Landscape and Urban Planning 217 (2022) 104261

Contents lists available at ScienceDirect

Landscape and Urban Planning

journal homepage: www.elsevier.com/locate/landurbplan

In the AI of the beholder: A comparative analysis of computer vision-assisted characterizations of human-nature interactions in urban green spaces

Andrea Ghermandi [a,b,*], Yaella Depietri [b,c], Michael Sinclair [a,b,d]

[a] Department of Natural Resources and Environmental Management, University of Haifa, 199 Aba Khushy, Haifa 3498838, Israel
[b] Natural Resources and Environmental Research Center, University of Haifa, 199 Aba Khoushy Ave, Mount Carmel Haifa 3498838, Israel
[c] Faculty of Architecture and Town Planning, Technion – Israel Institute of Technology, Haifa 3200003, Israel
[d] Urban Big Data Centre, School of Social and Political Sciences, University of Glasgow, Glasgow, UK

HIGHLIGHTS

- The content of online photographs offers insights on human-nature interactions.
- We apply three cloud-based computer vision services to outdoor photographs in Haifa.
- Clustering of 45 green and blue areas is affected by the choice of software.
- We find differences in the identification of activities, environment and feelings.
- The optimal choice of computer vision software depends on the intended application.

ARTICLE INFO

Keywords:
Big data
Cultural ecosystem services
Landscape planning
Passive crowdsourcing
Social media

ABSTRACT

Big data from photo-sharing platforms offer unique opportunities for the study of human-nature interactions and landscape planning. Research increasingly relies on computer vision in artificial intelligence to identify elements of interest in photographs and user preferences and sentiment towards them. Studies largely rely on pre-trained models from one of several available cloud-based, commercial image recognition services, but the extent to which findings depend on the implemented technology has not yet been explored. Here, we analyze ~ 10,000 outdoor photographs retrieved from three social media platforms and geolocated within green and blue spaces in Haifa (Israel) by means of machine tags from three popular cloud-based services. We find that clustering of the 45 investigated sites based on common characteristics of the photographs is considerably affected by the image recognition service chosen, especially for sites with limited data points (<80 photographs). Moreover, after associating the individual tags to specific aspects of the outdoor experience, we find substantial differences in the identification and ranking of outdoor recreational activities, characterization of the local biophysical environment (e.g., wildlife and vegetation), and feelings associated with the photographs. With no image recognition service clearly outperforming the others in all evaluation criteria, we argue that the optimal choice of image recognition service to rely on likely depends on the intended final application. Time and resource permitting, future studies should consider combining information from multiple sources for a characterization that is more nuanced and less prone to be affected by the idiosyncrasies of the individual technologies.

1. Introduction

The accumulation and accessibility to unprecedented amounts of digital data through social media platforms offer unique opportunities to understand how humans perceive and interact with the natural environment to inform sustainable environmental management and landscape planning (Ghermandi and Sinclair 2019; Calcagni et al., 2019). Considering that such data is generated and voluntarily shared by end-users for purposes other than scientific investigation, the process of retrieving and analyzing such data is generally referred to as passive

* Corresponding author at: Department of Natural Resources and Environmental Management, University of Haifa, 199 Aba Khushy, Haifa 3498838, Israel.
E-mail addresses: aghermand@univ.haifa.ac.il (A. Ghermandi), yaella.d@campus.technion.ac.il (Y. Depietri), msinclai@campus.haifa.ac.il (M. Sinclair).

https://doi.org/10.1016/j.landurbplan.2021.104261
Received 2 March 2021; Received in revised form 30 July 2021; Accepted 26 September 2021
Available online 6 October 2021

crowdsourcing (Connors et al., 2012).

The global penetration of smartphones and the integration of Global Positioning System (GPS) technology in various portable devices have led to vast amounts of photographs and text content, which are tagged with precise information about the time and location where they were created. After retrieval by researchers, generally relying on Application Programming Interfaces (APIs) (Toivonen et al., 2019), such data can be explored at broad spatial scales and at detailed temporal and spatial resolution, providing quasi real-time and site-specific knowledge otherwise unattainable without resource-intensive survey-based methods (Muñoz et al., 2019; Bubalo et al., 2019; Zielstra and Hochmair 2013; Heikinheimo et al., 2020; Sinclair et al., 2020). Photo-sharing platforms such as Flickr (http://flickr.com) and, to a lesser extent, Instagram (http://instagram.com) and Panoramio (which ceased operation in 2016) are among the sources that have been more extensively explored thus far in environmental studies (Ghermandi and Sinclair 2019), thanks to the rich window that the visual and textual analysis of photographs and the associated text allows into the users' experiences and preferences (Ghermandi et al., 2020a).

While early research in this field has focused primarily on the suitability of photo counts as proxies for number of recreational visits (Wood et al., 2013; Tenkanen et al., 2017; Donaire et al., 2014; Ghermandi 2016), a new wave of studies has more recently started to explore the visual content of photographs as a source of further insights into direct human-nature interactions. Such applications have focused on the identification of the cultural ecosystem service(s) captured in the photographs (Richards and Friess 2015; Oteros-Rozas et al., 2017), characterization of the factors contributing to the aesthetic value of landscapes and elements thereof (Van Berkel et al., 2018; Langemeyer et al., 2018; Foltête et al., 2020; Tenerelli et al., 2017), and visitors' preferences for specific aspects of a nature-based recreation experience (Tieskens et al., 2018; Ghermandi et al., 2020a; Hausmann et al., 2018). Partly due to the abundance of social media data, the visitation to urban parks (Donahue et al., 2018; Hamstead et al., 2018; Song et al., 2020; Zhang and Zhou 2018) and other urban (Alemu I et al., 2021; Depietri et al., 2021) and *peri*-urban green areas (Komossa et al., 2020; Ghermandi 2016) have received a particular attention in the literature.

Previous studies have frequently relied on manual classification of image content (Heikinheimo et al., 2017; Martínez Pastur et al., 2016; Angradi et al 2018; Guerrero et al., 2016; Ros Candeira et al., 2020; Amorim Maia et al., 2020; Tian et al., 2021). Such approach is however impractical for large databases. Recent studies thus increasingly turn to computer vision in artificial intelligence for automated image content analysis (Callau et al., 2019). In such analyses, pre-trained models from commercial cloud services are generally used to label photographs with content-related machine tags. The machine tags are then frequently used to cluster photographs, either based on the subject's theme (Richards and Tunçer, 2018), the cultural ecosystem services they reflect (Lee et al., 2019), or the natural sites under investigation based on shared typologies of human-nature interactions (Ghermandi et al., 2020a).

Among the best-known commercial providers that offer image recognition capabilities, one may count Google Cloud Vision (https://cloud.google.com/vision), Clarifai (https://www.clarifai.com/models/general-image-recognition), Microsoft Azure Computer Vision (https://azure.microsoft.com/en-us/services/cognitive-services/computer-vision/), IBM's Watson Visual Recognition (https://visual-recognition-code-pattern.ng.bluemix.net) and Amazon Rekognition (https://aws.amazon.com/rekognition/). Such services typically operate with a "freemium" strategy, which includes a plan for the free identification of elements of interest (through machine tagging) in a limited number of photographs or for a limited period. An example of image recognition software under a Creative Commons license for non-commercial use is SegNet (http://mi.eng.cam.ac.uk/projects/segnet/demo.php). Among the various options, those most frequently explored in assessments of landscape aesthetics and cultural ecosystem services are Google Cloud Vision (Alampi Sottini et al., 2019; Alemu I et al., 2021; Ghermandi et al., 2020a; Gosal and Ziv 2020; Runge et al., 2020; Richards and Tunçer 2018; Song et al., 2020), Clarifai (Depietri et al., 2021; Lee et al., 2019; Karasov et al., 2020) and Azure Computer Vision (Ruiz-Frau et al., 2020). Two exceptions are the studies by Seresinhe et al. (2018) and Payntar et al. (2020), which respectively rely on Places CNN and ResNet50.

Several studies have assessed the comparative accuracy of computer vision cloud-based services in detecting visual elements in photographs (Al-Omair and Huang 2018; Dodge and Karam 2016; Nilsson and Jönsson, 2019; Temel et al., 2019) and the accuracy of individual services for the identification of biophysical elements of the environment in photographs of nature-based recreation (Richards and Tunçer, 2018; Runge et al., 2020). However, to the best of our knowledge, no study has explored potential differences in how computer vision services interpret photographs reflecting direct human-nature interactions with landscapes and ecosystems, and how such differences might alter the conclusions drawn by researchers regarding visitors' cultural engagement with the sites and the preferences or feelings (i.e., conscious experiences of emotional reactions), that are associated with the experience.

This study investigates the application of three cloud computer vision services (Clarifai, Google Cloud Vision and Microsoft Azure Computer Vision) on almost ten thousand geotagged, outdoor social media photographs reflecting direct human-nature interactions in the entire network of green and blue spaces in the city of Haifa in Israel. Using two different clustering techniques, we first explore the extent to which the analysis of the machine tags retrieved from the different services leads to different grouping of the sites into clusters. Based on a classification of the tags into different categories reflecting various aspects of the outdoor experience, we then formally test whether the three services differ in the respective richness of vocabulary and frequency with which tags related to a specific aspect are retrieved. In this, we pay particular attention to the identification of the prevailing outdoor activities and of feelings associated with the experience captured in the photographs, in the investigated sites. We conclude by highlighting some of the implications of the findings for future studies assessing direct human-nature interactions for improved management and planning of landscapes and urban green areas.

2. Materials and methods

2.1. Study area

With an estimated population of about 280,000 inhabitants, Haifa is Israel's third largest city, located in the northwest of the country on the coast of the Mediterranean Sea (32°49′0″N 34°59′0″E) (Fig. 1). The climate is typically Mediterranean, with warm summers and mild, rainy winters. Average temperature ranges between 8.7℃ (in February) and 31.4℃ (in August), with high summer humidity levels. Precipitation averages 630 mm/year, almost all concentrated in the winter and spring. The city is built on the top and slopes of Mount Carmel (max elevation = 525 m a.s.l.). The built-up area is interspersed with ephemeral riverbeds ("wadis"), which are undeveloped, vegetated (often forested), corridors that run through the city from higher elevations to the coast. These open spaces host a rich vegetative community, including the common oak, terebinth, carob tree and mastic tree. Aleppo pines (*Pinus halepensis*) are also widespread in the area, primarily because of past tree planting campaigns. The wadis and green areas provide habitat for wildlife, such as wild boars, salamanders, golden jackals, porcupines, hyraxes, Egyptian mongooses, owls, and chameleons. Many of Haifa's wadis are marked by hiking trails, providing extensive recreational opportunities. Additional green open spaces host infrastructure of broader touristic interest, including beaches, historical sites, monasteries and churches, and a zoo. Of particular significance from a cultural and touristic point of view are the Baháí Gardens, a series of well-tended garden terraces, which descend from Mount Carmel offering a panoramic view on the downtown sections of the city, and

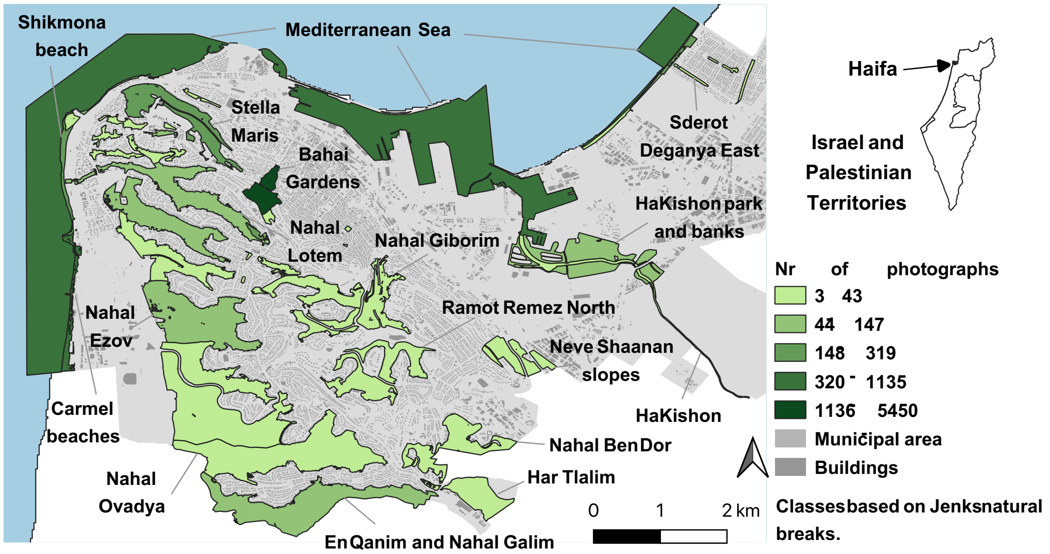

Fig. 1. City of Haifa with location of the investigated open spaces and choropleth map of retrieved geotagged photographs.

hosting the Shrine of the Báb, second holiest place for the Baháí religion. Overall, the mosaic of wadis, beaches and other open spaces give the city a green quality, especially in comparison to other urban agglomerations in Israel, a condition which is particularly valued by its inhabitants (Depietri and Orenstein 2020).

The location and spatial boundaries of green and blue open spaces within the city were obtained from Lotan et al. (2017) and imported in ArcGIS 10.6.1. We selected the layers of areas identified as gardens or parks, urban nature sites by the Society for the Protection of Nature in Israel (SPNI), and water bodies. Fenced areas, private gardens and ruderals were excluded. Overall, 45 individual sites were included in the analysis, for which at least one geotagged photograph could be retrieved.

2.2. Data collection and photo content analysis

Following Ghermandi et al. (2020a), we retrieved the metadata of photographs geotagged within the city's boundaries from a range of social media sites to reduce potential biases arising from the different use and composition of their respective user communities. Flickr and VKontakte photographs were retrieved using the respective APIs. For Flickr, they cover the years 2005–2018; for VKontakte, they cover the period since the launch of the service in 2006 to 26 November 2018. Data for Panoramio, which is no longer operational, were retrieved from Lotan et al. (2018), who used counts of Panoramio photographs as a proxy for recreational activities in open spaces in Israel. These data cover the period 2005–2014. A subset of the photographs was investigated by Depietri et al. (2021), who examined the relative strengths and weaknesses of social media data and participatory mapping (PPGIS) for the identification of the cultural ecosystem services provided by seven open spaces within the city.

The data for 13,547 photographs that were geotagged within the boundaries of the city's green and blue open spaces were extracted. These were manually scrutinized to identify indoor photographs, leading to the removal of 30% of the sample. A total of 9,454 outdoor photographs were suitable for further analysis. Similarly, Seresinhe et al. (2018) found indoor photographs to account for 35% of photographs in a similar investigation conducted for urban areas in Great Britain. The majority of photographs in the final sample (63%) were obtained from VKontakte, followed by Flickr (35%) and Panoramio (2%). The identified outdoor photographs were labeled using the pre-trained image labeling models available in Clarifai, Google Cloud Vision (henceforth "Google") and Microsoft Azure Computer Vision (henceforth "Azure") using the respective APIs. Up to 20 machine-generated tags describing the content of each of the images were retrieved from each model.

For the classification of tags according to specific aspects of the outdoor experience, we considered the framework proposed by Egorova (2020). This framework was specifically developed for application to online Volunteered Geographical Information (VGI), including data from photo-sharing platforms, and was considered more comprehensive and more oriented toward a functional distinction of elements of interest in outdoor photographs than alternative classifications in the literature (Callau et al., 2019; Buijs, 2009; Guerrero et al., 2016). Although created for the analysis of recreational activities, the model has more broad applicability to other aspects of outdoor nature-human interactions (such as aesthetic appreciation or casual interactions with the natural environment). Three domains of aspects related to nature-based experience are recognized (i.e., activity, environment, and feelings and cognition), which are further subdivided into 31 categories. Each tag that was retrieved at least twice from one of the computer vision services was associated with one of the 31 categories, or none of them if it was unrelated to human-nature interactions. The classification of tags was performed blindly as to what service the tag had been retrieved from.

2.3. Statistical analysis

2.3.1. Cluster analysis

Sites that were associated with a minimum of 50 tags (corresponding to at least three photographs), were grouped using two different clustering techniques: partitioning around medoids (PAM) (Maechler et al., 2019) and affinity propagation (Bodenhofer et al., 2011; Frey and Dueck, 2007). Two different techniques were used to assess the sensitivity of the findings to changes in methodology. The clustering relied on a dissimilarity matrix in which the dissimilarity between two sites is calculated as the proportion of different tags among the sets of the 50 tags most frequently associated with each of the sites (Richards and Tunçer, 2018). We used the average silhouette width (Kassambara and Mundt, 2019) to determine the optimal number of clusters. Both clustering techniques allow for the identification of exemplars (i.e., members of the input set that are best representative of the cluster to which they are assigned).

For the semantic characterization of the clusters, we selected the most representative tags in the cluster, as determined using the term frequency-inverse document frequency (TF-IDF) technique (Landauer et al., 1998). Such technique is often used in data mining to identify the words that best characterize a text within a collection of texts. In the present case, a TF-IDF score is calculated for each of the tags. The score increases proportionally to the number of times the tag appears among the top 50 tags for the sites in a specific cluster but is offset by the overall number of sites that contain the tag. Such correction is performed to avoid selecting tags that are very common across all sites (e.g., outdoors, people, nature) and thus of limited usefulness to identify the distinctive character of the clusters. For each of the clusters, the tags with the five highest TF-IDF scores were selected. We performed the TF-IDF both considering all the tags as well as only those which were present in at least 20% of the sites.

To evaluate the extent to which the sites are clustered in a similar way relying on different sets of tags, we used the Adjusted Rand Index (ARI). The index accounts for chance overlaps (see Table S1 in the Supplementary Materials for additional details). A value close to 1 indicates a high overlap between the partitions, whereby ARI = 1 indicates perfect overlap (Hubert and Arabie, 1985). We calculated ARI for the pairwise comparisons between partitions using the adj.rand.index function implemented in the pdfCluster in R (Azzalini and Menardi, 2014). ARI is calculated only for sites that are clustered based on all three sets of tags (i.e., sites for which at least 50 tags were retrieved from each computer vision service).

2.3.2. Distribution of aspect-specific tags

We considered 12 categories of human-nature interaction aspects from Egorova's (2020) model, which correspond to the second level of Egorova's classification, except for "biophysical environment" for which sub-categories at the third level were included in the analysis (Table 1).

A chi-square test with Bonferroni correction for multiple comparisons was used to evaluate differences across pairs of computer vision services with respect to both the number of unique tags retrieved and the frequency of occurrence of tags for each of the considered categories of human-nature interaction aspects (McHugh 2013). When the chi-square test rejected the hypothesis that the frequency of occurrence of aspects-specific tags does not vary across the examined computer vision services, a post-hoc comparison was performed, where the value of each category was tested against the sum of all others to determine which aspects are statistically significantly different (Latta et al., 2012).

As an additional level of analysis, for the "feelings" category we further investigated the accordance between the rankings of the nine sub-categories of feelings identified by Egorova (2020) across the different computer vision services. The nine sub-domains are aesthetic appreciation, sense of wilderness, relaxation, having fun, sense of adventure, sense of awe, experiencing something special, sense of (un) safety, and other feelings. For this analysis, we relied on Kendall's rank correlation (tau-b) as implemented in the Kendall package in R (McLeod 2011). All statistical analyses and charts were done in R, version 4.0.3 (2020-10-10) (R Core Team 2020).

Table 1
Categories of aspects of human-nature interactions investigated in this study.

Domain	2nd level categories	3rd level categories
Activity	**Activities and actions** **Activity-related artifacts** **Activity characteristics**[1]	
Environment	Biophysical environment	**Physical environment** **Wildlife and vegetation** **Environmental processes**
	Infrastructure **Social environment** **Cultural-historical environment**	
Feelings and cognition	**Feelings**	
	Sentiments **Knowledge**	

Notes: In **bold** are the 12 categories evaluated in this study; [1] "Activity characteristics" include duration, difficulty, and activity-related elements of space.

3. Results

The distribution of photographs within Haifa's open spaces is highly heterogeneous. Photographs are concentrated in touristic sites such as the Baháí Gardens and the coastal areas and beaches on the Mediterranean Sea (e.g., the Carmel and Shikmona beaches) (see Fig. 1). Except for Nahal Lotem (in the upper part of which is located the popular Haifa Educational Zoo), most wadis are associated with few photographs, confirming the relatively "wild" character of the sites. It is worth noting that such observations do not necessarily indicate a low visitation rate, considering that some outdoor activities and cultural values of locals may be under-represented (Depietri et al., 2021), and that the number of photographs, unlike photo-user-days (Wood et al., 2013), does not control for multiple photographs taken by individual users during one single visit (Ghermandi et al., 2020a).

Insofar as the computer vision services are concerned, a first difference emerges from the number of tags retrieved. Both Google and Azure were associated with less tags per photograph (respectively, 14.6 and 10.3) compared to Clarifai (20 tags per photograph). The largest number of unique tags was however retrieved from Google (2,376 tags), followed by Clarifai (1,654 tags) and Azure (1,379 tags).

3.1. Consistency of clustering results

Table 2 and Fig. 2 show the results of the cluster analysis obtained with the PAM technique and, for Table 2, the representative tags with the 20% cut-off (see Section 2.3.1). The representative tags from PAM

Table 2
PAM clusters, exemplars and most representative labels based on TF-IDF (with 20% cut-off) for urban open spaces in Haifa, obtained with three alternative sets of machine tags.

ID	Cluster exemplar	Sites in cluster	Representative tags	Proposed cluster label
Clarifai				
C1	Nahal Ovadya	20	Little, cute, smile, baby, child, fun	Family in nature
C2	Stella Maris	14	Cityscape, church, mountain, beach, sea	Coastal mountain city
C3	Shikmona beach	8	Surf, beach, sand, sea, seashore, sun, sunset	Sea and beaches
C4	Ramot Remez North	3	Church, cityscape, flora, flower, garden, leaf	Urban garden
Google				
G1	Baháí Gardens south slope	13	Cityscape, metropolis, daytime, coast, ocean	Coastal mountain city
G2	Shikmona beach	11	Coast, ocean, beach, shore, wave	Sea and beaches
G3	Nahal Ovadya	9	Branch, landscape, plant community, adaptation, botany, rock, vegetation, wilderness	Family in nature
G4	Sderot Deganya East	6	Selfie, hair, branch, fawn, sunset	Outdoor selfie
Azure				
A1	Nahal Amik	14	Playground, baby, dirt, little, toddler	Family in nature
A2	Stella Maris	6	Overlooking, hill, hillside, highland, jeans	Coastal mountain city
A3	Carmel beaches	6	Sun, wave, surfing, distance, sea, sunset, sandy	Sea and beaches
A4	Nahal Giborim	4	Stone, house, text, road, way, street	Urban life

Notes: TF-IDF is calculated on terms that are in the top 50 of most frequent labels for at least 20% of the investigated sites; only the tags with the five highest TF-IDF scores are included.

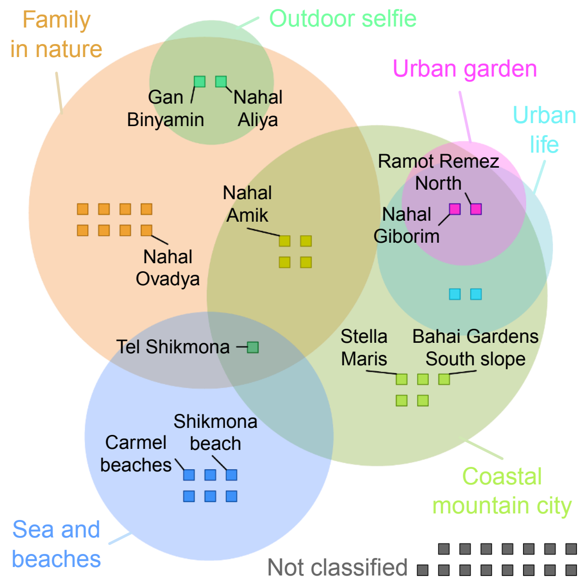

Fig. 2. Distribution of sites across PAM clusters obtained with the three computer vision services. Each square symbol in the Venn diagram represents one site. "Not classified" refers to sites that were not classified by Google and/or Azure because associated with < 50 tags.

clustering based on all tags are presented in Table S2, in the Supplementary Materials. The clustering results are quite consistent across the two investigated clustering techniques, thus in the reminder we will focus only on the results obtained with PAM, while the results of the cluster analysis based on affinity propagation and the comparison with PAM are presented and discussed in the Supplementary Materials (Tables S3 and S4).

For all three sets of machine tags, PAM identified four clusters of green and blue spaces in the city of Haifa. Notably, while Clarifai tags allowed to cluster 45 individual sites, the clustering of sites based on Google and Azure tags was limited to, respectively, 39 and 30 sites. This is due to the lower number of tags assigned to the photographs by the Google and Azure computer vision services, whereby the number of tags for the excluded sites dropped below the minimum number of 50 tags.

The analysis of the distribution of the sites across clusters, the exemplar sites and semantic interpretation of the representative tags reveal a fair degree of consistency across the three computer vision services, but also notable distinctions. Overall, the clusters capture well the character of the city's open spaces, which combine parks, gardens and wadis that are popular for family-oriented recreational activities, with coastal and mountain landscapes, and seaside activities.

Eight sites were consistently classified together in a cluster that we labeled "Family in nature". The representative tags for Clarifai and Azure for this cluster clearly stress the frequent presence of children in the photographs, a feature that is entirely missing from the Google tags, which rather focus on the "wild" character of the city's wadis. A large disproportion is also observed in the number of sites grouped under this cluster: while it is the largest cluster based on both Azure (47% of sites) and Clarifai tags (44% of sites), it only includes 23% of the sites classified based on Google tags.

The "Sea and beaches" cluster appears to be the most stable, with six sites consistently grouped together across the three computer vision services. Interestingly, all three image tagging services identify the association between coastal sites and "golden hour" photography (i.e., photography of sunset and sunrise), although for Google this only emerges from the TF-IDF analysis of all the tags (Table S2). This is consistent with the fact that Haifa's beaches, which are mostly oriented to face the West, are a prime location for the observation of sea sunsets.

A third, consistently identified cluster captures aspects related to cultural-historical buildings and the landscape elements of a "coastal mountain city". The clustering based on Azure tags puts a substantially lower number of sites in this cluster (20% of sites) compared to Google (33% of sites) and Clarifai (31% of sites). The fourth and last cluster for each set of machine tags appears to be more heterogenous in terms of composition and semantic domains associated with the representative tags. Accordingly, we assign in Fig. 2 a different name to the fourth cluster in each set: "Urban garden" for Clarifai, "Outdoor selfie" for Google, and "Urban life" for Azure.

While 63% (19 out of 30) of the sites that were clustered based on all three sets of machine tags are consistently grouped together (see Fig. 2), a certain overlap is observed in particular between the "Family in nature" and "Coastal mountain city" clusters, with four sites receiving a mixed classification depending on the set of machine tags used for analysis, as well as with the fourth cluster. The highest consistency is found for clusters derived from Clarifai and Azure (ARI Clarifai-Azure = 0.62), while lower consistency was found between Google and the other two services (ARI Google-Clarifai = 0.36; ARI Google-Azure = 0.42). Clustering consistency appears to be correlated with the number of photographs available for each individual site. Sites that are not clustered consistently across computer vision services tend to be associated with a smaller number of photographs (mean = 27.5, N = 11) than sites that are consistently grouped together (mean = 474.6, N = 19). However, a one-tailed *t* test fails to reject the null hypothesis of equal means (p-value = 0.066). Overall, sites with 80 photographs or more were all clustered consistently across computer vision services.

3.2. Comparative characterization of aspects of human-nature interactions

Among the most frequent tags for each of the three services, Google has the highest number of unique tags related to human-nature interactions (748 tags, compared to 486 for Clarifai and 399 for Azure). The three computer vision services differ in terms of their distribution of tags among the three aspect domains. The largest variety of tags related to the "activity" and "environment" domains is from Google (118 and 606 tags, respectively). Clarifai leads in the domain "feelings and cognition" (87 tags).

Fig. 3 shows the distribution of unique tags for the three services and for each of the sub-domains to which the tags were associated. No tag was associated with the sub-domain "activity characteristics". The sub-domain "sentiments" was also excluded since it was associated to only three tags (all retrieved from Clarifai). The chi-square test rejects the null hypothesis that there is no difference in the distribution of tags between Clarifai and both Google (χ^2 (df = 9) = 125.6, p < .001, N = 1471) and Azure (χ^2 (df = 9) = 66.73, p < .001, N = 1062). Clarifai is associated with a larger number of unique tags indicating feelings than both Google (χ^2 (df = 1) = 69.3, p < .001) and Azure (χ^2 (df = 1) = 24.9, p < .001), and more knowledge-related tags than Google (χ^2 (df = 1) = 8.2, p < .05). By contrast, the number of wildlife and vegetation-related tags in Clarifai is lower than both Google (χ^2 (df = 1) = 55.8, p < .001) and Azure (χ^2 (df = 1) = 36.0, p < .001). The null hypothesis is not rejected in the comparison between Google and Azure (χ^2 (df = 9) = 9.38, p > 0.05, N = 1387).

The overall frequency of occurrence of tags pertaining to individual aspects of direct human-nature interactions is presented in Fig. 4. The chi-square test rejects the null hypothesis of no difference in the distribution of tags across the computer vision services (p < 0.001). Most of the pairwise comparisons for specific aspects also reveal statistically significant differences (p < 0.001), with the exception of the comparison of Google and Azure for the sub-category "environmental processes" (χ^2 (df = 1) = 2.5, p > 0.05, N = 123,454) and the comparison of Google and Clarifai for "cultural-historical environment" (χ^2 (df = 1) = 3.2, p > 0.05, N = 173,208). Consistently with the larger average number of tags retrieved per photograph, Clarifai shows the highest frequency of tags for most categories (i.e., activities and actions, physical environment,

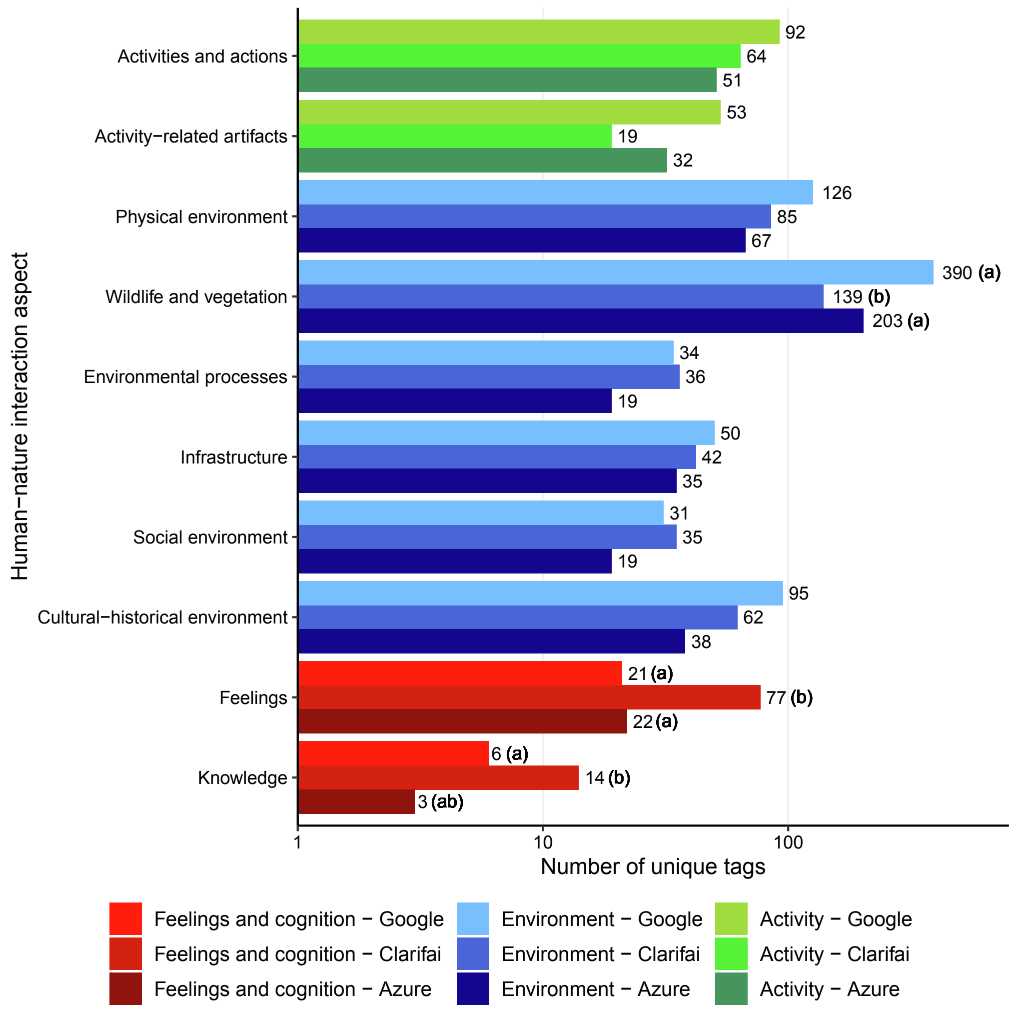

Fig. 3. Number of unique tags per human-nature interaction aspect, as retrieved from Clarifai (N = 573), Google (N = 898) and Azure (N = 489). Note: For each aspect, different letters denote a statistically significant difference at p-level < 0.05 or higher. No tag was associated with the sub-domain "activity characteristics". Three Clarifai tags associated with "sentiments" are not shown on the chart.

infrastructure, feelings, and social environment), but the lowest for activity-related artifacts, wildlife and vegetation. On the other hand, Google tags are most frequently associated with wildlife and vegetation, and knowledge, though the latter observation is based on only six tags, out of which 93% of the occurrences were for the "botany" tag. Azure shows the highest frequency of tags associated to activity-related artifacts. A list of the tags most frequently associated to each of the considered aspects of human-nature interaction is provided in Table S5 in the Supplementary Materials.

The differences among services are further highlighted in the lists of most frequent tags related to activities and actions (Fig. 5). All three services identify people posing for a photograph (either for a selfie or a portrait) as the most frequent activity captured in the photographs. Although the remaining actions are mostly shared by all three services, mainly reflecting – as expected – water-based sports and hiking/walking in nature, the relative ranking of the tags is quite different. Clarifai emphasizes surfing and walking, Azure highlights swimming and surfing, while Google identifies mostly passive leisure activities such as sitting and sun tanning. Notably, both Clarifai and Azure erroneously characterize photographs of the carefully maintained green lawns at the Baháí Gardens (and elsewhere) as photographs taken on a golf course. Consequently, golfing appears as one of the principal activities for both, while this is not a sport that is practiced in the city area.

Finally, we explored in further detail the category of feelings-related tags, due to its importance in characterizing the preferences of visitors and prevailing reactions associated to the outdoor experience. Fig. 6 shows the relative frequency of tags associated to photographs by each of the three computer vision services and amongst the nine sub-domains of feelings identified by Egorova (2020). Clarifai returned tags (>70) for each of the nine categories, while both Google and Azure did not provide any tag related to two categories: sense of (un)safety and experiencing something special from Google; sense of (un)safety and sense of awe from Azure. Less than ten tags were retrieved from Google for sense of awe, and from Azure for relaxation, sense of adventure and experiencing something special. The three computer vision services agree in characterizing sense of wilderness, having fun, and aesthetic appreciation as the three prevailing categories of feelings in the photographs. However, Google tags are dominated by the category "having fun" (48% of occurrences of feelings-related tags), while the most prominent category in both Clarifai and Azure is "sense of wilderness", albeit to a different degree (51% of feelings-related tags for Azure, 33% for Clarifai). Kendall's tau test with Bonferroni correction for multiple comparisons rejects the null hypothesis of independent rankings for all pairwise comparisons (Clarifai-Azure: tau = 0.873, p-value = 0.005; Google-Azure: tau = 0.800, p-value = 0.013; Clarifai-Google: tau = 0.704, p-value = 0.036), indicating that, despite the observed differences, the three computer vision services provide a comparable ranking of the feelings reflected in the investigated photographs.

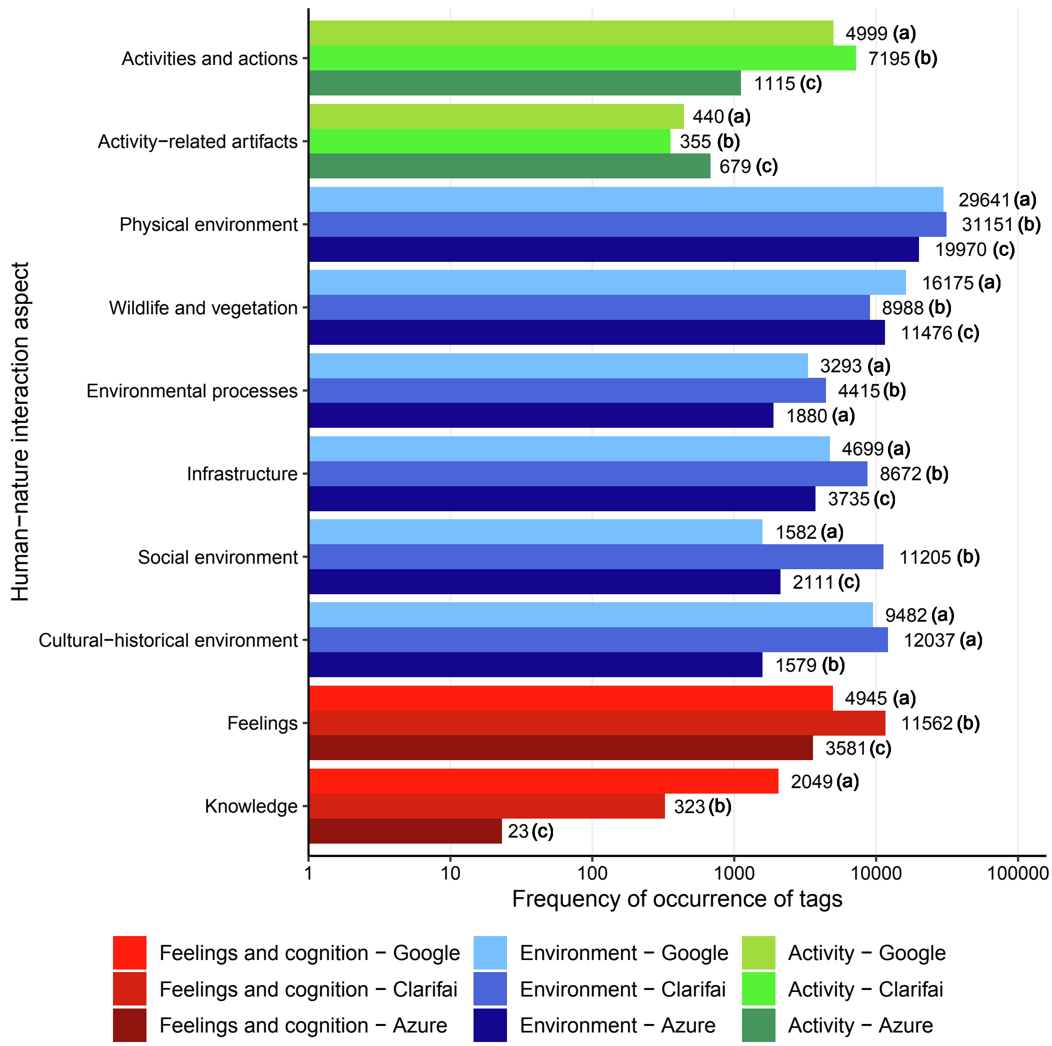

Fig. 4. Frequency of occurrence of tags pertaining to individual human-nature interaction aspects, as retrieved from Clarifai (N = 95,903), Google (N = 77,305) and Azure (N = 46,149). Note: For each aspect, different letters denote a statistically significant difference at p-level < 0.05 or higher.

4. Discussion and conclusion

Big data sources such as online photo-sharing platforms promise to deliver a transformative change in how researchers analyze and value direct human-nature interactions and cultural ecosystem services. Computer vision in artificial intelligence is an essential component in scaling up such efforts to large databases and large geographical and/or long temporal scales. Applications to landscape and urban planning may include, for instance, the planning and governance of urban green infrastructure (Guerrero et al., 2016), the valuation of the benefits of urban green and blue areas (Tian et al., 2021; Richards and Tunçer 2018; Song et al., 2020), the analysis of green gentrification in urban areas (Amorim Maia et al., 2020), and the characterization of perceptions of landscape aesthetics (Callau et al., 2019; Alampi Sottini et al., 2019; Gosal and Ziv 2020; Karasov et al., 2020; Seresinhe et al., 2018). To hold such promise analyses of social media photographs must rely on modelling, state-of-the-art solutions for the characterization of key aspects such as the presence of elements of interest, as well as the activities, preferences for elements of the landscape, and feelings reflected in the photographs.

With the development of custom image recognition models still in its infancy in environmental studies, to date most assessments turn to pre-trained, cloud-based, commercial services. The selection of the computer vision service on which to rely for the tagging of images has primarily been dictated by considerations of convenience and ease of access rather than accuracy and comprehensiveness of the retrieved information. Previous studies on the accuracy of cloud-based image recognition services have generally concluded that the leading services perform well especially for high-level concepts, albeit with a small margin of preference for Google Cloud Vision and with a general deterioration in performance in the presence of distortion, compression, blurriness, and rotation of the images (e.g., Nilsson and Jönsson, 2019). Still characterization of feelings and sentiments through image recognition models remains a subject of intense research (Islam and Zhang 2016; Truong and Lauw 2017).

Building on such insights, this study provides a comparative analysis of three leading cloud-based computer vision services in application to outdoor photographs taken in urban green and blue spaces in the city of Haifa in Israel. We find that the three investigated services substantially differ in terms of the average number of tags retrieved per photograph and that the clustering of the investigated sites based on common characteristics of the photographs is affected by the service chosen, especially for sites with a limited number of data points (<80 photographs).

Further analysis of the retrieved tags based on the different aspects of outdoor human-nature interactions reveals further differences. We find significant variation among the three services in the richness of vocabulary and frequency of retrieval for tags pertaining to activities,

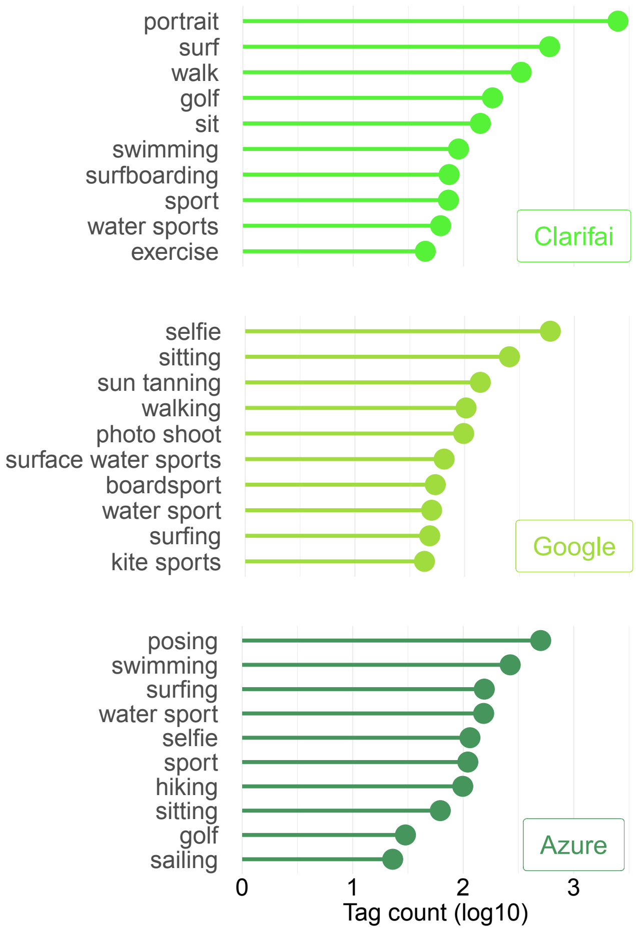

Fig. 5. Most frequent tags classified under the "activities and actions" sub-category for the three computer vision services. Frequently appearing but generic tags like "travel", "vacation" and "leisure" were not included in the chart.

environment, and feelings and cognition. Google is associated with the largest number of unique human-nature interactions-related tags and an advantage compared to the other services insofar as the tags related to wildlife and vegetation are concerned. Clarifai provided a larger number of tags per photograph and has an edge on the other services when it comes to the characterization of nuances of feelings, possibly indicating that it relies on a broader vocabulary for such aspects than the other services. Azure is associated with the lowest number of tags per photograph and number of unique tags related to human-nature interactions. In the comparative analysis, Azure was overshadowed by either Google or Clarifai for all aspects except for activity-related artifacts. Although we do not find substantial differences in the number and frequency of tags associated with recreational activities and actions, the three computer vision services provide different rankings for the activities most frequently represented in the photographs and we found at least one instance of mislabeling by Clarifai and Azure (i.e., golfing).

The observed differences among the three services suggest some complementarity between the qualitative information provided by each of them. For instance, a previous analysis of cultural ecosystem services in seven green and blue areas in Haifa, which compared the results obtained by participatory GIS (PGIS) and social media photographs (analyzed through Clarifai), concluded that the social media analysis was poorly tailored to capture the opportunities to interact with wildlife and vegetation in these areas (Depietri et al., 2021). The present analysis suggests that in this case, the additional use of Google could have provided additional and complementary information on the cultural importance of these elements of the landscape. Similar considerations can be done for identifying feelings associated with outdoor experience, whereby Azure might have been better able to capture the "sense of wilderness" which is associated with the city's wadis than Clarifai. Such observations support the idea that a more comprehensive understanding of the human-nature interactions taking place at the investigated sites could be gained combining insights from multiple photo recognition services.

Overall, based on the multiple criteria analyzed in this study, no image recognition service clearly outperforms or is inferior to the others in all evaluation criteria. Rather, drawing a parallel with what found by Ghermandi et al. (2020a) with regard to the benefits of using data derived from multiple photo-sharing platforms as opposed to only one, it seems reasonable to suggest that combining information retrieved from multiple computer vision services might provide a characterization that is more nuanced and less likely to be affected by the idiosyncrasies of the individual services. Similar suggestions were given by Kyriakou et al. (2019), who analyzed the issue of fairness of difference services (including Clarifai and Google) in the recognition of different social groups. With the understanding that such an option may not always be viable due to time and resource limitations, especially when it comes to the analysis of very large databases, we argue that the optimal choice of service will likely depend on the intended final application. Our results suggest, for instance, that Google might be favored in primarily biologically and ecologically oriented explorations, due to the broader vocabulary and frequent use of terms related to wildlife and vegetation. On the other hand, visual analyses of feelings and sentiments might benefit from relying on Clarifai's services.

Some of the limitations of the study require clarification. The investigation relies on a sample of photographs obtained from three social media sites. While this follows the good practice of considering multiple sources in an effort to reduce user and content biases (Ghermandi et al., 2020b), including data from other popular social networking sites (such as, for instance, Facebook and Instagram) might greatly enrich the analysis. This was not possible, however, due to the limitations these sites set on automated content retrieval and analysis. Moreover, the majority of photographs in the final sample were obtained from VKontakte, which suggest a bias towards overrepresenting the landscape experience of Russian-speaking locals and visitors. This should be taken into consideration when evaluating the overall distribution of tags, although it is not expected to significantly affect the results of the comparative analysis. It should also be noted that the present study does not address the issue of accuracy of the retrieved tags. Tags with a low confidence retrieved, for instance, from Clarifai or Azure (as opposed to Google, which only returns tags associated with an estimated confidence of 0.5 or higher) might have lower relevance and should be treated with caution. Finally, the results presented are derived from the analysis of a single case-study (the city of Haifa in Israel) and may not be straightforwardly transferable to other contexts.

Future research might further extend the exploration proposed in this study to include additional computer services. More importantly, we believe that there is scope for future assessments of the accuracy of computer vision services in environmental studies, especially in the characterization of feelings and sentiments in outdoor photographs and the associated cultural benefits, a dimension which is currently lacking in the literature. Specifically, there is a need for independent verification of the classification of images by experts or through supplementary social science studies, including the attribution of individual tags to specific categories. Moreover, further refinement of the classification of aspects of nature-based interactions used in this study, for instance to better characterize human interactions with biodiversity through subdivisions of the "wildlife and vegetation" category, would add depth to the analysis and resulting insights. Finally, we suggest that there is a strong potential for future assessments to benefit from combining visual content analyses such as the one proposed in this study with insights from the text associated to photographs' titles and tags, where available. The latter can provide additional insights into the photographer's

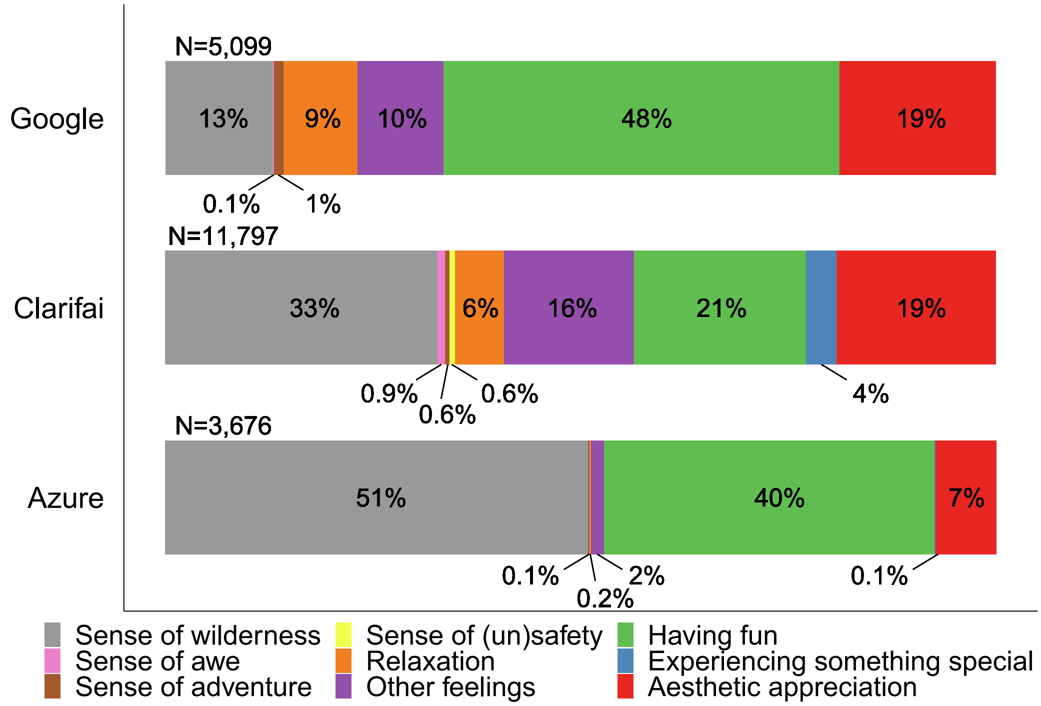

Fig. 6. Frequency of occurrence of tags associated with nine categories of feelings for the three computer vision services. Note: for each bar, the sum of all components may be slightly different from 100% due to rounding.

mental processes and personal conceptualization or describe the perceived relative importance of the visual elements (Ghermandi et al., 2020b; de Juan et al., 2007).

In conclusion, until specifically-developed image recognition models for cultural ecosystem services become commonplace (see Winder et al., 2021), we recommend that future studies exploring human-nature interactions through automated content analysis of social media photographs, for cultural ecosystem services assessments or landscape planning, will not treat the choice of the computer vision service to rely on as purely dictated by convenience considerations and rather either rely on multiple services or at least test the suitability of different services for the purpose at hand before full implementation.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.landurbplan.2021.104261.

References

Sottini, V. A., Barbierato, E., Bernetti, I., Capecchi, I., Fabbrizzi, S., & Menghini, S. (2019). Winescape perception and big data analysis: An assessment through social media photographs in the Chianti Classico region. *Wine Economics and Policy, 8*(2), 127–140.

Alemu I, J. B., Richards, D. R., Gaw, L.-F., Masoudi, M., Nathan, Y., & Friess, D. A. (2021). Identifying spatial patterns and interactions among multiple ecosystem services in an urban mangrove landscape. *Ecological Indicators, 121*, 107042. https://doi.org/10.1016/j.ecolind.2020.107042

Al-Omair, O. M., & Huang, S. (2018). A comparative study on detection accuracy of cloud-based emotion recognition services. In *In Proc. of 2018 International Conference on Signal Processing and Machine Learning* (pp. 142–148).

Amorim Maia, A. T., Calcagni, F., Connolly, J. J. T., Anguelovski, I., & Langemeyer, J. (2020). Hidden drivers of social injustice: Uncovering unequal cultural ecosystem services behind green gentrification. *Environmental Science & Policy, 112*, 254–263.

Angradi, T. R., Launspach, J. J., & Debbout, R. (2018). Determining preferences for ecosystem benefits in Great Lakes Areas of Concern from photographs posted to social media. *Journal of Great Lakes Research, 44*(2), 340–351.

Azzalini, A., & Menardi, G. (2014). Clustering via nonparametric sensity estimation: The R package pdfCluster. *Journal of Statistical Software, 57*(11), 1–26.

Bodenhofer, U., Kothmeier, A., & Hochreiter, S. (2011). APCluster: An R package for affinity propagation clustering. *Bioinformatics, 27*(17), 2463–2464.

Bubalo, M., van Zanten, B. T., & Verburg, P. H. (2019). Crowdsourcing geo-information on landscape perceptions and preferences: A review. *Landsc. Urban Plan., 184*, 101–111.

Buijs, A. (2009). Different theoretical approaches to study the human-nature relationship. *Public natures: Social representations of nature and local practices (Doctoral dissertation)*. The Netherlands: Wageningen University.

Calcagni, F., Maia, A. T. A., Connolly, J. J. T., & Langemeyer, J. (2019). Digital co-construction of relational values: Understanding the role of social media for sustainability. *Sustainability Science, 14*, 1309–1321.

Callau, A., Albert, M., Rota, J., & Giné, D. (2019). Landscape characterization using photographs from crowdsourced platforms: Content analysis of social media photographs. *Open Geosciences, 11*(1), 558–571.

Connors, J. P., Lei, S., & Kelly, M. (2012). Citizen science in the age of neogeography: Utilizing volunteered geographic information for environmental monitoring. *Annals of the Association of American Geographers, 102*(6), 1267–1289.

de Juan S, Ospina-Alvarez A, Villasante S, Ruiz-Frau A. 2020, Wide-scale assessment of cultural ecosystem services in coastal areas using graph theory on social media data. arXiv preprint arXiv:2007.14308. Jul 22.

Depietri, Y., & Orenstein, D. E. (2020). Managing fire risk at the wildland-urban interface requires reconciliation of tradeoffs between regulating and cultural ecosystem services. *Ecosystem Services, 44*, Article 101108.

Depietri Y, Ghermandi A, Campisi-Pinto S, Orenstein D (2019) Participatory GIS and geolocated social media data to assess cultural ecosystem services in Haifa (Israel): a methodological comparison. Proc. of 10th ESP World Conference, 21-25 October 2021, Hannover, Germany.

Dodge S, Karam L (2016) Understanding how image quality affects deep neural networks. In 2016 8th International Conference on Quality of Multimedia Experience (QoMEX), pp. 1-6. IEEE.

Donahue, M. L., Keeler, B. L., Wood, S. A., Fisher, D. M., Hamstead, Z. A., & McPhearson, T. (2018). Using social media to understand drivers of urban park visitation in the Twin Cities, MN. *Landscape Urban Planning, 175*, 1–10.

Donaire, J. A., Camprubí, R., & Galí, N. (2014). Tourist clusters from Flickr travel photography. *Tourism Manage Perspect, 11*, 26–33.

Egorova, E. (2020). Using textual VGI to model aspects of nature-based recreation activities in Aotearoa New Zealand. *Journal of Spatial Information Science*. DOI: 10.5311/JOSIS.0.0.687.

Foltête, J.-C., Ingensand, J., & Blanc, N. (2020). Coupling crowd-sourced imagery and visibility modelling to identify landscape preferences at the panorama level. *Landscape and Urban Planning, 197*, 103756. https://doi.org/10.1016/j.landurbplan.2020.103756

Frey, B. J., & Dueck, D. (2007). Clustering by passing messages between data points. *Science, 315*(5814), 972–976.
Ghermandi, A., Sinclair, M., Fichtman, E., & Gish, M. (2020). Novel insights on intensity and typology of direct human-nature interactions in protected areas through passive crowdsourcing. *Global Environmental Changes, 65*, 102189. https://doi.org/10.1016/j.gloenvcha.2020.102189
Ghermandi A, Camacho-Valdez V (2020b) Trejo-Espinosa H. Social media-based analysis of cultural ecosystem services and heritage tourism in a coastal region of Mexico. Tourism Management 77:104002.
Ghermandi, A., & Sinclair, M. (2019). Passive crowdsourcing of social media in environmental research: A systematic map. *Global Environmental Change, 55*, 36–47.
Ghermandi, A. (2016). Analysis of intensity and spatial patterns of public use in natural treatment systems using geotagged photos from social media. *Water Research, 105*, 297–304.
Gosal, A. S., & Ziv, G. (2020). Landscape aesthetics: Spatial modelling and mapping using social media images and machine learning. *Ecological Indicators, 117*, 106638. https://doi.org/10.1016/j.ecolind.2020.106638
Guerrero, P., Møller, M. S., Olafsson, A. S., & Snizek, B. (2016). Revealing cultural ecosystem services through Instagram images: The potential of social media volunteered geographic information for urban green infrastructure planning and governance. *Urban Planning., 1*(2), 1–7.
Hamstead, Z. A., Fisher, D., Ilieva, R. T., Wood, S. A., McPhearson, T., & Kremer, P. (2018). Geolocated social media as a rapid indicator of park visitation and equitable park access. *Computers, Environment and Urban Systems, 72*, 38–50.
Hausmann, A., Toivonen, T., Slotow, R., Tenkanen, H., Moilanen, A., Heikinheimo, V., & Di Minin, E. (2018). Social media data can be used to understand tourists' preferences for nature-based experiences in protected areas. *Conservation Letters, 11* (1), e12343. https://doi.org/10.1111/conl.12343
Heikinheimo, V., Tenkanen, H., Bergroth, C., Järv, O., Hiippala, T., & Toivonen, T. (2020). Understanding the use of urban green spaces from user-generated geographic information. *Landscape Urban Planning, 201*, 103845. https://doi.org/10.1016/j.landurbplan.2020.103845
Heikinheimo, V., Di Minin, E., Tenkanen, H., Hausmann, A., Erkkonen, J., & Toivonen, T. (2017). User-generated geographic information for visitor monitoring in a national park: A comparison of social media data and visitor survey. *ISPRS International Journal of Geo-Information, 6*(3), 85.
Hubert, L., & Arabie, P. (1985). Comparing partitions. *Journal of Classification, 2*(1), 193–218.
Islam J, Zhang Y (2016) Visual sentiment analysis for social images using transfer learning approach. In 2016 IEEE International Conferences on Big Data and Cloud Computing, Social Computing and Networking, Sustainable Computing and Communications, pp. 124-130. IEEE.
Karasov, O., Heremans, S., Külvik, M., Domnich, A., & Chervanyov, I. (2020). On how crowdsourced data and landscape organisation metrics can facilitate the mapping of cultural ecosystem services: An Estonian case study. *Land, 9*(5), 158.
Kassambara A, Mundt F. factoextra: Extract and visualize the results of multivariate data analyses R package version 1 2019 6 https://CRAN.R-project.org/package=factoextra (accessed 26 January 2021).
Komossa, F., Wartmann, F. M., Kienast, F., & Verburg, P. H. (2020). Comparing outdoor recreation preferences in peri-urban landscapes using different data gathering methods. *Landscape and Urban Planning, 199*, 103796. https://doi.org/10.1016/j.landurbplan.2020.103796
Kyriakou, K., Barlas, P., Kleanthous, S., & Otterbacher, J. (2019). Fairness in proprietary image tagging algorithms: A cross-platform audit on people images. *ICWSM, 13*, 313–322.
Landauer, T. K., Foltz, P. W., & Laham, D. (1998). An introduction to latent semantic analysis. *Discourse Processes, 25*(2-3), 259–284.
Langemeyer, J., Calcagni, F., & Baró, F. (2018). Mapping the intangible: Using geolocated social media data to examine landscape aesthetics. *Land Use Policy, 77*, 542–552.
Latta, S. C., Howell, C. A., Dettling, M. D., & Cormier, R. L. (2012). Use of data on avian demographics and site persistence during overwintering to assess quality of restored riparian habitat. *Conservation Biology, 26*, 482–492.
Lee, H., Seo, B., Koellner, T., & Lautenbach, S. (2019). Mapping cultural ecosystem services 2.0–Potential and shortcomings from unlabeled crowd sourced images. *Ecological Indicators, 96*, 505–515.
Lotan, A., Kost, R., Mandelik, Y., Peled, Y., Chakuki, D., Zemah Shamir, S., & Ram, Y. (2018). National scale mapping of ecosystem services in Israel – genetic resources, pollination and cultural services. *One Ecosyst., 3*, e25494. https://doi.org/10.3897/oneeco.3.e25494
Lotan, A., Safriel, U., & Feitelson, E. (Eds.). (2017). *Israel's National Nature Assessment Program. The Steinhardt Museum of Natural History*. Tel-Aviv University (in Hebrew).
Martínez Pastur, G., Peri, P. L., Lencinas, M. V., García-Llorente, M., & Martín-López, B. (2016). Spatial patterns of cultural ecosystem services provision in Southern Patagonia. *Landscape Ecology, 31*(2), 383–399.
Maechler, M., Rousseeuw, P., Struyf, A., Hubert, M., Hornik, K. (2019) cluster: Cluster analysis basics and extensions R package version 2 1 2019. https://cran.r-project.org/web/packages/cluster/cluster.pdf (accessed 31 December 2019).
McHugh, M. L. (2013). The Chi-square test of independence. *Biochemia Medica, 23*(2), 143–149.
McLeod AI (2011) Kendall: Kendall rank correlation and Mann-Kendall trend test. R package version 2.2. https://CRAN.R-project.org/package=Kendall.
Muñoz, L., Hausner, V. H., & Monz, C. A. (2019). Advantages and limitations of using mobile apps for protected area monitoring and management. *Society and Natural Resources, 32*, 473–488.
Nilsson, K., & Jönsson, H. E. (2019). A comparison of image and object level annotation performance of image recognition cloud services and custom Convolutional Neural Network models (Dissertation). Retrieved from http://urn.kb.se/resolve?urn=urn:nbn:se:bth-18074.
Oteros-Rozas, E., Martin-Lopez, B., Fagerholm, N., Bieling, C., & Plieninger, T. (2017). Using social media photos to explore the relation between cultural ecosystem services and landscape features across five European sites. *Ecological Indicators, 94* (2), 74–86.
Payntar, N. D., Hsiao, W.-L., Covey, R. A., & Grauman, K. (2020). Learning patterns of tourist movement and photography from geotagged photos at archaeological heritage sites in Cuzco. *Peru. Tourism Management, 82*, 104165. https://doi.org/10.1016/j.tourman.2020.104165
R Core Team R: A language and environment for statistical computing 2020 R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.
Richards, D. R., & Tunçer, B. (2018). Using image recognition to automate assessment of cultural ecosystem services from social media photographs. *Ecosystem Services, 31*, 318–325.
Richards, D. R., & Friess, D. A. (2015). A rapid indicator of cultural ecosystem service usage at a fine spatial scale: Content analysis of social media photographs. *Ecological Indicators, 53*, 187–195.
Ros Candeira, A., Moreno Llorca, R., Alcaraz Segura, D., Bonet García, F. J., & Vaz, A. S. (2020). Social media photo content for Sierra Nevada: A dataset to support the assessment of cultural ecosystem services in protected areas. *Nature Conservation, 38*, 1–12.
Ruiz-Frau, A., Ospina-Alvarez, A., Villasante, S., Pita, P., Maya-Jariego, I., & de Juan, S. (2020). Using graph theory and social media data to assess cultural ecosystem services in coastal areas: Method development and application. *Ecosystem Services, 45*, 101176. https://doi.org/10.1016/j.ecoser.2020.101176
Runge, C. A., Hausner, V. H., Daigle, R. M., & Monz, C. A. (2020). Pan-Arctic analysis of cultural ecosystem services using social media and automated content analysis. *Environmental Research Communications, 2*(7), 075001. https://doi.org/10.1088/2515-7620/ab9c33
Seresinhe, C. I., Moat, H. S., & Preis, T. (2018). Quantifying scenic areas using crowdsourced data. *Environment and Planning B: Urban Analytics and City Science, 45* (3), 567–582.
Sinclair, M., Mayer, M., Woltering, M., & Ghermandi, A. (2020). Using social media to estimate visitor provenance and patterns of recreation in Germany's national parks. *Journal of Environment Management, 263*, Article 110418.
Song, X. P., Richards, D. R., He, P., & Tan, P. Y. (2020). Does geo-located social media reflect the visit frequency of urban parks? A city-wide analysis using the count and content of photographs. *Landscape and Urban Planning, 203*, 103908. https://doi.org/10.1016/j.landurbplan.2020.103908
Temel, D., Lee, J., & AlRegib, G. (2019). Object recognition under multifarious conditions: A reliability analysis and a feature similarity-based performance estimation. In *In 2019 IEEE International Conference on Image Processing (ICIP)* (pp. 3033–3037).
Tenerelli, P., Püffel, C., & Luque, S. (2017). Spatial assessment of aesthetic services in a complex mountain region: Combining visual landscape properties with crowdsourced geographic information. *Landscape Ecology, 32*(5), 1097–1115.
Tenkanen, H., Di Minin, E. D., Heikinheimo, V., Hausmann, A., Herbst, M., & Kajala, L. (2017). Toivonen T (2017) Instagram, Flickr, or Twitter: Assessing the usability of social media data for visitor monitoring in protected areas. *Scientific Reports, 7*, 1–11.
Tian, T., Sun, L., Peng, S., Sun, F., & Che, Y. (2021). Understanding the process from perception to cultural ecosystem services assessment by comparing valuation methods. *Urban Forestry and Urban Greening, 57*, 126945. https://doi.org/10.1016/j.ufug.2020.126945
Tieskens, K. F., Van Zanten, B. T., Schulp, C. J. E., & Verburg, P. H. (2018). Aesthetic appreciation of the cultural landscape through social media: An analysis of revealed preference in the Dutch river landscape. *Landscape and Urban Planning, 177*, 128–137.
Toivonen, T., Heikinheimo, V., Fink, C., Hausmann, A., Hiippala, T., Järv, O., … Di Minin, E. (2019). Social media data for conservation science: A methodological overview. *Biological Conservation, 233*, 298–315.
Truong, Q. T., & Lauw, H. W. (2017). Visual sentiment analysis for review images with item-oriented and user-oriented CNN. In *In Proceedings of the 25th ACM international conference on Multimedia* (pp. 1274–1282).
Van Berkel, D. B., Tabrizian, P., Dorning, M. A., Smart, L., Newcomb, D., Mehaffey, M., … Meentemeyer, R. K. (2018). Quantifying the visual-sensory landscape qualities that contribute to cultural ecosystem services using social media and LiDAR. *Ecosystem Services, 31*, 326–335.
Winder S, Lee H, Seo B, Lia E, Wood S (2021) Landscape-scale insights into recreational activities derived from image content and an open-source classifier. Proceedings of the 3rd ESP Europe Conference, 7-10 June 2021, Tartu, Estonia.
Wood, S. A., Guerry, A. D., Silver, J. M., & Lacayo, M. (2013). Using social media to quantify nature-based tourism and recreation. *Scientific Reports, 3*, 1–7.
Zhang, S., & Zhou, W. (2018). Recreational visits to urban parks and factors affecting park visits: Evidence from geotagged social media data. *Landscape and Urban Planning, 180*, 27–35.
Zielstra, D., & Hochmair, H. H. (2013). Positional accuracy analysis of Flickr and Panoramio images for selected world regions. *Journal of Spatial Science, 58*, 251–273.