How to address data privacy concerns when using social media data in conservation science

Di Minin, Enrico

Special Section: Contributed Paper

Exploring human–nature interactions in national parks with social media photographs and computer vision

Tuomas Väisänen [ID],[1,2] * Vuokko Heikinheimo [ID],[1,2] Tuomo Hiippala [ID],[1,3,2] and Tuuli Toivonen [ID][1,4,2]

[1]Digital Geography Lab, Department of Geosciences and Geography, University of Helsinki, Helsinki, 00014, Finland
[2]Helsinki Institute of Sustainability Science, University of Helsinki, Helsinki, 00014, Finland
[3]Department of Languages, University of Helsinki, Helsinki, 00014, Finland
[4]Conservation Science Group, University of Cambridge, Cambridge, CB2 3EJ, U.K.

Abstract: Understanding the activities and preferences of visitors is crucial for managing protected areas and planning conservation strategies. Conservation culturomics promotes the use of user-generated online content in conservation science. Geotagged social media content is a unique source of in situ information on human presence and activities in nature. Photographs posted on social media platforms are a promising source of information, but analyzing large volumes of photographs manually remains laborious. We examined the application of state-of-the-art computer-vision methods to studying human–nature interactions. We used semantic clustering, scene classification, and object detection to automatically analyze photographs taken in Finnish national parks by domestic and international visitors. Our results showed that human–nature interactions can be extracted from user-generated photographs with computer vision. The different methods complemented each other by revealing broad visual themes related to level of the data set, landscape photogeneity, and human activities. Geotagged photographs revealed distinct regional profiles for national parks (e.g., preferences in landscapes and activities), which are potentially useful in park management. Photographic content differed between domestic and international visitors, which indicates differences in activities and preferences. Information extracted automatically from photographs can help identify preferences among diverse visitor groups, which can be used to create profiles of national parks for conservation marketing and to support conservation strategies that rely on public acceptance. The application of computer-vision methods to automatic content analysis of photographs should be explored further in conservation culturomics, particularly in combination with rich metadata available on social media platforms.

Keywords: computer vision, deep learning, feature extraction, Flickr, human–nature interaction, national parks, object recognition, photography, preferences, visitor monitoring

Exploración de las Interacciones Humano-Naturaleza en los Parques Nacionales por Medio de Fotografías en Redes Sociales y Visión por Computadora

Resumen: La comprensión de las actividades y preferencias de los visitantes es crucial para el manejo de las áreas protegidas y la planeación de las estrategias de conservación. La culturomia de la conservación promueve el uso del contenido en línea generado por usuarios en las ciencias de la conservación. El contenido de redes sociales etiquetado geográficamente es una fuente única de información in situ sobre la presencia humana y sus actividades

en la naturaleza. Las fotografías publicadas en las redes sociales son una fuente prometedora de información, aunque el análisis manual de grandes volúmenes de fotografías sigue siendo laborioso. Evaluamos la aplicación de algunos métodos de punta de visión por computadora para estudiar las interacciones humano-naturaleza. Usamos agrupaciones semánticas, clasificación de escenas y detección de objetos para analizar automáticamente las fotografías tomadas por visitantes domésticos e internacionales dentro de los parques nacionales finlandeses. Nuestros resultados mostraron que las interacciones humano-naturaleza pueden extraerse de fotografías generadas por usuarios mediante la visión por computadora. Los diferentes métodos se complementaron unos a otros al revelar temas visuales generalizados relacionados con el nivel del conjunto de datos, fotogeneidad del paisaje y las actividades humanas. Las fotografías geoetiquetadas revelaron unos perfiles regionales distintos para los parques nacionales (p. ej.: preferencias en los paisajes y las actividades), que son potencialmente útiles para el manejo de los parques. El contenido fotográfico difirió entre los visitantes domésticos y los internacionales, lo cual indica diferencias en sus actividades y preferencias. La información extraída automáticamente de las fotografías puede ayudar a identificar las preferencias entre los grupos diversos de visitantes, lo cual puede usarse para crear un perfil de cada parque nacional para su uso en el mercadeo de la conservación y para apoyar a las estrategias de conservación que dependen de la aceptación pública. La aplicación de los métodos de visión por computadora al análisis automático de contenido de las fotografías debería explorarse mucho más en la culturomia de la conservación, particularmente en combinación con la riqueza de metadatos disponibles en las plataformas sociales.

Palabras Clave: aprendizaje profundo, extracción de características, Flickr, fotografía, interacción humano-naturaleza, monitoreo de visitantes, parques nacionales, preferencias, reconocimiento de objetos, visión por computadora

摘要： 了解游客的活动及喜好对保护区管理和保护策略的制定至关重要。保护文化组学提倡在保护科学中使用用户生成发布的在线内容, 而有地理标签的社交媒体内容正是人们在自然中出现和活动的现场信息的独特来源。虽然发布在社交媒体平台上的照片是潜在的信息来源, 但人工分析大量照片仍十分费力。本研究探索了最先进的计算机视觉方法在研究人与自然互动方面的应用。我们使用语义聚类、情景分类和目标检测等方法对国内外游客在芬兰国家公园拍摄的照片进行了自动分析, 结果表明可以用计算机视觉从用户生成发布的照片中提取人与自然的互动信息。不同方法通过揭示与数据集水平、景观摄影效果和人类活动相关的广泛的视觉主题而相互补充。带有地理标记的照片展示了国家公园不同区域的情况 (如人们对景观和活动的偏好), 这可以用于国家公园的管理。国内外游客摄影内容的差异也体现了他们活动内容和喜好的差异。从照片中自动提取的信息可以帮助确定不同游客群体的偏好, 这可以用来构建国家公园的资料以用于保护宣传, 还可以支持依赖于公众接受的保护策略。我们认为, 保护文化组学应进一步探索计算机视觉方法在自动分析照片内容中的应用, 特别是与社交媒体平台上丰富的元数据相结合。**[翻译: 胡怡思; 审校: 聂永刚]**

关键词： 国家公园, Flickr网站, 计算机视觉, 深度学习, 特征抽取, 物体识别, 人与自然的互动, 偏好, 游客监控, 摄影

Introduction

Protected areas are considered a cornerstone for protecting species and ecosystems (Watson et al. 2014). In many countries, iconic national parks act as the flagships of the protected-area network. Historically, the establishment of national parks originated from the desire to preserve scenic landscape areas of national or regional importance (Lee 1972; Schullery & Whittlesey 2003). Even today, protected-area visitor rates are associated with access to scenic landscapes, available visitor activities, and biodiversity values (Neuvonen et al. 2010; Siikamäki et al. 2015; Hausmann et al. 2017). Recreational use of conservation areas may directly or indirectly help fund conservation on-site and gain political support for conservation (Di Minin et al. 2013; Whitelaw et al. 2014; Balmford et al. 2015), even if the relationship between recreational visits and nature conservation is sometimes complex (Bateman & Fleming 2017; Buckley 2018).

To develop the recreational use of protected areas in line with conservation goals, protected-area organizations in many countries actively gather visitor informa-tion. Depending on the organization, visitor information may be collected using registration forms at the park entrance, placing counters along the paths, or by conducting surveys or interviews on-site or online (Pietilä & Fagerholm 2019). Information about different groups of visitors may then be used to guide national park management and marketing actions, as well as conservation strategies (Kruger et al. 2017).

Although it is acknowledged that understanding the human dimensions of environmental issues supports nature conservation (Bennett et al. 2017; Sutherland et al. 2018), traditional on-site approaches for collecting information are time-consuming and costly. Therefore, user-generated online content is increasingly used as an information source in conservation science under the emerging subfield of conservation culturomics (Arts et al. 2015; Di Minin et al. 2015; Ladle et al. 2016) and the interest is also increasing among practitioners.

Social media data are a particularly interesting source of information for understanding human–nature interactions because they provide spatially and temporally explicit data on visits, together with rich textual and

visual content (Toivonen et al. 2019). Although the textual content analysis can provide useful insights on nature conservation (Ladle et al. 2016), there have been calls for increased attention to visual communication in conservation culturomics research (Sherren et al. 2017; Ghermandi & Sinclair 2019). Focusing on photographs as the source of information allows the challenges arising from language and limitations of textual analysis to be avoided (Carter et al. 2013).

Social media photographs have already proven useful, for example, for obtaining information on visitor preferences or activities in national parks (Heikinheimo et al. 2017; Hausmann et al. 2018) and on cultural ecosystem services across landscapes (Richards & Friess 2015; Van Berkel et al. 2018; Pickering et al. 2020). Laborious manual analyses of photographs are now complemented by automated visual content analysis methods. State-of-the art computer-vision methods allow for information to be extracted from large volumes of photographs by classifying the content into predefined classes (such as landscapes), by recognizing discrete objects (such as species), or by grouping together similar images for human analysts. These approaches have recently been used to monitor species (Sharma et al. 2018) and to examine aesthetic preferences (Seresinhe et al. 2017, 2018) and human activities and preferences (Richards & Tunçer 2018; Gosal et al. 2019; Koylu et al. 2019).

We aimed to contribute to the application of computer-vision methods to visual content analysis in protected-area visitor monitoring. We evaluated the applicability of 3 computer-vision methods for extracting information on human–nature interactions in national parks with social media photographs. We aimed to answer questions that are typically analyzed by visitor surveys, such as the preferences of different visitor groups or geographical differences of activities. Our study area was Finnish national parks, and we used Flickr data for our exploration. We sought to answer the following questions: What information can state-of-the-art computer-vision methods extract from social media photographs? Do different visitor groups share different types of photographs from national parks? How does photographic content vary between different types of national parks?

To answer to our questions, we collected geotagged Flickr data from the 20 most popular national parks in Finland. We classified the users into national and international visitors based on their profile information. We applied t3 computer-vision methods to the photographs, namely, semantic clustering of photographic content, scene classification, and instance-level object detection and evaluated their applicability to visitor monitoring of protected areas. Using our findings, we considered the potential and challenges of using social media photographs and computer-vision methods to understand the use of and values associated with protected areas and in conservation more broadly.

Methods

Study Area

Finland has 40 national parks located from hemiboreal coastal zone to the tundra of the northernmost parts of Lapland. Visitor numbers are rising steadily. In 2019, the parks received more than 3.2 million visitors (https://www.metsa.fi/web/en/visitationnumbers). In 2020, the numbers have surged due to the COVID-19 crisis and people wishing to visit nature. The parks are managed by Parks and Wildlife Finland (*Metsähallitus*). The organization has systematically collected information on national park use, activities, and preferences for several decades (Kajala et al. 2007) and established profiles of the parks (broader description in Appendix S1). Because the natural and seminatural landscapes are relatively similar throughout Finland, we wanted to see if the computer-vision methods used could reveal differences between national parks located across different landscape regions. We focused our analysis on 20 popular national parks based on the availability of Flickr photographs. We grouped the selected national parks into 4 broad landscape categories for further analysis (Fig. 1).

Downloading Flickr Data

Flickr is a social media platform for sharing images and video, and it is particularly popular among professional photographers and nature enthusiasts (Di Minin et al. 2015). The Flickr API allows open access to Flickr content in compliance with the restrictions set by photo owners (https://www.flickr.com/help/terms/api). Geotagged Flickr posts correspond relatively well to the popularity of Finnish national parks (Tenkanen et al. 2017). We used data from Flickr instead of other platforms (such as Twitter or Instagram) because its terms of service allow for the application of computer-vision methods to analyze the visual content of photographs (Toivonen et al. 2019).

First, we searched the Flickr API (https://www.flickr.com/services/api/) for all geotagged Flickr posts located within 500 m of all Finnish national parks ($n = 40$) in January 2019. This returned 14,585 geotagged posts uploaded by 969 unique users from 2002 to 2019. Second, we downloaded the images at their original size up to the highest available size allowed by the application programing interface (1024×768 pixels). In total, 13,363 images were available for download. Finally, we selected 20 parks with the highest Flickr post counts (>100) as our final data set for content analysis: 12,759 images uploaded by 824 unique users. The amount of Flickr data and Finnish national park visitor counts are available in Appendix S2.
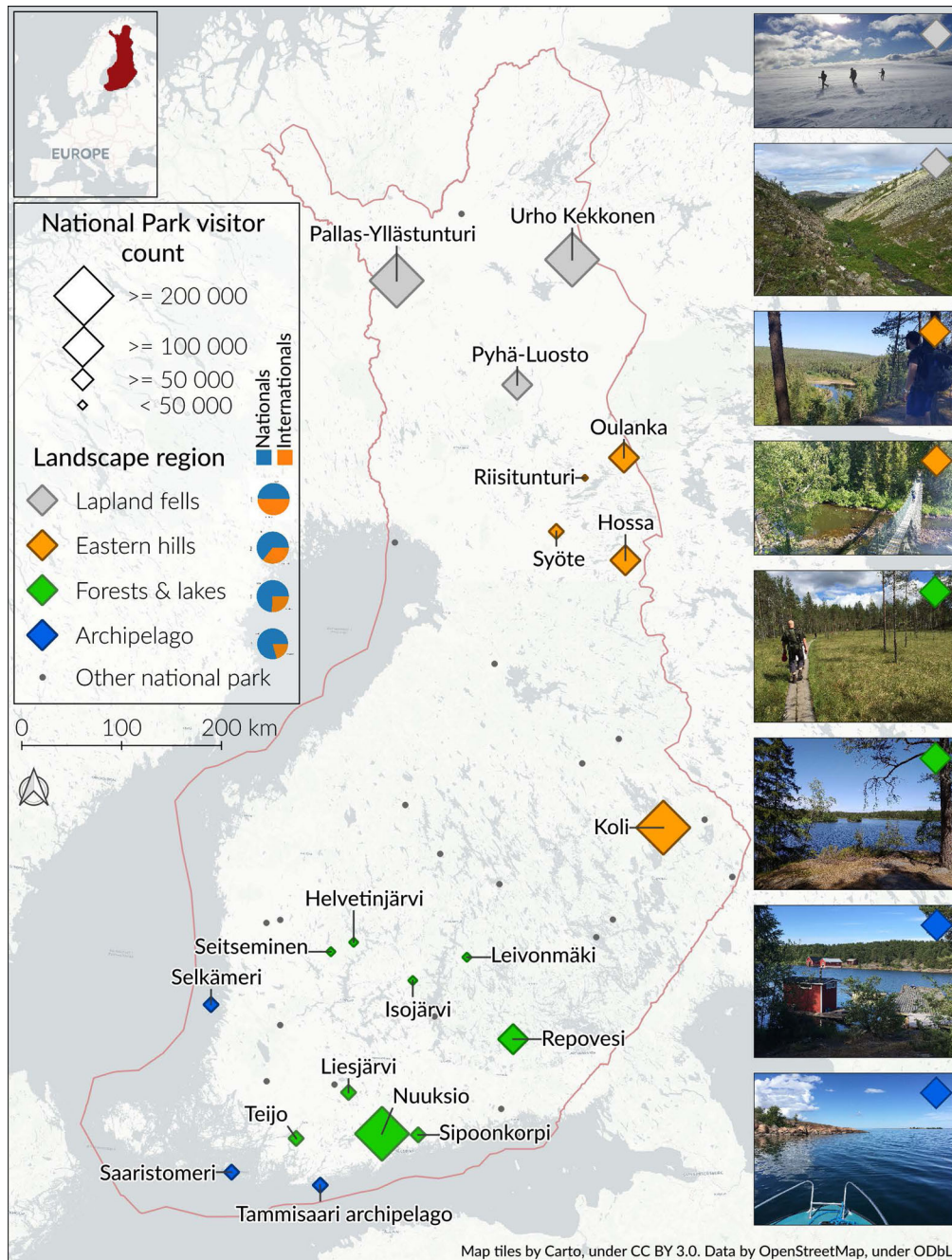
Figure 1. Twenty most popular Finnish national parks on Flickr, general landscape region of these parks, other national parks (dots), official annual visitor counts, and proportion of Flickr posts made by local and international visitors.

User Classification

We manually classified the 824 unique users in the data set as national (from Finland) or international and by gender based on the information available in public profiles. We detected the probable country of residence for each user, primarily based on the self-reported home location in the user profile. If the user had not reported their home location, we combined information from the user's name, profile descriptions, and linked websites to determine the country of residence. In some cases, forenames and surnames can give a good indication of the geographic region of origin (Longley et al. 2015), particularly when combined with other information. If profile information was not sufficient, we also considered the geographic distribution of photographs for determining the home location. For example, we classified users as locals if they mentioned a Finnish hometown in the profile description, used the

Table 1. Computer-vision methods, training data sets, description of output, and purpose of each in examining human–nature interactions in photographic content from 20 Finnish national parks posted by Flickr users.

Method	Algorithms	Training data set	Output	Purpose
Semantic clustering	ResNeXt101 (Xie et al. 2017)	ImageNet (Russakovsky et al. 2015)	2048-dimensional vector representing the semantic content of the photograph, which is then reduced to 2 dimensions with UMAP	clustering of photos that are semantically similar with each other (see Results & Fig. 2)
Scene classification	VGG16 (Simonyan & Zisserman 2015)	Places365 (Zhou et al. 2018)	probability distribution over 365 classes in Places365 for the input image; class label with the highest probability used	landscape or scene classification of each photograph (see Results, Fig. 4, & Table 2)
Instance-level object detection	Mask R-CNN (He et al. 2017)	MSCOCO (Lin et al. 2014)	probability distribution over 80 classes in MS COCO and a polygon indicating the location and shape of each object detected in the photograph only objects detected with a probability of 70% or higher included	objects identified in images (see Results, Table 2, & Appendix S11)

Finnish language, had a distinctively Finnish name, or had their personal web page under a Finnish domain. Similarly, we classified users as internationals if the profile information referred to a place of residence outside of Finland. We recorded gender as male or female based on the username, profile picture, and other available information. For some users, it was not possible to detect the home location or gender due to limited or ambiguous information.

Automating Content Analysis with Computer Vision

We used 3 computer-vision methods for automatic visual content analysis of photographs taken at national parks. These methods use deep neural networks, a family of machine learning algorithms (LeCun et al. 2015). Semantic clustering involves using a pretrained neural network to extract a high-dimensional feature vector that represents the semantic content of the photograph, whose dimensionality is then reduced to enable plotting low-dimensional representations to explore similarities and differences between photographs and their contents. We used a neural network trained to classify objects into 1000 categories as a feature extractor. Scene classification involves classifying photographs into predefined categories, providing a set of potential category labels and their associated probabilities (Zhou et al. 2018). Instance-level object detection detects individual instances of objects belonging to predefined categories and their locations in the photograph (He et al. 2017). This method returns the predicted label of the object, its associated probability, and its predicted location in the photograph. The computer-vision methods are summarized in Table 1.

All images required preprocessing because the computer-vision methods used required the input size to be of fixed dimensions. Images were resized to 224×224 pixels for feature extraction and scene classification, and to 512×512 pixels for instance-level object detection. Because most images did not have an aspect ratio of 1:1 (equal height and width), we resized the images to a fixed height of 224 or 512 pixels before cropping 224 or 512 pixels in the middle of the image. This kind of center crop, which assumes that the most important content is centered in the photograph, preserves the shape of objects in the image because the aspect ratio is not altered, although some objects at the edges of the photograph may be lost during preprocessing.

Semantic Clustering

We evaluated several neural network architectures and pretrained models for semantic clustering. The neural network architectures included VGG16 (Simonyan & Zisserman 2015), NASNet (Zoph et al. 2018), Xception (Chollet 2017), ResNet50 (He et al. 2016), and ResNeXt101 (Xie et al. 2017), which were trained to classify images into the 1000 object categories in the ImageNet Large Scale Visual Recognition Challenge (Russakovsky et al. 2015). We evaluated the performance of each architecture and model qualitatively by extracting high-dimensional feature vectors, the size of which ranged from 512 to 2048 dimensions. We then used UMAP (uniform manifold approximation and projection) (McInnes et al. 2018), a dimensionality reduction algorithm, to reduce the feature vectors to 2 dimensions for visualization. The UMAP algorithm reduces dimensionality by learning to map points between high- and low-dimensional spaces, while attempting to preserve the structure of the high-dimensional graph (McInnes et al. 2018). Based on a qualitative evaluation of the results, we chose the ResNeXt101 model for feature extraction
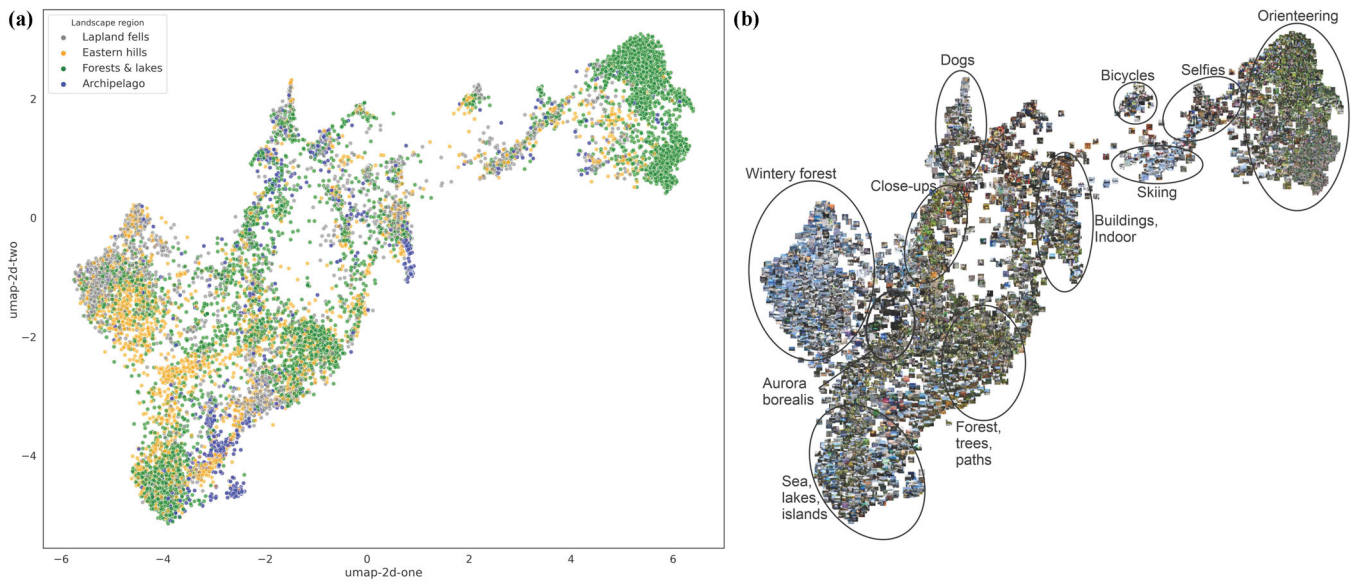
Figure 2. Flickr photos placed in UMAP (unifom manifold approximation and projection)-plots (a) as dots color-coded by their landscape regions (see Fig. 1) (images similar in their vector representation are placed close to each other) and (b) as thumbnails and ellipses delineating clusters of similar photo content (labeled by the authors after visual examination). See Appendix S5 for a larger version of (b).

to be used together with UMAP to semantically cluster the photographs based on their content.

Scene Classification

For scene classification, we used a neural network with the VGG16 architecture (Simonyan & Zisserman 2015) trained on the Places365 data set (Zhou et al. 2018) and implemented by Kalliatakis (2017). Places365 is a data set that contains 1.8 million images belonging to 365 indoor and outdoor scene categories. The data set was developed for scene classification, that is, the task of recognizing the type of a visual scene represented in an image. Places365 is a subset of the larger Places2 database, which contains 10 million images for 434 scene categories. For each image, we retrieved the top-3 predicted labels and their associated probabilities.

Instance-Level Object Detection

For instance-level object detection, we used a neural network with the Mask R-CNN architecture (He et al. 2017) trained on the Microsoft COCO (common objects in context) data set, which features 80 object categories consisting of everyday objects, such as persons, household items, and animals (Lin et al. 2014). We used a Mask R-CNN implementation by Abdulla (2017). Mask R-CNN provides each object detected and segmented from an image with a probability that reflects the confidence of the model about the prediction. To improve the results, we included only object instances detected with a confidence of 0.7 (70%) or higher.

Results

User Groups

Out of the 824 unique users, we identified 62% as locals and 33% as internationals. Visitors were mostly from Europe, the United States, and Japan. For 5% of the users, who contributed 2% of the photographs, the exact country of residence could not be determined, but they were counted as international visitors in the final classification. Gender classification revealed a strong gender bias; 79% of users were men. Only 10% of the users were classified as women, who contributed 4% of the photographs. We could not determine gender for 11% of the user profiles. These profiles either had no clear indication of gender or they were organizational profiles. Overall, local men had posted 72% of all photographs. (Details on origin and gender classification in Appendix S3.)

Automatic Visual Content Analysis with Computer Vision

Semantic clustering with 2-dimensional UMAP representations of original images showed that photographs with similar semantic content were clustered (Fig. 2), indicating that the neural network could extract distinctive high-level semantic information from the photographs. Individual clusters feature photographs of seasonal activities, such as skiing and orienteering; objects, such as dogs, plants, and humans; and landscapes, such as sky with auroras or shoreline views. Photographs with human activities are clustered together, whereas landscapes form their own clusters. Furthermore, photographs of forests during winter and summer, as well as seascapes,
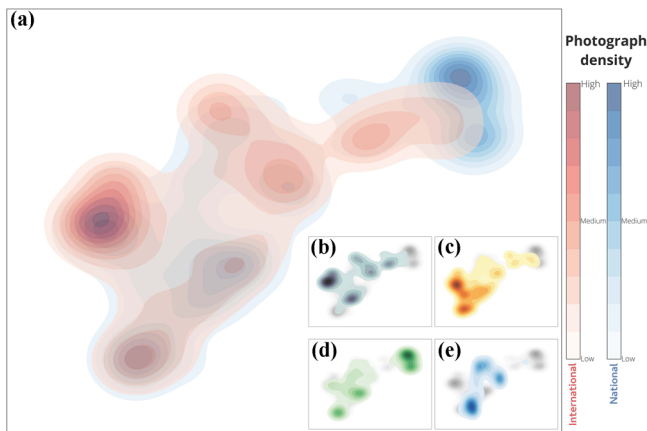
Figure 3. Kernel density estimation plots of image features from 20 most popular national parks presented by user origin and landscape regions: (a) image feature plots for national and international visitors and image feature plots for (b) Lapland Fells, (c) Eastern Hills, (d) Forests and Lakes, and (e) Archipelago regardless of user origin (the darker the color, the denser the cluster of semantically similar photographs).

form individual clusters, to name just a few examples. The photographic content also seems to cluster according to the landscape region. Further visualizations allow drawing comparisons between popular photographic content among national and international visitors, as well as different landscape regions (Fig. 3) and seasons (Appendix S6).

Scene classification predicted a scene category for each photograph (Fig. 4). The classifier identified 325 unique scene categories in the data set. Natural and seminatural scene categories, for example, forest path (9%), broadleaf forest (7%), snowfield (4%), and tundra (4%), were the most common. The 10 most common scene categories were found in 41% of the photographs (Fig. 4). Validating the classification results manually verified that they were mostly meaningful. Distinct scene classes were positioned in distinct areas in the 2-dimensional UMAP plot providing further validation (Appendix S8). The average confidence for the predictions ranged from 0.48 for the most likely category to 0.16 for the second and 0.084 for the third categories. Poorest confidence values were associated with close-ups and portraits of humans, which is not surprising given that the neural network was trained to classify visual scenes, not objects.

Instance-level object detection found objects in 7821 photographs. The most common object was a person, which was detected in approximately 37% of all photographs and in 60% of photographs that were predicted to contain some object. Most photographs contained a single person (49%) followed by 2 (20%) and 3 (10%) persons. Most photographs (4058) contained only

one unique detected object, whereas the maximum number of unique detected objects in a single photograph was 11. To obtain some indication of the activities of users, we looked at the most common objects other than person recognized with instance-level object detection (Table 2 & Appendix S11). Many objects are related to activities, including backpack (present in 13% of photographs with objects), bench (9%), bird (7%), and boat (6%). Objects directly related to sport activities included bicycle (3%), skis (3%), sports ball (2%), frisbee (2%), and kite (1%).

Differences between National and International Visitors

The results revealed differences between national and international visitors to the parks. Photographs taken by the 2 groups largely overlapped each other in the visualization in Fig. 3, which suggests that both national and international visitors take photographs with largely similar content, but certain differences between these 2 groups may be identified by comparing Figs. 2 and 3. For example, almost all photographs of orienteering were taken by national visitors and in the same landscape region (Appendix S13). Photographs of forests taken in the summer were more common among the locals, whereas international visitors shared photographs of forests in the winter, selfies, and skiing. Similar differences appeared in the scene classification results when looking at the most confidently identified scene categories: national visitors post more photographs belonging to forest path and broadleaf forest categories, whereas international visitors shared photographs of ski slope, snowfield, and tundra (Fig. 4).

Objects were detected in 4938 photographs (61%) uploaded by national visitors and 1741 photographs (59%) by international visitors. We identified the most common objects from photographs taken by both groups (Table 2 & Appendix S11). A person was the most common category, present in 3706 photographs (37%) by national and in 1023 photographs (37%) by international visitors. On the average, photographs taken by international visitors featured more people than national visitors, regardless of the season or landscape region, although national visitors feature more persons during summers and in the Forests and Lakes landscape region (Appendix S14). Common objects detected among both visitor groups were related to physical activities (backpack, bicycle, and skis) and eating and picnicking (bench and dining table). Category dog reflects both dog walking and dog sleigh riding. The latter is a popular activity primarily among international visitors (Appendix S12). Other categories that reflect nature photography (bird and potted plant) were more popular among national visitors.

A permutational analysis of variance (PERMANOVA) test with 999 permutations revealed a statistically significant difference ($p < 0.001$) in the semantic clustering of

Table 2. Landscape regions, related social media user counts, and distribution of posts among the most popular scene categories and object classes (excluding person) per region.

Landscape region	Users (% national, international)	Photos (national, international)	Top 5 scene categories, all (%)	Top 5 scene categories, national visitors (%)*	Top 5 scene categories, international visitors (%)	Top 5 objects, all (%)*	Top 5 objects, national visitors (%)	Top 5 objects, international visitors (%)
Lapland Fells	283 (49, 51)	3031 (57, 43)	tundra (10), snowfield (9), ski slope (8), desert or sand (4), broadleaf forest (4)	tundra (7), snowfield (5), ski slope (3), broadleaf forest (3), desert or sand (2)	snowfield (5), ski slope (5), tundra (3), ski resort (2), desert or sand (2)	backpack (6), bench (6), car (6), skis (5), dining table (4)	car (3), bench (3), backpack (3), bicycle (3), skis (3)	backpack (3), bench (3), skis (2), car (2), dog (2)
Eastern Hills	180 (64, 36)	2544 (69, 31)	creek (8), broadleaf forest (8), snowfield (7), forest path (7), ski slope (5)	creek (7), broadleaf forest (6), snowfield (6), forest path (5), ski slope (3)	broadleaf forest (2), forest path (2), ski slope (2), tree farm (2), snowfield (1)	bench (6), backpack (5), boat (4), bird (4), bicycle (3)	bench (4), boat (3), bicycle (3), bird (3), backpack (2)	backpack (3), bench (3), boat (1), bird (1), skis (1)
Forests & Lakes	375 (74, 26)	6158 (91, 9)	forest path (16), broadleaf forest (9), park (5), natural lake (5), swamp (4)	forest path (15), broadleaf forest (8), park (5), natural lake (4), athletic field or outdoor (3)	broadleaf forest (1), forest path (1), natural lake (1), swamp (<1)	backpack (11), bench (6), bird (6), potted plant (4), car (2)	backpack (11), bench (5), bird (3), potted plant (3), Frisbee (2)	bench (1), potted plant (<1), boat (<1), bird (<1), car (<1)
Archipelago	107 (79, 21)	900 (88, 12)	tundra (7), harbor (5), beach (4), lighthouse (3), ocean (3)	tundra (6), harbor (4), lighthouse (3), sky (2), beach (2)	beach (2), lagoon (1), tundra (1), harbor (<1), ocean (<1)	boat (20), bird (12), dining table (5), bench (5), potted plant (5)	boat (17), bird (11), dining table (5), bench (4), potted plant (4)	boat (5), potted plant (1), backpack (1), bird (1), bowl (<1)

*Percentages in scene categories and objects indicate proportion out of all photos in each landscape region.
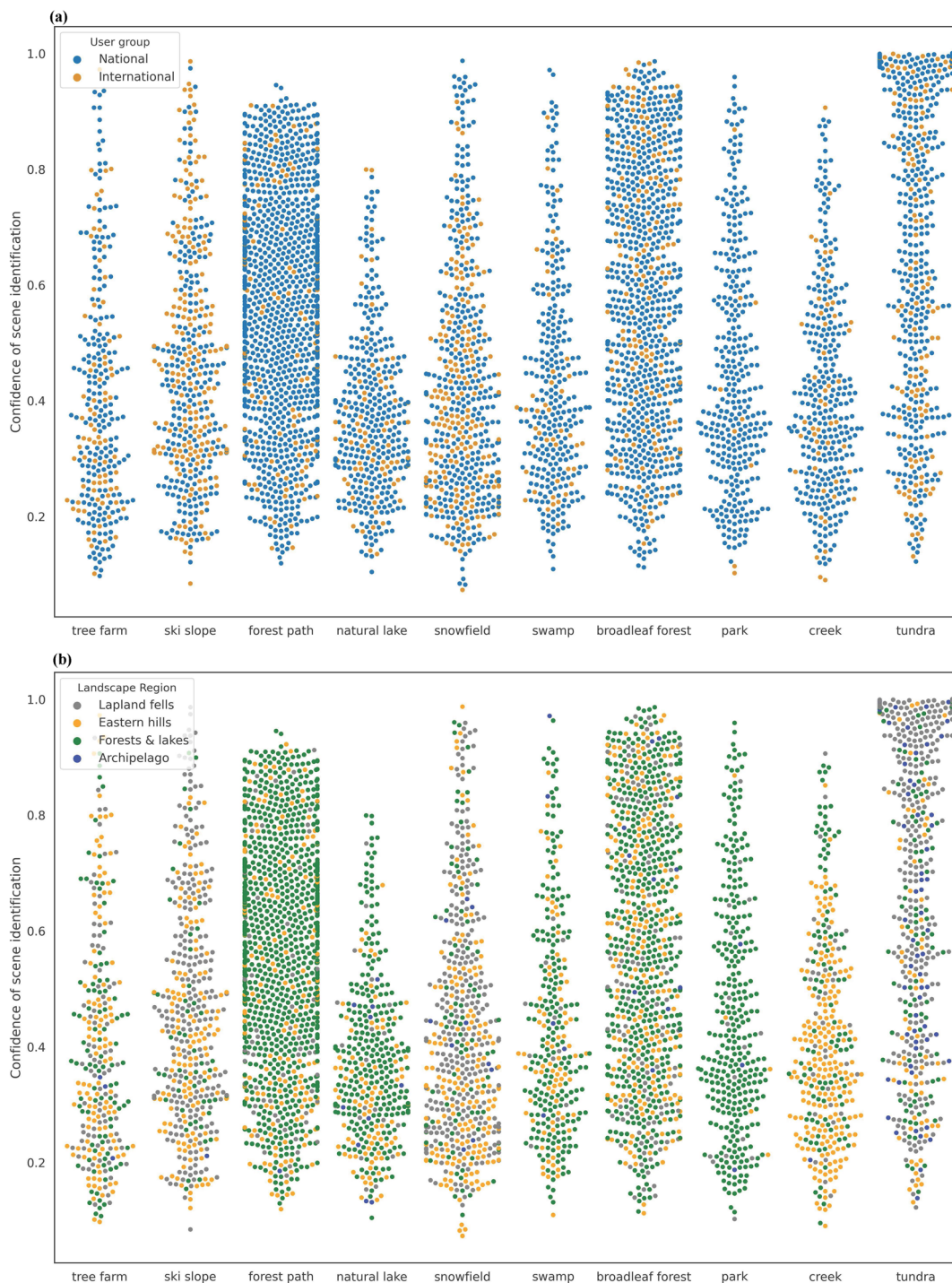
Figure 4. Ten most common and best-identified scene categories in the Flickr photographs from the 20 most popular Finnish national parks: (a) photos taken by national and international visitors and (b) landscape region (Fig. 1) of photos (x-axis, 10 most common classes; y-axis, confidence of scene classification; points, individual images). The classification is based on VGG16 neural network architecture trained on the Places365 data set. For the photo categories, 5276 photos were from the 20 most popular national parks in Finland.

photographs taken by international and national visitors (Appendix S7). Mann–Whitney U test revealed statistically significant ($p < 0.001$) differences between nation-

als and internationals for 3 object categories: dog, backpack, and dining table (Appendix S9), and for a single scene category, forest path ($p < 0.05$) (Appendix S10).

Differences between National Parks in Different Landscape Regions

The plots for each landscape region showed distinct clusters of photographs from each region (Fig. 2). A closer visual examination of clusters featuring orienteering and forest paths showed that they came mostly from the national parks in the Forest and Lakes region in southern and central Finland. Winter photographs were mostly taken in the Lapland Fells or Eastern Hills. Sea and lakeside photographs were predominantly from more southern landscape regions (Archipelago and Forest and Lakes). Many activity photographs with dogs, skis, or bikes were distributed across landscape regions. The cluster for orienteering (Fig. 2) overlapped largely with the Forests and Lakes region, forest path, and park scene categories and coincided temporally with orienteering events (Appendix S13). The results for scene classification revealed a similar trend (Fig. 4). Photographs classified to the forest path category came mostly from Forest and Lakes region, whereas most photographs classified as tundra or ski slope were taken in the Lapland Fells. Due to the visual similarity of certain landscapes in Finland, some photographs have clearly been misclassified, for example, tundra in the archipelago shores (Table 2 & Appendix S8).

Discussion

We used 3 computer-vision methods to automate the visual content analysis of photographs from national parks, and to evaluate their usability in understanding differences between regions and visitor groups. To support the application of these methods in practice, we concentrated on models that were available off-the-shelf and pretrained to perform a given task. In other words, their application does not require provision of manually labeled data or advanced in-house programing. Our results showed that each of the methods provided a view of the photo content and could be useful for a range of information needs in protected-area user monitoring and management.

Many photographs taken in Finnish national parks featured landscapes, and we used scene classification to classify photographs into predefined categories (Zhou et al. 2018). The model predicted scene categories that fit the landscape regions defined for Finnish national parks: tundra and ski slope were commonly predicted for photographs taken at Lapland Fells, forest path in the Forest and Lakes region, and creek in the Eastern Hills region, which featured prominent river landscapes. Although some of these predictions may sound trivial, they confirm that scene classification produces meaningful results and provides quantifications of the representation of these landscapes in the photograph content. In our

case, scene classification provided information on the most photogenic landscapes in each landscape region and separately for national and international visitors. If photographs represent landscape values (van Zanten et al. 2016), our results suggest that the international visitors value winter landscapes and activities like skiing and dog sledging, whereas Finnish visitors appreciate summer forests, autumn colors, and activities like orienteering, biking, and cooking. Finding international visitors, for example, valuing snow and using commercial services more than the local visitors is in accordance with individual park-level visitor surveys (see https://julkaisut.metsa.fi/), but social media photo analysis provides more nuances at broader geographical scales and higher temporal resolution.

Instance-level object detection predicts instances of predefined object categories and their locations in the photograph (He et al. 2017). We found this approach to be useful for separating landscape photographs from close-up photographs and their combinations. In earlier works, visitor activities have been classified manually based on the contents of social media photographs (Heikinheimo et al. 2018). Identifying the objects present in photographs automatically contributed to this need. The most common objects (e.g., backpack, skis, boat, or bird) can be directly associated with activities that have been identified as the most popular in visitor surveys. Instance-level object detection can also be used to select photographs for further analysis. To exemplify, inspecting photographs with dogs revealed a major difference between national and international visitors in our data. Both groups share photographs of dogs, but almost all photographs of dogs taken by international visitors were taken on organized dog sleigh safaris in the Lapland Fells, whereas national visitors shared photographs of dogs mainly from forest walks (Appendix S12). This illustrates that analyzing the results of automatic visual content analysis can reveal differences between park activities and visitor groups and how visitors use services provided by the local economy.

Unlike the first 2 methods, semantic clustering does not assign photographs or objects detected in them into predefined categories. Rather, it is useful for automatically organizing large volumes of photographs without any prior knowledge of their content. This enables a rapid overview of visual content posted across all protected areas by revealing meaningful clusters of photographs featuring different landscapes, animals, and human activities. This information may provide protected-area managers with rapid situational awareness. In the case of Finnish national parks, semantic clustering enabled identifying subtle differences between visitor groups and national parks across the entire data set. We propose that this method can be used to obtain an overall understanding of the photographic content posted from even broad areas of interest.

All 3 computer-vision methods provided complementary perspectives to the automatic analysis of social media photography. In our case, automatic content analysis of photographs confirmed previous insights from visitor surveys, such as preferred activities, but also provided a completely new level of detail compared with traditional visitor surveys. These insights include emerging or event-type activities (e.g., orienteering in some parks), differing preferences between Finnish and foreign visitors (e.g., interest in dog sledging and other commercially organized wintertime activities among foreigners), and differing seasonality in visual content among visitor groups. In well-managed parks, the local park management is often familiar with their most popular activities. The proposed methods can provide an equally detailed understanding of visitor activities at national and regional scales yet provide a fine-grained view at a temporal resolution of individual events. Considering high costs involved with traditional visitor surveying, our positive experiences suggest that these methods may considerably improve understanding of visits to protected areas and human–nature interaction in general, particularly in areas where detailed monitoring of visitors is not feasible.

Like many recent studies on green areas (Sherren et al. 2017; Ghermandi & Sinclair 2019; Toivonen et al. 2019), we used Flickr as our data source in this study. Other social media platforms, such as Instagram, may capture a broader variety of human activities (Hausmann et al. 2018), but are not available for download or allowing computer-vision analysis (Toivonen et al. 2019). Therefore, despite the biases in the user base and the more limited content, Flickr continues to be a relevant source of data for visual content analysis, particularly when applying automated methods. Because the most photographed object on social media platforms is often people, both analysis and reporting of results must follow appropriate ethical practices (Zook et al. 2017; Di Minin et al. 2021). Compared with manual analyses of the photographs, the application of computer-vision methods may be less intrusive because individual photographs are not viewed by a human except when verifying the output from algorithms.

Beyond visitor monitoring and social media analyses, computer-vision methods are broadly interesting to various needs of conservation science. They may make it easier, for example, to analyze phenological changes (Correia et al. 2020), observe the occurrence of species (Willi et al. 2019), or track illegal wildlife trade (Di Minin et al. 2019). These methods hold much potential for further development in terms of combining semantic representations of content with other sources of information. For example, semantic clustering could be enriched by combining semantic representations of photographs with metadata related to time, place, user profile, and camera type, allowing the resulting visualizations to incorporate information about both photographs and their context.

Futhermore, analyzing the combinations of textual and visual content would likely provide an even more comprehensive picture of visitor preferences and activities in nature.

Our findings suggest that applying the computer-vision methods to social media photographs is a useful addition to the visitor monitoring toolkit in protected areas. Different methods provide complementary views to large collections of user-generated photographs by identifying landscapes or objects that stand in for specific activities or simply by organizing large volumes of photographs based on their semantic content. The proposed methods improve constantly as new architectures, models, and data sets are developed and made openly available, which allows them to be rapidly incorporated into the analysis workflows of conservation science. We thus propose that the application of computer-vision methods to social media data should be explored further under the umbrella of conservation culturomics.

Acknowledgments

Supporting Information

Additional information is available online in the Supporting Information section at the end of the online article. The authors are solely responsible for the content and functionality of these materials. Queries (other than absence of the material) should be directed to the corresponding author. The scripts used in this article are available at https://github.com/DigitalGeographyLab/someparkbaskets.

Literature Cited

Abdulla W. 2017. Mask R-CNN for object detection and instance segmentation on Keras and TensorFlow. Github. Available from https://github.com/matterport/Mask_RCNN. (accessed May 2020).

Arts K, van der Wal R, Adams WM. 2015. Digital technology and the conservation of nature. Ambio **44:**661–673.

Balmford A, Green JMH, Anderson M, Beresford J, Huang C, Naidoo R, Walpole M, Manica A. 2015. Walk on the wild side: estimating the global magnitude of visits to protected areas. PLOS Biology **13:**e1002074.

Bateman PW, Fleming PA. 2017. Are negative effects of tourist activities on wildlife over-reported? A review of assessment methods and empirical results. Biological Conservation **211:**10–19.

Bennett NJ, et al. 2017. Conservation social science: understanding and integrating human dimensions to improve conservation. Biological Conservation **205:**93–108.

Buckley R. 2018. Tourism and natural world heritage: a complicated relationship. Journal of Travel Research **57:**563–578.

Carter S, Weerkamp W, Tsagkias M. 2013. Microblog language identification: overcoming the limitations of short, unedited and idiomatic text. Language Resources and Evaluation **47:**195–215.

Chollet F. 2017. Xception: deep learning with depthwise separable convolutions. Pages 1800–1807 in Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition. IEEE, New York. https://doi.org/10.1109/CVPR.2017.195.

Correia DLP, Bouachir D, Gervais D, Pureswaran D, Kneeshaw DD, De Grandpré L. 2020. Leveraging artificial intelligence for large-scale plant phenology studies from noisy time-lapse images. IEEE Access **8:**13151–13160.

Di Minin E, MacMillan DC, Goodman PS, Escott B, Slotow R, Moilanen A. 2013. Conservation businesses and conservation planning in a biological diversity hotspot. Conservation Biology **27:**808–820.

Di Minin E, Tenkanen H, Toivonen T. 2015. Prospects and challenges for social media data in conservation science. Frontiers in Environmental Science **3** https://doi.org/10.3389/fenvs.2015.00063.

Di Minin E, Fink C, Hiippala T, Tenkanen H. 2019. A framework for investigating illegal wildlife trade on social media with machine learning. Conservation Biology **33:**210–213.

Di Minin E, Fink C, Hausmann A, Kremer J, Kulkarni R. 2021. How to address data privacy concerns when using social media data in conservation science. Conservation Biology **35**.

Ghermandi A, Sinclair M. 2019. Passive crowdsourcing of social media in environmental research: a systematic map. Global Environmental Change **55:**36–47.

Gosal AS, Geijzendorffer IR, Václavík T, Poulin B, Ziv G. 2019. Using social media, machine learning and natural language processing to map multiple recreational beneficiaries. Ecosystem Services **38:**100958.

Hausmann A, Toivonen T, Heikinheimo V, Tenkanen H, Slotow R, Di Minin E. 2017. Social media reveal that charismatic species are not the main attractor of ecotourists to sub-Saharan protected areas. Scientific Reports **7:**763.

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:**e12343.

He K, Gkioxari G, Dollár P, Girshick RB. 2017. Mask R-CNN. Pages 2980–2988 in Proceedings of the 2017 IEEE International Conference on Computer Vision. IEEE, New York. https://doi.org/10.1109/ICCV.2017.322.

He K, Zhang X, Ren S, Sun J. 2016. Deep residual learning for image recognition. Pages 770–778 in Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition. IEEE, New York. https://doi.org/10.1109/CVPR.2016.90.

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:**85.

Heikinheimo V, Tenkanen H, Hiippala T, Toivonen T. 2018. Digital imaginations of national parks in different social media: a data exploration. Pages 45–52 in Proceedings of PLATIAL'18: Workshop on Platial Analysis. Zenodo, Genova. https://doi.org/10.5281/zenodo.1472745.

Kajala L, et al. 2007. Visitor monitoring in nature areas – a manual based on experiences from the Nordic and Baltic countries. TemaNord **2007:**534.

Kalliatakis G. 2017. Keras-VGG16-Places365. GitHub. Available from https://github.com/GKalliatakis/Keras-VGG16-places365. Accessed 2nd of May 2020.

Koylu C, Zhao C, Shao W. 2019. Deep neural networks and kernel density estimation for detecting human activity patterns from geotagged images: a case study of birdwatching on Flickr. ISPRS International Journal of Geo-Information **8:**45.

Kruger M, Viljoen A, Saayman M. 2017. Who visits the Kruger National Park, and why? Identifying target markets. Journal of Travel and Tourism Marketing **34:**312–340.

Ladle RJ, Correia RA, Do Y, Joo G-J, Malhado ACM, Proulx R, Roberge J-M, Jepson P. 2016. Conservation culturomics. Frontiers in Ecology and the Environment **14:**269–275.

LeCun Y, Bengio Y, Hinton G. 2015. Deep learning. Nature **521:**436–444.

Lee RF. 1972. Family tree of the national park system: a chart with accompanying text designed to illustrate the growth of the national park system. Eastern National Park & Monument Association, Fort, Washington.

Lin TY, Maire M, Belongie S, Hays J, Perona P, Ramanan D, Dollár P, Zitnick CL. 2014. Microsoft COCO: common objects in context. Pages 740–755 in Proceedings of the 2014 European Conference on Computer Vision. Springer International Publishing, Cham. https://doi.org/10.1007/978-3-319-10602-1_48.

Longley PA, Adnan M, Lansley G. 2015. The geotemporal demographics of Twitter usage. Environment and Planning A **47:**465–484.

McInnes L, Healy J, Saul N, Grossberger L. 2018. UMAP: uniform manifold approximation and projection for dimension reduction. Journal of Open Source Software **3:**861.

Neuvonen M, Pouta E, Puustinen J, Sievänen T. 2010. Visits to national parks: effects of park characteristics and spatial demand. Journal for Nature Conservation **18:**224–229.

Pickering C, Walden-Schreiner C, Barros A, Rossi SD. 2020. Using social media images and text to examine how tourists view and value the highest mountain in Australia. Journal of Outdoor Recreation and Tourism **29**. https://doi.org/10.1016/j.jort.2019.100252.

Pietilä M, Fagerholm N. 2019. A management perspective to using public participation GIS in planning for visitor use in national parks. Journal of Environmental Planning and Management **62:**1133–1148.

Richards DR, Friess DA. 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.

Richards DR, Tunçer B. 2018. Using image recognition to automate assessment of cultural ecosystem services from social media photographs. Ecosystem Services **31:**318–325.

Russakovsky O, et al. 2015. ImageNet large scale visual recognition challenge. International Journal of Computer Vision **115:**211–252.

Schullery P, Whittlesey LH. 2003. Myth and history in the creation of Yellowstone National Park. University of Nebraska Press, Lincoln.

Seresinhe CI, Moat HS, Preis T. 2018. Quantifying scenic areas using crowdsourced data. Environment and Planning B: Urban Analytics and City Science **45:**567–582.

Seresinhe CI, Preis T, Moat HS. 2017. Using deep learning to quantify the beauty of outdoor places. Royal Society Open Science **4:**170170.

Sharma N, Scully-Power P, Blumenstein M. 2018. Shark detection from aerial imagery using region-based CNN, a study. Pages 224–236 in Lecture notes in computer science (including subseries lecture notes in artificial intelligence and lecture notes in bioinformatics). Springer, Cham.

Sherren K, Smit M, Holmlund M, Parkins JR, Chen Y. 2017. Conservation culturomics should include images and a wider range of scholars. Frontiers in Ecology and the Environment **15:**289–290.

Siikamäki P, Kangas K, Paasivaara A, Schroderus S. 2015. Biodiversity attracts visitors to national parks. Biodiversity and Conservation **24:**2521–2534.

Simonyan K, Zisserman A. 2015. Very deep convolutional networks for large-scale image recognition. Proceedings of the 2015 International Conference on Learning Representations. ICLR, La Jolla, California. Available from https://arxiv.org/abs/1409.1556. (accessed May 2020).

Sutherland WJ, et al. 2018. A 2018 horizon scan of emerging issues for global conservation and biological diversity. Trends in Ecology & Evolution **33:**47–58.

Tenkanen H, Di Minin E, Heikinheimo V, Hausmann A, Herbst M, Kajala L, Toivonen T. 2017. Instagram, Flickr, or Twitter: assessing the usability of social media data for visitor monitoring in protected areas. Scientific Reports **7:**17615.

Toivonen T, Heikinheimo V, Fink C, Hausmann A, Hiippala T, Järv O, Tenkanen H, Di Minin E. 2019. Social media data for conservation science: a methodological overview. Biological Conservation **233:**298–315.

Van Berkel DB, Tabrizian P, Dorning MA, Smart L, Newcomb D, Mehaffey M, Neale A, Meentemeyer RK. 2018. Quantifying the visual-sensory landscape qualities that contribute to cultural ecosystem services using social media and LiDAR. Ecosystem Services **31:**326–335.

van Zanten BT, Van Berkel DB, Meentemeyer RK, Smith JW, Tieskens KF, Verburg PH. 2016. Continental-scale quantification of landscape values using social media data. Proceedings of the National Academy of Sciences of the United States of America **113:**12974–12979.

Watson JEM, Dudley N, Segan DB, Hockings M. 2014. The performance and potential of protected areas. Nature **515:**67–73.

Whitelaw PA, King BEM, Tolkach D. 2014. Protected areas, conservation and tourism – financing the sustainable dream. Journal of Sustainable Tourism **22:**584–603.

Willi M, Pitman RT, Cardoso AW, Locke C, Swanson A, Boyer A, Veldthuis M, Fortson L. 2019. Identifying animal species in camera trap images using deep learning and citizen science. Methods in Ecology and Evolution **10:**80–91.

Xie S, Girshick R, Dollár P, Tu Z, He K. 2017. Aggregated residual transformations for deep neural networks. Pages 5987–5995 in Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition. IEEE, New York. https://doi.org/10.1109/CVPR.2018.00907.

Zhou B, Lapedriza A, Khosla A, Oliva A, Torralba A. 2018. Places: a 10 million image database for scene recognition. IEEE Transactions on Pattern Analysis and Machine Intelligence **40:**1452–1464.

Zook M, et al. 2017. Ten simple rules for responsible big data research. PLOS Computational Biology **13:**e1005399.

Zoph B, Vasudevan V, Shlens J, Le Q V. 2018. Learning transferable architectures for scalable image recognition. Pages 8697–8710 in Proceedings of the 2018 IEEE Conference on Computer Vision and Pattern Recognition. IEEE, New York. https://doi.org/10.1109/CVPR.2018.00907.