Spat. Inf. Res. (2020) 28(2):241–255
https://doi.org/10.1007/s41324-019-00285-x

# Understanding tourists' urban images with geotagged photos using convolutional neural networks

**Dongeun Kim[1] · Youngok Kang[1] · Yerim Park[1] · Nayeon Kim[1] · Juyoon Lee[1]**

Received: 4 March 2019 / Revised: 21 July 2019 / Accepted: 25 July 2019 / Published online: 31 July 2019
© The Author(s) 2019

**Abstract** This study aims to track down representative images and elements of sightseeing attractions by analyzing the photos uploaded on Flickr by Seoul tourists with the image mining technique. For this purpose, we crawled the photos uploaded on Flickr and classified users into residents and tourists; drew 11 region of attractions (RoA) in Seoul by analyzing the spatial density of the photos; classified the photos into 1000 categories and then 14 categories by grouping 1000 categories by utilizing Inception V3 model; analyzed the characteristics of the photo image by RoA. Key findings of this study are that tourists are interested in old palaces, historical monuments, stores, food, etc. and those key elements are distinguished from the major sightseeing attractions in Seoul. More specifically, tourists are more interested in palaces and cultural assets in Jongno and Namsan, food and restaurants in Shinchon, Hongdae, Itaewon, Yeouido, Garosu-gil, and Apgujeong, war monuments or specific artifacts in War Memorial and the National Museum of Korea, facilities, temples, and pictures of cultural properties in Samsung Station, and toyshops in Jamsil. This study is meaningful in three folds: first, it tries to analyze urban image through the photos posted on SNS by tourists. Second, it uses deep learning technique to analyze the photos. Third, it classifies and analyzes the whole photos posted by Seoul tourists while most of other researches focus on only specific objects. However, this study has a limitation because the Inception v3 model which has been used in this research is a pre-trained model created by training the ImageNet data. In future research, it is necessary to classify photo categories according to the purpose of tourism and retrain the model by creating new training data set focusing on elements of Korea.

**Keywords** Image data mining · Geotagged photos · Flickr · Convolutional neural network · Inception v3 model

## 1 Introduction

Today people prefer to share the posts such as texts, images, and videos via Social Network Services (SNS) with others without regard to time and location. Moreover, the geo-tagged photos uploaded on the site by tourists display the perception and the action of tourists as well as the images that tourists feel about the sightseeing attractions [1]. As the images of touristic sites are closely associated with the tourists' attraction and intention, they serve as a reference for other tourists who seek to travel to those sites [2]. In addition, as the touristic images on SNS can be continually produced and reproduced, we are able to ascertain the perceptions and the trends of representative sightseeing elements and locations by analyzing the images uploaded on SNS. Furthermore, this process contributes to the basic research on tourism for discovering, developing, and improving sightseeing attractions [3].

We think that it is possible to conduct in-depth analysis with the extracted information in tandem with pre-existing methodologies of spatial data analysis because geo-tagged photos contain locational information. Especially we can make better use of Flickr data because they contain the information on location and time and are automatically affiliated with photo metadata. Previous studies which have utilized geotagged data on SNS have mostly explored the

✉ Youngok Kang
ykang@ewha.ac.kr

[1] Department of Social Studies, Ewha Womans University, Seoul, South Korea

location that users occupied [4–6], the patterns of movement [7, 8] and the texts of uploaded photos [9–16]. However, as the image analysis using deep learning technology becomes available, the studies using the photos posted on SNS keep increasing recently. Examples of researches on analyzing the photos posted on the SNS include classification of food [11], analysis of bird observations between experts and ordinary people [17], estimation of weather preference by visiting specific places [18]. Most of the studies are focused on analyzing the photos which contain specific objects. There have been no studies to analyze the image of tourists in the area by classifying the whole photos posted by the tourists who visit the specific area.

The purpose of this study is to analyze representative images and elements of sightseeing attractions by analyzing the photos uploaded on Flickr by Seoul tourists. For this purpose, first, we crawled the photos uploaded on Flickr, which is one of Social Network Service (SNS) platforms that people can share geotagged photos, and classified users into residents and tourists. Second, we drew 11 region of attractions (RoA) in Seoul by analyzing the spatial density of the photos uploaded by tourists. Third, we classified the photos into 1000 categories and then 14 categories by grouping 1000 categories by utilizing Inception V3 model, which is one of the convolutional neural networks (CNN) with deep learning capability. Finally, we analyzed the characteristics of photo image by RoA.

## 2 Research on image data mining via convolutional neural networks

Image data mining is the process of extracting information or knowledge from image data [19]. Recently, with the increase in the volume of image data as well as the improvement of training algorithm, techniques of image data mining using artificial neural networks have been applied to various fields such as medicine, environmental studies, information science, and computer graphics [20]. Convolutional neural network (CNN) which is one of artificial neural networks has been developed based on neurological knowledge surrounding the visual cortex of humans and animals [21]. As CNN has been shown to be effective in distinguishing and categorizing the photo images, it has become a trend to make use of it in most image data mining research. CNN is basically composed of three layers such as a convolutional layer, a pooling layer, and a fully connected layer. One can not only produce a variety of models by changing the CNN configurations, but also train the CNN through the scan of the image characteristics.

Researches on classification of images by category using CNN method have been actively conducted in the field of medicine. Krishnan et al. [22] categorized liver diseases surfaced on the images of ultrasonic inspection. Sawant et al. [23] detected brain cancer through MRI, and Motlagh et al. [24] distinguished breast cancer from the images of histopathological samples. Further, CNN method has been applied in other fields of image mining. Park and Shim [25] established a model of discerning the genre from the images of movie posters, taking inspiration from the thought that elements such as title font and chroma of images of movie posters can differ according to the genre of the movies. Lee and Lee [26] created the model which could recognize the characters in the animation of 'The Simpson', and Xu et al. [27] conducted a research on distinguishing geo-tagged land images by the conditions of land coverage.

The studies that have executed the image data mining using the images on SNS are as follows: Kaneko and Yanai [12] researched to track down event photos such as festivals, sports game, earthquake and fires by analyzing geotagged photos on Tweeter. Deng et al. [16] compared the images of Shanghai seen by tourists from the East and the West through the tags of photos uploaded on Flickr by tourists. These studies did not analyze the image itself, but categorized the images through tags included in the photos. Meanwhile, Okuyama and Yanai [14] selected representative images of designated locations after extracting the locations from the photos uploaded on Flickr. This study has applied Speeded-Up Robust Features (SURF) technique out of various image data mining techniques. The SURF technique is good at recognizing a specific object because it extracts and recognizes the features of an image. However, it is being evaluated that it is difficult to recognize a photographic background and it is low in view of the accuracy of classification [9, 28].

Recently, the CNN method is mainly used for analyzing the photos posted on SNS. Jang and Cho [9] proposed a method of extracting tags automatically from the images posted on Instagram. Hong and Shin [10] proposed a method of recommending followers (information providers) by extracting the categories with the huge number of images uploaded after categorizing the images posted by Instagram users. Kagaya and Aizawa [11] distinguished the images that actually contained food from those that did not among the populated photos when searching "#food" on Instagram. On the other hand, Koylu et al. [17] analyzed the spatial distribution of birds' photos posted on Flickr by comparing the birds' photos uploaded by ordinary people with the photos uploaded by birds' experts and found the difference in viewing birds between ordinary people and birds' experts. Chu et al. [18] estimated the weather when a user visited each landmark through the appearance of a sky

or a cloud in the landmark photos posted on Flickr. In recent years, the studies on the use of CNN to analyze the photos posted on SNS keep increasing continuously. However, there are only a few researches on analyzing the images of a region by classifying all photos posted on SNS while there are many researches which focus on specific objects.

## 3 Method of analysis and procedures

### 3.1 Data collection and extraction of RoA

We have crawled the photos from Flickr using the open API in the spatial range of 37.4°–37.8° latitude, 126.8°–127.2° longitude, which represents spatial range of Seoul and in the temporal range from January 1, 2015 to December 31, 2017. We have collected a total of 86,304 photos uploaded by 1974 users. Then we have distinguished residents from tourists among 1974 users. We divided 1974 users into 868 users who have filled in owner location on Flicker metadata and 1106 users who either have not filled in owner location or difficult to distinguish accurately. Furthermore we divided 868 users into 689 users who were classified as tourists and 179 users as residents of Seoul. In addition, we divided 1106 users into 319 users who are residents of Seoul and 787 users who are tourists. If the time difference between the first photograph and the last photograph taken in the Seoul area during the study period exceeds 30 days, users are categorized as residents; otherwise users are categorized as tourists. As a result a total of 1476 users were discerned as tourists after sorting out 689 users who have filled in and 787 users who have not filled in their location, respectively. Finally, we analyzed the image of Seoul based on a total of 39,157 photos on Flickr uploaded by a total of 1476 tourists [29].

Based on data collection, we extracted RoA from the 39,157 Flickr data uploaded by tourists through the use of Density Based Spatial Clustering Application with Noise (DBSCAN) algorithm. DBSCAN is a density-based clustering algorithm which can be used to identify clusters of any shape in a data set containing noise and outliers [30]. It groups together the points that are closely packed together, marking as outliers the points that lie alone in low-density regions. DBSCAN requires two parameters which are ε (eps) and the minimum number of points to form a dense region. In order to adopt the optimal combination, various pairs of parameters were applied to the experiment. The minimum number of points was set between 200 and 2000 and the minimum search radius was set between 300 and 1000 m. As a result of the experiment, 11 RoA were derived by adopting the combination of the minimum number of points of 350 and the minimum search radius of 250 m, which appropriately include major tourist attractions in Seoul [29]. We named the RoA by referring to "The survey of the current state of foreign tourists" conducted by the Korea Tourism Organization in 2017 [31]. We found that 11 RoA derived from this study are coincide with the major tourist attractions in Seoul reported by "The survey of the current state of foreign tourists". Table 1 and Fig. 1 show the information on each RoA and Fig. 2 illustrates analytical method and procedure of this study.

### 3.2 Method of image data mining

We conducted image data mining with 38,691 photos among 39,156 photos posted by 1476 users, excluding 465 photos deleted by users. For our analysis we applied Python version 3.6 and Tensor flow, open source machine learning library. We applied Inception v3 model of Google Net, which is one of various CNN models, in the photo data mining. Inception v3 is a model "trained" with ImageNet's image data set, which comprises of 14,197,122 images divided into 1000 categories [32]. The images in ImageNet are divided into 27 primary categories and 1000 secondary categories. In case of classifying the images with the Inception v3 model, the model returns the category name that resembles most with the input image among 1000 categories and its accuracy value. In addition to GoogleNet, there are also LeNet-5, AlexNet, and ResNet, which are various variations of the basic structure of neural networks. The Inception module, a subnetwork included with GoogleNet, has a deep structure and makes GoogleNet use parameters more effectively than other models [20]. Among the various models of GoogleNet using the Inception module, Inception v3 model provides low-error rates and its source codes are widely available.

As Inception v3 model uses Tensor Flow to operate, it is necessary to pre-process the photos into appropriate formats before analyzing photo data. As data crawled from Flickr's API are in the format of image URL, we downloaded them into BMP format and then converted them into size of 299 * 299 RGB, which can be used in the Inception v3 model. It is not easy to derive the meaning by comparing 1000 categories when each image is categorized into one of 1000 categories. Moreover, the 27 primary categories in ImageNet were also not easily applicable to the category for tourism. Given these constraints, we generated 14 new categories that were suitable for the field of tourism based on the values resulting from the categorization of the 38,691 images. Basically, 14 categories have been created by referring to the categories of major activities on "The survey of the current state of foreign tourists" conducted by the Korea Tourism Organization in 2017. These categories are as follows: "food," "entertainment," "shopping," "transportation," "cityscape," "facilities,"

**Table 1** RoA in Seoul

| Name | Region of attraction | Number of photos |
|---|---|---|
| Jongro, Namsan | Samcheong-dong, palaces (Gyeongbokgung, Deoksugung, Changgyeonggung), Cheonggye Stream, Sejong Center for the Performing Arts, City Hall, Namdaemun Marketplace, Seoul Station, Insa-dong, Myeong-dong, Hanyang Wall Course 3, Namsangol Park, Namsan Tower, DDP, Gwangjang Market, Daehakro | 21,323 |
| Shinchon, Hongdae | Ewha Women's University, Yeonse-ro, Yeonnam-dong, Hongdae Station area, Mecenatpolis | 2607 |
| War Memorial of Korea | War Memorial of Korea | 1017 |
| National Museum of Korea | National Museum of Korea | 970 |
| Samsung Station, Bongeunsa Temple, Coex Mall | Bongeunsa Temple, Coex Mall | 876 |
| Jamsil | Lotte World, Lotte World Tower | 849 |
| Itaewon | Hannam-dong, Itaewon Station, Gyeongridan Road | 842 |
| Gangnam Station | Gangnam Station Street | 752 |
| Yeouido | IFC Mall | 430 |
| Garosu-gil | Garosu-gil Road | 421 |
| Apgujeong | K-Pop Street | 306 |

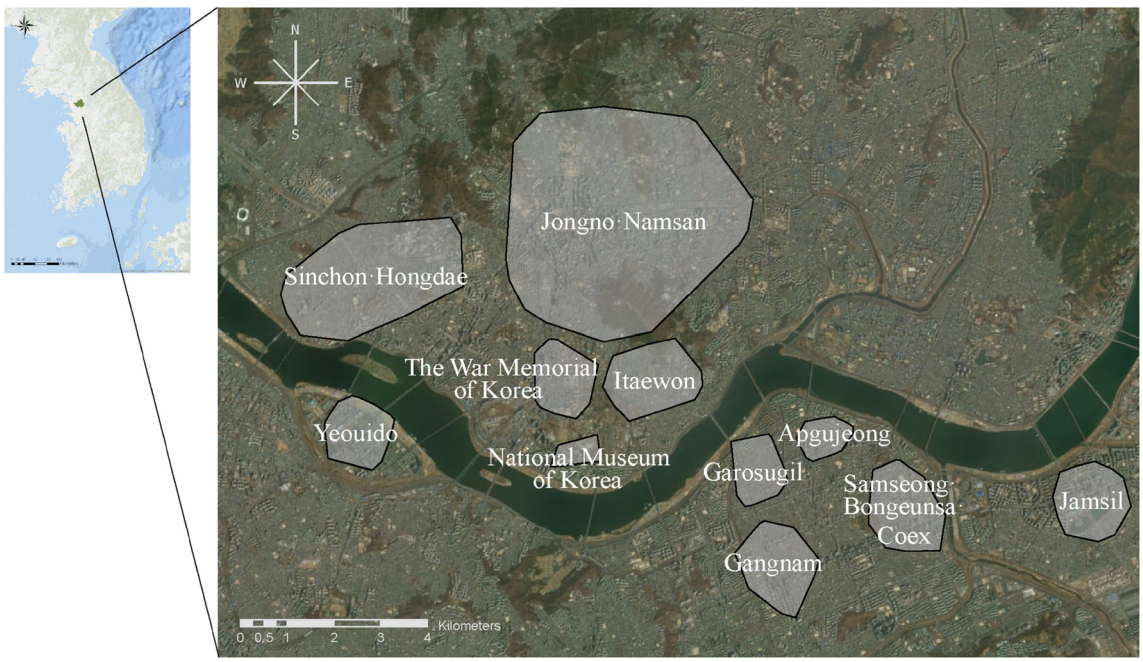



**Fig. 1** RoA in Seoul

"residence," "natural views/flora and fauna," "people," "religion," "clothing," "palace/historical monuments/cultural properties," "objects/miscellaneous," and "exhibits/sculptures" as shown in Table 2.

## 4 Results of analysis

### 4.1 Image of Seoul

We were able to produce 858 categories out of 1000 categories of ImageNet by classifying the 38,691 photos uploaded by Seoul tourists. The categories with a proportion of 1% or above among 858 categories are shown in Fig. 3. When looking at the category into details, there are

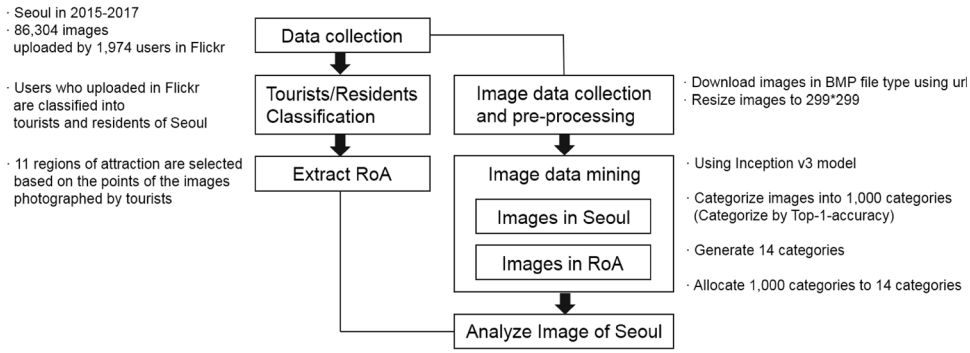



**Fig. 2** Research flow

**Table 2** Newly produced 14 primary categories and corresponding secondary categories

| Primary categories | Examples of secondary categories |
|---|---|
| Food | Bakery, bakeshop, bake house/coffee mug/restaurant, eating house, eating place, eatery, etc. |
| Entertainment | Beer bottle/horizontal bar/wine bottle, etc. |
| Shopping | Barbershop/cinema, movie theatre, movie house, picture palace/confectionery, confectionery, candy store, etc. |
| Transportation | Ambulance/bicycle-built-for-two, tandem bicycle, tandem/canoe, etc. |
| Cityscape | Cab, hack, taxi, taxicab/spotlight, spot/volcano |
| Facilities | Beacon, lighthouse, beacon light, Pharos/greenhouse, nursery, glasshouse/fountain, etc. |
| Residence | Mobile home, manufactured home/prison, prison house |
| Natural views/flora and fauna | Trench, Tinca tinca/goldfish, Carassius auratus/lakeside, lakeshore, etc. |
| People | Ballplayer, baseball player/groom, bridegroom/scuba diver |
| Religion | Altar/church, church building/stupa, tope |
| Clothing | Gown/kimono/neck brace, etc. |
| Palace/historical monuments/cultural properties | Abacus/bell cote, bell cot/palace, etc. |
| Objects/miscellaneous | Accordion, piano accordion, squeeze box/ashcan, trash can, garbage can, waste bin, ash bin, ash bin, ashbin/candle, taper, wax light, etc. |
| Exhibits/sculptures | Balloon/pedestal, plinth, footstall/totem pole, etc. |

usually images of front gate for "palace", roof tiles for "bell cote" and "tile roof", and interior gardens for "patio, terrace". Like this, we can get an idea of the representative images of palaces that tourists have in mind when visiting Seoul. In the category of food, "plate" includes traditional Korean cuisine, sashimi, and pasta, "restaurant" does barbeque house, café and inner interiors, "food market" does the images of markets such as supermarkets and street markets, traditional market and street food, "hot pot" does the images of soup and "menu" does menu list. The "toyshop" contains the images of not only actual toy stores but also objects including certain characters and interiors of various shops, such as variety stores and hardware stores. The "movie theatre" includes the images of shop exteriors such as clothing stores and restaurant. The "stage" includes the images of building interiors and those that emphasize equipment. Both "taxi" and "traffic light" include the images of streets, cars parked along the road, and neon signage decorating the outside of buildings. The "prison" and "monastery" include the images of crowded residential areas, museums, and the like. "Lakeside," on the other hand, includes images of natural views with not only lakes or rivers but also trees or sky, while "pier" does the images of rivers, streams, ponds, college campus, and so on. To sum it all up, we can deduce the tourists' perception of Seoul in which owns palaces, food, buildings, and facilities.

Table 3 shows the results of assigning 858 categories to 14 primary categories for analysis by subjects. Figure 4 shows the results of the top five primary categories by extracting and examining their secondary categories. We can see that tourists who visit Seoul are generally interested in palaces, historical monuments, cultural properties, objects, food, facilities, natural views, and flora and fauna. More specifically, when looking into the category in details, "palace/historical monuments/cultural properties",

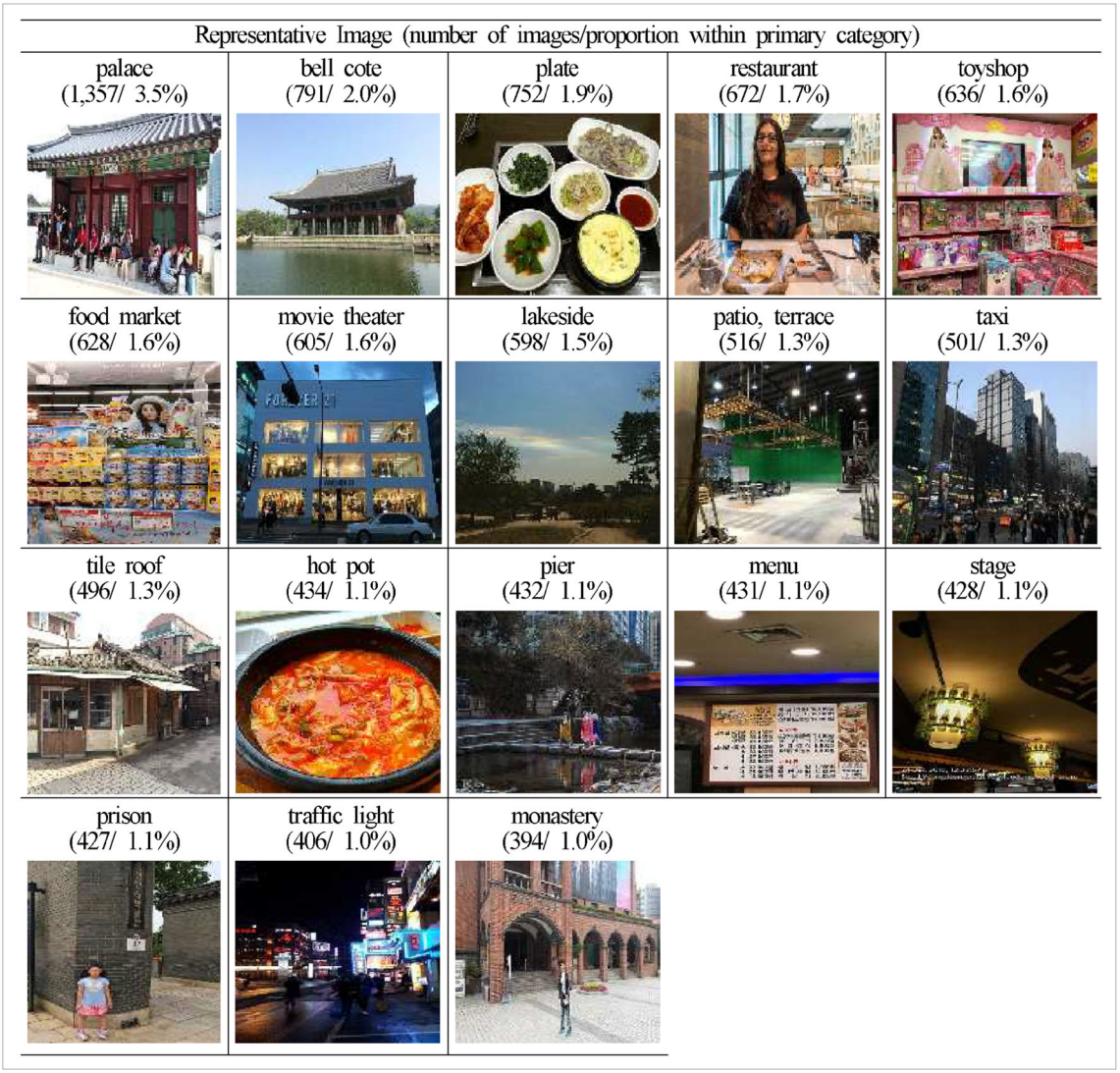



**Fig. 3** Results of classifying photos by 1000 categories (category name, number of photo, photo proportion)

"palace" and "bell cote" contain the images of palaces, tile-roofed houses, and Korean-style houses, respectively. "Patio and terrace" contain the images of courtyards and "tile roof" contains the images of rafters. From this we can deduce that a good number of tourists seem to consider the palaces and the traditional houses as representative images of Seoul. "Umbrella" which belongs to a subcategory of "objects/miscellaneous" includes the images of not only actual umbrellas but also silhouettes that resemble the shape of an umbrella. Similarly, "tray" contain not only some images of food on tray, but also the images of objects that resemble a tray. "Book jacket" has the images of historical monuments and exhibits. As mentioned before, this is probably due to the lack of adequate categories to properly categorize the images taken by tourists. "Plate" which belongs to a subcategory of "food" has numerous images of food such as traditional Korean cuisine and sashimi. "Restaurant" has the images of restaurants and coffee shops. "Food market" has the images of big supermarkets and traditional street markets. "Hot pot" has the images of food such as rice cake in hot sauce, soups, and teppanyaki. "Pier" which belongs to a subcategory of "facilities" contains the images of Cheonggye Stream and the ECC building of Ewha Women's University. "Planetarium" does the images of landmarks such as Dongdaemun Design Plaza while a subcategory of "natural views/flora and fauna" contains the images mostly of sky, the Han River, and mountains.

**Table 3** Classification results of photos by 14 primary categories

| Categories | Number of photos | Proportion (%) |
|---|---|---|
| Palaces/historical monuments/cultural properties | 6627 | 17.1 |
| Objects/miscellaneous | 6211 | 16.1 |
| Food | 5899 | 15.2 |
| Facilities | 5607 | 14.5 |
| Natural views/flora and fauna | 3149 | 8.1 |
| Shopping | 2316 | 6.0 |
| Clothing | 2273 | 5.9 |
| Transportation | 2204 | 5.7 |
| Urbanscape | 1469 | 3.8 |
| Exhibits/sculptures | 1452 | 3.8 |
| Religion | 502 | 1.3 |
| Residence | 465 | 1.2 |
| Entertainment | 296 | 0.8 |
| People | 221 | 0.6 |
| Total | 38,691 | 100.0 |

### 4.2 Comparison of image by RoA

We categorized the photos into 11 RoA[1] in Seoul to compare their different characteristics. Table 4 shows the number of photos and the proportions included in the photos of 11 RoA. For example, there are 20,987 photos in Jongro and Namsan, which make up 54.2% of all photos, and there are 2584 photos in Shinchon and Hongdae, which make up 6.7%. Uploaded photos of other locations were generally similar in number. "Appendix 1" show the results that photos of RoA are classified into 1000 categories.

The photos of Jongro and Namsan included specific elements such as palace facades, palace gates, walls, and other structures, while the photos of War Memorial and National Museum of Korea included various kinds of cultural properties and historical monuments. The photos of Shinchon, Hongdae, and Itaewon included not only food itself but also the interiors of restaurants and other shops, especially the photos of Itaewon did alcohol such as beers and cocktails. The photos of Samsung Station, Bongeunsa Temple, Coex Mall, Jamsil, Gangnam Station, Apgujeong, and Garosu-gil included various stores and sculptures. More specifically, there were photos of temples in Samsung Station, Bongeunsa Temple, and Coex Mall. The photos of Jamsil, Gangnam Station and Garosu-gil/Apgujeong included ponds and amusement parks, urban scape, food, respectively. Meanwhile, the photos of Yeouido appeared to include not only food and restaurants but also Han River.

Figure 5 shows the results of assigning 858 categories to 14 primary categories for every RoA. We can see that the tourists who visit Jongro, Namsan, War Memorial of Korea, and National Museum of Korea usually are interested in "palace/historical monuments/cultural properties," "facilities," and "objects/miscellaneous." As the images for National Museum of Korea categorized as "objects/miscellaneous" are mostly of historical monuments or cultural properties, we can see that the tourists who visit this area have the images of palaces, historical monuments, and cultural properties in common. Meanwhile, the tourists who visit Shinchon, Hongdae, Itaewon, Gangnam Station, Garosu-gil, and Apgujeong have the images of "food", those who visit Samsung Station, Bongeunsa Temple, Coex Mall, Jamsil, and Yeouido have the images of "facilities", and those who visit Garosu-gil, Jamsil, Gangnam Station, Itaewon, Shinchon, Hongdae, and Apgujeong have the images of "shopping". While the images of Gangnam Station are related to "urban scape," the images of Jongro, Namsan, Samsung Station, Bongeunsa Temple, Coex Mall, and Yeouido are related to "natural views/flora and fauna." Figure 6 shows a map with the 14 primary categories and representative photos of 11 RoA.

[1] In Jongno, which is located in the center of Seoul, there are many palaces and historical sites in addition to a connection passage to Namsan since it had served as a royal palace during Joseon Dynasty. At the War Memorial Hall, visitors can experience exhibitions related to the Korean War. At the National Museum of Korea, visitors can enjoy historical relics of Korea and participate in various educational and cultural events. In Shinchon and Hongdae, there are many universities, bars, restaurants, and shopping centers on the streets. In Samsungyeog, Bongeunsa, KoEX, there are many sky scrapers, a temple called Bongeunsa and a complex mall where tourists can see the largest aquarium in Seoul, respectively. In Gangnamyeog, Garosugil, Abgujeong, there are many luxury shops and spaces where big events are held. In Jamsil, there are amusement parks and lake parks. In Itaewon, there are many good places where foreigners gathered because there is the US Forces in Korea. In Yeouido there is Han River Park as the load of Han River.

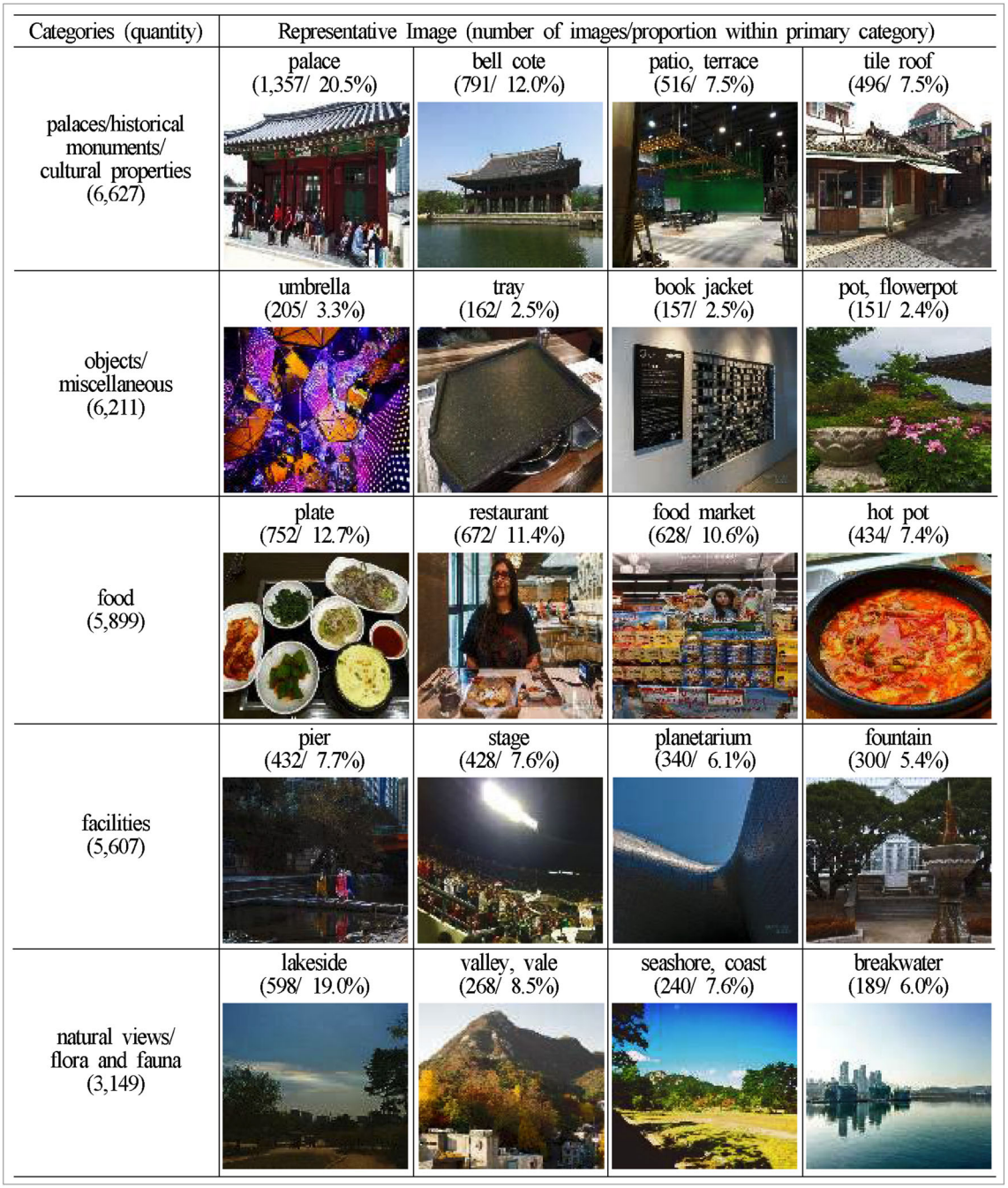



**Fig. 4** Top five primary and corresponding subcategories of photos (category name, number of photo, photo proportion)

### 4.3 Accuracy assessment

The Inception v3 model is a pre-trained model trained by 14,197,122 photographs in ImageNet which are classified into 1000 categories. Therefore, when we input the photos collected in this study, the Inception v3 model returns the category with the highest similarity value (%) among 1000 categories. It means that the result of this model can be different from the actual category of the photos. Therefore, we conducted manual labeling for 38,691 photos used in this study to evaluate the accuracy of the result of Inception v3 model. If the result of Inception v3 model matches the result of manual labeling, it is determined to be "True". Otherwise, it is determined to be "False".

Out of the total 38,691 photos, 10,807 photos were matched and 27,884 photos were mismatched, that is, the overall accuracy ratio was about 27.93%. Table 5 shows

**Table 4** Number of photos per RoA

| RoA | Number of photos | Proportion (%) |
|---|---|---|
| Jongro, Namsan | 20,987 | 54.2 |
| Shinchon, Hongdae | 2584 | 6.7 |
| War Memorial of Korea | 1008 | 2.6 |
| National Museum of Korea | 957 | 2.5 |
| Samsun Station, Bongeunsa Temple, Coex Mall | 872 | 2.3 |
| Jamsil | 840 | 2.2 |
| Itaewon | 829 | 2.1 |
| Gangnam Station | 744 | 1.9 |
| Yeouido | 428 | 1.1 |
| Garosu-gil | 419 | 1.1 |
| Apgujeong | 306 | 0.8 |
| Outliers | 8717 | 22.5 |
| Total | 38,691 | 100.0 |

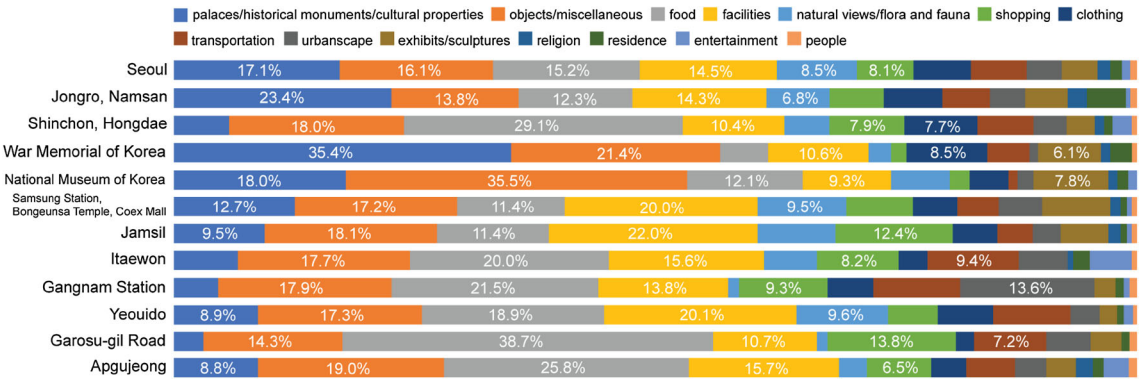



**Fig. 5** Results of classifying photos into 14 primary categories by RoA

the accuracy ratio between Inception v3 model and manual labeling by category in which the number of photos is more than 1%. The highest matching categories are 'plate', 'tile roof', 'restaurant', 'hot pot', while the lowest matching categories are 'monastery', 'prison', 'bell cote', and 'movie theater'. In case of 'plate', 'tile roof', 'restaurant', and 'hot pot', the accuracy ratio is high because there are little difference by country. On the other hand, in case of 'monastery', 'prison', 'bell cote', and 'movie theater', the accuracy ratio is low because building types are different by country and culture.

Figure 7 shows an example of the category 'palace' with a high matching ratio. While the photos of Gyeongbok Palace, Changdeok Palace, and Deoksugung are classified as 'palace' correctly, the photos of Yongsan War Memorial, university building, Hanok, residential area in front of Cheonggyecheon, and pavillion are misclassified as 'palace'. In the case of the palace, Western style building in Korea was misclassified as 'palace' because the model is based on the ImageNet's data. Figure 8 shows an example of a photo classified as 'pier'. The photos of Han River Park, Hangang Bridge, and Cheonggyecheon are classified as 'pier' correctly while Ewha Womans University ECC building, Jongno Tower cloud building, Myeongdong shopping mall building and Dongdaemun design plaza building are misclassified as 'pier'. If the sky or the base of building is widely photographed, it is misclassified as a 'pier' because it can be recognized as a river.

Through the accuracy evaluation process, we could find some implications. First, it is necessary to classify the photos posted by tourists for the purpose of tourism. Because 1000 categories in ImageNet are not intended for

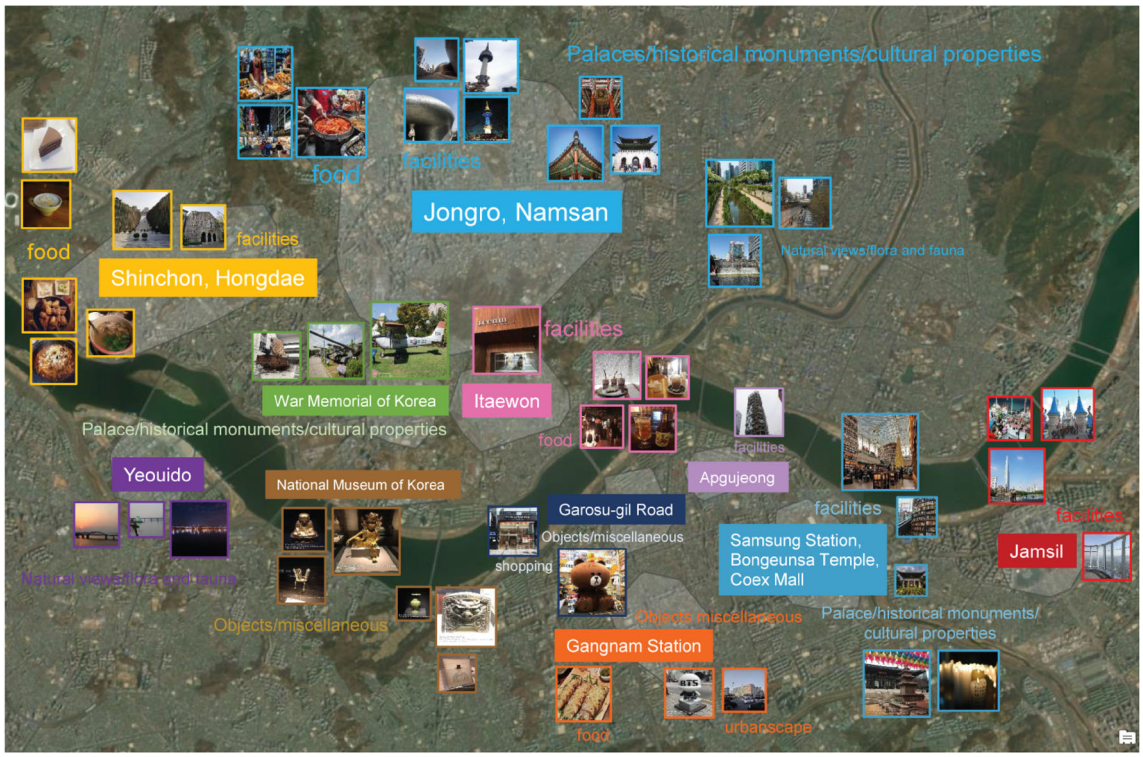



**Fig. 6** Representative categories and images by RoA

**Table 5** Accuracy ratio by categories based on Inception v3 model and manual labelling

| Predicted categories | Total | True | False | Accuracy ratio (%) |
|---|---|---|---|---|
| Palace | 1357 | 1070 | 287 | 78.85 |
| Bell cote, bell cot | 791 | 25 | 766 | 3.16 |
| Plate | 752 | 750 | 2 | 99.73 |
| Restaurant, eating house, eating place, eatery | 672 | 450 | 222 | 66.96 |
| Toyshop | 636 | 201 | 435 | 31.60 |
| Grocery store, grocery, food market, market | 628 | 317 | 311 | 50.48 |
| Cinema, movie theater, movie theatre, movie house, picture palace | 605 | 37 | 568 | 6.12 |
| Lakeside, lakeshore | 598 | 168 | 430 | 28.09 |
| Patio, terrace | 516 | 140 | 376 | 27.13 |
| Cab, hack, taxi, taxicab | 501 | 84 | 417 | 16.77 |
| Tile roof | 496 | 415 | 81 | 83.67 |
| Hot pot, hotpot | 434 | 290 | 144 | 66.82 |
| Pier | 432 | 216 | 216 | 50.00 |
| Menu | 431 | 134 | 297 | 31.09 |
| Stage | 428 | 185 | 243 | 43.22 |
| Prison, prison house | 427 | 16 | 411 | 3.75 |
| Traffic light, traffic signal, stoplight | 406 | 139 | 267 | 34.24 |
| Monastery | 394 | 2 | 392 | 0.51 |
| Total | 38,691 | 10,807 | 27,884 | 27.93 |

**Fig. 7** Example of the category 'palace': **a** photos that the result of inception v3 model matches the manual labelling, **b** photos that the result of Inception v3 model mismatches the manual labelling

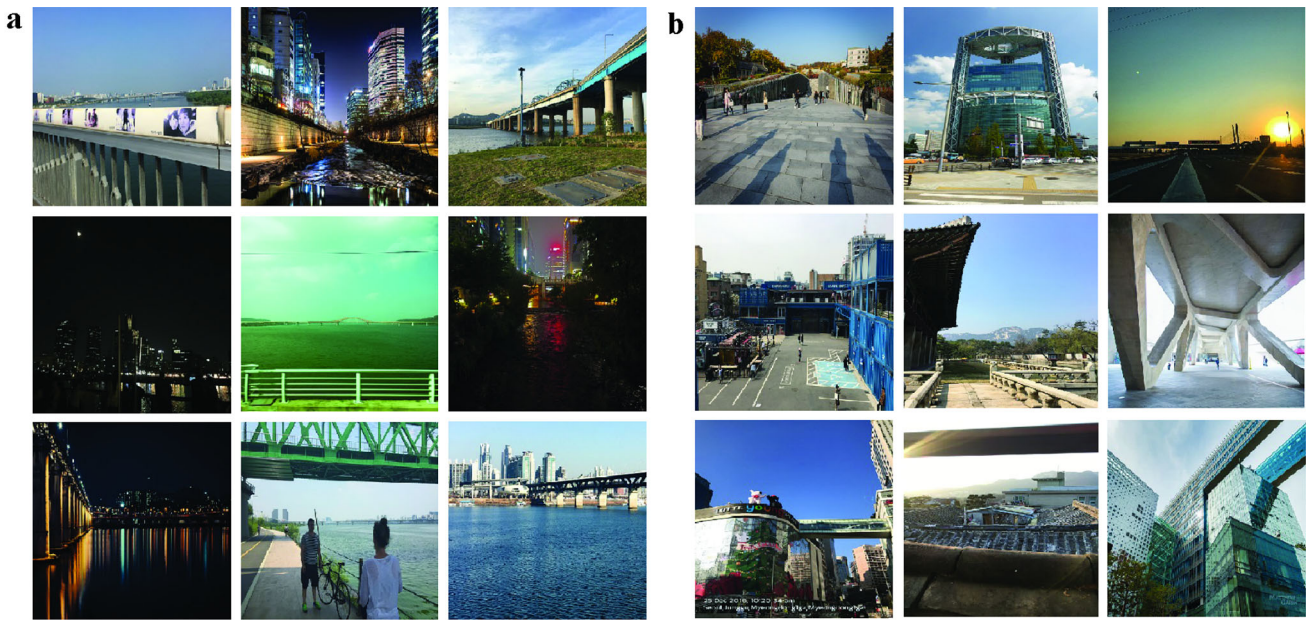



**Fig. 8** Example of the category 'pier': **a** photos that the result of inception v3 model matches the manual labelling, **b** photos that the result of Inception v3 model mismatches the manual labelling

tourism analysis, the revised categories need to be created for tourism purposes. Secondly, it is necessary to create a new training data set and retrain the Inception v3 model to suit the actual situation of the representative objects of Seoul, such as cultural assets and the symbols preferred by tourists.

## 5 Conclusion

In this study we aim to analyze the tourists' images of Seoul by analyzing the photos uploaded on Flickr by Seoul tourists. We were able to find out that tourists have a strong image of palaces, historical monuments, traditional food, and restaurants, etc. These characteristics are distinguished from one RoA to another. The images that tourists feel about Jongro and Namsan are palace and cultural properties while the images of Shinchon, Hongdae, Itaewon, Yeouido, Garosugil, and Apgujeong are food and restaurants. The images of War Memorial of Korea and National Museum of Korea are the monuments that could be photographed on site and the artifacts that were displayed in the museum. Moreover, the images of Samsung Station are a combination of facilities, temples, and cultural properties while the images of Jamsil are toyshops and amusement park.

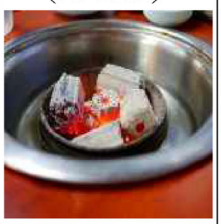
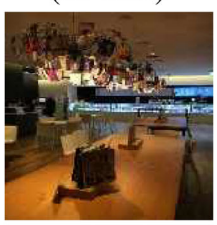
This study is meaningful in three folds: first, it analyzes urban image with the photos posted on the SNS by tourists. Second, it uses deep learning technique to analyze the photos. Third, it classifies and analyzes the whole photos posted by Seoul tourists while most researches focus on only specific objects.

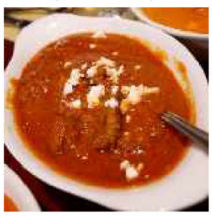
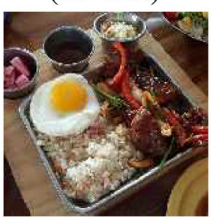
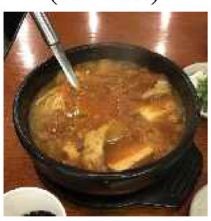
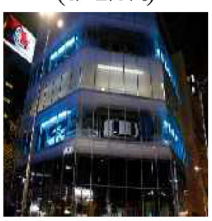
On the other hand, we could find out new research topics that further studies are needed. We recognized that the Inception v3 model that we applied in this study had a limitation because it is a pre-trained model using ImageNet's data set which does not include Korea's characteristics. It was not possible to accurately categorize certain iconic landmarks of Korea such as Namsan Tower, Dongdaemun Design Plaza, and Hanbok that are not widely known in the world. The photos related to palaces and Hanok villages were scattered in the categories such as 'Palace', 'bell cote' and 'terrace'. In future research it is necessary to create traing data set and retrain Inception v3 model based on the photos posted by tourists who visit Seoul and make a category suitable for the purpose of tourism.

**Acknowledgements** This Research was supported by the Technology Advancement Research Program funded by Ministry of Land, Infrastructure and Transport of Korean government (Grant No. 19CTAP-C151886-01).

**Open Access** This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (http://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

## Appendix 1: Secondary categories and representative images per RoA

| RoA (Quantity) | Representative Images (Quantity/Proportion within RoA) | | | | |
|---|---|---|---|---|---|
| Jongro, Namsan (20,987) | palace (1240/ 5.9%) | bell cote (688/ 3.3%) | tile roof (441/ 2.1%) | food market (435/ 2.1%) | patio/terrace (402/ 1.9%) |
| Shinchon, Hongdae (2,584) | plate (133/ 5.1%) | restaurant (68/ 2.6%) | movie theater (66/ 2.6%) | hot pot (66/ 2.6%)  | menu (49/ 1.9%) |
| War Memorial of Korea (1,008) | warplane (66/ 6.5%) | tank (47/ 4.7%) | military uniform (40/ 4.0%) | memorial tablet (33/ 3.3%) | rifle (25/ 2.5%) |
| National Museum of Korea (957) | memorial tablet (54/ 5.6%) | mask (38/ 4.0%) | menu (35/ 3.7%) | vase (32/ 3.3%) | book jacket (22/ 2.3%) |
| Samsun Station, Bongeunsa Station, Coex Mall (872) | stage (26/ 3.0%)  | pedestal (26/ 3.0%) | restaurant (21/ 2.4%) | bell cote (21/ 2.4%) | cinema (18/ 2.1%) |

| | | | | | |
|---|---|---|---|---|---|
| Jamsil (840) | toyshop (58/ 6.9%) | stage (27/ 3.1%) | lakeside (23/ 2.7%) | carousel (18/ 2.1%) | cinema (17/ 2.0%) |
| Itaewon (829) | restaurant (35/ 4.2%) | cinema (24/ 2.9%) | beer glass (21/ 2.5%) | plate (20/ 2.4%) | streetcar (16/ 1.9%) |
| Gangnam Station (744) | taxi (64/ 8.6%) | plate (38/ 5.1%)  | cinema (21/ 2.8%) | street car (18/ 2.4%) | traffic light (18/ 2.4%) |
| Yeouido (428) | plate (21/ 4.9%) | pier (13/ 3.0%) | lakeside (12/ 2.8%) | restaurant (10/ 2.3%) | stage (10/ 2.3%) |
| Garosu-gil (419) | plate (26/ 6.2%)  | restaurant (21/ 5.0%) | toyshop (20/ 4.8%) | hot pot (17/ 4.0%)  | cinema (13/ 3.1%) |
| Apgujeong (306) | plate (23/ 7.5%) | balance beam (8/ 2.6%) | trash can (7/ 2.3%) | restaurant (6/ 2.0%) | toyshop (6/ 2.0%)  |

## References


1. Donaire, J., Camprubi, R., & Gali, N. (2014). Tourism clusters from Flickr travel photography. *Tourism Management Perspectives, 11,* 26–33.
2. Park, J., Yoon, H., Kwon, H., Jeong, W., & Park, J. (2012). Geovisualization of city image: Focusing on the evaluation of representative image components of Seoul. *Seoul Studies, The Seoul Institute, 13*(1), 167–180.
3. Saito, N., Ogawa, T., Asamizu, S., & Haseyama, M. (2018). Tourism category classification on image sharing services through estimation of existence of reliable results. In *Proceedings of the 2018 ACM on international conference on multimedia retrieval* (pp. 493–496). ACM.
4. Chen, M., Arribas-Bel, D., & Singleton, A. (2019). Understanding the dynamics of urban areas of interest through volunteered geographic information. *Journal of Geographical Systems, 21*(1), 89–109.
5. Kádár, B. (2014). Measuring tourist activities in cities using geotagged photography. *Tourism Geographies, 16*(1), 88–104.
6. Sun, Y., Fan, H., Helbich, M., & Zipf, A. (2013). Analyzing human activities through volunteered geographic information: Using Flickr to analyze spatial and temporal pattern of tourist accommodation. In J. Krisp (Ed.), *Progress in location-based services* (pp. 57–69). Lecture notes in geoinformation andcartography. Springer, Berlin.
7. Yuan, Y., & Medel, M. (2016). Characterizing international travel behaviour from geotagged photos: A case study of Flickr. *PLoS ONE, 11*(5), e0154885.
8. Zheng, Y., Zha, Z., & Chua, T. (2012). Mining travel patterns from geotagged photos. *ACM Transactions on Intelligent Systems and Technology, 3*(3), 56–73.
9. Jang, H., & Cho, S. (2016). Automatic tagging for social images using convolution neural networks. *Journal of KIISE. Korean Institute of Information Scientists and Engineers, 43*(1), 47–53.
10. Hong, T., & Shin, J. (2016). Recommendation method of SNS following category classification of image and text information. *Smart Media Journal. Korean Institute of Smart Media, 5*(3), 54–61.
11. Kagaya, H., & Aizawa, K. (2015). Highly accurate food/non-food image classification based on a deep convolutional neural network. In *International conference on image analysis and processing* (pp. 350–357). Springer.
12. Kaneko, T., & Yanai, K. (2013). Visual event mining from geotweet photos. In *2013 IEEE international conference on multimedia and expo workshops (ICMEW)* (pp. 1–6). IEEE.
13. Kisilevich, S., Rohrdantz, C., Maidel, V., & Keim, D. (2013). What do you think about this photo? A novel approach to opinion and sentiment analysis of photo comments. *International Journal of Data Mining, Modelling and Management, 5*(2), 138–157.
14. Okuyama, K., & Yanai, K. (2013). A travel planning system based on travel trajectories extracted from a large number of geotagged photos on the web. In J. Jesse et al. (Eds.), *The era of interactive media* (pp. 657–670). Springer, New York.
15. Hollenstein, L., & Purves, R. (2010). Exploring place through user-generated content: Using Flickr tags to describe city cores. *Journal of Spatial Information Science, 2010*(1), 21–48.
16. 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.
17. Koylu, C., Zhao, C., & Shao, W. (2019). Deep neural networks and kernel density estimation for detecting human activity patterns from geo-tagged images: A case study of birdwatching on Flickr. *ISPRS International Journal of Geo-Information, 8*(1), 45.
18. Chu, W. T., Zheng, X. Y., & Ding, D. S. (2017). Camera as weather sensor: Estimating weather information from single images. *Journal of Visual Communication and Image Representation, 46,* 233–249.
19. Deepak, S., & Chavan, M. (2012). Content-based image retrieval: Review. *International Journal of Emerging Technology and Advanced Engineering, 2*(9), 2250–2459.
20. Géron, A. (2017). *Hands-on machine learning with Scikit-Learn and TensorFlow*. Sebastopol: O'Reilly Media.
21. Ciresan, D., Meier, U., Masci, J., Maria Gambardella, L., & Schmidhuber, J. (2011). Flexible, high performance convolutional neural networks for image classification. *IJCAI Proceedings of the International Joint Conference on Artificial Intelligence, 22*(1), 1237–1242.
22. Raghesh Krishnan, K., Midhila, M., & Sudhakar, R. (2018). Tensor flow based analysis and classification of liver disorders from ultrasonography images. In D. Hemanth & S. Smys (Eds.), *Computational vision and bio inspired computing* (Vol. 28, pp. 734–743). Lecture notes in computational vision and biomechanics. Springer, Cham.
23. Sawant, A., Bhandari, M., Yadav, R., Yele, R., & Bendale, S. (2018). Brain cancer detection from MRI: A machine learning approach (TENSORFLOW). *International Research Journal of Engineering and Technology, 5*(4), 2089–2094.
24. Motlagh, N., Jannesary, M., Aboulkheyr, H., Khosravi, P., Elemento, O., Totonchi, M., et al. (2018). Breast cancer histopathological image classification: A deep learning approach. bioRxiv. https://doi.org/10.1101/242818.
25. Park, S., & Shim, H. (2017). Movie poster classification into genres via convolutional neural network. *Conference proceedings of the Korean Institute ofInformation Scientists and Engineers*. KIISE, June, pp. 890–892.
26. Lee, T., & Lee, I. (2017). Animation character recognition using deep learning. *Conference proceedings of Korea Computer Graphics Society*. KCGS, pp. 37–38.
27. Xu, G., Zhu, X., Fu, D., Dong, J., & Xiao, X. (2017). Automatic land cover classification of geo-tagged field photos by deep learning. *Environmental Modelling and Software, 91,* 127–134.
28. Arroyo, R., Alcantarilla, P. F., Bergasa, L. M., & Romera, E. (2016). Fusion and binarization of CNN features for robust topological localization across seasons. In *2016 IEEE/RSJ international conference on intelligent robots and systems (IROS)* (pp. 4656–4663). IEEE.
29. Kim, N., Kang, Y., Lee, J., Kim, D., & Park, Y. (2019). Tourists hot spot analysis in Seoul using geotagged photos. *Seoul Studies, 20*(1), 81–96.
30. Ester, M., Kriegel, H., Sander, J., & Xu, X. (1996). A density-based algorithm for discovering clusters in large spatial databases with noise. In E. Simoudis, J. Han & U. M. Fayyad (Eds.), *Proceedings of the 2nd international conference on knowledge discovery and data mining (KDD-96)* (pp. 226–231). AAAI Press. CiteSeerX 10.1.1.121.9220. ISBN 1-57735-004-9.
31. Ministry of Culture, Sports, and Tourism. (2017). *The survey of the current state of foreign tourists*. Korea Tourism Organization.
32. http://image-net.org/.


**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.