Urban Recreational Ecosystem Services Investigation Based on Social Media Images

by

Wei Hu

A thesis submitted
in partial fulfillment of the requirements
for the degree of
Master of Science
(Environment and Sustainability)
in the University of Michigan
April 2022

Thesis Committee:

Assistant Professor Derek B. Van Berkel, Chair
Assistant Professor Neil Carter
Assistant Professor Mark Lindquist

ABSTRACT

Recreational ecosystem services (RES) are understood as the benefits that people derive from landscapes and natural environments through recreational activities. The growing social media datasets have contributed to overcoming limitations of spatial and temporal coverage for RES studies that traditional survey-based approaches have. Related RES research using social media such as photo-sharing platforms has primarily focused on natural and ecological areas outside cities at regional or national scale and utilized geotagged photographs as reliable proxies for empirical access rates. The urban dimension of RES is under-explored, and potential information about the environmental composition and user preferences in photos is overlooked. Using data retrieved from the photo-sharing platform Flickr, we explore the potential role of computer vision (CV) in understanding RES related to environmental composition and human activities. After that, we assess RES for the urban outdoor environment of Ann Arbor. Specifically, by manual validation of recognition results for 1,500 Flickr photographs, we evaluate whether scene recognition algorithms and models pre-trained with three different labeling systems on a standard CV dataset can be applied to tackle complex visual tasks in realistic urban scenarios. Contrary to consistent outstanding performance on standard CV datasets, we find substantial changes in the performance of recognizing physical environmental composition and human activities depending on the semantic scale the model uses for labeling. Via recognition results, we further study people's preferences for environmental composition and outdoor activities and their associations, then detect popular RES places for different recreational usages in Ann Arbor with a high spatial resolution. This article concludes with the feasibility of applying pre-trained CV models for urban RES studying. Time and resource permitting, future studies should consider combining information from multiple sources for a more accurate evaluation of RES characteristics, thus can be integrated with decision making, planning, and management to enhance city planning and human well-being.

TABLE OF CONTENTS

ABSTRACT ii

LIST OF FIGURES vi

LIST OF TABLES vii

CHAPTER

I. Introduction 1

II. Methodology 8

2.1 Study area 8
2.2 Data collection and pre-processing 9
2.3 Computer vision for RES study 10
2.4 Photo content extraction 14
2.4.1 Content recognition with different semantic scales . 14
2.4.2 Assessment of recognition content 16
2.5 Analysis of urban recreational ecosystem services 17
2.5.1 Correlation analysis 17
2.5.2 Spatial distribution 18

III. Result 19

3.1 Overview of available Flickr data 19
3.2 Assessment of recognition content 20
3.3 Overview of recognition content 23
3.4 Correlation Analysis 25
3.5 Spatial distribution 28

IV. Discussion and Conclusion 33

BIBLIOGRAPHY 36

APPENDIX 44

LIST OF FIGURES

Figure

3.1 Attribute frequency map 24

3.2 Correlation matrix for RES study (left) 26

3.3 Correlation matrix for RES study (right) 27

3.4 Ann Arbor RES map 28

3.5 Ann Arbor biking map 30

3.6 Ann Arbor boating map 30

3.7 Ann Arbor exercise map 31

3.8 Ann Arbor hiking map 31

3.9 Ann Arbor picnicking map 32

A.1 Distribution of user uploading volume 48

A.2 Photo volume changes over years 48

A.3 Photo volume changes over months 49

A.4 Indoor/outdoor recognition result over years 50

A.5 Correlation matrix for the SUN attribute (left) 51

A.6 Correlation matrix for the SUN attribute (right) 52

LIST OF TABLES

Table

2.1 Notation for calculating r_Φ 18

3.1 Recognition accuracy of the environment type 21

3.2 Recognition accuracy of the Places scene 21

3.3 Recognition accuracy of the SUN attribute 22

A.1 Category and attributes of the SUN dataset (all) 46

A.2 Overall criteria for the SUN attribute recognition 47

CHAPTER I

Introduction

The concept of ecosystem services (ES) connects the natural environment to benefits we derive, directly or indirectly, from ecosystem functions (*Costanza et al.*, 1997). Identified as an essential class of cultural ecosystem services (CES), recreation ecosystem services (RES) benefit people through providing leisure and opportunities for recreational activities such as eco-tourism, sport fishing, and hiking (*Haines-Young and Potschin*, 2012). Over the last decade, crowd-sourced geospatial data such as social media data has provided researchers with a new measure of human behavior for RES studies (*Wood et al.*, 2013; *Donaire et al.*, 2014; *Ghermandi*, 2016; *Sessions et al.*, 2016; *Tenkanen et al.*, 2017; *Sinclair et al.*, 2020). Such data have wide spatial and temporal coverage compared to traditional survey-based data and allow for high fidelity understanding of human interaction within the environment. This is especially relevant for urban regions, where the study of RES requires spatial resolutions that match the fine scale of urban nature interactions (e.g., small parks, local trees), the diversity of activities across large urban populations, and the complexity and fragmentation of landscape composition that far exceeds natural environments (*Welch*, 1982; *Herold et al.*, 2003). This research uses the city of Ann Arbor as a case study to explore the technical feasibility of such innovations for RES investigation using social media data from the photo-sharing platform Flickr in the urban context.

The importance of urban RES is related to the huge impact they have on improving the well-being of residents(*Van Kamp et al.*, 2003). According to the United Nations, 68% of the world population is expected live in urban areas within 30 years (*Desa*, 2018). Urbanization fundamentally alters ecological and social functioning (*Alberti and Marzluff*, 2004; *Haase et al.*, 2014), impacting city-dwellers' health due to factors such as overcrowding and air quality related to traffic and the heat island effect, crime, living stressors (e.g., food insufficiency and stressful neighborhood conditions) (*Rose*, 1999; *Latkin and Curry*, 2003), and reduced social support (*Srivastava*, 2009). Ailments connected with stressful and sedentary living are common among urban populations, partly due to the structure and composition of the urban environment (*Clarke*, 1972). For example, diabetes, heart disease, respiratory illness, and mental health issues have in part been linked to environmental factors, including greenspace accessibility (*Beyer et al.*, 2014). Research has shown that exposure to nature has emotional and cognitive advantages, thereby contributing to physical and psychological benefits that may be deprived as urbanization continues (*Bratman et al.*, 2015; *Twohig-Bennett and Jones*, 2018). By providing leisure and outdoor recreation opportunities, RES often serve as a prime vehicle for reconnecting people with the natural environment for social well-being (*Hermes et al.*, 2018). Without adequate understanding and characterization, there is a high risk that the natural area that provides RES will not receive the appropriate protection and conservation to ensure future ability to support human well-being.

The majority of people are exposed to nature in urban areas. National parks and protected reserves are often inaccessible, yet they are often the main focus of the study of RES. Studies have shown that distance to natural areas, and accessibility directly influences the number of close-to-home outings among city residents (*Neuvonen et al.*, 2007; *De Valck et al.*, 2017). Short distance to green areas is significantly associated with less stress and a lower likelihood of obesity (*Nielsen and*

Hansen, 2007). In an urban context, RES include opportunities for walking through a park, hiking, picnicking, and biking and often requires specific landscape elements and human-made infrastructure. RES is frequently enhanced by the landscape's aesthetic quality and amenities associated with natural environments, such as pavements and benches. Thus, understanding recreational usages, landscape composition of the natural environment, and their relationships would aid in monitoring and conserving unique places that benefit people recreationally, as well as prioritizing maintenance of existing areas and potential development of new sites that provide these essential RES (*Hermes et al.*, 2018).

Traditional methods to investigate the impact of urban RES rely on surveys, which need to be performed on-site in parks, on trails, and other natural spaces. Related research investigates landscapes compositions, infrastructure constructions, and individuals' recreational preferences under various environmental scenarios (*Goossen and Langers*, 2000; *Bartczak et al.*, 2008; *Ode et al.*, 2009; *Brown*, 2010; *Kienast et al.*, 2012; *Pleasant et al.*, 2014; *Schägner et al.*, 2016; *Fischer et al.*, 2018). However, urban RES sites are often spatially disbursed, consisting of complex and diverse landscapes and providing various recreational opportunities. These methods are expensive, time-consuming, labor-intensive, and have limited spatial and temporal coverage, posing a challenge to investigating and assessing urban RES covering the whole city area timely (*Kajala*, 2007).

An emerging approach, which uses the locations of social media photography to investigate users' attitudes and behavior with regard to the environment, may overcome some of the limitations mentioned above. Over the last decades, along with the rapid development of information and communications technology and the availability of new sensors, new ways of collecting data have emerged, bringing potential new data sources for landscapes and the natural environment investigation. In an era of ever-increasing user-generated content, data acquired by the public, volunteered geographic information (VGI), has led to increased availability of spatial information. Unlike authoritative datasets such as remote sensing, these new data types expand and enhance geographic data in terms of thematic variety, free availability, and user centrality (*Sester et al.*, 2014). Recognized as critical VGI, social media contents, including text, photos, videos, etc., are voluntarily published online by users, expressing their perceptions, interests, requirements, and behaviors tied to specific locations. Furthermore, social media data are collected unobtrusively, and users are not generally constrained when generating information (*Quercia et al.*, 2015), which eliminates the Hawthorne effect that objects alter their behavior when realizing they are being observed (*McCarney et al.*, 2007). Beneficial from a bias-alleviated and user-centric vision of diverse aspects of society, economics, environment, and culture (*Martí et al.*, 2019), social media has been widely acknowledged as a potential resource for advancing urban research from specific human aspects, such as mobility (*Noulas et al.*, 2012), human behavior (*Hochman and Manovich*, 2013), and land use and human activities (*García-Palomares et al.*, 2018; *Lu*, 2019) at large scales and with limited resources.

Among various types of social media data, geotagged photographs from photo-sharing platforms such as Flickr are highly suitable for exploring the urban RES due to their reflection on people's interest in locations and the surrounding environment. In fact, metadata contained within geotagged Flickr photographs has been used primarily to understand interactions between humans and the natural environment at wide spatial scales (*Wood et al.*, 2013). Early related studies only use geotagged images as crowd-sourced information that can serve as a reliable proxy for empirical visitation rates (*Girardin et al.*, 2008; *Wood et al.*, 2013; *Gliozzo et al.*, 2016). This potentially leaves out valuable information regarding environmental components that can be analyzed to infer how and why people interact with nature (*Dorwart et al.*, 2009). Most users are likely to post more photos of areas they deem beautiful, which partly reflect public preferences of landscape and its corresponding recreational usages. However, what is captured in the photo represents a subjective choice, and there is a certain user bias in understanding the environment through the photo.

Notwithstanding the biases in capturing photographs, recent work has attempted to integrate the visual content of photographs for further insights into human-nature interactions. Such explorations include CES identification across country or continent (*Richards and Friess*, 2015; *Oteros-Rozas et al.*, 2018), characterization of the factors contributing to the aesthetic value of landscapes (*Van Berkel et al.*, 2018; *Tenerelli et al.*, 2017), and visitors' preferences for specific aspects of a nature-based recreation experience in protected areas (*Hausmann et al.*, 2018; *Ghermandi et al.*, 2020b). All these attempts focus on natural and ecological areas at regional or national scales outside cities, such as national sites, rural areas, and protected reserves; the urban dimension is underexposed. When shifting focus back to urban, researchers mainly study green space with explicit planning boundaries such as urban parks (*Donahue et al.*, 2018; *Hamstead et al.*, 2018), lacking coverage of semi-natural areas such as campus.

Technically, previous studies have frequently relied on manual classification and interpretation of image content (*Guerrero et al.*, 2016; *Martínez Pastur et al.*, 2016; *Heikinheimo et al.*, 2017), which is impractical for large databases such as social media. Therefore, as computer vision services arise in artificial intelligence applications, several studies have applied pre-trained models from commercial cloud services to label photographs with content-related tags. In these studies, tags generated from photographs are used for hierarchical clustering based on expert-defined tags' classes, to explore themes of photographs (*Richards and Tunçer*, 2018), CES photographs reflect such as landscape aesthetics (*Lee et al.*, 2019), the profile of user preferences in natural sites on different shared platforms (*Ghermandi et al.*, 2020b), how machine tags from different services lead to different interpretation with landscapes and ecosystems (*Ghermandi et al.*, 2022). Commercial services such as Google Cloud Vision and Azure Computer Vision can return content tags for thousands of recognizable objects, which are granular, and thus can hardly form scene-level semantic understanding. When re-organized in these studies, hierarchical clusters defined are rough, focusing on object categories (e.g., people, plants, animals, transport) or geographic landscapes at large scale (e.g., rural, desert, coast, tundra, broad leaf forest, glacier, etc.).

However, when studying RES, characterizing urban environmental composition using large-scale geographic landscapes is not applicable. For example, all vegetation in a city may be classified as tundra after large-scale geographic landscape identification. The urban landscape follows regional geographical characteristics on the macro-level while diverse and complex on the micro-level. Moreover, RES investigation in the urban context needs a scene-level description of the environment rather than objects detection in photographs, fine-grained interpretation of direct human-nature interactions such as biking, hiking, and picnicking, rather than a general description as ‘recreational activities’. To the best of our knowledge, no study has refined computer vision in artificial intelligence applications on RES investigation down to the urban context, explored the potential differences in accuracy when computer vision interpret photographs in different semantic scales, and how such interpretations might reflect people’s preference for natural landscape compositions and infrastructure for different recreational activities in urban regions. The objectives of the study are:

1. To summarize potential usages of different computer vision tasks in the urban context.

2. To assess scene recognition accuracy for three labeling systems with different semantic scales.

3. To recognize natural environmental features, infrastructures, and existing or

potential human recreational activities for urban RES investigation from photo-sharing platforms Flickr, using scene recognition with suitable semantic scales.

4. To explore different natural environmental features and infrastructure needs for various recreational activities in the area of the city of Ann Arbor.

CHAPTER II

Methodology

In this chapter, we first present a brief introduction to our study area, the city of Ann Arbor in Section 2.1. Then, in Section 2.2, we introduce how we collect data for the research area and the pre-processing preparation for our work. After that, we conceptually investigate and summarize different computer vision tasks and their potential suitability for the urban RES study in Section 2.3. Based on this investigation, we apply three labeling systems to extract photo content and propose assessment methods in Section 2.4. Last, we provide two methods for the urban RES analysis in Section 2.5.

2.1 Study area

With an estimated population of about $124,000$ (*United States Census Bureau*, 2020), Ann Arbor is the fifth-largest city in the Michigan state, which is in the Great Lakes region of the upper Midwestern United States. It is located ($42°16' N$, $83°44' W$) in the southeastern part of Michigan, included in the larger Greater Detroit Combined Statistical Area. The city has a total area of 74.33 square kilometers (km^2): 72.08 km^2 is land and 2.25 km^2 is water, much of which is part of the Huron River. Influenced by the Great Lakes, Ann Arbor has a typically Midwestern humid continental climate. There are four distinct seasons: winters are cold and snowy, with average highs around 1°C. Summers are warm to hot and humid, with average highs around 27°C and with slightly more precipitation. Spring and autumn are transitional between the two. Snowfall, which usually occurs from November to April but occasionally starts in October, averages 147 centimeters per season. The landscape of Ann Arbor consists of hills and valleys. The city contains more than $50,000$ trees along its streets and 157 municipal parks ranging from small neighborhood green spots to large recreation areas.

2.2 Data collection and pre-processing

Flickr is a photograph-sharing platform with over 112 million users and over 5 billion photos of geotagged photographs, which has become a commonly-used source of social-media photographs for assessing cultural ecosystem services (*Wood et al.*, 2013; *Richards and Friess*, 2015; *Van Zanten et al.*, 2016; *Lee et al.*, 2019; *Ghermandi et al.*, 2020b, 2022). As one of the largest sources of geotagged photographs available online, the Flickr application programming interface (API) is available for non-commercial use (`https://www.flickr.com/services/api/`). Using package *photosearcher* in R (*Fox et al.*, 2020) to interact with the Flickr API, metadata was extracted from photographs taken between 1 January 2005 and 31 December 2019, within the boundary of the city of Ann Arbor, which data is available from the city of Ann Arbor's service website (`https://www.a2gov.org/services/data/Pages/default.aspx`).

Public metadata retrieved includes unique image id, user id, image url, photograph coordinates, and a time stamp from when the picture was taken. It is worth noting that users have different habits of taking and uploading photos. Some prefer to take and upload one at a time, while some as photography enthusiasts prefer to capture and upload multiple photos at the same site. To balance the impact of different usage preferences of Flickr on analysis, we create a new attribute called user-date-location (UDL). For each image, we combine the user id, location, and taken time.

To ensure the uniqueness of UDL, we randomly retain a single photo taken by the same user for each date at the exact location. The location data, longitude and latitude, are in six decimal precision, which means photos taken within approximately 0.5 square meters (m^2) are treated as taken at the same location. In this way, we exclude instances where users take multiple photographs at the same location, while retaining photographs along their walking route, which reveal behavioral dynamics and concurrent appreciated landscapes' locations.

2.3 Computer vision for RES study

Here we explain how computer vision algorithms can help understand the outdoor environment and recreational characteristics from Flickr photographs giving an overview of its definition, function, and applications. Computer vision can be described as the task of learning and deciphering the visual elements of digital imagery in order to quantify them (*LeCun et al.*, 2015). This class of computer algorithms functions like eyesight, detecting visual objects based on a cognitive understanding of a scene gained from a task-specific sample of supplied images or frames of images (*Viola and Jones*, 2001). Due to deep learning methods and the accessibility of large datasets, computer vision can now deal with a variety of challenges and analyze images more precisely and effectively across a wide range of applications in realistic settings (*Lin et al.*, 2014; *Russakovsky et al.*, 2015; *Cordts et al.*, 2016). Deep learning models can be trained differently depending on the types of visual work, with distinct layer and algorithm settings. Computer vision algorithms can be grouped into eight basic tasks: image classification, segmentation and localization, tracking, action recognition, perception, generative models, clustering, and decision-making (*Guo et al.*, 2016). Other frameworks and operations can be constructed and built based on these fundamental tasks.

Among all the above tasks, the most commonly used in urban studies are classification, segmentation and localization, and action recognition (*Ibrahim et al.*, 2020). Image Classification, one of the core problems in computer vision, aims to assign an input image a single label from various classes. Many other seemingly distinct tasks (e.g., object detection, segmentation) can be reduced to image classification. Models such as AlexNet (*Krizhevsky et al.*, 2012), VGGNet (*Simonyan and Zisserman*, 2014), GoogLeNet (*Szegedy et al.*, 2015), ResNet (*He et al.*, 2016) have been built to recognize visual objects in large repositories of images. One such famous images dataset is ImageNET, which contains 15 million images, with $22,000$ different classes (*Russakovsky et al.*, 2015). Classification models usually return a probability measure for each class, indicating the certainty of the classes models assign (e.g., for an image containing a tree, it may return a vector that the position representing the tree gets a score of 0.9, and the position representing the grass gets a score 0.6). Segmentation and localization are the processes of identifying multiple objects in a single image, which can be subdivided into semantic segmentation, object detection, and instance segmentation. Semantic segmentation is pixel-based and performed for the entire image; each pixel in the image is assigned a class, while object detection and instance/object segmentation focus on particular objects. Unlike object detection, which has the output of a series of bounding boxes with object annotation, the result for object segmentation is also pixel-based. Therefore, the concept of semantic segmentation and object segmentation are mixed for those cases when the photograph has no specific foreground objects or content for the entire photo is the focus. Action recognition detects motion and actions from a still image, and relies on the concept of the triplet inputs (object, verb, target), generating real-world events and behaviors (*Girdhar et al.*, 2018).

For urban RES study, interpretation of a scene background can offer context about the photographic location, key elements captured and activities taking place in the social media image. Rather than merely identifying specific objects, analysis of the entire photo and understanding the place where the photo is taken are valuable for gaining RES related information. Semantic segmentation is the central part of the solution for computer vision applications in urban RES study; many advances in computer vision semantic segmentation have been achieved in this application context. Researchers have been focusing on developing models of inferring land use and land cover from satellite imagery both in rural and urban areas (*Talukdar et al.*, 2020; *Karpatne et al.*, 2016; *Buslaev et al.*, 2018; *Hamaguchi and Hikosaka*, 2018), investigate urban greenspace and human activities (*Lu*, 2019), estimating pedestrian volume (*Chen et al.*, 2020), comparing and discerning differences in the micro built environment characteristics between drug activities and street robberies (*Biljecki and Ito*, 2021). Specifically, semantic segmentation provides a better understanding of the elements of an urban scene, such as sky, ground, road, building, vegetation, etc. (*Ordonez and Berg*, 2014). Each pixel of an image is assigned a separate class label, concluding a list of semantic classes that describes the element composition of a single image.

The above semantic recognition describes local appearance separately using objects/elements (the list of semantic classes concluded) within the imagery. However, an accurate understanding of an image requires the joint consideration of local appearance, semantic information, and global scene context. Most current state-of-the-art visual recognition algorithms are only able to discern object categories or their spatial relationships. For example, a scene with a TV, a sofa, and a person is likely to occur within a dwelling; however, these same objects could occur in the hotel room of a beach resort (*Zhou et al.*, 2014b). Shared objects suggest similar locational contexts, despite huge differences that result in the class overlap. Attempt to ameliorate these issues belong to an emerging class of computer vision research on scene recognition that attempts to define the context of photographs from both objects and scenes. Research has found that object detectors emerge in the last layer of convolutions neural networks (CNN) for scene recognition (*Zhou et al.*, 2014a). In other words, the scene classification network automatically discovers meaningful objects detectors, representative of the learned scene categories, as scenes are composed of objects. Without being explicitly taught the notion of objects, the same network can perform both scene recognition and object localization, as object detectors emerge as a result of learning to recognize scenes. Technically, an extra layer added to neural architecture for object segmentation can achieve scene recognition; scene recognition is a new task of classification based on object segmentation. Thus, the latest advancements in deep learning methods for scene recognition are motivated by the availability of large and exhaustive datasets and hardware that allow network training, enabling scene understanding in an uncontrolled and realistic environment. Famous scene-centric datasets contain Scene15 (*Lazebnik et al.*, 2006), MIT Indoor67 database (*Quattoni and Torralba*, 2009), the scene understanding (SUN) database (*Xiao et al.*, 2010), and Places (*Zhou et al.*, 2017).

Another task component of computer vision applicable for the study of urban RES is action recognition. Research has found that the global average pooling (GAP) layer (*Johansson et al.*, 2016) explicitly enables the CNN to have remarkable localization ability despite being trained on image-level labels (*Zhou et al.*, 2016). Using GAP, a class activation map (CAM) can be generated to highlight local discriminated parts of the target object by investigating the contribution of hidden units to the output of a classification network. This means CNNs can perform object localization without any bounding box annotations. As the concept model of triplet inputs (object, verb, target) revealed, particular object and target are capable of identifying the action (verb in the model), suggesting action can be recognized in a simple still image by using CNNs model trained for classification (*Girdhar et al.*, 2018).

2.4 Photo content extraction

In this section, we first explore three labeling systems to extract photo content at different semantic scales using ResNet pre-trained on the Places dataset. Then, we proposes assessment criteria for content recognized using these labeling systems.

2.4.1 Content recognition with different semantic scales

To fully understand RES in the urban context, we obtained semantic descriptions regarding environmental elements (e.g., landscape composition), global scene, and existing or potential human activities from our dataset. As discussed in Section 2.3, computer vision tasks that might extract these include differing semantic scales: object, scene abstraction, and action. All these recognition results can be attained using the same training network by extracting results from different layers or slightly changing the network's structure. In addition to significant advances in deep neural network frameworks, a major contributing factor to the success of top-performing methods in computer vision is the availability of large-scale, public training datasets such as ImageNet (*Russakovsky et al.*, 2015) and Microsoft COCO (*Lin et al.*, 2014), while both are object-centric.

However, no existing dataset adequately captures the complexity of real-world urban scenes for semantic outdoor urban scene understanding. Considering the need for recognizing scenes, potential pre-trained models capable of interpreting urban environments should be trained on scene-centric datasets. The Places database contains a quasi-exhaustive repository of around 8 million scene photographs. With 365 semantic scene categories, this database comprises about 98% of the type of places a human can encounter in the world (*Zhou et al.*, 2017). This multi-million-item dataset allows data-hungry machine learning algorithms to reach near-human semantic classification performance at visual object and scene recognition tasks.

Based on the Places dataset, we explore three different labeling systems to predict scene tags including the environment type (indoor and outdoor), the scene description (e.g., grass, vegetation, man-made, etc.) (*Xiao et al.*, 2010), and an estimate of the place (e.g., beach, forest, field, etc.). Labeling systems for an estimate of the place are from original semantic scene categories in the Places dataset (*Zhou et al.*, 2017). There are 365 scene categories in the Places dataset (visualization website: `http://places2.csail.mit.edu/scene_hierarchy.html`). All scene categories are organized in 3 hierarchical levels including 159 'indoor', 80 'outdoor natural', and 159 'outdoor man-made' categories. The outdoor natural and constructed categories overlap and we combined them as the general outdoor types with 206 categories that are disjoint with indoor types. We extracted this first level result and assign them to the second labeling systems: environment types, which include indoor and outdoor environments. The third labeling system used, scene attributes from the SUN dataset, contains 102 discriminative attributes; we re-organize them into six categories (number of attributes): functions (36), materials (38), lighting (3), surface properties (10), spatial envelope (12), and moods (3). Details for attributes in each group are shown in supplementary material Table A.1. It is worth noting that the SUN attribute is summarized from crowd-sourced human studies, which is suitable for the interpretation of the environment and human-environment interaction in the urban context. Attributes in the category of functions are human activities of high possibility in the urban area (e.g., driving, biking, transporting, hiking, climbing, picnicking, reading, etc.). Attributes in the category of materials are related to semantic objects, making up the urban appearance, including natural compositions (e.g., trees, grass, shrubbery, foliage, soil, flowers, water, ice, snow, etc.) and man-made materials of infrastructures (e.g., wire, railroad, asphalt, pavement, etc.).

We use ResNet, a pre-trained convolution neural network (CNN) model based on open source classified photographic content (*He et al.*, 2016). The CNN model returns three possible labels categorizing outdoor/indoor type, the Places scene (labels with the 5 highest confidence scores), and the SUN attribute (highest 10).

2.4.2 Assessment of recognition content

Computer vision models are trained by a standard dataset consisting of high-quality images. However, their performances when deployed in real-life applications can include errors and misclassifications. To give an impression of the accuracy of the recognition, we manually assessed and compared the outdoor/indoor classes, the top 5 Places semantic scene categories and top 10 SUN scene attribute results for each image 2.4.2. We randomly selected a subset of $1,500$ photographs from collected data, and compared these with recognition results. We build assessment criteria for each labeling systems separately.

For result of outdoor/indoor classes, we measured percentage error and agreement between manual and classification using weighted Cohen's kappa, an index of agreement between different classifications commonly used in psychology (*Landis and Koch*, 1977).

Assessments of the Places scene and the SUN attributes are conducted based on the result of environment type, as this study focuses on the outdoor environment. For the Places scene, we calculate Top-1 and Top-5 accuracy for recognition results, similar to the validation in the CV area. The Top-1 accuracy is the percentage of the images where manual verification agrees with the top predicted label. The Top-5 accuracy is the percentage of testing images where manual verification agrees with the top-ranked 5 predicted labels given by an algorithm. As some ambiguity exists between some scene categories, the Top-5 accuracy is a more suitable criterion for measuring scene classification performance.

Criteria for assessing the SUN attribute results is divided into two parts: one is for the 6 attribute categories we defined in Section 2.4.1, and the other is for the overall. The observer gives an overall grade ranging from 1 to 5 based on the feeling after reading the 10 tags generated from the SUN attribute, with 1 meaning totally wrong and 5 meaning precisely correct and no vital information omitted. Detailed criteria description can be found in supplementary material Table A.2. For each attribute category, the observer gives a grade from $\{-1, 0, 1, 2\}$. The grade is 0 if tags generated are not in the certain attribute category; -1 if wrong; 1 if partly right; and 2 if precisely correct.

2.5 Analysis of urban recreational ecosystem services

We selected photographs recognized as 'outdoor' to map and analyze the urban recreational ecosystem services, and used SUN scene classes describing recreational potentials to characterized frequented outdoor recreation locations. The SUN scene attribute are based on broad perception of environmental preference (attributes in category materials) related to outdoor human activities (attributes in category functions).

2.5.1 Correlation analysis

To understand what attributes occur within the same photo, we investigate the co-occurrence of SUN attributes for the whole city area. The correlation among words reflects the relationship between photographic attributes, which helps understand the association between urban RES. Regarding text, correlation among words is measured in a binary form: either the words appear together or they do not. Such binary correlation is measured by coefficient r_Φ (*Cramér*, 2016):

$$r_\Phi = \frac{n_{11}n_{00} - n_{10}n_{01}}{\sqrt{n_{1\cdot}n_{0\cdot}n_{\cdot 0}n_{\cdot 1}}}, \tag{2.1}$$

where frequencies n for two words X and Y are defined in Table 2.1.

	Has Y	No Y	Total
Has X	n_{11}	n_{10}	$n_{1\cdot}$
No X	n_{01}	n_{00}	$n_{0\cdot}$
Total	$n_{\cdot 1}$	$n_{\cdot 0}$	

Table 2.1: Definition of frequencies n for two words X and Y in Equation 2.1.

2.5.2 Spatial distribution

We investigate several of the most common recreational activities in Ann Arbor based on the SUN attributes, which were reclassified into possible functions (e.g., boating, eating). We assume that each photo observation is as a site measurement of enjoying an urban activity that can be fulfilled by that location and surrounding areas after being filtered by UDL as described in Section 2.2. We map this recreational suitability using kernel density estimation (KDE) and hexagonal binning (500 meters) to assess the spatial distribution of popular areas for various activities (*Silverman*, 2018).

CHAPTER III

Result

This chapter presents the experiment results of our study. First, we give an overview of available Flickr data in the research area in Section 3.1. Then, in Section 3.2, we report the validation result for the three labeling systems we discuss in the previous Chapter II. After the assessment, we apply chosen labeling systems to the whole data and provide an overview of content recognition results in Section 3.3. Finally, we generate the investigation of urban RES in Section 3.4 and in Section 3.5.

3.1 Overview of available Flickr data

We retrieved 157,511 photographs uploaded to Flickr by 2,302 photographers from 2005 to 2019 within Ann Arbor. The median number of photographs each user took is 5, while the mean is 68, and 10.82% of photographers uploaded more than 50 photographs. The standard deviation of the number of photographs by each user is 68.42. The maximum number of photographs uploaded by a single user is 88,434, suggesting that the data is highly unbalanced. After filtering for high-frequency photography at a single 0.5-meter location using our UDI definition, we were left with 38,628 photographs, partially solving this data imbalance (supplementary material Figure A.1). This filtered dataset resulted in a median of 3, a mean of 16, and a reduction of high-frequency photographs bias our sample (e.g., 6.21% of photographers uploaded more than 50 photographs). Still, there is a standard deviation of 96.21 for all photographers, and a single user contributing 3,981 suggests some bias. While this might skew the visitation analysis, our interest is broader, and keeping these photographs provides more insights into the city's aesthetics related to recreational activities.

The number of overall photos retrieved has increased rapidly since 2013. In contrast, the number of filtered images by UDL showed a downward trend since 2013 (supplementary material Figure A.2). To a certain extent, the trend in the number of filtered photographs corresponds to the platform's popularity. The growth in original photo volume after 2013 may be due to (1) changes in the overall behavior pattern of users capturing and uploading images, and (2) the contributions of photography enthusiasts. The monthly changes in photo proportion in a year are relatively consistent before and after UDL filtering, except for a significant increase in July and a decrease in January and June. After filtering, the largest number of photos were taken in July, and the lowest was in December (supplementary material Figure A.3).

3.2 Assessment of recognition content

The overall (observed) accuracy of the environment type recognition is 94.07%. As shown in Table 3.1, the user accuracy of the indoor environment type is 85.22%. The user accuracy of the outdoor environment type is 98.06%. In other words, among the photos recognized as the outdoor environment type, 98.06% of them are indeed taken outdoor. Additionally, the producer accuracy of the outdoor environment is 93.62%, suggesting 6.38% of the ground truth outdoor photos are incorrectly classified as the outdoor. The F1 score for the outdoor category is 0.958. The weighted Cohen's kappa of the agreement between the manual and automated classification was 0.858, which is considered perfect in the original paper (*Landis and Koch*, 1977). Overall, the agreement manual and computer vision recognition of the environment type is pretty high; thus, we can use computer vision as the primary tool to identify whether the photo is taken indoors or outdoor.

	Ground truth					
Class	Indoor	Outdoor	Total	Producer acc.	User acc.	F1
Indoor	398	69	467	95.22%	85.22%	0.899
Outdoor	20	1013	1033	**93.62%**	**98.06%**	**0.958**
Total	418	1082	1500			

Table 3.1: Recognition accuracy of the environment type for Flickr data by the model Places365-ResNet: confusion matrix, producer accuracy, user accuracy and F1 score. Accuracy assessed by 1,500 photographs.

Accuracy assessments of the Places scene and the SUN attribute are conducted for photos classed as 'outdoor' by the computer vision model, 1,033 in total. Table 3.2 indicates low classified accuracy within the highest probability class (21.30%), and somewhat better accuracy with the Top-5 probability classes (53.34%). This is considerably less accurate when comparing to the original study of ResNet training on the Places365 dataset. The poor performance is likely due to difficulty in assigning Places labels in the urban context. For example, many photographs with bare earth were classified as tundra, while with snow and ice are classed as glaciers. Due to this location misclassification, we exclude the Places scene recognition result for following analyses in 3.4 and 3.5.

	Top-1 acc.	Top-5 acc.
Places365-ResNet (*Zhou et al.*, 2017)	54.65%	85.07%
Flickr-ResNet (Ann Arbor)	**21.30%**	53.34%

Table 3.2: Recognition accuracy of the Places scene for Flickr data: a comparison of Top-1 and Top-5 accuracy with results on the original testing dataset. Accuracy assessed by 1,033 photographs.

Assessing the accuracy of the SUN attribute recognition results using our criteria score (Score 1, totally wrong; Score 5 totally correct, which is shown in supplementary material Table A.2) resulted in an overall score of 4.14 with a standard deviation of 0.90. Over 75% of the testing cases achieve over 4 points, and over 25% achieve 5 points. This indicates high accuracy in predicting SUN attributes for our sample photographs. Table 3.3 shows descriptive statistics of manual grading and accuracy of recognition for each category of the SUN attribute. Notably, the second line of each grade is the count proportion for that grade of the count of photos with a related category tag. It is worth noting that materials, lighting, and spatial categories have high prediction accuracy, with precisely right over about 80% and partly and precisely right over about 95%. Validation samples relating to the moods category are small, which might not be sufficient for discussion. The accuracy for the surface category is average. Besides, the functions category criteria is slightly different: if there exist human actions in the photo and tags are correct, we give 2; if there are no human actions in the photo and tags show potential activities, we give 1 point. Therefore, the recognition accuracy of functions also reaches 80%.

	Functions	Materials	Lighting	Spatial	Surface	Moods
no related tags (0)	633	4	2	3	651	1016
wrong (-1)	68	44	16	26	86	4
or not suitable	17.00%	4.28%	1.55%	2.52%	22.51%	23.53%
partly right (1)	165	135	18	218	150	1
or suitable	41.25%	13.12%	1.75%	21.17%	39.27%	5.88%
precisely right (2)	167	850	997	796	136	12
or exist	41.75%	82.60%	96.70%	77.28%	35.60%	70.59%

Table 3.3: Recognition accuracy of the SUN attribute for Flickr data: a descriptive statistics of manual grading for each category. Accuracy assessed by 1,033 photographs, is the count proportion for a certain grade of the count of photos with a related category tag.

To conclude, models pre-trained with the environment type and the SUN attribute labeling systems achieve notable recognition accuracy when applied to Flickr data in the urban area. In contrast, the model's performance with Places scene labeling systems was less accurate. The environment type recognition result shows good performance in filtering out photos taken indoors. The SUN attribute recognition result delivers a suitable description of the image scene, including environment compositions and existing or potential human activities, which provides much information for following urban RES analyses.

3.3 Overview of recognition content

As discussed in the previous Section 3.2, we apply the environment recognition result to filtered photos, resulting in 22,795 outdoor photographs with a unique UDL. Around 59% of the UDL-filtered photos were taken outdoors, compared to 65% before filtering. The rapid growth of original photo volume after 2013 is mainly due to a large number of repeated outdoor photos (supplementary material Figure A.4).

Figure 3.1 summarizes the relative frequency of SUN attributes in Ann Arbor for both original and UDL-filtered Flickr photos. For UDL-filtered photos, the most frequently assigned tag is ‘natural light’ (assigned to 22,465 photos, 98.6%), followed by ‘open area’ (assigned to 21,501 photos, 94.3%), ‘no horizon’ (assigned to 19,017 photos, 83.4%) and ‘man-made’ (assigned to 17,043 photos, 74.8%). Among environmental composition, the most frequently assigned tag is ‘foliage’ (assigned to 9,039 photos, 40.0%), followed by ‘vegetation’ (assigned to 8,530 photos, 37.4%), ‘trees’ (assigned to 8,276 photos, 36.3%), ‘leaves’ (assigned to 7,268 photos, 31.9%), and ‘grass’ (assigned to 4,665 photos, 20.5%). Among activities, the most frequently assigned tag is ‘driving’ (assigned to 3,304 photos, 14.5%), followed by ‘competing’ (assigned to 2,397 photos, 10.5%), ‘sports’ (assigned to 2,191 photos, 9.6%), ‘transporting’ (assigned to 1,514 photos, 6.7%), and ‘congregating’ (assigned to 1,452 photos, 6.4%). After UDL filtering, tags related with urban increase, such as ‘man-made’ (from 49.2% to 74.7%), ‘asphalt’ (from 4.1% to 11.1%), ‘metal’ (from 7.9% to 10.0%), ‘glass’ (from 3.4% to 9.0%) and ‘pavement’ (from 3.1% to 7.7%). Common recreational activities are ‘biking’, ‘exercise’, ‘picnicking’, ‘touring’ and ‘boating’.

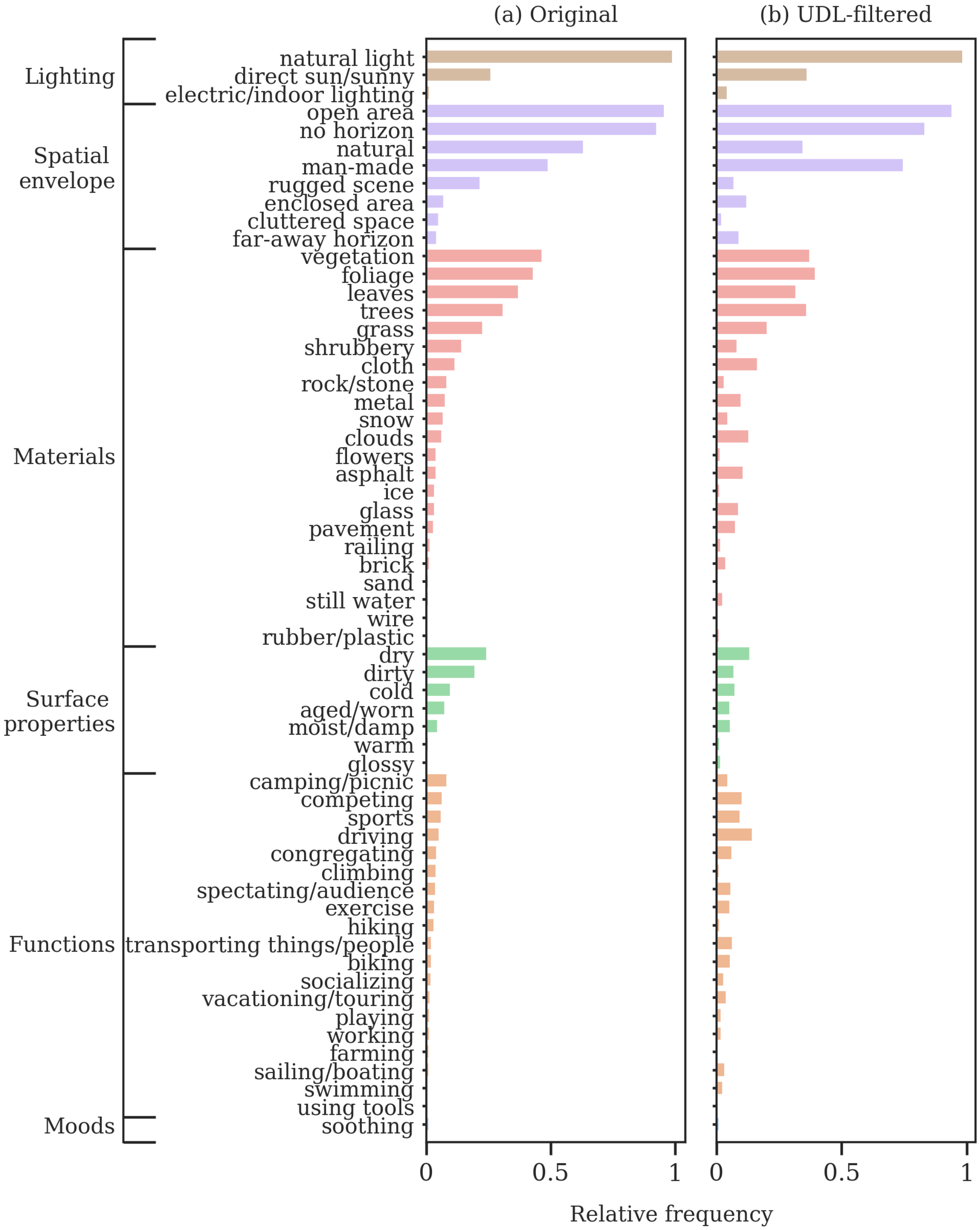


Figure 3.1: Relative frequency of attributes for Flickr photos in Ann Arbor (from 2005 to 2019). (a) is for original photos retrieved from Flickr, (b) is for UDL-filtered photos. Frequency is scaled by the total number of photos, 108,649 and 22,795, respectively. The map only show attribute with frequency number larger than 500 and 100.

3.4 Correlation Analysis

The co-occurrence of tags in photos is visualized via a correlation matrix relating the urban environmental composition tags and human activities (Figure 3.2 and 3.3). Among recreational actives, ‘picnicking’ is highly relevant to the natural environment, having a correlation of 0.275 with ‘grass’ and ‘vegetation’, 0.185 with ‘foliage’, and 0.150 with ‘leaves’. ‘boating’ and ‘swimming’ has a strong correlation with ‘still water’ (0.605 and 0.702 respectively). ‘biking’ has higher requirements for the construction of urban infrastructure roads compared with other recreational actives, with a correlation of 0.618 with ‘asphalt’ and 0.493 with ‘pavement’. For ‘sports’, ‘exercise’ and ‘playing’, except for a positive correlation with ‘grass’, all correlation with other natural composition such as ‘vegetation’, ‘foliage’, ‘trees’, and ‘leaves’ are negative. ‘climbing’ is correlated with ‘rock/stone’. Besides co-occurrence of urban environmental compositions and human activity tags, a complete correlation matrix of all SUN attributes is in the supplementary material Figure A.5 and A.6.

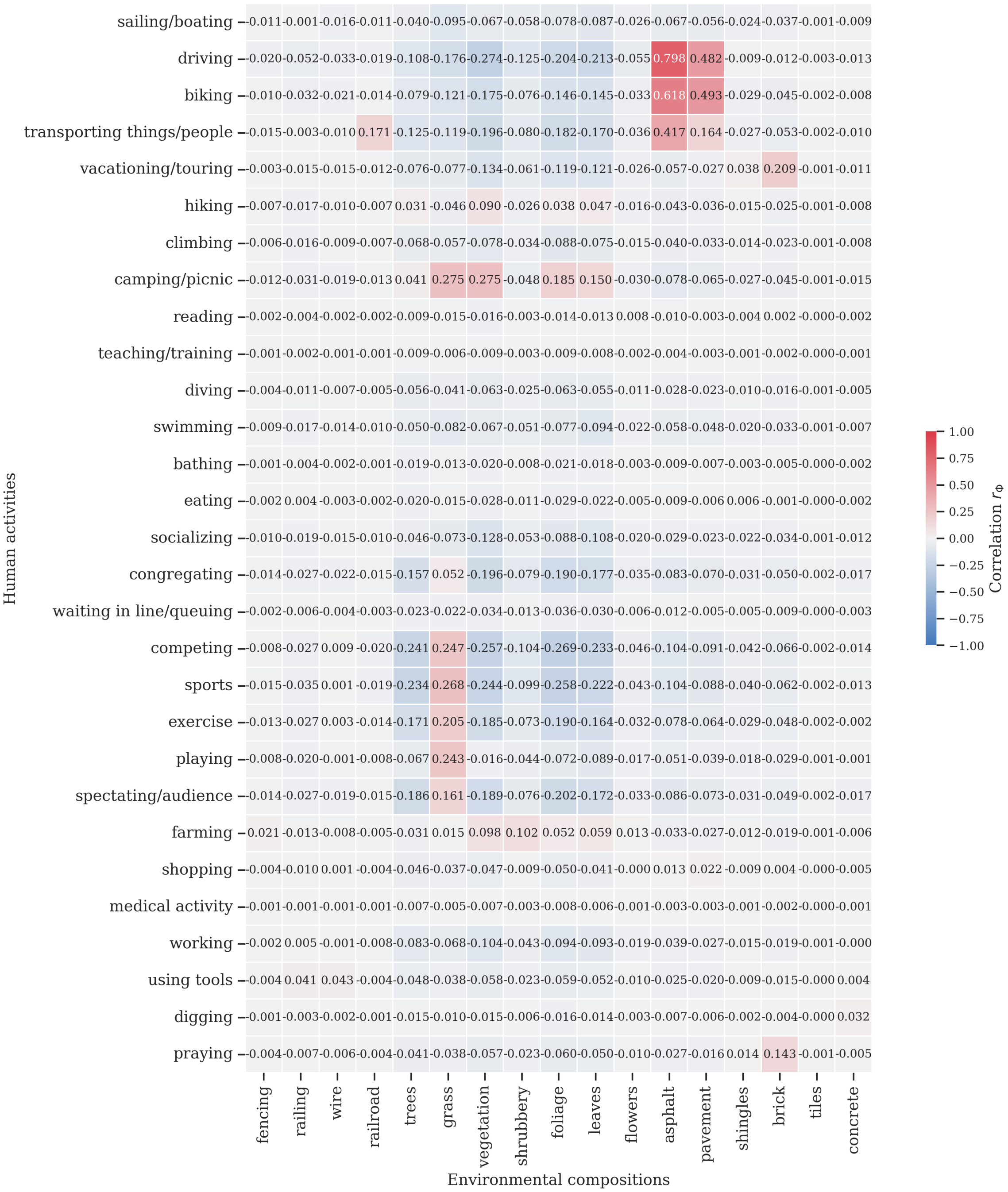
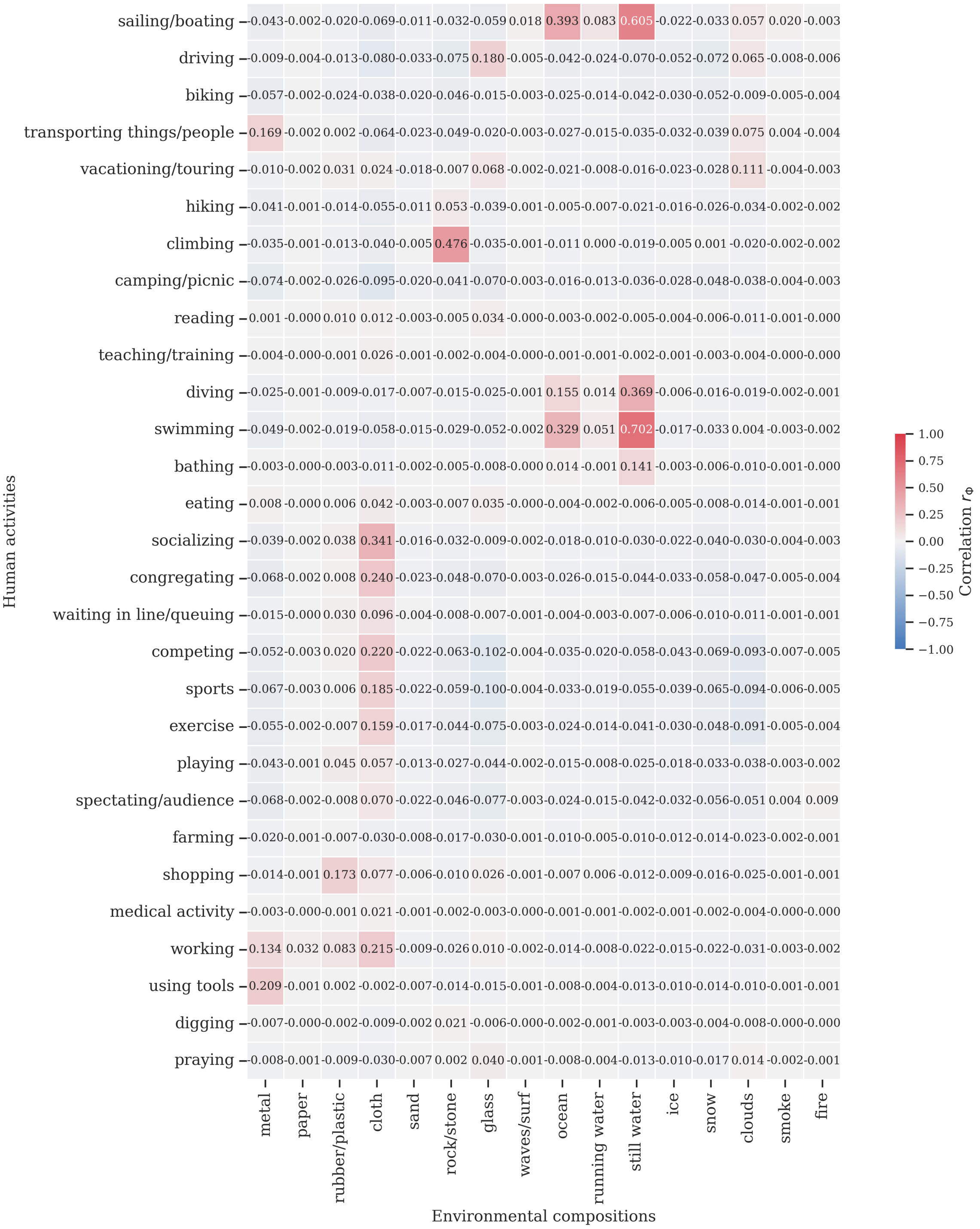


Figure 3.2: Correlation matrix of urban environmental composition and human activities for Flickr photos from 2005 to 2019 in Ann Arbor (left).

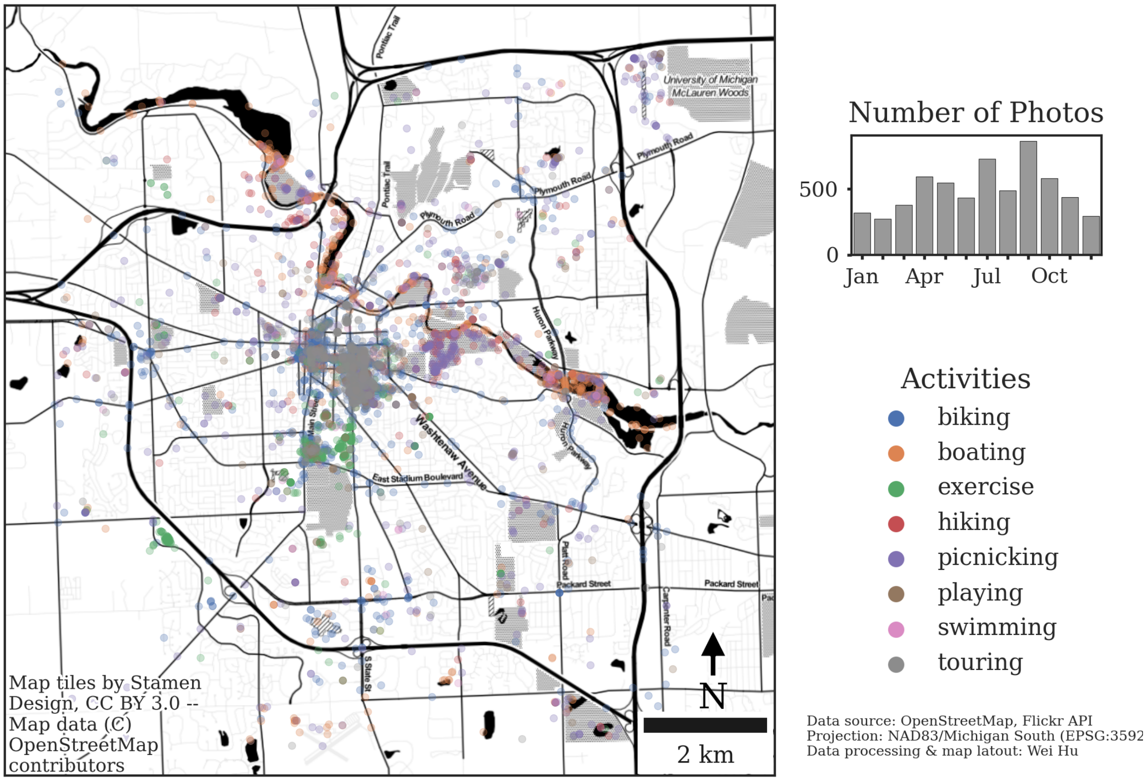


	metal	paper	rubber/plastic	cloth	sand	rock/stone	glass	waves/surf	ocean	running water	still water	ice	snow	clouds	smoke	fire
sailing/boating	-0.043	-0.002	-0.020	-0.069	-0.011	-0.032	-0.059	0.018	0.393	0.083	0.605	-0.022	-0.033	0.057	0.020	-0.003
driving	-0.009	-0.004	-0.013	-0.080	-0.033	-0.075	0.180	-0.005	-0.042	-0.024	-0.070	-0.052	-0.072	0.065	-0.008	-0.006
biking	-0.057	-0.002	-0.024	-0.038	-0.020	-0.046	-0.015	-0.003	-0.025	-0.014	-0.042	-0.030	-0.052	-0.009	-0.005	-0.004
transporting things/people	0.169	-0.002	0.002	-0.064	-0.023	-0.049	-0.020	-0.003	-0.027	-0.015	-0.035	-0.032	-0.039	0.075	0.004	-0.004
vacationing/touring	-0.010	-0.002	0.031	0.024	-0.018	-0.007	0.068	-0.002	-0.021	-0.008	-0.016	-0.023	-0.028	0.111	-0.004	-0.003
hiking	-0.041	-0.001	-0.014	-0.055	-0.011	0.053	-0.039	-0.001	-0.005	-0.007	-0.021	-0.016	-0.026	-0.034	-0.002	-0.002
climbing	-0.035	-0.001	-0.013	-0.040	-0.005	0.476	-0.035	-0.001	-0.011	0.000	-0.019	-0.005	0.001	-0.020	-0.002	-0.002
camping/picnic	-0.074	-0.002	-0.026	-0.095	-0.020	-0.041	-0.070	-0.003	-0.016	-0.013	-0.036	-0.028	-0.048	-0.038	-0.004	-0.003
reading	0.001	-0.000	0.010	0.012	-0.003	-0.005	0.034	-0.000	-0.003	-0.002	-0.005	-0.004	-0.006	-0.011	-0.001	-0.000
teaching/training	-0.004	-0.000	-0.001	0.026	-0.001	-0.002	-0.004	-0.000	-0.001	-0.001	-0.002	-0.001	-0.003	-0.004	-0.000	-0.000
diving	-0.025	-0.001	-0.009	-0.017	-0.007	-0.015	-0.025	-0.001	0.155	0.014	0.369	-0.006	-0.016	-0.019	-0.002	-0.001
swimming	-0.049	-0.002	-0.019	-0.058	-0.015	-0.029	-0.052	-0.002	0.329	0.051	0.702	-0.017	-0.033	0.004	-0.003	-0.002
bathing	-0.003	-0.000	-0.003	-0.011	-0.002	-0.005	-0.008	-0.000	0.014	-0.001	0.141	-0.003	-0.006	-0.010	-0.001	-0.000
eating	0.008	-0.000	0.006	0.042	-0.003	-0.007	0.035	-0.000	-0.004	-0.002	-0.006	-0.005	-0.008	-0.014	-0.001	-0.001
socializing	-0.039	-0.002	0.038	0.341	-0.016	-0.032	-0.009	-0.002	-0.018	-0.010	-0.030	-0.022	-0.040	-0.030	-0.004	-0.003
congregating	-0.068	-0.002	0.008	0.240	-0.023	-0.048	-0.070	-0.003	-0.026	-0.015	-0.044	-0.033	-0.058	-0.047	-0.005	-0.004
waiting in line/queuing	-0.015	-0.000	0.030	0.096	-0.004	-0.008	-0.007	-0.001	-0.004	-0.003	-0.007	-0.006	-0.010	-0.011	-0.001	-0.001
competing	-0.052	-0.003	0.020	0.220	-0.022	-0.063	-0.102	-0.004	-0.035	-0.020	-0.058	-0.043	-0.069	-0.093	-0.007	-0.005
sports	-0.067	-0.003	0.006	0.185	-0.022	-0.059	-0.100	-0.004	-0.033	-0.019	-0.055	-0.039	-0.065	-0.094	-0.006	-0.005
exercise	-0.055	-0.002	-0.007	0.159	-0.017	-0.044	-0.075	-0.003	-0.024	-0.014	-0.041	-0.030	-0.048	-0.091	-0.005	-0.004
playing	-0.043	-0.001	0.045	0.057	-0.013	-0.027	-0.044	-0.002	-0.015	-0.008	-0.025	-0.018	-0.033	-0.038	-0.003	-0.002
spectating/audience	-0.068	-0.002	-0.008	0.070	-0.022	-0.046	-0.077	-0.003	-0.024	-0.015	-0.042	-0.032	-0.056	-0.051	0.004	0.009
farming	-0.020	-0.001	-0.007	-0.030	-0.008	-0.017	-0.030	-0.001	-0.010	-0.005	-0.010	-0.012	-0.014	-0.023	-0.002	-0.001
shopping	-0.014	-0.001	0.173	0.077	-0.006	-0.010	0.026	-0.001	-0.007	0.006	-0.012	-0.009	-0.016	-0.025	-0.001	-0.001
medical activity	-0.003	-0.000	-0.001	0.021	-0.001	-0.002	-0.003	-0.000	-0.001	-0.001	-0.002	-0.001	-0.002	-0.004	-0.000	-0.000
working	0.134	0.032	0.083	0.215	-0.009	-0.026	0.010	-0.002	-0.014	-0.008	-0.022	-0.015	-0.022	-0.031	-0.003	-0.002
using tools	0.209	-0.001	0.002	-0.002	-0.007	-0.014	-0.015	-0.001	-0.008	-0.004	-0.013	-0.010	-0.014	-0.010	-0.001	-0.001
digging	-0.007	-0.000	-0.002	-0.009	-0.002	0.021	-0.006	-0.000	-0.002	-0.001	-0.003	-0.003	-0.004	-0.008	-0.000	-0.000
praying	-0.008	-0.001	-0.009	-0.030	-0.007	0.002	0.040	-0.001	-0.008	-0.004	-0.013	-0.010	-0.017	0.014	-0.002	-0.001

Figure 3.3: Correlation matrix of urban environmental composition and human activities for Flickr photos from 2005 to 2019 in Ann Arbor (right).

3.5 Spatial distribution

The spatial distribution of RES in Ann Arbor is presented using: (1) points pattern of all sites that provide outdoor recreational function, (2) spatial distribution characterization for individual recreational functions (biking, boating, exercise, hiking, and picnicking) in Ann Arbor. Figure 3.4 shows the spatial distribution of all Flickr photos with recreational tags. Users were more active in April, May, July, September and October. Photos were mainly distributed on the campus of the University of Michigan and near the parks that scatter around the city. Two special cases are exercise and boating. Photos with tag exercise concentrated in the Michigan Stadium and several sports centers in Ann Arbor and tag boating are distributed along the Huron river.



Map tiles by Stamen Design, CC BY 3.0 -- Map data (C) OpenStreetMap contributors

Data source: OpenStreetMap, Flickr API
Projection: NAD83/Michigan South (EPSG:3592
Data processing & map latout: Wei Hu

Figure 3.4: Spatial distribution of Flickr photos with recreational activity tags in Ann Arbor and distribution of photos over months (from 2005 to 2019).

Figures 3.5 to 3.9 show kernel density heat maps of biking, boating, exercise, hiking, and picnicking, respectively. In Ann Arbor, the most active month for cycling is July. Popular sites mainly concentrate in urban center areas, and some urban natural areas such as parks. People tend to ride where there is asphalt, pavement, etc., with urban road infrastructure. The most active month for boating is July and October. Popular sites are along with the Huron river, mainly distributed in the Argo nature area, Fuller park, and Gallup park. People tend to boat in areas with open views, clear water, and gentle currents. The most active month for exercise is September. Popular sites are Michigan Stadium and several sports centers scattered around the city. As the largest stadium in the United States, Michigan Stadium has a great influence on the distribution of sports venues and September is also the month when the new football season begins. Hiking activity is more evenly distributed across the months, with the exception of February and August, which are the coldest and hottest times in Ann Arbor, respectively. The spatial distribution of hiking also shifts from the main university campus to the botanical gardens and surrounding parks along the Huron river. People prefer to avoid man-made structures and direct sunlight and go deeper into natural areas with abundant vegetation. Picnicking is mainly distributed from April to October of the year, which also avoids the snow season in Ann Arbor. Popular sites are distributed in large urban parks and small community green spaces. People avoid too many man-made elements during Picnicking and prefer places with grass and vegetation. At the same time, it is hoped that the site will be more open and have a beautiful view.

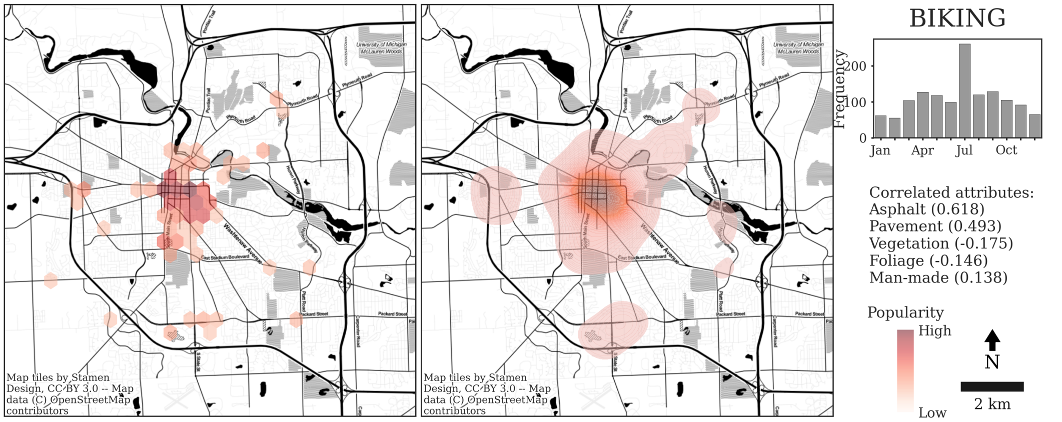

BIKING
Frequency
0
100
200
Jan
Apr
Jul
Oct
Correlated attributes:
Asphalt (0.618)
Pavement (0.493)
Vegetation (-0.175)
Foliage (-0.146)
Man-made (0.138)
Popularity
High
Low
N
2 km
Map tiles by Stamen Design, CC BY 3.0 -- Map data (C) OpenStreetMap contributors
Map tiles by Stamen Design, CC BY 3.0 -- Map data (C) OpenStreetMap contributors

Figure 3.5: Spatial distribution of Flickr photos with biking tag in Ann Arbor and distribution of photos over the months (from 2005 to 2019).

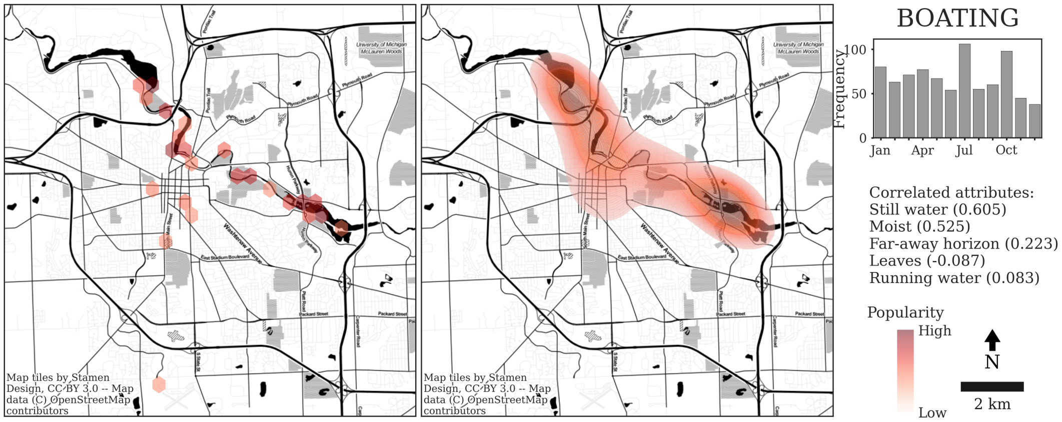


Figure 3.6: Spatial distribution of Flickr photos with boating tag in Ann Arbor and distribution of photos over the months (from 2005 to 2019).

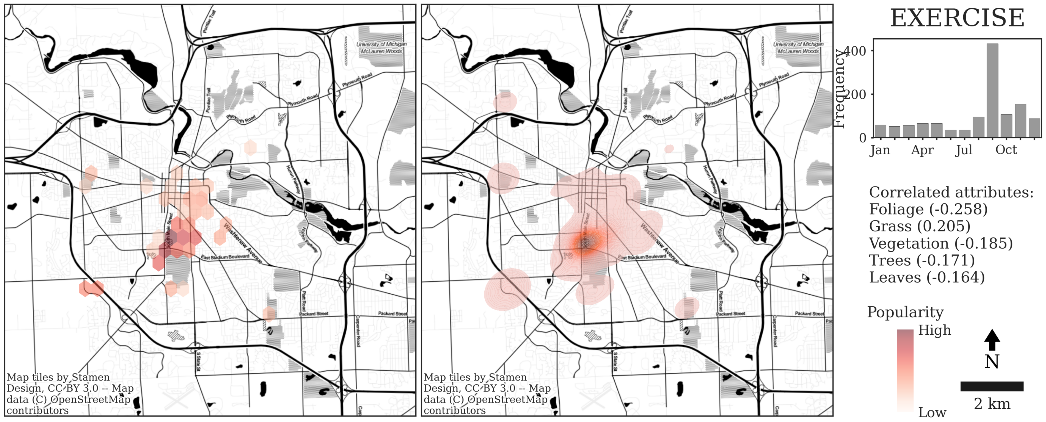

Map tiles by Stamen Design, CC BY 3.0 -- Map data (C) OpenStreetMap contributors
Map tiles by Stamen Design, CC BY 3.0 -- Map data (C) OpenStreetMap contributors
EXERCISE
Frequency
400
200
0
Jan
Apr
Jul
Oct
Correlated attributes:
Foliage (-0.258)
Grass (0.205)
Vegetation (-0.185)
Trees (-0.171)
Leaves (-0.164)
Popularity
High
Low
N
2 km

Figure 3.7: Spatial distribution of Flickr photos with exercise tag in Ann Arbor and distribution of photos over the months (from 2005 to 2019).

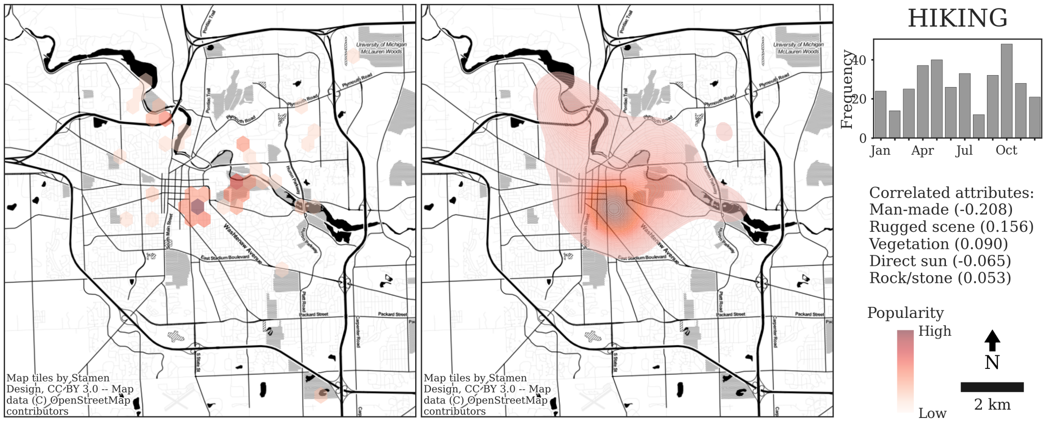


Figure 3.8: Spatial distribution of Flickr photos with hiking tag in Ann Arbor and distribution of photos over the months (from 2005 to 2019).

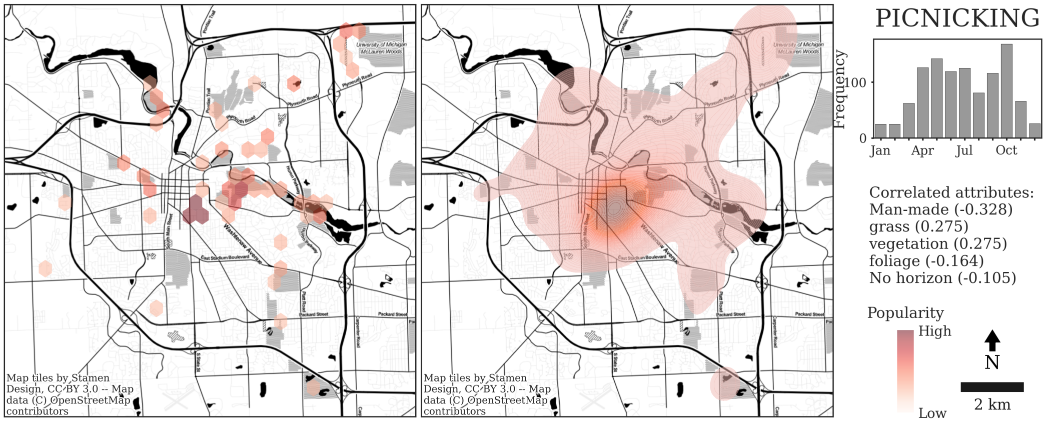


Figure 3.9: Spatial distribution of Flickr photos with biking tag in Ann Arbor and distribution of photos over the months (2005 - 2019).

CHAPTER IV

Discussion and Conclusion

Geospatial big data retrieved from social media brings new opportunities for researchers to investigate and analyze human-nature interactions and RES. CV in artificial intelligence opens up a feasible way to analyze the contents of a large volume of photos from photo-sharing platforms such as Flickr. To date, most related RES assessments have relied on commercial cloud-based services and focused on natural and ecological areas outside cities on a regional or national scale. While not a huge expense, the cost of analyzing images using commercial services (e.g., $1.50 per 1,000 images for Google Cloud Vision) may constrain the use of the approach at very large scales. This thesis offers new insights on applying open-sourced CV models to RES investigation and refines the investigation perspective down to the urban context.

Accurate investigation of RES can be enhance by characterization of associated geographic and location characteristics, such as the environmental composition and landscape preferences for activities, which can be inferred from social media photographs. Although computer vision models trained on different datasets and labeling systems perform well on standard test datasets, their performance in recognizing the content of complex images in real scenarios (e.g., distortion, compression, blurriness, and rotation of the images) is unclear. In addition, it is unclear whether the labeling system can provide recognition results that are suitable for urban RES study. Building on such insights, this study assesses three labeling systems with different semantic scales: the environment type (indoor or outdoor), the SUN attributes, and Places scene. We find that the three investigated labeling systems differ in the accuracy of recognition results for Flickr photos in urban regions. Models with the environment type and the SUN attributes labeling systems perform well in photo recognition. In contrast, the model with the Places scene performs exceptionally poorly compared to the standard dataset. This may be due to: (1) the scene labels in Places describe photos at a large geographic scale, such as deserts, glaciers, deciduous broad-leaved forests, etc.. Such scene labels are difficult to apply to describe urban environments. (2) Tags in the Places scene are overall descriptions of the photo scene, and each photo can be assigned one Places scene tag, while it is hard for a single tag to hit the photo scene accurately.

We investigate RES for Ann Arbor by combining recognition results using the environment type and the SUN attributes. We are able to identify specific popular recreational activities, such as biking, boating, exercise, hiking, and picnicking, and their corresponding environment composition and spatial distribution. The content analysis of more than 20,000 geotagged photographs taken in Ann Arbor offers an insight into the richness of information that can be retrieved from such crowd-sourced data, well beyond the simple estimation of visitation intensity. Such information may provide valuable insights to city managers in setting priorities for developing recreational infrastructure, identifying undervalued areas, or spotting sites that require major maintenance during a particular season of high visiting interest. This study shows how passive crowd-sourced data can be a cost-effective tool to complement and extend the currently available information.

Some of the limitations of the study require clarification. The investigation relies on a sample of photographs obtained from one social media platform. In this article, we consider the occurrence of nature photographs as a general indicator of the public interest in nature, while motivations for people to take photographs of nature vary; some record positive attributes of the environment that they find appealing, while others record negative attributes of the environment or social event they are involved. Any analysis of social media data is limited by the biases inherent in the source dataset, particularly as younger people most commonly use social media and different demographic groups use different platforms. Multiple sources need to be considered to reduce user and content biases (*Ghermandi et al.*, 2020a). Another limitation is the exclusive reliance on automatically assigned tags to analyze photographic content. It should be noted that algorithms are inconsistent in their recognition accuracy for specific labels. We did not investigate whether they are significantly different and, if so, how to account for these inconsistencies in further analysis. This should be considered when evaluating the overall distribution of tags, although it is not expected to significantly affect the comparative analysis results. The analyses performed in this study are exploratory and require further robustness tests in the future.

We investigate environmental composition and landscape properties purely based on automatic tags; the study dataset can be further expanded with data from other photo-sharing platforms such as Panoramio, and standard geographic datasets such as land cover. Besides, there is a potential to enrich point data sources with other types of crowd-sourced information, such as GPS-based trajectories from sports or hiking applications. We believe that there is scope for future improvement of the accuracy of computer vision services in environmental studies. There are three sentiment-related tags in the SUN attribute. More emotion-related tags provide a new direction for RES study, especially in the characterization of feelings and sentiments in outdoor photographs and the associated cultural benefits. Further research could also characterize the interests and preferences of different subsets of the population, such as local citizens and tourists: young people, the elder, and people with kids. This information would enable a better representation of RES for landscape planning and spatial analysis.

BIBLIOGRAPHY

Alberti, M., and J. M. Marzluff (2004), Ecological resilience in urban ecosystems: linking urban patterns to human and ecological functions, *Urban ecosystems*, *7*(3), 241–265.

Bartczak, A., H. Lindhjem, S. Navrud, M. Zandersen, and T. Żylicz (2008), Valuing forest recreation on the national level in a transition economy: The case of poland, *Forest Policy and Economics*, *10*(7-8), 467–472.

Beyer, K. M., A. Kaltenbach, A. Szabo, S. Bogar, F. J. Nieto, and K. M. Malecki (2014), Exposure to neighborhood green space and mental health: evidence from the survey of the health of wisconsin, *International journal of environmental research and public health*, *11*(3), 3453–3472.

Biljecki, F., and K. Ito (2021), Street view imagery in urban analytics and gis: A review, *Landscape and Urban Planning*, *215*, 104,217.

Bratman, G. N., J. P. Hamilton, K. S. Hahn, G. C. Daily, and J. J. Gross (2015), Nature experience reduces rumination and subgenual prefrontal cortex activation, *Proceedings of the national academy of sciences*, *112*(28), 8567–8572.

Brown, K. M. (2010), *Assessing future recreation demand*, Scottish Natural Heritage.

Buslaev, A., S. Seferbekov, V. Iglovikov, and A. Shvets (2018), Fully convolutional network for automatic road extraction from satellite imagery, in *Proceedings of the IEEE conference on computer vision and pattern recognition workshops*, pp. 207–210.

Chen, L., Y. Lu, Q. Sheng, Y. Ye, R. Wang, and Y. Liu (2020), Estimating pedestrian volume using street view images: A large-scale validation test, *Computers, Environment and Urban Systems*, *81*, 101,481.

Clarke, J. F. (1972), Some effects of the urban structure on heat mortality, *Environmental research*, *5*(1), 93–104.

Cordts, M., M. Omran, S. Ramos, T. Rehfeld, M. Enzweiler, R. Benenson, U. Franke, S. Roth, and B. Schiele (2016), The cityscapes dataset for semantic urban scene understanding, in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 3213–3223.

Costanza, R., et al. (1997), The value of the world's ecosystem services and natural capital, *nature*, *387*(6630), 253–260.

Cramér, H. (2016), Mathematical methods of statistics (pms-9), volume 9, in *Mathematical Methods of Statistics (PMS-9), Volume 9*, Princeton university press.

De Valck, J., D. Landuyt, S. Broekx, I. Liekens, L. De Nocker, and L. Vranken (2017), Outdoor recreation in various landscapes: Which site characteristics really matter?, *Land Use Policy*, *65*, 186–197.

Desa, U. (2018), Revision of world urbanization prospects, *UN Department of Economic and Social Affairs*, *16*.

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

Donaire, J. A., R. Camprubí, and N. Galí (2014), Tourist clusters from flickr travel photography, *Tourism management perspectives*, *11*, 26–33.

Dorwart, C. E., R. L. Moore, and Y.-F. Leung (2009), Visitors' perceptions of a trail environment and effects on experiences: A model for nature-based recreation experiences, *Leisure Sciences*, *32*(1), 33–54.

Fischer, L. K., et al. (2018), Recreational ecosystem services in european cities: Sociocultural and geographical contexts matter for park use, *Ecosystem services*, *31*, 455–467.

Fox, N., T. August, F. Mancini, K. E. Parks, F. Eigenbrod, J. M. Bullock, L. Sutter, and L. J. Graham (2020), "photosearcher" package in r: An accessible and reproducible method for harvesting large datasets from flickr, *SoftwareX*, *12*, 100,624.

García-Palomares, J. C., M. H. Salas-Olmedo, B. Moya-Gomez, A. Condeco-Melhorado, and J. Gutierrez (2018), City dynamics through twitter: Relationships between land use and spatiotemporal demographics, *Cities*, *72*, 310–319.

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.

Ghermandi, A., V. Camacho-Valdez, and H. Trejo-Espinosa (2020a), Social media-based analysis of cultural ecosystem services and heritage tourism in a coastal region of mexico, *Tourism Management*, *77*, 104,002.

Ghermandi, A., M. Sinclair, E. Fichtman, and M. Gish (2020b), Novel insights on intensity and typology of direct human-nature interactions in protected areas through passive crowdsourcing, *Global Environmental Change*, *65*, 102,189.

Ghermandi, A., Y. Depietri, and M. Sinclair (2022), In the ai of the beholder: A comparative analysis of computer vision-assisted characterizations of human-nature interactions in urban green spaces, *Landscape and Urban Planning*, *217*, 104,261.

Girardin, F., F. Calabrese, F. Dal Fiore, C. Ratti, and J. Blat (2008), Digital footprinting: Uncovering tourists with user-generated content, *IEEE Pervasive computing*, *7*(4), 36–43.

Girdhar, R., G. Gkioxari, L. Torresani, M. Paluri, and D. Tran (2018), Detect-and-track: Efficient pose estimation in videos, in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 350–359.

Gliozzo, G., N. Pettorelli, and M. Haklay (2016), Using crowdsourced imagery to detect cultural ecosystem services: a case study in south wales, uk, *Ecology and Society*, *21*(3).

Goossen, M., and F. Langers (2000), Assessing quality of rural areas in the netherlands: finding the most important indicators for recreation, *Landscape and urban planning*, *46*(4), 241–251.

Guerrero, P., M. S. Møller, A. S. Olafsson, and B. Snizek (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), 1–17.

Guo, Y., Y. Liu, A. Oerlemans, S. Lao, S. Wu, and M. S. Lew (2016), Deep learning for visual understanding: A review, *Neurocomputing*, *187*, 27–48.

Haase, D., et al. (2014), A quantitative review of urban ecosystem service assessments: concepts, models, and implementation, *Ambio*, *43*(4), 413–433.

Haines-Young, R., and M. Potschin (2012), Common international classification of ecosystem services (cices, version 4.1), *European Environment Agency*, *33*, 107.

Hamaguchi, R., and S. Hikosaka (2018), Building detection from satellite imagery using ensemble of size-specific detectors, in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, pp. 187–191.

Hamstead, Z. A., D. Fisher, R. T. Ilieva, S. A. Wood, T. McPhearson, and P. Kremer (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., T. Toivonen, R. Slotow, H. Tenkanen, A. Moilanen, V. Heikinheimo, and E. Di Minin (2018), Social media data can be used to understand tourists' preferences for nature-based experiences in protected areas, *Conservation Letters*, *11*(1), e12,343.

He, K., X. Zhang, S. Ren, and J. Sun (2016), Deep residual learning for image recognition, in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 770–778.

Heikinheimo, V., E. D. Minin, H. Tenkanen, A. Hausmann, J. Erkkonen, and T. Toivonen (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.

Hermes, J., D. Van Berkel, B. Burkhard, T. Plieninger, N. Fagerholm, C. von Haaren, and C. Albert (2018), Assessment and valuation of recreational ecosystem services of landscapes.

Herold, M., M. E. Gardner, and D. A. Roberts (2003), Spectral resolution requirements for mapping urban areas, *IEEE Transactions on Geoscience and remote sensing*, *41*(9), 1907–1919.

Hochman, N., and L. Manovich (2013), Zooming into an instagram city: Reading the local through social media, *First Monday*.

Ibrahim, M. R., J. Haworth, and T. Cheng (2020), Understanding cities with machine eyes: A review of deep computer vision in urban analytics, *Cities*, *96*, 102,481.

Johansson, F., U. Shalit, and D. Sontag (2016), Learning representations for counterfactual inference, in *International conference on machine learning*, pp. 3020–3029, PMLR.

Kajala, L. (2007), *Visitor monitoring in nature areas: A manual based on experiences from the Nordic and Baltic countries*, Nordic Council of Ministers.

Karpatne, A., Z. Jiang, R. R. Vatsavai, S. Shekhar, and V. Kumar (2016), Monitoring land-cover changes: A machine-learning perspective, *IEEE Geoscience and Remote Sensing Magazine*, *4*(2), 8–21.

Kienast, F., B. Degenhardt, B. Weilenmann, Y. Wäger, and M. Buchecker (2012), Gis-assisted mapping of landscape suitability for nearby recreation, *Landscape and Urban Planning*, *105*(4), 385–399.

Krizhevsky, A., I. Sutskever, and G. E. Hinton (2012), Imagenet classification with deep convolutional neural networks, *Advances in neural information processing systems*, *25*.

Landis, J. R., and G. G. Koch (1977), The measurement of observer agreement for categorical data, *biometrics*, pp. 159–174.

Latkin, C. A., and A. D. Curry (2003), Stressful neighborhoods and depression: a prospective study of the impact of neighborhood disorder, *Journal of health and social behavior*, pp. 34–44.

Lazebnik, S., C. Schmid, and J. Ponce (2006), Beyond bags of features: Spatial pyramid matching for recognizing natural scene categories, in *2006 IEEE computer society conference on computer vision and pattern recognition (CVPR'06)*, vol. 2, pp. 2169–2178, IEEE.

LeCun, Y., Y. Bengio, and G. Hinton (2015), Deep learning, *nature*, *521*(7553), 436–444.

Lee, H., B. Seo, T. Koellner, and S. Lautenbach (2019), Mapping cultural ecosystem services 2.0–potential and shortcomings from unlabeled crowd sourced images, *Ecological Indicators*, *96*, 505–515.

Lin, T.-Y., M. Maire, S. Belongie, J. Hays, P. Perona, D. Ramanan, P. Dollár, and C. L. Zitnick (2014), Microsoft coco: Common objects in context, in *European conference on computer vision*, pp. 740–755, Springer.

Lu, Y. (2019), Using google street view to investigate the association between street greenery and physical activity, *Landscape and Urban Planning*, *191*, 103,435.

Martí, P., L. Serrano-Estrada, and A. Nolasco-Cirugeda (2019), Social media data: Challenges, opportunities and limitations in urban studies, *Computers, Environment and Urban Systems*, *74*, 161–174.

Martínez Pastur, G., P. L. Peri, M. V. Lencinas, M. García-Llorente, and B. Martín-López (2016), Spatial patterns of cultural ecosystem services provision in southern patagonia, *Landscape ecology*, *31*(2), 383–399.

McCarney, R., J. Warner, S. Iliffe, R. Van Haselen, M. Griffin, and P. Fisher (2007), The hawthorne effect: a randomised, controlled trial, *BMC medical research methodology*, *7*(1), 1–8.

Neuvonen, M., T. Sievänen, S. Tönnes, and T. Koskela (2007), Access to green areas and the frequency of visits–a case study in helsinki, *Urban Forestry & Urban Greening*, *6*(4), 235–247.

Nielsen, T. S., and K. B. Hansen (2007), Do green areas affect health? results from a danish survey on the use of green areas and health indicators, *Health & place*, *13*(4), 839–850.

Noulas, A., S. Scellato, R. Lambiotte, M. Pontil, and C. Mascolo (2012), A tale of many cities: universal patterns in human urban mobility, *PloS one*, *7*(5), e37,027.

Ode, Å., G. Fry, M. S. Tveit, P. Messager, and D. Miller (2009), Indicators of perceived naturalness as drivers of landscape preference, *Journal of environmental management*, *90*(1), 375–383.

Ordonez, V., and T. L. Berg (2014), Learning high-level judgments of urban perception, in *European conference on computer vision*, pp. 494–510, Springer.

Oteros-Rozas, E., B. Martín-López, N. Fagerholm, C. Bieling, and T. Plieninger (2018), Using social media photos to explore the relation between cultural ecosystem services and landscape features across five european sites, *Ecological Indicators*, *94*, 74–86.

Pleasant, M. M., S. A. Gray, C. Lepczyk, A. Fernandes, N. Hunter, and D. Ford (2014), Managing cultural ecosystem services, *Ecosystem Services*, *8*, 141–147.

Quattoni, A., and A. Torralba (2009), Recognizing indoor scenes, in *2009 IEEE conference on computer vision and pattern recognition*, pp. 413–420, IEEE.

Quercia, D., R. Schifanella, L. M. Aiello, and K. McLean (2015), Smelly maps: the digital life of urban smellscapes, in *Ninth International AAAI Conference on Web and Social Media*.

Richards, D. R., and D. A. Friess (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, D. R., and B. Tunçer (2018), Using image recognition to automate assessment of cultural ecosystem services from social media photographs, *Ecosystem services*, *31*, 318–325.

Rose, D. (1999), Economic determinants and dietary consequences of food insecurity in the united states, *The Journal of nutrition*, *129*(2), 517S–520S.

Russakovsky, O., et al. (2015), Imagenet large scale visual recognition challenge, *International journal of computer vision*, *115*(3), 211–252.

Schägner, J. P., L. Brander, J. Maes, M. L. Paracchini, and V. Hartje (2016), Mapping recreational visits and values of european national parks by combining statistical modelling and unit value transfer, *Journal for Nature Conservation*, *31*, 71–84.

Sessions, C., S. A. Wood, S. Rabotyagov, and D. M. Fisher (2016), Measuring recreational visitation at us national parks with crowd-sourced photographs, *Journal of environmental management*, *183*, 703–711.

Sester, M., J. Jokar Arsanjani, R. Klammer, D. Burghardt, and J.-H. Haunert (2014), Integrating and generalising volunteered geographic information, in *Abstracting geographic information in a data rich world*, pp. 119–155, Springer.

Silverman, B. W. (2018), *Density estimation for statistics and data analysis*, Routledge.

Simonyan, K., and A. Zisserman (2014), Very deep convolutional networks for large-scale image recognition, *arXiv preprint arXiv:1409.1556*.

Sinclair, M., M. Mayer, M. Woltering, and A. Ghermandi (2020), Using social media to estimate visitor provenance and patterns of recreation in germany's national parks, *Journal of Environmental Management*, *263*, 110,418.

Srivastava, K. (2009), Urbanization and mental health, *Industrial psychiatry journal*, *18*(2), 75.

Szegedy, C., W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich (2015), Going deeper with convolutions, in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 1–9.

Talukdar, S., P. Singha, S. Mahato, S. Pal, Y.-A. Liou, A. Rahman, et al. (2020), Land-use land-cover classification by machine learning classifiers for satellite observations—a review, *Remote Sensing*, *12*(7), 1135.

Tenerelli, P., C. Püffel, and S. Luque (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., E. Di Minin, V. Heikinheimo, A. Hausmann, M. Herbst, L. Kajala, and T. Toivonen (2017), Instagram, flickr, or twitter: Assessing the usability of social media data for visitor monitoring in protected areas, *Scientific reports*, *7*(1), 1–11.

Twohig-Bennett, C., and A. Jones (2018), The health benefits of the great outdoors: A systematic review and meta-analysis of greenspace exposure and health outcomes, *Environmental research*, *166*, 628–637.

United States Census Bureau (2020), Quickfacts: Ann arbor city, michigan, https://www.census.gov/quickfacts/fact/table/annarborcitymichigan/POP010220.

Van Berkel, D. B., P. Tabrizian, M. A. Dorning, L. Smart, D. Newcomb, M. Mehaffey, A. Neale, and R. K. Meentemeyer (2018), Quantifying the visual-sensory landscape qualities that contribute to cultural ecosystem services using social media and lidar, *Ecosystem services*, *31*, 326–335.

Van Kamp, I., K. Leidelmeijer, G. Marsman, and A. De Hollander (2003), Urban environmental quality and human well-being: Towards a conceptual framework and demarcation of concepts; a literature study, *Landscape and urban planning*, *65*(1-2), 5–18.

Van Zanten, B. T., D. B. Van Berkel, R. K. Meentemeyer, J. W. Smith, K. F. Tieskens, and P. H. Verburg (2016), Continental-scale quantification of landscape values using social media data, *Proceedings of the National Academy of Sciences*, *113*(46), 12,974–12,979.

Viola, P., and M. Jones (2001), Rapid object detection using a boosted cascade of simple features, in *Proceedings of the 2001 IEEE computer society conference on computer vision and pattern recognition. CVPR 2001*, vol. 1, pp. I–I, Ieee.

Welch, R. (1982), Spatial resolution requirements for urban studies, *International Journal of Remote Sensing*, *3*(2), 139–146.

Wood, S. A., A. D. Guerry, J. M. Silver, and M. Lacayo (2013), Using social media to quantify nature-based tourism and recreation, *Scientific reports*, *3*(1), 1–7.

Xiao, J., J. Hays, K. A. Ehinger, A. Oliva, and A. Torralba (2010), Sun database: Large-scale scene recognition from abbey to zoo, in *2010 IEEE computer society conference on computer vision and pattern recognition*, pp. 3485–3492, IEEE.

Zhou, B., A. Khosla, A. Lapedriza, A. Oliva, and A. Torralba (2014a), Object detectors emerge in deep scene cnns, *arXiv preprint arXiv:1412.6856*.

Zhou, B., A. Lapedriza, J. Xiao, A. Torralba, and A. Oliva (2014b), Learning deep features for scene recognition using places database, *Advances in neural information processing systems*, *27*.

Zhou, B., A. Khosla, A. Lapedriza, A. Oliva, and A. Torralba (2016), Learning deep features for discriminative localization, in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 2921–2929.

Zhou, B., A. Lapedriza, A. Khosla, A. Oliva, and A. Torralba (2017), Places: A 10 million image database for scene recognition, *IEEE transactions on pattern analysis and machine intelligence*, *40*(6), 1452–1464.

APPENDIX A

Supplementary Material for Urban Recreational Ecosystem Services Investigation Based on Social Media Images

Category	Attributes
Functions	sailing/boating, driving, biking, transporting things/people, sunbathing, vacationing/touring, hiking, climbing, camping/picnic, reading, studying/learning, teaching/training, research, diving, swimming, bathing, eating, cleaning, socializing, congregating, waiting in line/queuing, competing, sports, exercise, playing, gaming, spectating/audience, farming, constructing/building, shopping, medical activity, working, using tools, digging, conducting business, praying (36)
Materials	fencing, railing, wire, railroad, trees, grass, vegetation, shrubbery, foliage, leaves, flowers, asphalt, pavement, shingles, carpet, brick, tiles, concrete, metal, paper, wood (not part of a tree), vinyl/linoleum, rubber/plastic, cloth, sand, rock/stone, dirt/soil, marble, glass, waves/surf, ocean, running water, still water, ice, snow, clouds, smoke, fire (38)
Lighting	natural light, direct sun/sunny, electric/indoor lighting (3)
Surface properties	aged/worn, glossy, matte, sterile, moist/damp, dry, dirty, rusty, warm, cold (10)
Spatial envelope	natural, man-made, open area, semi-enclosed area, enclosed area, far-away horizon, no horizon, rugged scene, mostly vertical components, mostly horizontal components, symmetrical, cluttered space (12)
Moods	scary, soothing, stressful (3)

Table A.1: Category of the SUN attributes (all). It is worth noting that the SUN attribute is summarized from crowd-sourced human studies. Attributes in the category of functions are human activities of high possibility in the urban area (e.g., driving, biking, transporting, hiking, climbing, picnicking, reading, etc.). Attributes in the category of materials are related to semantic objects, which form the urban appearance, including both natural compositions (e.g., trees, grass, shrubbery, foliage, soil, flowers, water, ice, snow, etc.) and man-made materials of infrastructures (e.g., wire, railroad, asphalt, pavement, etc.).

Score	Criteria
1	Totally wrong.
2	Partly correct, with important information omitted.
3	Partly correct, without important information omitted.
4	Exactly correct, with important information omitted.
5	Exactly correct, without important information omitted.

Table A.2: Criteria for overall score of ten predicted SUN attributes for each images. Score ranges from 1 to 5 based on the feeling of the observer after reading the 10 tags generated from the SUN attribute.

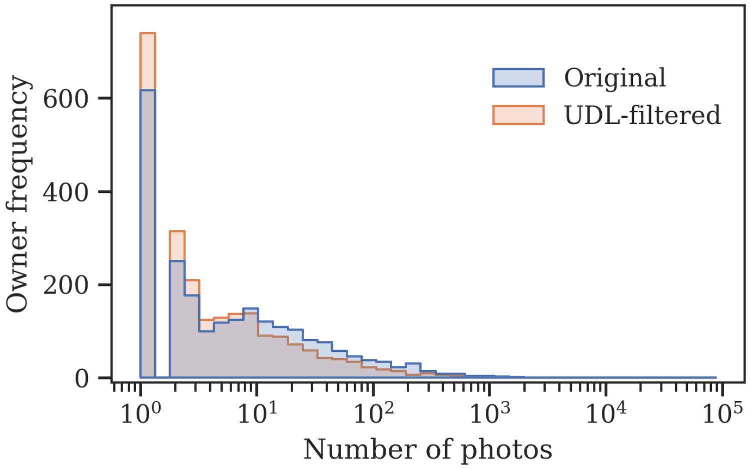


Figure A.1: Distribution comparison of user photo uploading volume between original repository of Flickr and UDL-filtered result in Ann Arbor from 2005 to 2019.

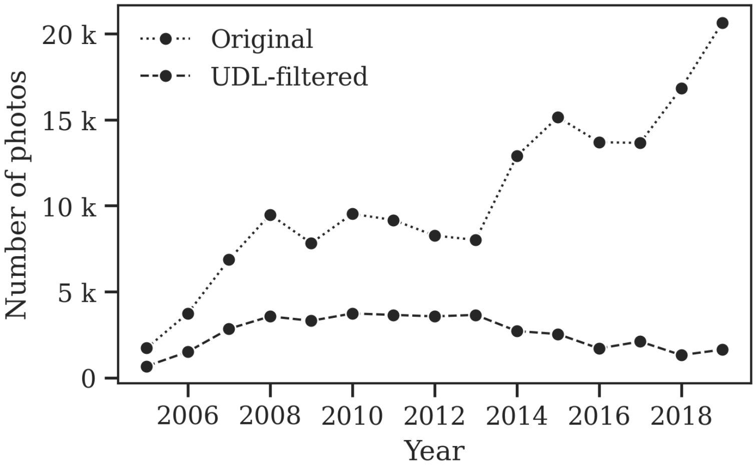


Figure A.2: Comparison of photo volume changes over years (from 2005 to 2019) between original repository of Flickr and UDL-filtered result in Ann Arbor.

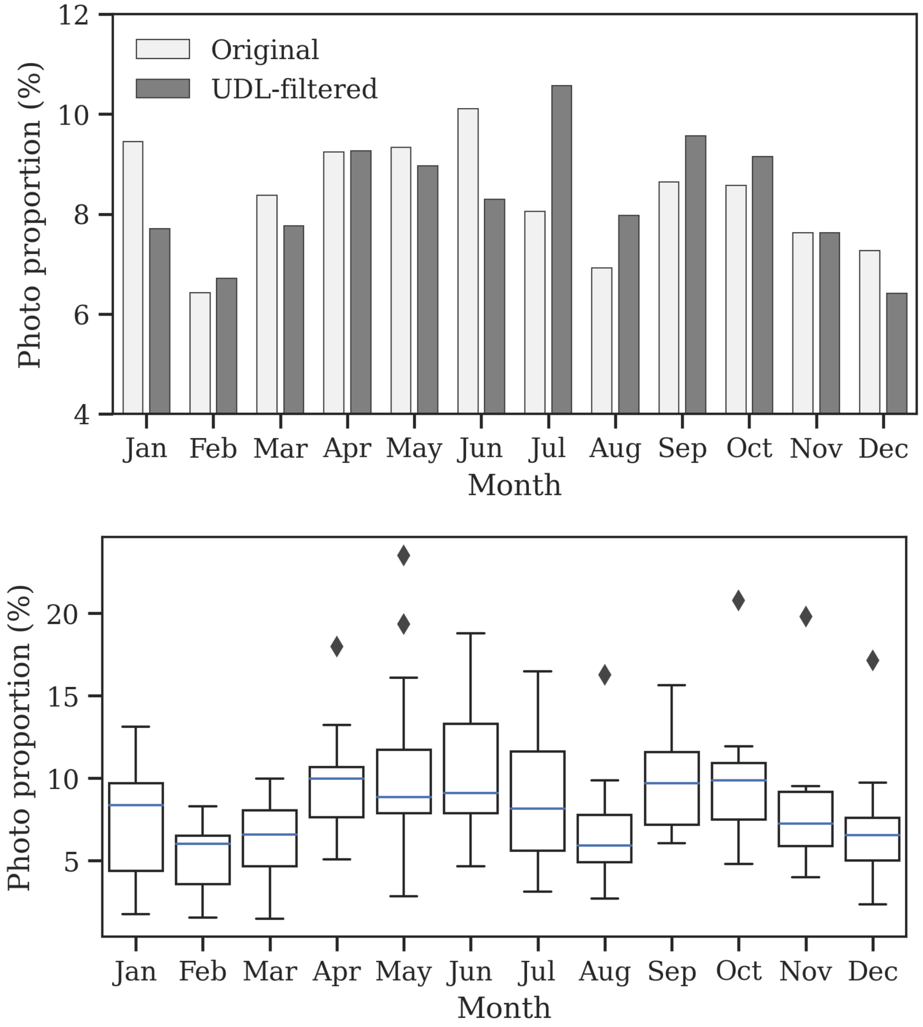


Figure A.3: Photo volume changes over months (from 2005 to 2019) in Ann Arbor. Comparison of total proportion between original repository of Flickr and UDL-filtered data (top) and annual proportion of photo volume for UDL-filtered result (bottom).

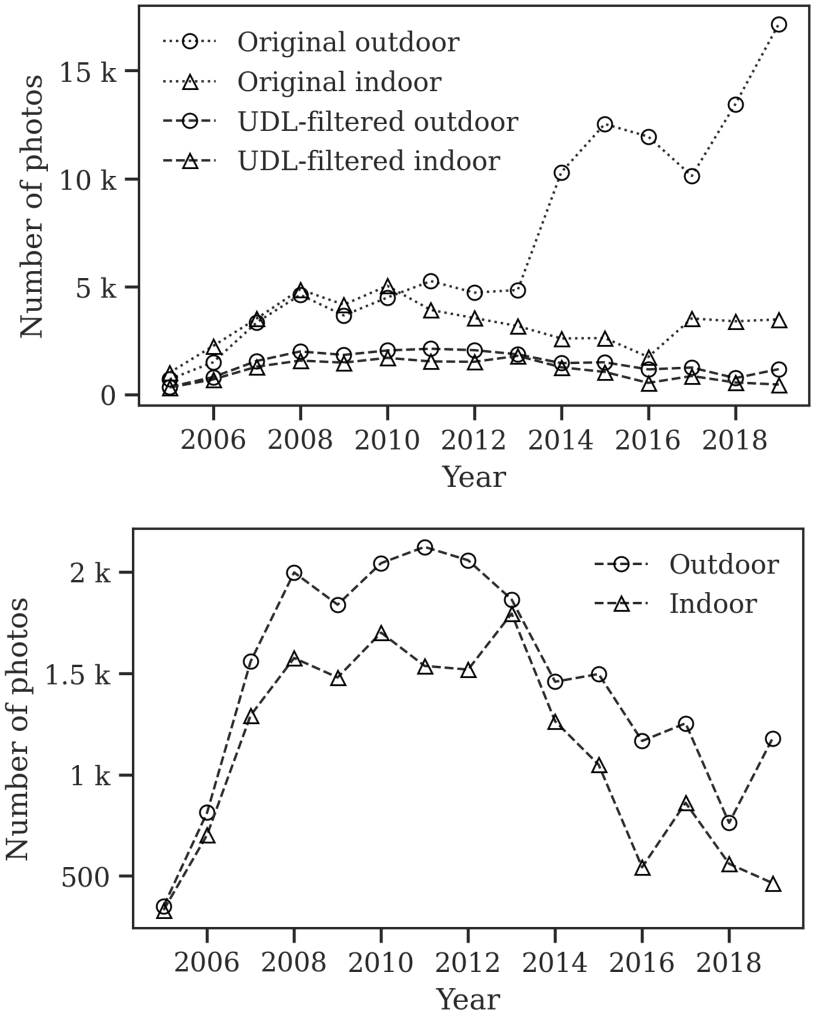


Figure A.4: Comparison of environment type recognition result between original repository of Flickr and UDL-filtered data in Ann Arbor from 2005 to 2019.

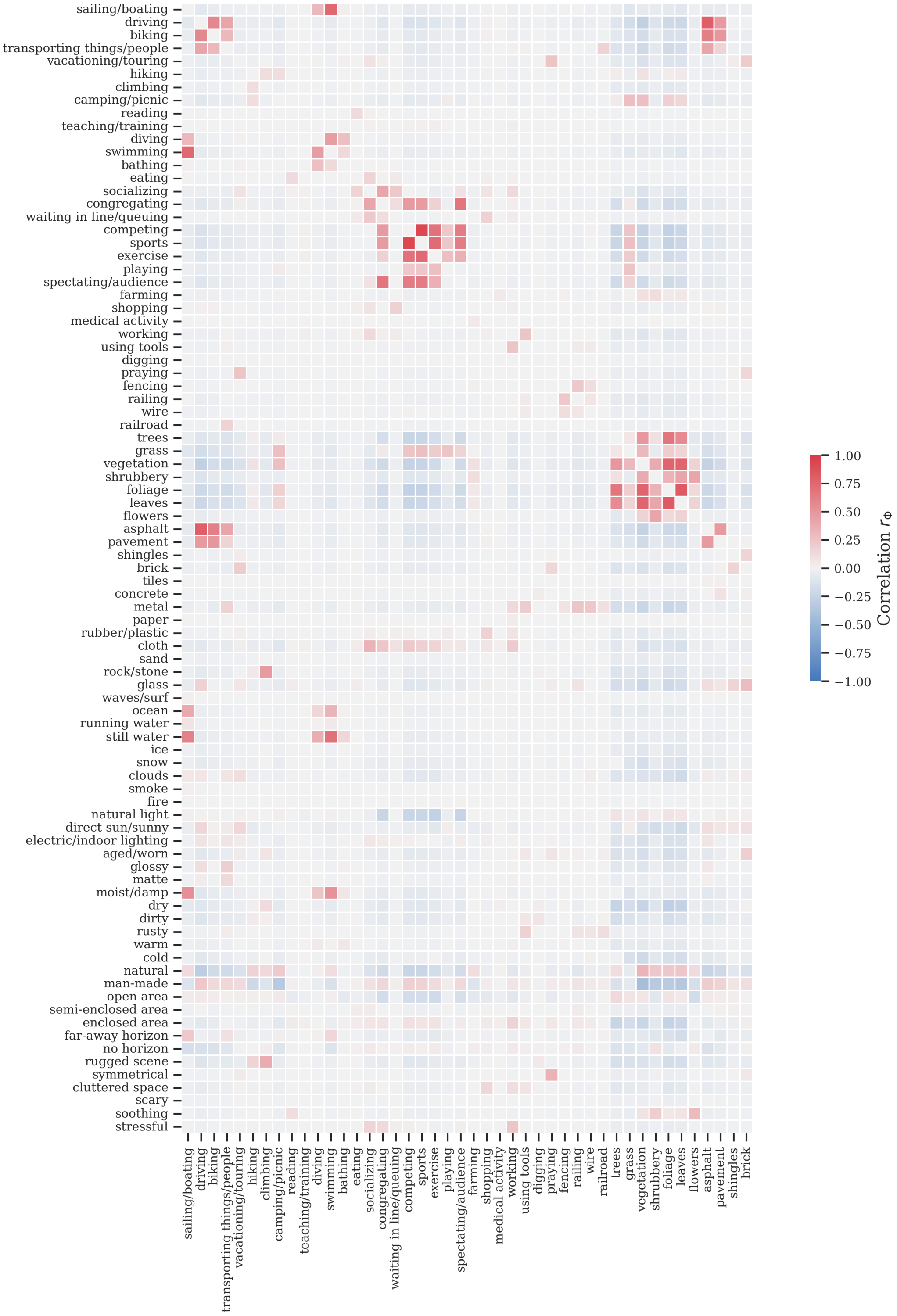


Figure A.5: Correlation matrix for the SUN attribute recognition of Flickr data in Ann Arbor. (left).

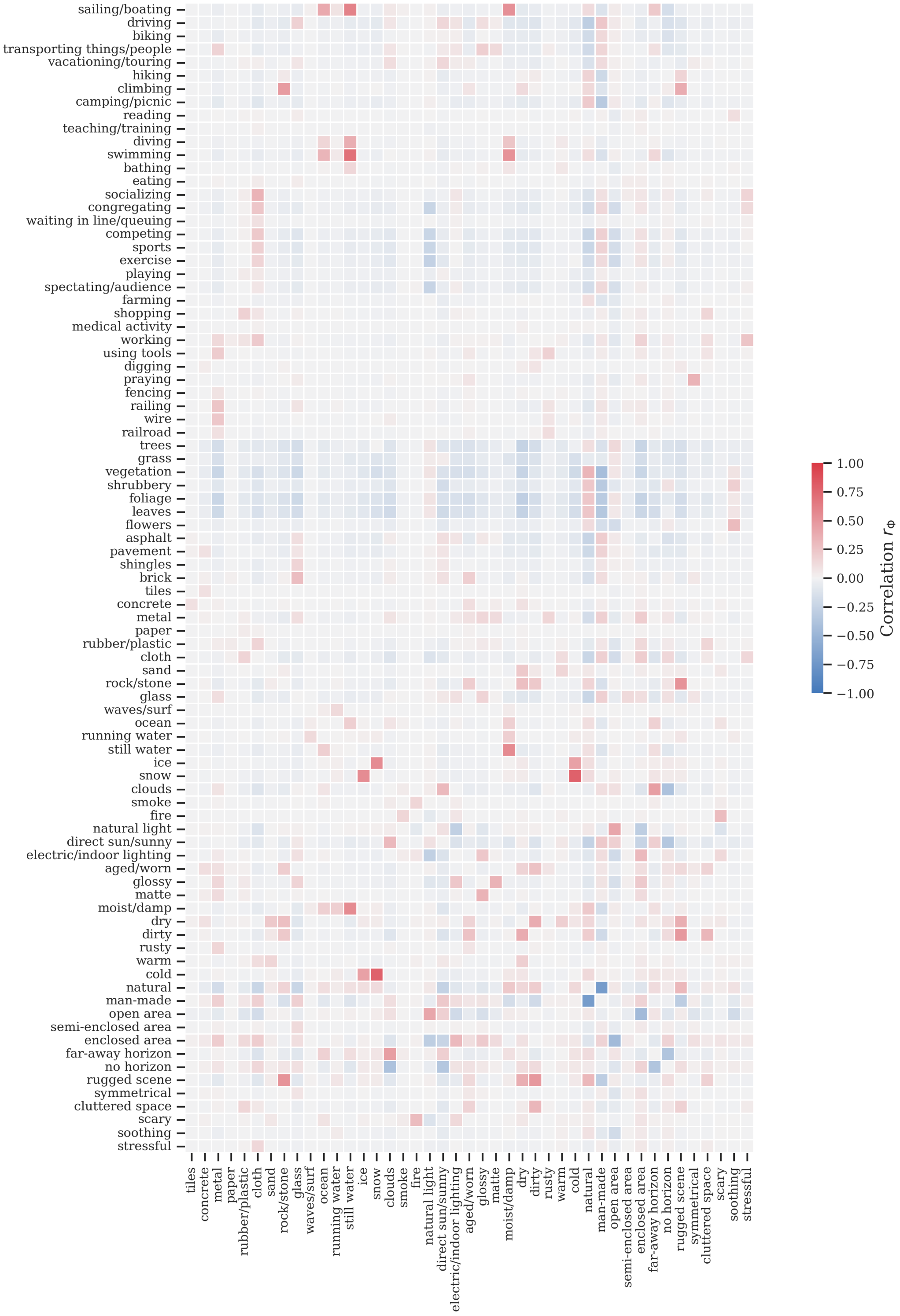


Figure A.6: Correlation matrix for the SUN attribute recognition of Flickr data in Ann Arbor. (right).