Mapping destination images and behavioral patterns from user-generated photos: A computer vision approach

Dr. Kun Zhang
College of Tourism and Service Management
Nankai University
No. 38 Tongyan Road
Jinnan District
Tianjin 300350, China
Email: zxzhangkun@163.com

Professor Ye Chen*
College of Tourism and Service Management
Nankai University
No. 38 Tongyan Road
Jinnan District
Tianjin 300350, China
Email: chenye@nankai.edu.cn
* corresponding author

Dr. Zhibin Lin
Durham University Business School
Mill Hill Lane
Durham
DH1 3LB, UK
Email: zhibin.lin@durham.ac.uk

Abstract

Destination image studies were traditionally based on questionnaire surveys, but the recent rise of user-generated content and social media big data analytics provide new opportunities for advancing tourism research. This study adopts one of the latest artificial intelligence computer vision technologies to identify the differences in the perceived destination image and behavioral patterns between residents and tourists from user-generated photos. Data were mined from Flickr, which yields 58,392 relevant geotagged photos taken in Hong Kong. The findings reveal that the perceptual differences between the two groups lay on seven types of perceptions. The differences in spatial distribution and behavioral trajectory were visualized through a series of maps. This study provides new insights into the destination image which has implications for the tourism promotion and spatial development of the destination.

Keywords: Destination image; Behavioral pattern; Social media; User-generated photo; Computer vision; Deep learning.

Acknowledgement:
This study is partially funded by National Natural Science Foundation of China (No.51608278).

1. Introduction

The rapid growth of social media in the form of collaboratively created contents such as user-generated photos present new opportunities for tourism research (Deng &Li, 2018; Nikjoo &Bakhshi, 2019). Supported by integrated information and data technologies, content mining and data analytics enable new value creation opportunities for both tourists and tourism organizations (Xiang et.al, 2015). Social media data such Flickr photos are freely available, reducing the costs of market research, development and communications (Li et.al, 2018). The diversity of various groups of people including both tourists and residents contribute contents in social media makes the insights generated from data analytics particularly conducive for tourism innovations. Social media big data offer new, real-time, and accurate ways to understand people's perceptions and behaviors (Li et.al, 2018; Xiang et.al, 2015).

Massive numbers of photos are generated every day in social media, which represent a wealth of interesting information about residents and tourists' memories, attitudes, and behaviors. Photos have become an important carrier for destination image (Lo et.al, 2011). Most previous studies employing visual content analysis were restricted to a small number of sample photos (Nikjoo &Bakhshi, 2019), due to the limits of data mining and visual content analysis technologies (Li et.al, 2018). Today's computer vision technology, a branch of artificial intelligence has developed quickly. It is now possible to perform analysis over a large volume of photos' visual content by applying a computer deep learning model.

The aim of this study is to adopt one of the latest artificial intelligence computer vision technologies to perform big data analysis on user-generated, geo-tagged photos to uncover the perceptional and behavioral differences between residents and tourists. Hong Kong was used as an empirical case study. We first created a dataset of the tourists and residents' photos through data mining the social network Flickr. We then applied a computer deep learning model on the data which recognized the visual contents of the photos and organized into 103 scenes, and subsequently reclassified into 13 categories for destination image perception analysis. After that, we compared the perceptional and behavioral preferences of tourists and with those of residents. The trajectories of tourists and residents were visualized in maps and the differences between the two groups were analyzed.

This study contributes to tourism research in three aspects: method application, research process, and theoretical framework. Firstly, the computer deep learning model in the field of computer vision was applied for visual content analysis in this study, which is one of the few emerging studies that adopt artificial intelligence technology for advancing tourism research in the era of big data (Deng &Li, 2018). The central component of a smart tourism destination is big data, which are collected, integrated, and processed from a variety of sources such as sensors, internet of things, transaction routes, as well as social media. Big data help to accurately predict resident and tourist needs and demands, which opens up new avenues for innovation and collaboration across various organizations at the destination (Buonincontri &Micera, 2016; Xiang et.al, 2017). Secondly, in comparison to the traditional questionnaire survey methods, the research process of the current study combines the statistical analysis and spatial pattern analysis together, and provides a series of visual outputs for tourism spatial planning. Thirdly, according to the output of computer scene recognition, a perceptional category with 103 attributes was established which enriches the theoretical framework of tourism destination image study (Crompton, 1979; Echtner &Ritchie, 1993; Stylidis et.al, 2017).

2. Literature review

2.1. User-generated photos in social media

User-generated photos are freely available in a variety of social media platforms such as Instagram, Facebook, Twitter, and Flickr. Embraced with metadata and enjoyed high popularity, Flickr is becoming a powerful database for tourism research (Angus et.al, 2010; Donaire et.al, 2014; Kennedy et.al, 2007; Miah et.al, 2017). User-generated photos and travel are intrinsically linked, as the photos record the user's experience and memory of the destination (Larsen, 2008). A photo can capture the information of the destination more efficiently than other forms of messaging, such as text, sound, and others (Kim et.al, 2014). As such user-generated photos provide a valuable data source for tourism research (Li et.al, 2018).

In recent years, with the breakthrough of data mining technology, using online photos for the interpretation of the tourism phenomenon has received increasing research attention. For example, Mariani et al (2016) have explored the use of photos on Facebook to promote a destination. Other scholars also advocate the use of photos as a powerful tool for a destination's marketing and promotion (Balomenou &Garrod,

2019; Molinillo et.al, 2018; Pan et.al, 2014).

A user-generated photo in social media usually contains two major types of information: a) the geographic and temporal information attached to the photo, and b) the visual content of the photo itself. The rich information contained in user-generated photos provides valuable sources for generating tourism insights.

2.1.1. Geographic and temporal information attached to photos

With the advancement of the Global Positioning System (GPS) technology, geographical information can be stored in a photo, i.e. the so-called geotagged photos. In addition, temporal information is also attached to a photo. The geographic and temporal information make user-generated photos a powerful source for the study of tourism. Many studies have explored the movement patterns of tourists since the Geographic Information System (GIS) was firstly introduced (Lau et.al, 2006). For example, Hu et al (2015) present a framework for extracting and identifying areas of interest (AOI) based on geo-tagged photos (Hu et.al, 2015). Chua et al (2016) describe an approach to analyzing geotagged social media data from Twitter to understand tourist flows (Chua et.al, 2016). Vu, Li, Law, and Ye (2015) used 29,443 geo-tagged photos collected from Flickr to visualize tourists' movement trajectories (Vu et.al, 2015). Li & Wang (2018) use geotagged photos collected from Flickr to analyze the spatial distributions of tourists and residents (Li &Zhou et.al, 2018). Moreover, studies of photos have crossed the fields of tourism, geography, and computer science (East et.al, 2017; Oku et.al, 2015; Zhang et.al, 2020; Zheng &Liao, 2019; Zheng et.al, 2017).

2.1.2. The visual content of photos

The central information of a photo is its visual content. Generally, there are three crucial steps involved in visual content analytics: a) visual data collection and screening, b) data decision, and c) visual content analysis. Most of the extant studies of user-generated photos apply big data mining technology in the first step. Researchers then need to decide whether to analyze all the photos collected or just a sample of them. For the former, a computer deep learning model needs to be used, as for the latter, the small sample of photos could always be analyzed in a manual way. As computing technology is required for visual content analysis, prior studies are largely based on the manual processing of a small sample of the data. For example, Stepchenkova and Zhan (2013) collected thousands of photos from the official website and Flickr respectively by using big data mining technology, and the identified 20 attributes for visual content analysis by manual recognition. The research challenge now is to replace manual analysis with the latest computer vision technology to analyze the entire collected big data of user-generated photos.

2.2. Computer vision technology for visual content analysis

Artificial intelligence, machine learning and deep learning are concepts that emerged in the 1950s, 1980s and 2010s in sequence. Their scope is gradually narrowing down. Deep learning is a subset of machine learning methods. The breakthrough of deep learning promotes the prosperity of artificial intelligence. Computer vision is the field of dominantly applying deep learning, which is a science about how to make machines "see" things, which includes tasks such as acquiring, processing, analyzing and interpreting the visual contents. Computer vision is built on artificial intelligence systems that can extract "information" from images.

Face recognition is one of the most well-known applications of computer vision. In addition to face recognition, computer vision mainly can be applied to conduct image classification/scene recognition, target detection, image semantic segmentation, image generation, video classification and so on. Among these study directions, image classification/ scene recognition is to distinguish different kinds of images according to their semantic information, which is a crucial foundation in computer vision and provides support for high-level visual tasks, such as object detection, image segmentation, object tracking, behavior analysis, face recognition and so on.

Pioneer research using scene recognition technology has been carried out in the urban study. Liu et al. (2016) categorized photos of 26 cities in the world (an average of 100,000 photos per city) into 102 scenes and 7 perceptions by scene recognition model in computer vision, and compared different city images by analyzing the characteristics of photos' visual content in each city (Liu et.al, 2016). Zhang et.al (2019) have conducted a study about tourists' perceptions and behaviors by applying a scene recognition model for the visual content analysis of tourists' photos.

In addition to scene recognition model, sematic segmentation model can be used to semantically segment visual content into several components (normal 8-12 components) for application in various areas, such as urban greening and perceived naturalness (Hyam, 2017; Li &Ratti et.al, 2018), neighborhood walkability and Bikeability (Evans-Cowley &Akar, 2014; Yin et.al, 2015).

2.3. Perceptual and behavioral differences between tourists and residents

Tourists and residents interact and share spaces and services in the destination (Zhang et.al, 2006), and tourist attractions are increasingly woven into residents' everyday living spaces (Li &Zhou et.al, 2018). Tourists are not only interested in the destination itself, but also its local people. Learning about local people's daily life is seen by tourists as an important factor in the quality of their traveling experience (Choe &Kim, 2018; Lee et.al, 2018; Notzke, 1999; Sengel et.al, 2015). The image of the destination as perceived by its residents has a significant effect on incoming tourist arrival (Nunkoo &Gursoy, 2012; Tsaur et.al, 2006). As such, understanding the perceptions and experiences of both tourists and residents is important for sustainable city planning and tourism development(Deng et.al, 2019).

The majority of the earlier destination image studies focused on the tourists' perspective, however, recently the importance of residents' perceptions for tourism development has been increasingly recognized (Ku &Mak, 2017; Stylidis et.al, 2017). According to stakeholder theory (Freeman, 1984), any group or people who influence or are influenced by tourism development should be considered in tourism development. The two major groups of tourism stakeholders are tourists and residents. Local people, their culture, attitude, friendliness and hospitality are key attributes of a destination image (Gallarza et.al, 2002), and their support is vital for sustainable tourism development (Young et.al, 2007). Moreover, compared with tourists, residents have an insider understanding of the destination, and their perception and behaviors could provide influential information for tourist's activities in the destination (Arsal et.al, 2010; Monterrubio, 2019). Researchers have advocated studying residents' perceptions and attitudes into tourism development planning (Lin et.al, 2017; Ruiz et.al, 2019; Sharpley, 2014; Stylidis et.al, 2017).

Previous studies were mostly based on questionnaire surveys. Zhang & Chan (2016) compared residents and tourists' perceptions towards natural-based tourism in Hong Kong, and as can be expected, they found that tourists are interested in travel products, logistics, and quality, while residents are concerned about nature conservation, environmental education, facilities, and government policy (Zhang &Chan, 2016). Ku and Mak (2017) analyzed the differences of destination image perception along 30 attributes between residents and tourists visiting scenic areas in Hualien, Taiwan. They found that residents considered recreational activities, relaxation, and natural environment were important attributes, while tourists regarded facilities and space planning were important (Ku &Mak, 2017). Valek and Williams (2018) examined the difference of their perceptions towards the same destination in Abu Dhabi, and reported that tourists rated "sea, sun and sand" as the highest among the destination image attributes, while locals rated "friendly people" as the highest destination attribute (Valek &Williams, 2018).

For the research of the behavior difference between residents and tourists, photo attached geographic and temporal information was served as accurate information for orientating tourists' behavioral activities and mapping their traveling trajectory. Online photographs have become an important data resource comparing with the conventional data collection methods such as questionnaires and interviews.

2.4. The attributes/categories of a destination image

Destination image refers to the sum of knowledge, belief and emotions when one perceives of a destination (Crompton, 1979). It is the total impression of a destination in the minds of tourists and residents (Echtner &Ritchie, 1993). The overall image of a destination consists of both the cognitive and affective components (Stylidis et.al, 2017). The cognitive component consists of those attributes that are related to the resources of a tourist destination, such as historical and cultural attractions, the scenery, climate, transport infrastructure, restaurants, and accommodation (Lin et.al, 2017), while the affective component is usually measured along with semantic scales such as unpleasant-pleasant, distressing-relaxing, gloomy-exciting, and sleepy-arousing (Chew &Jahari, 2014). In the information age, tourists' perceived destination image would be affected by more different information resources like the Internet, virtual environments, virtual experiences, artificial technology, etc. The destination image tends to be complex, relativistic and dynamic (Gallarza et al., 2002), and it is hard to measure with common multiattribute quantitative measurement(Govers &Go, 2003). Therefore, this research adopts a deep learning-based multiattribute method to improve the research method.

As shown in Table 1, previous studies have various ways of selecting and organizing the attributes, ranging from 7 to 20 categories (Stepchenkova &Zhan, 2013). The major cognitive attributes include nature and landscape, architecture/buildings, people, traditional clothing, art object, tourism facilities, urban landscape, domesticated animals, plants, leisure activities, food, transport/infrastructure (Stepchenkova &Zhan, 2013). In addition to cognition, there are also affective ones described using adjectives such as pleasant, relaxing, exciting, unpleasant, gloomy, sleepy, or distressing (Mak, 2017).

[Table 1 about here]

To accurately analyze a destination, the selection of attributes in the analytical framework is essential to under the relationship between the attributes and the overall entity, i.e. the overall destination image. As mentioned earlier, most of the existing attribute research on destination image had used a questionnaire to collect data, while it could generate missing or inaccurate information given by the hysteresis of tourists' memories. Some other studies used texts and reviews from the Internet or social media (Mak, 2017; Nikjoo &Bakhshi, 2019), while a few scholars used artificial image analysis (Stepchenkova &Zhan, 2013).

3. Methodology

3.1. Study framework

This study aims to compare the differences between tourists' and residents' perceptions and behaviors in the tourism destination. In the perceptual aspect, we adopted the attribute-holistic framework proposed by Echtner and Ritchie (1993) as a theoretical basis. Referring to our newly created attribute categories, we took tourists' and residents' photos as evidence for the perceptions and compared the photos' visual presence. For the behavioral element, we focus on the spatial distribution- area of interest (AOI), itinerary, and pattern, which were visualized in the geographic information system. Figure 1 shows a framework of the methodological procedure of the study, which consists of three major steps. The first step is data acquisition and screening (Section 3.2.); the second step is visual content analysis – scene recognition and reclassification (Section 3.3), which includes a deep learning model of scene recognition and our new induced category for perceptual attributes; and Step 3 is spatial analysis (Section 3.4), which consists of the distribution of tourists and residents' footprints, behavioral trajectories, and spatial behavioral patterns.

[Figure 1 about here]

3.2. Data Collection-Flickr YFCC 100M

The data used for this research are from the Yahoo Flickr Creative Commons 100 Million Dataset (YFCC100M), which is the largest public multimedia collection ever released, with approximately 99.2 million photos uploaded to Flickr between 2004 and 2014. Each metadata in the dataset usually contains the following information: its Flickr identifier, the user who created it, the camera that took it, the time it was taken and uploaded, the location where it was taken (if available), and the textural information attached by photos (title, description, and tags) (Thomee et.al, 2016). In this study, four types of information were employed: a) the photo/video and user identifier (pID/uID), used for identifying and counting the photo or user; b) the date and that the photo was taken, used for tracking tourists' trajectory; c) the longitude and latitude, used for matching geographical coordinates; and d) the photo/video download URL, used for download the original photos.

In total, YFCC 100M contains 48,366,323 photos with geographic information (Deng &Li, 2018). To support academic research, 2,171,162 geo-tagged photos located in China were made available to scholars by Beijing City Lab (BCL). With the help of software ArcGIS, all the points (one single point stand for one photo) appeared within the boundary of Hong Kong were select from the BCL released data. In all, 85903 photos located in Hong Kong were collected, which is about 4% of the total number of geotagged photos in China. In order to distinguish tourists and residents, the user's birthplace, city and country were retrieved by invoking API (Application Programming Interface) data in Flickr referring to the information of uID. Consequently, 1753 users' home location information was identified, and 58392 photos uploaded by them were selected for this research.

3.3. Visual content analysis: Scene recognition model and the induced category

Convolutional Neural Networks (CNN) is a kind of feedforward neural network with deep structure and convolution calculation, which is one of the representative algorithms of deep learning and widely been used in the field of computer vision. The well-known CNN structures include ResNet, WRN, DPN, SENet, etc. In this study, ResNet is employed for the visual content analysis and task-scene recognition which is found to greatly improve gradient flow, thus allowing training of much deeper models with tens or even hundreds of layers (He et.al, 2015). A demonstrated structure is shown in Figure 2. To train the dataset, we used over ten thousand photos with 103 types of scenes' labels and built up the model of scene recognition, with each photo has a single label about the scene. Technologically, 103 is nearly the maximum number of scenes that deep learning models can recognize in the field of computer vision. The scenes covered almost all of the scenes captured in a camera. Following this train set, an efficient and robust deep learning structure for scene recognition was established, which had a recall rate that achieved 90%, and the false recognition rate was 0.1%. A total of 58,392 photos were input into the learning model, and 103 scenes are recognized as the output.

[Figure 2 about here]

The 103 scenes show much information about tourists and residents' perceptions and behaviors. To summarize and extract the primary characteristics from the information, we need to theoretically induce the analytical dimensions and organize the fragmented information into a simplified frame. Referring to tourists and residents' activities, the 103 scenes are reclassified into 13 categories (Table 2), which covered most of the attributes of destination image (Lin, Vlachos, & Ollier, 2018). The frequency of photos in each scene and category were calculated and the difference between the residents and tourists would be compared by the statistical distribution of photos' frequency.

[Table 2 about here]

3.4. Spatial analysis: ArcGIS as the tool

ArcGIS is adopted as the instrument for spatial analysis in the study, specifically, the tools of kernel density analysis and tracking analyst were employed. The kernel density analysis in ArcGIS could calculate the density of point elements around each output raster pixel, which is an ideal option for visualizing the aggregation characteristics of tourists' behavior and perceptions. By the orientation of geographic coordinates, the distribution of two groups' footprint can be mapped. By saving the scene information of tourists and residents' photos into ArcGIS, the spatial distribution of two groups' perceptions in the destination could be visualized. The tool of tracking analyst could draw the moving route of tourists and residents by tracking the spatial and temporal information attached by the uploaded photos.

3.5. The tourist destination for the case study: Hong Kong

Hong Kong, which is located on the eastern side of the Pearl River estuary in southern China, is a highly prosperous international metropolis and world-renowned international destination. Hong Kong has the world's largest number of skyscrapers. Traditional Chinese culture and Western culture are fused in this city. Beautiful and modern urban landscape and unique historical and cultural identity have made Hong Kong an attractive place for tourists. Hong Kong has become an important inbound tourism market, attracting international tourists from all over the world. According to the Travel and Tourism Competitiveness Report 2017 released by the World Economic Forum, Hong Kong ranks the 11th among 136 major tourism destinations worldwide, the second in Asia after Japan. In addition, according to World Tourism Organization (UNWTO) statistics, Hong Kong international inbound tourists has reached 26.69 million in 2017, the contribution of the tourism industry to GDP in Hong Kong reached 25020.7million, accounting for nearly 8% of total GDP. The tourism industry is an important part of Hong Kong's economy. According to data released by the Hong Kong Tourism Board, about 65.15 million visitors visited Hong Kong in 2018, of which about 14 million were visitors from overseas.

4. Results

4.1. Descriptive analysis

People with different geographic, social and cultural backgrounds tend to have different travel preferences (Vu et.al, 2015), and naturally, tourists show different perceptions and behaviors form residents in a destination. In this study, users who have their home location recorded in Flickr outside Hong Kong are treated as tourists and those who show their home location in Flickr as Hong Kong are treated as residents. Consequently, 1430 tourists and 323 Hong Kong residents were identified. In total 58,392 photos were collected, of which 29,311 were uploaded by residents, and 29,081 were uploaded by tourists (see Table 3). The average number of photos uploaded by the residents is far greater than the inbound tourists. Figures 3 and 4 present the distribution of tourist' numbers and the number of their uploaded photos by country or region. The top 5 countries (regions) in terms of tourist numbers are US, UK, Taiwan, Australia, and Canada, and the top 5 countries (regions) in terms of photo numbers are US, UK, Taiwan, Germany and Canada.

[Table 3 about here]

[Figure 3 about here]

[Figure 4 about here]

4.2. Perceptional difference between tourists and residents

The specific differences between tourists and residents were obtained by statistically comparing the number of photos of different scenes in the same category (Figure 5). The general perceptional analysis revealed several apparent differences, which included architecture, plant, water, mountain, food, culture, and shopping.

i. Architecture: Tourists showed greater preferences than residents for each type of scene that belongs to architecture. Specifically, their choice for the old buildings, overlooks, skyscrapers, and the corridors is more evident than for other scenes.

ii. Plant: Residents observed more green plants and flowers than tourists.

iii. Water: Residents tended to visit more of the beaches than tourists.

iv. Mountain: Residents were more interested in the mountains than tourists.

v. Food: Tourists preferred to show the food in the destination than residents, and they also pay more attention to the restaurants.

vi. Culture: Residents attended more cultural events, e.g., fireworks and shows in a stage.

vii. Shopping. Tourists had a greater preference for mall courtyards than residents.

[Figure 5 about here]

4.3. Mapping the difference between tourists and residents

4.3.1. Spatial distribution of tourists and residents' footprints

The spatial result of kernel density analysis shows that the footprint of residents is much larger than that of tourists. As can be expected, tourists' footprint is mainly distributed around the major well-known tourist attractions, airports and shopping areas in the city center. Areas of interest (AOI) for tourists include Tian Tan Buddha Statue, Disneyland, The Peak Tower, Hong Kong Central, Tsim Sha Tsui Area and Mong Kok. The densest areas ranked at the top two are Tsim Sha Tsui and Central, which are located on both sides of Victoria Harbor. The densest areas of residents' footprints are Causeway Bay, Tsim Sha Tsui and Central. Hung Hom is an area of interest only for residents. As an extension of the urban central area aside Victoria Harbor, hotspots of residents' footprint were also found at Tsuen Wan and Sha Tin in New territories.

[Figure 6 & 7 about here]

4.3.2. Spatial distribution of tourists and residents' trajectory

The information on tourists' movement in tourism destinations is of great significance for the traffic management and the arrangement of various infrastructures, especially for metropolitan areas with a high density of traffic (Vu et.al, 2015). In order to explore the differences in the geographical movement of tourists and residents, we used the tracking analysis tool in ArcGIS to depict their movement trajectories. As there are too many lines overlapping each other, kernel density analysis for lines was conducted twice. The output mapping shown in Figure 8 aids our observation of the spatial structure of tourists and residents' moving patterns. The result shows that the trajectory is highly oriented by the hotspot where tourists or residents gather. For tourists, there are three moving routes: one is between Hong Kong International Airport and Hong Kong Central, one is from Tian Tan Buddha Statue to Hong Kong Central, and the last is the most frequent route between Tsim Sha Tsui and Central which cross the Victoria Harbor. For residents, five main moving routes could be recognized, the first is between Central to Tung Lo Wan alongside the water, the next two are among airport, Central and Tsim Sha Tsui which are the same as tourists. The other two can be interpreted as the routes for commuting between central, Tsuen Wan and Sha Tin.

[Figure 8 about here]

4.3.3. Distribution of tourists and resident's behaviors in destination

By logging the scene information to the attributes of shapefile in ArcGIS, we can observe the activities of tourists and residents, as displayed in Figure 9. The spatial features of these two groups' behavioral activities are described as followed.

[Figure 9 about here]

Tourists' cultural activities are highly concentrated in Disneyland, which includes watching stage shows, firework and dragon-lion dance. Residents' cultural activities are highly concentrated in Hung Hom, as there is a famous stadium for concerts. Furthermore, the traditional firework normally happens around the area of Victoria Harbor, both tourists and residents' participation in cultural activities are found there. The popular shopping places for tourists in rank order are Central, Tsim Sha Tsui, Tsuen Wan, Yau Ma Tei and the airport. In contrast, the favorite place of residents for shopping is Tsuen Wan, then followed by Central, Tsim Sha Tsui and many places in New Territories. Most of the sightseeing behaviors for the architecture of tourists are found in Hong Kong Central and The Peak Tower, the former area has the greatest number of high-rise buildings and the latter is the best space for overlooking the whole Victoria Harbor. While residents' recorded places for architecture sightseeing have a wider range than tourists, which extended from central to many other areas in Hong Kong. The behavior of observation for insects and animals by tourists are found in urban open space, such as Hong Kong City Park and the Southern Ocean Park. In contrast, residents' behavior of observation for insects and animals are dispersed in the rural areas, for example, Ma On Shan Country Park and Pat Sin Leng Country Park.

Both of tourists and residents’ transiting behaviors are centralized in Airports and the busiest area around Victoria Harbor, residents have a wider distribution of transiting behaviors as their living and working space are scattered around Hong Kong. Both groups' eating behaviors are oriented in four places- Central, Tsim Sha Tsui, Mang Kok and Causeway Bay. The differences between these two groups are that most tourists eat at Central and then Tsim Sha Tsui, while residents prefer to eat at Causeway Bay and then Central.

The activities of visiting mountain by tourists and residents happened not all in the same places. Lantau Island is the recreational mountainous area for both two groups. The difference is that tourists prefer to visit Victoria Peak in the city and residents prefer to visit the mountainous area, for example, Ma On Shan. The water area visited by residents could be found almost place near to the sea or lakes. However, the water area visited by tourists were restricted at Victoria Harbor. There is much textual information that has been read by both tourists and residents in Central and Tsim Sha Tsui, as shopping brands are full of the street in these two areas. In addition, many photos about tourism interpretation were shot by residents in Ma on Shan Country Park which indicates that it is an important place for residents' climbing activities.

The decision of taking part in leisure and entertainment activities for tourists and residents is totally different: Disneyland is the hotspot of tourists, while residents more prefer to relax in the area of Tsuen Wan. Most of the plant observation of tourists and residents happen in green space in Hong Kong. Tourists prefer the green space around Tai Mo Shan and residents prefer the open space in rural areas. When a major part of the photo is the sky, it would be assorted into the natural scene. According to the distributions of the natural phenomenon scene, Victoria Harbor is the place for both tourists and residents to gaze at the sky, while the content of the sky also appeared in the residents' photos taken in many other places in Hong Kong.

5. Discussion

The visual content of massive photos uploaded by tourists and residents has the potential to provide a comprehensive understanding of their perceived destination image and the attached geographical and temporal information in photos can effectively portray the spatial pattern and moving trajectory. Unlike most of the previous studies adopt the manual technique, this study adopts the emerging machine learning technology to uncover the insights from user-generated content, which provide great opportunities for value creation in a new, “smart” way (Brandt et.al, 2017; Buhalis &Amaranggana, 2015). The conclusions show that the destination image of 13 scenes attributes forms a more specific and comprehensive image picture. Supported by deep learning scene recognition model in computer vision technology, this study performs both perceptional and spatial behavior pattern analysis by adopting a novel and advanced computer vision deep learning approach to analyzing user-generated geo-tagged photos of Hong Kong in Flickr. The findings of this study provide clear insights into the differences of behavioral footprints between tourists and residents, which is not possible in the previous studies that rely on questionnaire surveys (Ku &Mak, 2017; Slak Valek &Williams, 2018).

5.1. Tourists' perception of the destination image.

Natural phenomenon scenes about whether, sky and diurnal variation is the fundamental content in photos, the "night" has been perceived frequently by tourists which indicates that sightseeing and doing other activities at night play an important role in tourists' traveling in Hong Kong. Architecture and food are the primary things that tourists perceived in Hong Kong, which highly coincides with Hong Kong's city characteristics. Hong Kong is one of the most famous cities featured by skyscrapers in the world and provides diverse food. Many photos containing the content of the plant reflect that Hong Kong is a green and ecological city, the plentiful perception of transportation represents Hong Kong has an efficient transportation system. While the rank of tourists' perception of cultural activities falls behind plant and transportation, it indicates that Hong Kong has the cultural potential in the prospects of tourism development.

5.2. Perceptional comparison between tourists and residents

The prominent difference is that residents pay more attention to the natural environment than tourists, such as plants, animal and insects. Secondly, it should be noted that the perceptional priority of residents for cultural activities is higher than tourists. In addition, local people have a stronger perception of natural landscapes than tourists, for example, water and hill. In the future, tourism development should consider the mountain and water resources in Hong Kong for recreational planning.

5.3. Behavioral comparison between tourists and residents

It is evident that the area around Victoria Harbor is the most active area for both the tourists and residents. Most of the perceptions and behaviors about food, architecture, shopping, tourism interpretation, and natural scenes are found there. Specific hot spots for both two groups include Tsim Sha Tsui, Central, Causeway Bay and Mong Kok, the difference is the local people always give a high priority to Causeway Bay. In terms of cultural and recreational activities, tourists mainly experience cultural and recreational activities in Disneyland Park. While locals' cultural activities mainly happen around Victoria Harbor, and most of their recreational activities take place in Tsuen Wan, which is a residential area. In addition, the rural parks are more important places for the local people, while tourists visit less, all this evidence could provide a basis for the spatial layout and development of the leisure industry in Hong Kong.

5.4. Overall distribution and behavioral trajectory

The active area of residents is much larger than that of tourists, especially tourist visit less in the northern part of the New Territories. The footprint of tourists is mainly distributed around Airport, Tian Tan Buddha Statue, Disneyland, The Peak Tower, Hong Kong Central, Tsim Sha Tsui Area and Mong Kok. As for the moving trajectory, the Victoria Harbor area is the most intensive area for both tourists and residents.

While, the lines from the central area to Tsuen Wan and Sha Tin are local commuting routes, and the route between central to Lantau Island is specially served for tourists. This result of the moving trajectory of residents and tourists in Hong Kong is of great significance for coordinating the residents' daily traffic with the tourism traffic.

6. Conclusion and future research

In this study, we adopted a computer deep learning model to analyze user-generated-photos of a tourist destination to generate new insights into the differences of destination image perception and travel trajectories between tourists and residents. The findings have important implications for destination marketing and management. First, the perceptual characteristics of tourists and residents identified in this study serve as a reference for identifying and promoting the perceived tourism destination image. Second, destination managers could consider using the revealed overlapping recreational spaces between tourists and residents for the organization of a series of activities for social communication and value co-creation between tourists and residents, thus improve the host-guest interaction and enhance tourist attachment to the destination. Finally, the differences in the spatial trajectory between tourists and residents revealed in this study give the destination managers a reference for optimizing public infrastructure and services that tailor to the specific needs of the tourists and residents.

Future research will provide promising results that overcome the limitations in the current study. The findings of this study are limited to the use of one social media platform Flickr only. Future research could use data mining technology to collect data across various platforms. Researchers could generate greater insights by adding a new dimension of time in their analysis to understand how a destination's image changes over time. With the advancement of new algorithms and computer power, visual contents could be processed more accurately and efficiently. The potential of applying computer vision for analyzing destination image is enormous.

References:

Angus, E., Stuart, D., Thelwall, M. (2010). Flickr' s potential as an academic image resource: An exploratory study. *Journal of Librarianship and Information Science, 42*(4), 268-278.

Arsal, I., Woosnam, K. M., Baldwin, E. D., Backman, S. J. (2010). Residents as Travel Destination Information Providers: An Online Community Perspective. *Journal of Travel Research, 49*(4), 400-413.

Balomenou, N., Garrod, B. (2019). Photographs in tourism research: Prejudice, power, performance and participant-generated images. *Tourism Management, 70*, 201-217.

Brandt, T., Bendler, J., Neumann, D. (2017). Social media analytics and value creation in urban smart tourism ecosystems. *Information & Management, 54*(6), S1008926563.

Buhalis, D., Amaranggana, A. (2015). *Smart Tourism Destinations Enhancing Tourism Experience Through Personalisation of Services.* Paper presented at the Enter.

Buonincontri, P., Micera, R. (2016). The experience co-creation in smart tourism destinations: a multiple case analysis of European destinations. *Information Technology & Tourism*, *16*(3), 1-31.

Chew, E. Y. T., Jahari, S. A. (2014). Destination image as a mediator between perceived risks and revisit intention: A case of post-disaster Japan. *Tourism Management, 40*, 382-393.

Choe, J. Y. J., Kim, S. S. (2018). Effects of tourists' local food consumption value on attitude, food destination image, and behavioral intention. *International Journal of Hospitality Management, 71*, 1-10.

Chua, A., Servillo, L., Marcheggiani, E., Moere, A. V. (2016). Mapping Cilento: Using geotagged social media data to characterize tourist flows in southern Italy. *Tourism Management, 57*, 295-310.

Crompton, J. L. (1979). Motivations for pleasure vacation. *Annals of Tourism Research, 6*(4), 408-424.

Deng, N., Li, X. R. (2018). Feeling a destination through the "right" photos: A machine learning model for DMOs' photo selection. *Tourism Management, 65*, 267-278.

Deng, N., Liu, J., Dai, Y., Li, H. (2019). Different cultures, different photos: A comparison of Shanghai's pictorial destination image between East and West. *Tourism Management Perspectives, 30*, 182-192.

Donaire, J. A., Camprubí, R., Galí, N. (2014). Tourist clusters from Flickr travel photography. *Tourism Management Perspectives, 11*, 26-33.

East, D., Osborne, P., Kemp, S., Woodfine, T. (2017). Combining GPS & survey data improves understanding of visitor behaviour. *Tourism Management*, *61*, 307-320.

Echtner, C. M., Ritchie, J. R. B. (1993). The Measurement of Destination Image: An Empirical Assessment. *Journal of Travel Research, 31*(4), 3-13.

Evans-Cowley, J. S., Akar, G. (2014). StreetSeen Visual Survey Tool for Determining Factors That Make a Street Attractive for Bicycling. *Transportation Research Record Journal of the Transportation Research Board, 2468*(-1), 19-27.

Freeman, R. E. (1984). *Strategic Management: A Stakeholder Approach*. Boston: Pitman Publishing.

Gallarza, M. G., Saura, I. G., Garcı a, H. C. (2002). Destination image: Towards a Conceptual Framework. *Annals of Tourism Research, 29*(1), 56-78.

Govers, R., Go, F. M. (2003). Deconstructing destination image in the information age. *Information Technology & Tourism, 6*(1), 13-29.

He, K., Sun, J., Ren, S., Zhang, X. (2015). Deep Residual Learning for Image Recognition., 770-778.

Hernández-Mogollón, J. M., Duarte, P. A., Folgado-Fernández, J. A. (2018). The contribution of cultural events to the formation of the cognitive and affective images of a tourist destination. *Journal of Destination Marketing & Management, 8,* 170-178.

Hu, Y., Gao, S., Janowicz, K., Yu, B., Li, W., Prasad, S. (2015). Extracting and understanding urban areas of interest using geotagged photos. *Computers Environment & Urban Systems, 54*, 240-254.

Hyam, R. (2017). Automated Image Sampling and Classification Can Be Used to Explore Perceived Naturalness of Urban Spaces. *Plos One, 12*(1), e169357.

Kennedy, L., Naaman, M., Ahern, S., Nair, R., Rattenbury, T. (2007). *How flickr helps us make sense of the world: context and content in community-contributed media collections.* Paper presented at the ACM International Conference on Multimedia.

Kim, S. B., Kim, D. Y., Wise, K. (2014). *The effect of searching and surfing on recognition of destination images on Facebook pages*: Elsevier Science Publishers B. V.

Ku, G. C. M., Mak, A. H. N. (2017). Exploring the discrepancies in perceived destination images from residents' and tourists' perspectives: a revised importance – performance analysis approach. *Asia Pacific Journal of Tourism Research, 22*(11), 1124-1138.

Larsen, J. (2008). Practices and Flows of Digital Photography: An Ethnographic Framework. *Mobilities, 3*(1), 141-160.

Lau, G., Mckercher, B., Tribe, J. (2006). Understanding tourist movement patterns in a destination: a GIS approach. *Tourism & Hospitality Research, 7*(1), 39-49.

Lee, T. H., Chao, W. H., Lin, H. (2018). Cultural inheritance of Hakka cuisine: A perspective from tourists' experiences. *Journal of Destination Marketing & Management, 7*, 101-111.

Li, D., Zhou, X., Wang, M. (2018). Analyzing and visualizing the spatial interactions between tourists and locals: A Flickr study in ten US cities. *Cities, 74,* 249-258.

Li, J., Xu, L., Tang, L., Wang, S., Li, L. (2018). Big data in tourism research: A literature review. *Tourism Management, 68,* 301-323.

Li, X., Ratti, C., Seiferling, I. (2018). Quantifying the shade provision of street trees in urban landscape: A case study in Boston, USA, using Google Street View. *Landscape & Planning, 169*.

Lin, Z., Chen, Y., Filieri, R. (2017). Resident-tourist value co-creation: The role of residents' perceived tourism impacts and life satisfaction. *Tourism Management, 61*, 436-442.

Liu, L., Zhou, B., Zhao, J., Ryan, B. D. (2016). C-IMAGE: city cognitive mapping through geo-tagged photos. *Geojournal, 81*(6), 817-861.

Lo, I. S., McKercher, B., Lo, A., Cheung, C., Law, R. (2011). Tourism and online photography. *Tourism Management, 32*(4), 725-731.

Mak, A. H. N. (2017). Online destination image: Comparing national tourism organisation's and tourists' perspectives. *Tourism Management, 60,* 280-297.

Miah, S. J., Vu, H. Q., Gammack, J., McGrath, M. (2017). A Big Data Analytics Method for Tourist Behaviour Analysis. *Information & Management, 54*(6), 771-785.

Molinillo, S., Liébana-Cabanillas, F., Anaya-Sánchez, R., Buhalis, D. (2018). DMO online platforms: Image and intention to visit. *Tourism Management, 65*, 116-130.

Monterrubio, C. (2019). Hosts and guests' social representations of nudism: A mutual gaze approach. *Annals of Tourism Research, 75,* 18-28.

Nikjoo, A., Bakhshi, H. (2019). The presence of tourists and residents in shared travel photos. *Tourism Management, 70*, 89-98.

Notzke, C. (1999). Indigenous tourism development in the arctic. *Annals of Tourism Research, 26*(1), 55-76.

Nunkoo, R., Gursoy, D. (2012). Residents' support for tourism. *Annals of Tourism Research, 39*(1), 243-268.

Oku, K., Hattori, F., Kawagoe, K. (2015). Tweet-mapping Method for Tourist Spots Based on Now-Tweets and Spot-photos. *Procedia Computer Science, 60*, 1318-1327.

Pan, S., Lee, J., Tsai, H. (2014). Travel photos: Motivations, image dimensions, and affective qualities of places. *Tourism Management, 40*(1), 59-69.

Ruiz, E., de la Cruz, E., Vázquez, F. (2019). Sustainable Tourism and Residents' Perception towards the Brand: The Case of Malaga (Spain). *Sustainability, 11*(1), 292.

Sengel, T., Karagoz, A., Cetin, G., Dincer, F. I., Ertugral, S. M., Balık, M. (2015). Tourists' Approach to Local Food. *Procedia - Social and Behavioral Sciences, 195*, 429-437.

Sharpley, R. (2014). Host perceptions of tourism: A review of the research. *Tourism Management, 42*, 37-49.

Slak Valek, N., Williams, R. B. (2018). One place, two perspectives: Destination image for tourists and nationals in Abu Dhabi. *Tourism Management Perspectives, 27*, 152-161.

Stepchenkova, S., Zhan, F. (2013). Visual destination images of Peru: Comparative content analysis of DMO and user-generated photography. *Tourism Management, 36*, 590-601.

Stylidis, D., Shani, A., Belhassen, Y. (2017). Testing an integrated destination image model across residents and tourists. *Tourism Management, 58*, 184-195.

Thomee, B., Elizalde, B., Shamma, D. A., Ni, K., Friedland, G., Poland, D., Borth, D., Li, L. (2016). YFCC100M. *Communications of the Acm, 59*(2), 64-73.

Tsaur, S., Lin, Y., Lin, J. (2006). Evaluating ecotourism sustainability from the integrated perspective of resource, community and tourism. *Tourism Management, 27*(4), 640-653.

Valek, N. S., Williams, R. B. (2018). One place, two perspectives: Destination image for tourists and nationals in Abu Dhabi. *Tourism Management Perspectives, 27*, 152-161.

Vu, H. Q., Li, G., Law, R., Ye, B. H. (2015). Exploring the travel behaviors of inbound tourists to Hong Kong using geotagged photos. *Tourism Management, 46*, 222-232.

Xiang, Z., Du, Q., Ma, Y., Fan, W. (2017). A comparative analysis of major online review platforms: Implications for social media analytics in hospitality and tourism. *Tourism Management, 58*, 51-65.

Xiang, Z., Schwartz, Z., Gerdes, J. H., Uysal, M. (2015). What can big data and text analytics tell us about hotel guest experience and satisfaction? *International Journal of Hospitality Management, 44*, 120-130.

Yin, L., Cheng, Q., Wang, Z., Shao, Z. (2015). 'Big data' for pedestrian volume: Exploring the use of Google Street View images for pedestrian counts. *Applied Geography, 63*, 337-345.

Young, C. A., Corsun, D. L., Baloglu, S. (2007). A taxonomy of hosts visiting friends and relatives. *Annals of Tourism Research, 34*(2), 497-516.

Zhang, J., Inbakaran, R. J., Jackson, M. S. (2006). Understanding Community Attitudes Towards Tourism and Host-Guest Interaction in the Urban-Rural Border Region. *Tourism Geographies, 8*(2), 182-204.

Zhang, K., Chen, D., Li, C. (2020). How are Tourists Different? - Reading Geo-tagged Photos through a Deep Learning Model. *Journal of Quality Assurance in Hospitality & Tourism, 21*(2), 234-243.

Zhang, K., Chen, Y., Li, C. (2019). Discovering the tourists' behaviors and perceptions in a tourism destination by analyzing photos' visual content with a computer deep learning model: The case of Beijing. *Tourism Management, 75*, 595-608.

Zhang, S., Chan, C. (2016). Nature-based tourism development in Hong Kong: Importance –

Performance perceptions of local residents and tourists. *Tourism Management Perspectives*, *20*, 38-46.

Zheng, W., Liao, Z. (2019). Using a heuristic approach to design personalized tour routes for heterogeneous tourist groups. *Tourism Management*, *72*, 313-325.

Zheng, W., Liao, Z., Qin, J. (2017). Using a four-step heuristic algorithm to design personalized day tour route within a tourist attraction. *Tourism Management*, *62*, 335-349.

Table 1. Selected studies of the attributes/categories of a destination image

Authors	Study and methods	Destination image attributes
(Stepchenkova &Zhan, 2013).	Compares between the projected image and the accepted image of the tourist destination. The sample is selected from the official websites and Flickr respectively. The image content is analyzed manually, using an analysis framework of 20 attributes.	Nature & landscape, people, archaeological sites, way of life, traditional clothing, architecture/buildings, outdoor/adventure, wild life, art object, tourism facilities, urban landscape, domesticated animals, plants, festivals & rituals, leisure activities, food, country landscape, transport/infrastructure, tour, and others (20 categories of cognitive image)
(Ku &Mak, 2017).	Compares the perceptual difference between residents and tourists on destination image using questionnaire survey.	Recreational activity, relaxation, space planning and management, transportation, environmental hygiene, natural scenery, facility maintenance and management, store management (8 categories of cognitive image)
(Mak, 2017)	Compares the projected image and perceived image of a tourist destination through website keyword search.	Natural environment, infrastructure, culture & art, specific activities, food & beverage, flora & fauna, people, transportation, information, accommodation, tourist attraction, pleasant, relaxing, exciting, unpleasant, gloomy, sleepy, distressing (18 categories of cognitive image and affective image).
(Hernández-Mogollón et.al, 2018)	Examines the influence of cultural events on the cognitive and affective images of tourist destinations using questionnaire survey.	Pleasant-unpleasant, relaxing-distressing, exciting-depressing, cognitive image, infrastructure, easy access to the area, socioeconomic environment, prices and quality, reasonable cost of restaurants, friendly local people, good place for children and families, welcome centers, good weather, safe and secure environment (13 categories of cognitive and affective image).
(Valek &Williams, 2018)	Identifies the differences in image perceptions of tourists and locals using questionnaire survey.	Quality of accommodation and service, cultural attractions (museums, galleries), authentic Emirati culture, cuisine, customs, shopping and entertainment facilities, scenery and natural attractions (desert, sunsets, …), sun, sand, and sea (6 categories of cognitive image)

Table 2. Categories developed from 103 recognized scenes

Category	Scene
Culture	*Dragon dance, fireworks, the Chinese character 福 (happiness), library, lion dance, red envelope, stage, the Chinese character 喜 (joy).*
Entertainment	*Badminton court, bar, baseball court, billiard room, bowling alley, chess, football court, go, indoor basketball court, mahjong, ping-pong court, playground, swimming pool, tennis court*
Food	*Dining room, food, McDonald's, restaurant*
Insect & animal	*Bee, butterfly, camel, cat, deer, dog, dragonfly, elephant, giraffe, kangaroo, ladybug, leopard, lion, ornamental fish, panda, peacock, penguin, rabbit, rhinoceros, tiger, tortoise*
Mountain	*Mountain*
Natural phenomenon	*Night, overcast, snow, sunset, blue sky*
Plant	*Fallen leave, flower, green planet*
Shopping	*Mall courtyard, supermarket*
Interpretation	*Map, text*
Traffic	*Aircraft, bicycle, cabin, car, helicopter, in car, motorcycle, ship, station, train*
Architecture	*Cathedral hall, corridor, European buildings, Islam buildings, old buildings, overlook, skyscraper, wide-angle tower, work site*
Water	*Beach, bridge, waterfall, waterside, water surface*
Others	*air conditioner, bedroom, camera, classroom, dinosaur, glasses, high heels, keyboard, kitchen, little pony, living room, meeting room, office, teddy bear, the Smurfs, transformer, washing machine, washroom, watch*

Table 3. Residents, tourists and their uploaded photos

Group	Number of users	Number of photos	Number of photos per user
Tourists	1430	29081	20.34
Locals	323	29311	90.75
Total	1753	58392	33.31

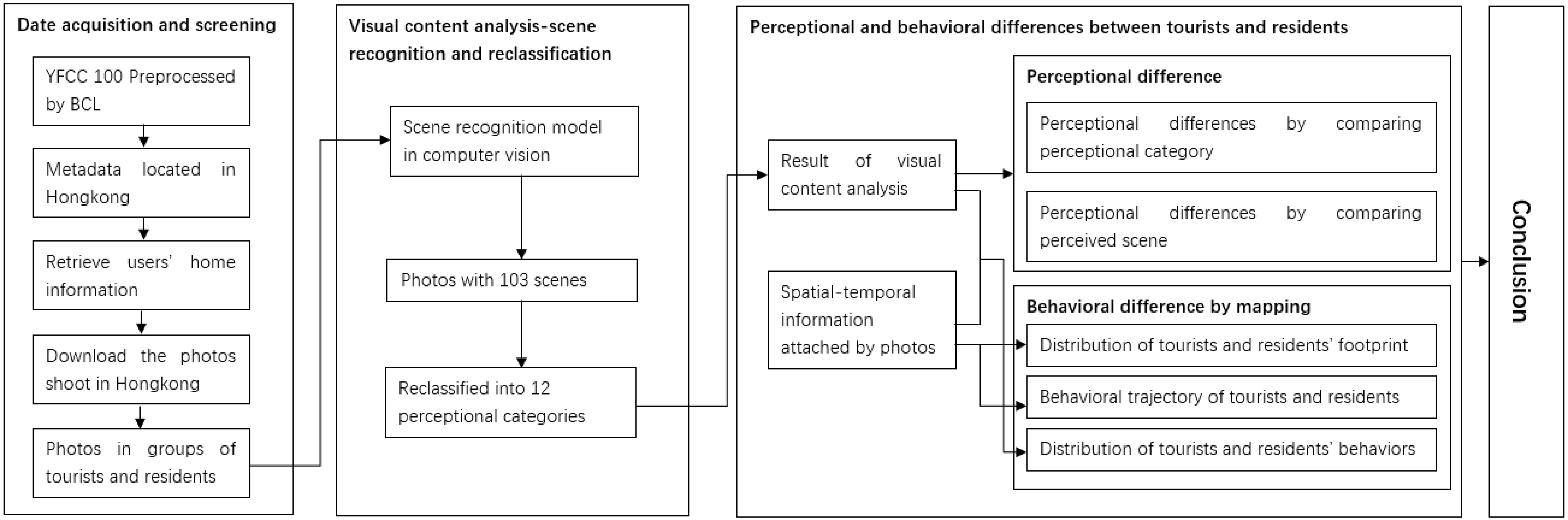

Figure 1 Study framework

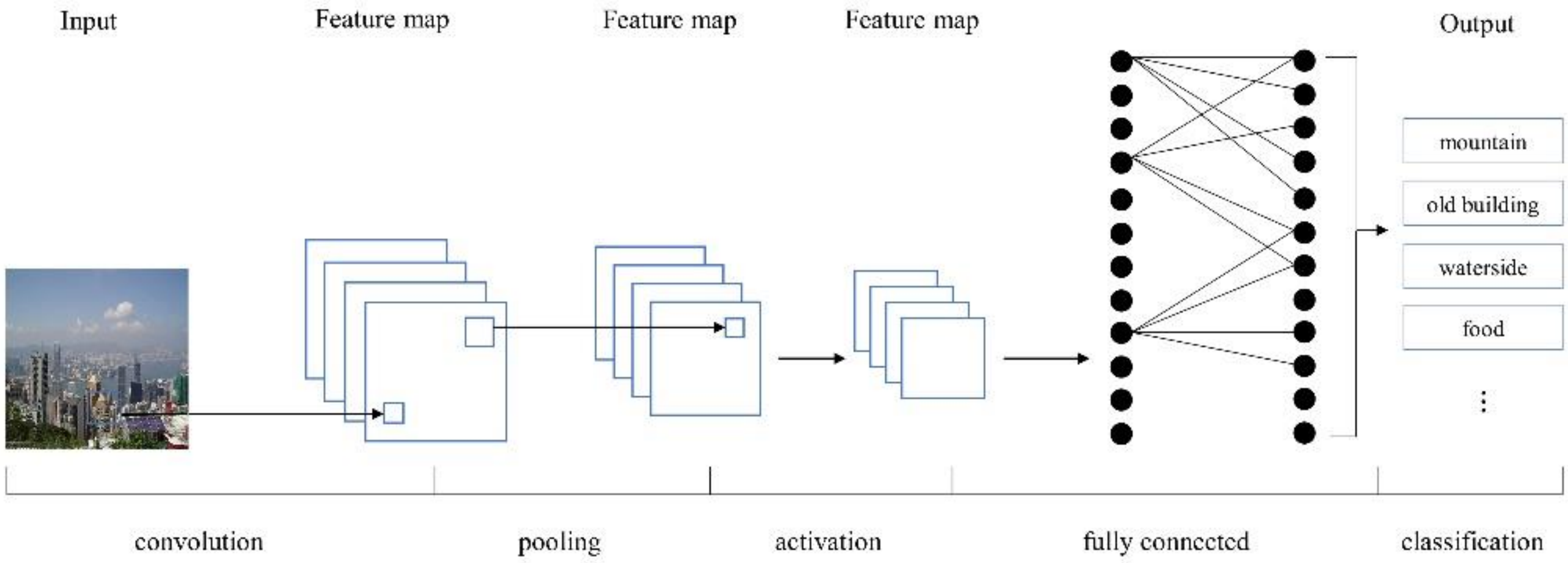

Figure 2. Structure of CNN

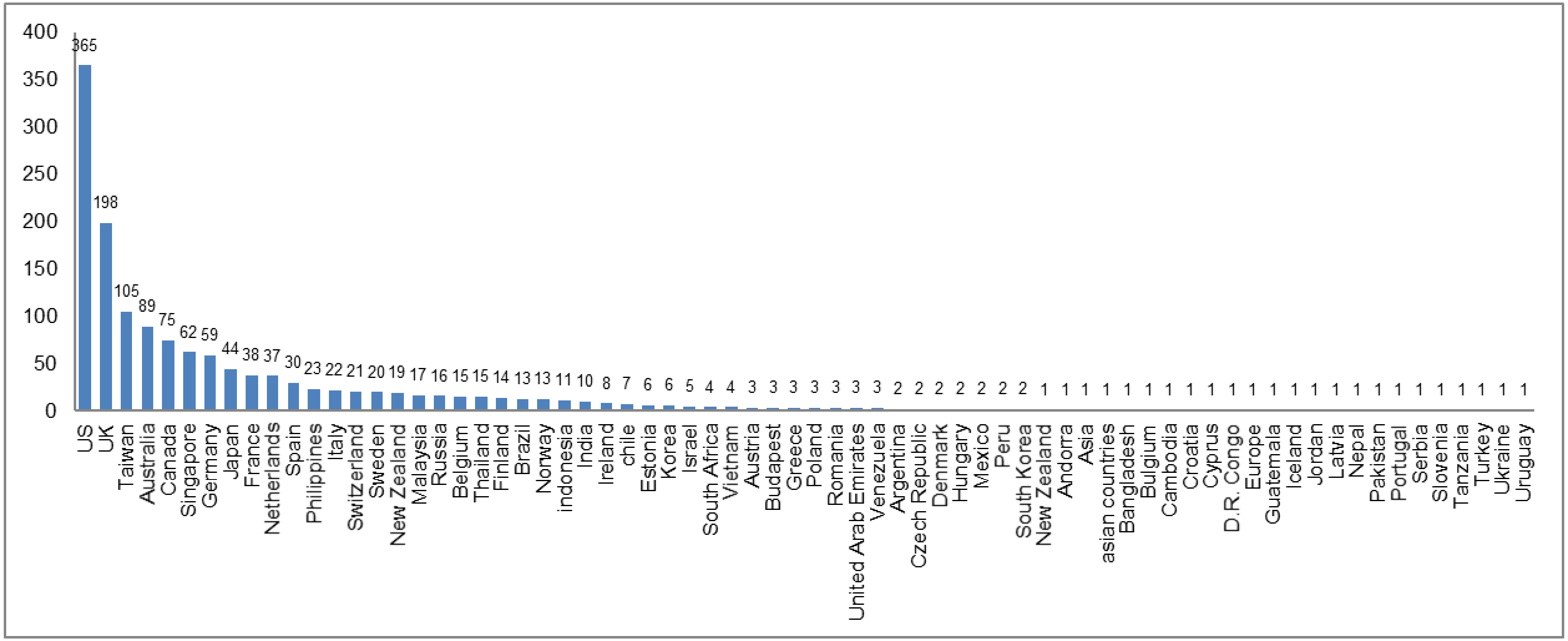

Figure 3. Distribution of the number of tourists by country or region

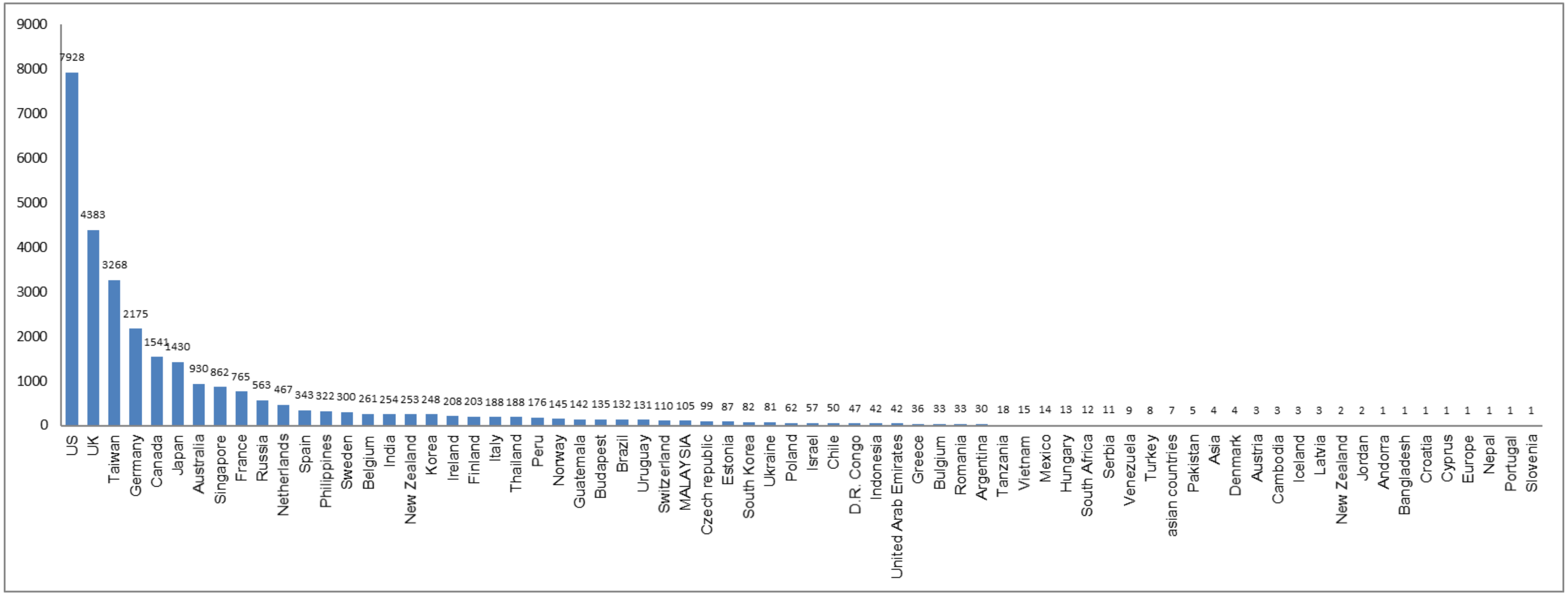

Figure 4. Distribution of the number of tourists uploaded photos by country or region

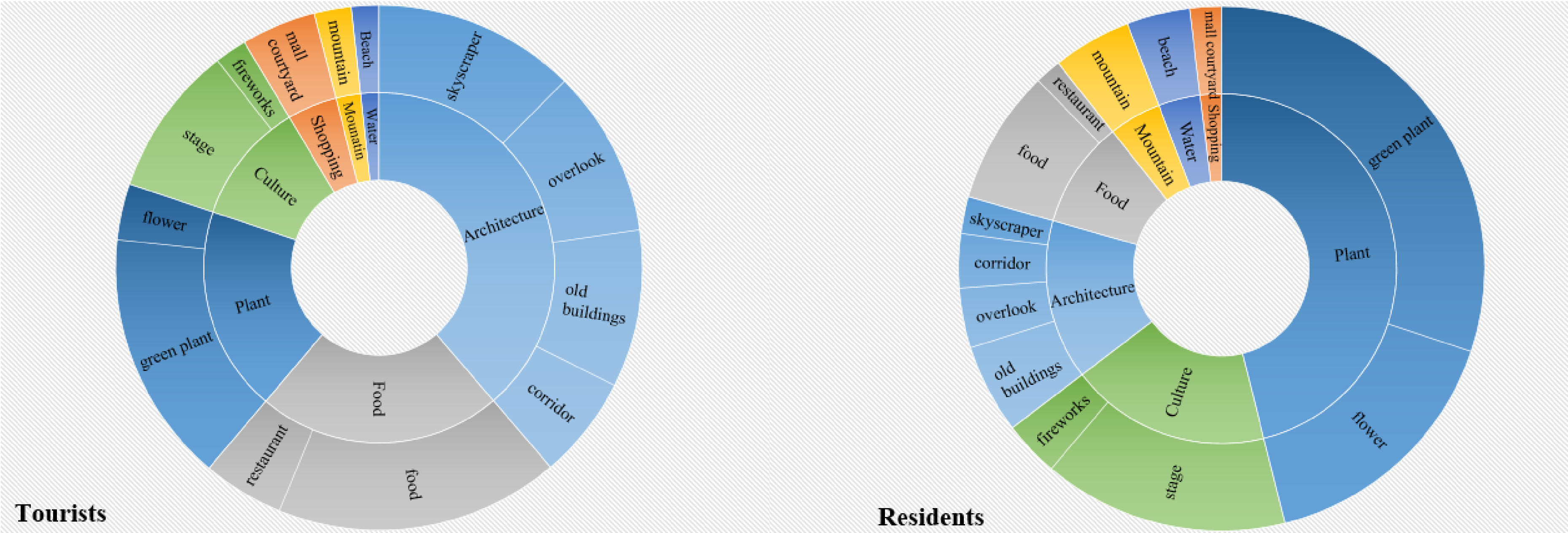

Figure 5. Specific perceptional difference between tourists and residents

Figure 6. The most important districts and places in Hong Kong

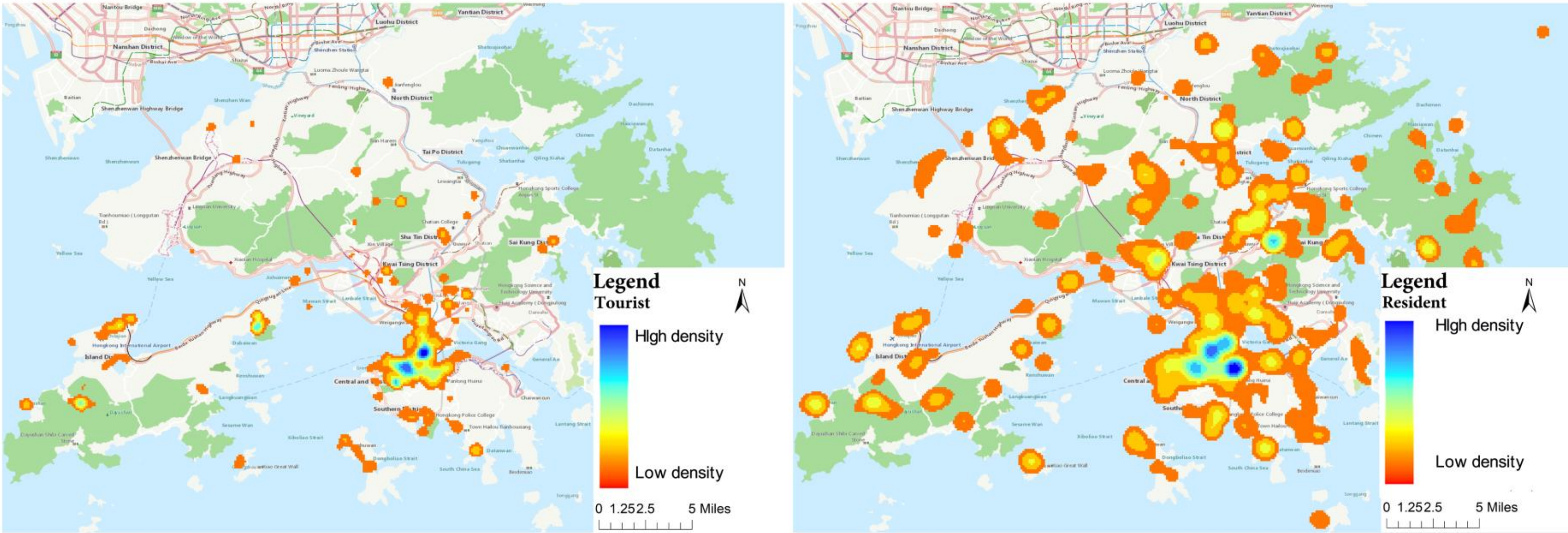

Figure 7. Distribution of tourists and residents' footprint

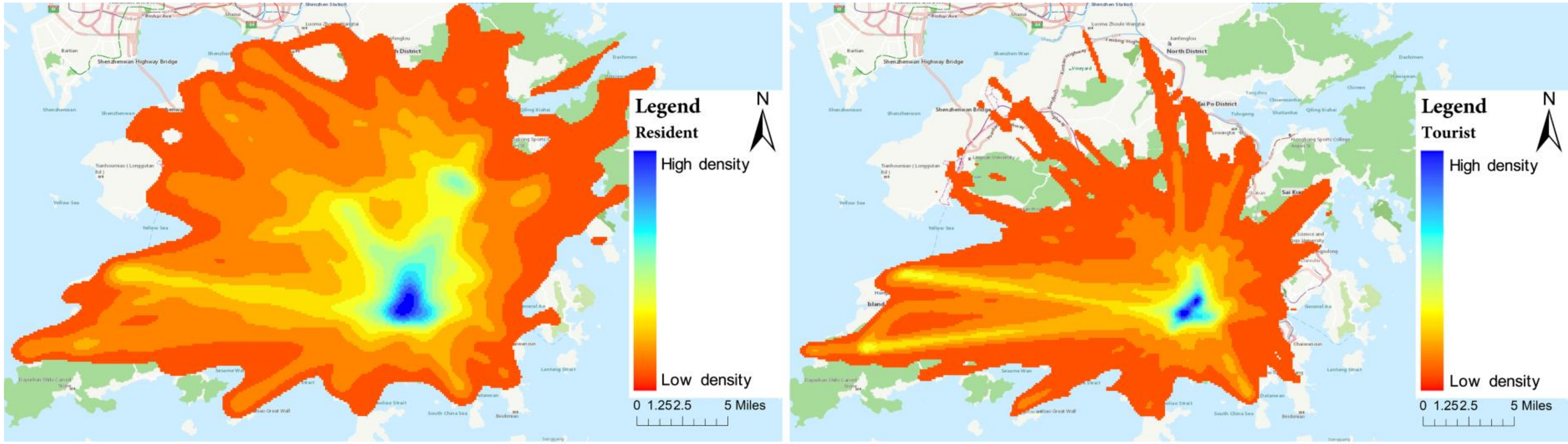

Figure 8. Density distribution of residents and tourists' trajectory

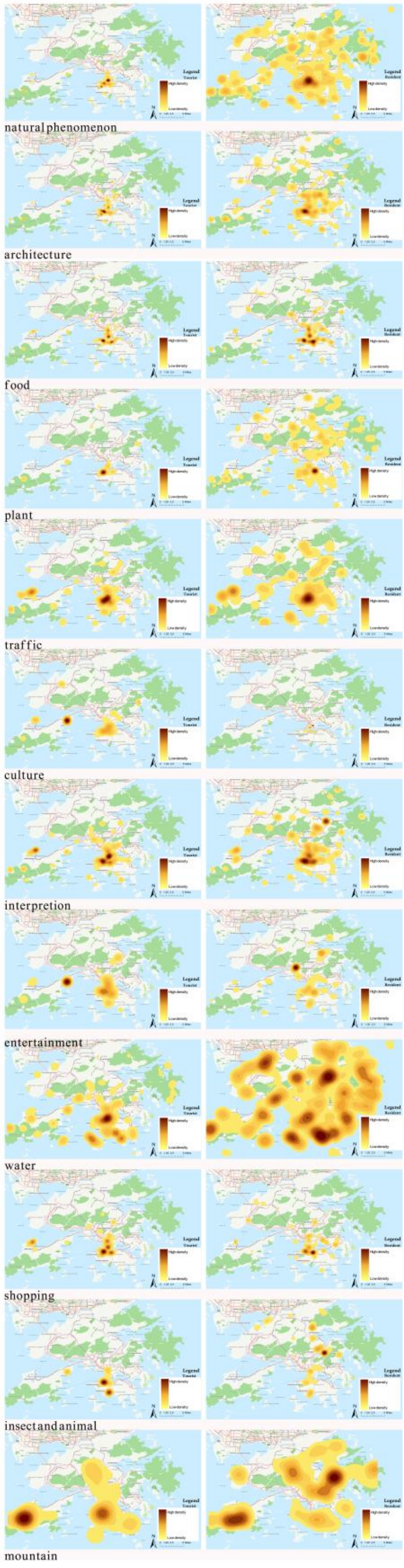

Figure 9. Density distribution of tourists and residents' behavioral activities