese Tourists' Online Representations of an Iconic Landscape Journey." *Journal of Destination Marketing & Management*, vol. 4, no. 1, 2015, pp. 24–35, [doi:10.1016/j.jdmm.2014.11.001](https://doi.org/10.1016/j.jdmm.2014.11.001).

- Pearce, Philip L., and Zhe Wang. "Human Ethology and Tourists' Photographic Poses." *Annals of Tourism Research*, vol. 74, 2019, pp. 108–20, [doi:10.1016/j.annals.2018.11.002](https://doi.org/10.1016/j.annals.2018.11.002).
- Popescu, Adrian, et al. "Mining Tourist Information from User-Supplied Collections." *Proceeding of the 18th ACM Conference on Information and Knowledge Management*, 2009, [doi:10.1145/1645953.1646211](https://doi.org/10.1145/1645953.1646211).
- Robinson, Mike, and David Picard. "Moments, Magic and Memories: Photographing Tourists, Tourist Photographs and Making Worlds." *The Framed World: Tourism, Tourists and Photography*, edited by David Picard and Mike Robinson, Routledge, 2009, pp. 1–38.
- Rossi, Luca, et al. "Venice through the Lens of Instagram: A Visual Narrative of Tourism in Venice." *Companion Proceedings of the The Web Conference 2018*, 2018, [doi:10.1145/3184558.3191557](https://doi.org/10.1145/3184558.3191557).
- Scarles, Caroline. "Becoming Tourist: Renegotiating the Visual in the Tourist Experience." *Environment and Planning D: Society and Space*, vol. 27, no. 3, 2009, pp. 465–88, [doi:10.1068/d1707](https://doi.org/10.1068/d1707).
- Sessions, Carrie, et al. "Measuring Recreational Visitation at U.S. National Parks with Crowd-Sourced Photographs." *Journal of Environmental Management*, vol. 183, 2016, pp. 703–11, [doi:10.1016/j.jenvman.2016.09.018](https://doi.org/10.1016/j.jenvman.2016.09.018).
- Sharif Razavian, Ali, et al. "CNN Features Off-the-Shelf: An Astounding Baseline for Recognition." *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, 2014, [doi:10.1109/cvprw.2014.131](https://doi.org/10.1109/cvprw.2014.131).
- "Social 2018 Main Findings." *We Are Flint*, Feb. 2018, <https://weareflint.co.uk/main-findings-social-media-demographics-uk-usa-2018/>.
- Sontag, Susan. *On Photography*. Farrar, Straus and Giroux, 1977.
- Stylianou-Lambert, Theopisti. "Tourists with Cameras: Reproducing or Producing?" *Annals of Tourism Research*, vol. 39, no. 4, 2012, pp. 1817–38, [doi:10.1016/j.annals.2012.05.004](https://doi.org/10.1016/j.annals.2012.05.004).
- Su, Shiliang, et al. "Characterizing Geographical Preferences of International Tourists and the Local Influential Factors in China Using Geo-Tagged Photos on Social Media." *Applied Geography*, vol. 73, 2016, pp. 26–37, [doi:10.1016/j.apgeog.2016.06.001](https://doi.org/10.1016/j.apgeog.2016.06.001).
- Talukdar, J., et al. "Transfer Learning for Object Detection Using State-of-the-Art Deep Neural Networks." *2018 5th International Conference on Signal Processing and Integrated Networks (SPIN)*, 2018, [doi:10.1109/spin.2018.8474198](https://doi.org/10.1109/spin.2018.8474198).
- Tenkanen, Henriikki, et al. "Instagram, Flickr, or Twitter: Assessing the Usability of Social Media Data for Visitor Monitoring in Protected Areas." *Scientific Reports*, vol. 7, no. 1, 2017, pp. 1–11, [doi:10.1038/s41598-017-18007-4](https://doi.org/10.1038/s41598-017-18007-4).
- "The Three Fairies Sculpture." *The Rollright Trust*, Nov. 2017, <https://www.rollrightstones.co.uk/news/story/faeries-sculpture>.
- Urry, John. *Leisure and Travel in Contemporary Societies*. Sage Publications, 1990.
- Urry, John, and Jonas Larsen. *The Tourist Gaze 3.0*. Sage Publications, 2011, [doi:10.4135/9781446251904](https://doi.org/10.4135/9781446251904).
- Vu, Huy Quan, et al. "Exploring the Travel Behaviors of Inbound Tourists to Hong Kong Using Geotagged Photos." *Tourism Management*, vol. 46, 2015, pp. 222–32, [doi:10.1016/j.tourman.2014.07.003](https://doi.org/10.1016/j.tourman.2014.07.003).
- Wood, Spencer A., et al. "Using Social Media to Quantify Nature-Based Tourism and Recreation." *Scientific Reports*, vol. 3, no. 1, 2013, pp. 1–7, [doi:10.1038/srep02976](https://doi.org/10.1038/srep02976).
- Xiao, Tete, et al. "Unified Perceptual Parsing for Scene Understanding." *Proceedings of the European Conference on Computer Vision (ECCV)*, 2018, pp. 432–48, [doi:10.1007/978-3-030-01228-1_26](https://doi.org/10.1007/978-3-030-01228-1_26).

Zhou, Xiaolu, et al. "Detecting Tourism Destinations Using Scalable Geospatial Analysis Based on Cloud Computing Platform." *Computers, Environment and Urban Systems*, vol. 54, 2015, pp. 144–53, [doi:10.1016/j.compenvurbsys.2015.07.006](https://doi.org/10.1016/j.compenvurbsys.2015.07.006).

SUPPLEMENTARY MATERIALS

Supplementary Information for Heritage site-seeing through the visitor's lens on Instagram

Download: <https://culturalanalytics.org/article/38966-heritage-site-seeing-through-the-visitor-s-lens-on-instagram/attachment/102990.pdf>
