e-Review of Tourism Research (eRTR), Vol. 17, No. 2, 2019
http://ertr.tamu.edu

Kun Zhang, Qinghui Li, Hongyu Wang, Dongzhi Chen
Nankai University

Exploring the Presence of Tourists' Photos Through Algorithmic Visual Content Analysis

Big pictorial data is a significant data resource for discovering tourists' behaviors and perceptions. Innovatively, this study adopted two deep learning models, namely scene recognition and semantic segmentation for uncovering the presence of tourist photos in a tourism destination. In all, 36497 photos shared by oversea tourists in Beijing were screened out by data mining and taken for visual content analysis. By developing two types of categories, the perceived destination attractions by tourists were analyzed. Theoretically, this study contributes to the establishment of a smart approach for understanding tourist preference through big pictorial data.

Keywords: big pictorial data, deep learning model, visual content analysis, tourism destination, user-generated photography

Kun Zhang
College of Tourism and Service Management
Nankai University
No.38 Tongyan Road, Haihe Education Park, Jinnan District
Tianjin, 300350
People's Republic of China
Email: zxzhangkun@163.com

Qinghui Li
College of Tourism and Service Management
Nankai University
No.38 Tongyan Road, Haihe Education Park, Jinnan District
Tianjin, 300350
People's Republic of China
Email: 474704507@qq.com

Hongyu Wang
College of Tourism and Service Management
Nankai University
No.38 Tongyan Road, Haihe Education Park, Jinnan District
Tianjin, 300350
People's Republic of China
Email: 798742526@qq.com

Dongzhi Chen
College of Tourism and Service Management
Nankai University
No.38 Tongyan Road, Haihe Education Park, Jinnan District
Tianjin, 300350
People's Republic of China

Email: chendongzhi@mail.nankai.edu.cn

Dr. Kun Zhang is a lecturer in the College of Tourism and Service Management at Nankai University. Her research interests are tourism planning, with the focus on tourism destination image and AI technology, tourists' demands, and sustainable tourism planning.

Qinhui Li is the Ph.D student of the college of Tourism and Service Management at Nankai University. Qinghui's research interests include international tourism cooperation and sharing economy.

Hongyu (Prudent) Wang is a Postdoctoral candidate in the College of Tourism and Service Management at the NanKai University. Her research interests include touristic learning behaviour, educational tourism development and innovation of tourism education.

Dongzhi Chen is the Ph.D candidate of the College of Tourism Service and Management at Nankai University. Her research focuses on tourism policy and destination image.

Introduction

Big pictorial data has become a significant data resource for understanding tourists' behaviours and perceptions (Kim et al., 2014; Li et al., 2018; Xiang et al., 2015). However, natural intelligence is disabled to handle and analyze such a large number of the photos, a more effective and smart approach for interpreting the user-generated photographs is highly required. With the development of computer vision technology, a number of robust deep learning models enable to analyse the visual content of the photos in several different dimensions (Baró et al., 2009; Pantic et al., 2007; You et al., 2015), which provides an essential technical support for the study of tourists' behaviour in destination. In this study, intending to explore a smart way for understanding tourist behaviours through big pictorial data, it adopted two deep learning models, namely scene recognition and semantic segmentation for uncovering the presence of tourist photos in a tourism destination.

Literature Review

The photo is a pictorial representation of the tourism destination image (Hunter, 2016), and they well represent tourists' perception, preference, and choices (Pan et al., 2014; Henderson et al., 2010; Choi et al., 2007). Referring to the theory of tourist gaze (Urry, 1990), a lot of research and discussion have been carried out. In 2003, a hermeneutic circle of representation was described by Jenkins as it passes from the tourist to the media to the potential consumer, thence to the destination, and finally back to the tourist (Jenkins, 2003). Recently, such a circle of representation was revised by Balomenou and Garrod, and the photographs took by tourists are viewed as the icons which contribute to the projected image (Balomenou & Garrod, 2019). In this sense, the exploration for the presence of tourists' photos holds a significant meaning to the holistic promotion of a destination.

Currently, the visual content analysis of the phots is attribute-based and make the main focal themes in the pictures to be identified (Stepchenkova & Zhan, 2013). In which, the categories are the crucial framework for classifying and detecting the photo's presence (Bell, 2001). Several previous studies provide various ways of assorting the attributes of photos (Valek & Williams, 2018; Ku & Mak, 2017; Mak, 2017). By synthesizing these studies, the significant attributes may include nature landscape, architecture, people, tourism facilities, urban landscape, tourism activities, food, transport/infrastructure, etc. (Stepchenkova & Zhan, 2013; Zhang et al., 2019). With a concerning to the social interaction and human activity, Hunter (2008) summarized the visual representation of photos into two subclasses, which are "tourism representations by space" and "tourism representations by subject". The former subclass is divided as natural landscapes, cultivated landscapes, heritage and material culture, tourism products, and the latter one is composed of no human subject, tourist, host, tourist and host (Hunter, 2008). Nikjoo & Bakhshi (2019) have analyzed tourists' photos with the categories of the tourists, the hosts, the tourists with hosts, and no human presence. Although most of the previous studies are conducted through a way of manual analysis, their classifying methods of visual content analysis inspire a lot for this study.

Methodology

The research process mainly includes four parts, namely data mining and screening, visual content analysis by deep learning models, the design of categories, and the findings of statistical analysis. The part of data mining and screening was delivered in the next section.

In the part of visual content analysis, two deep learning models, namely scene recognition and semantic segmentation, are adopted. We employ widely spread ResNet-101 for the scene classification task (He et al., 2015). A demonstrated flow process is shown in figure 1. As the output, 102 scenes are distinguished. And for the task of semantic segmentation, we used the DeepLab model, which is one of the state-of-the-art methods for semantic segmentation (Chen et al., 2017). Some samples of the output are shown in figure 2. As the output, eight main semantic elements are detected, and all the other elements could not be figured out are defined as the background, and the areas of the eight elements are calculated.

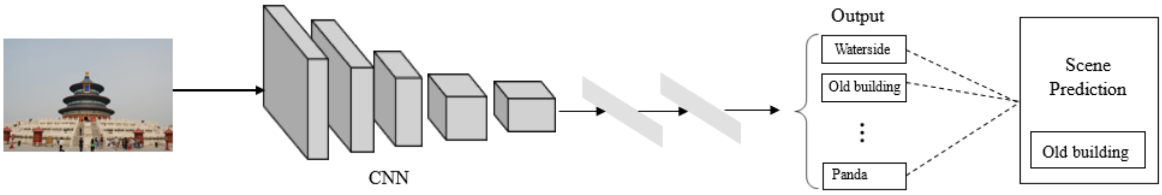


Figure 1. The flow of the deep learning model for scene classification

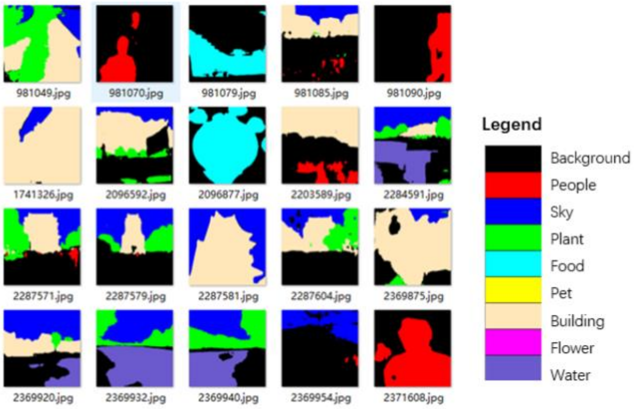


Figure 2. Samples of output from the deep learning model of semantic segmentation

For the design of categories, two frameworks were developed according to the two models. Firstly, a framework with two levels of categories for assorting the photos' 102 scenes was established (Table 1), which was referring to the typologies mentioned in the literature review (Hunter, 2008; Valek & Williams, 2018; Ku & Mak, 2017; Mak, 2017). At the same time, by concerning to the percentage of the people's area in the photo, the other categories were created, which included "the proportion of people is more than 10%", "there are no people" and "the proportion of people is less than 10%". The decision for such a partition is referring to the position of people in the photo. By checking four types of photos, which are "the proportion of people is more than 1%, 3%, 5%, and 10%", it is concluded that only when the proportion of people are more than 10%, people are the central subjects in the photo, and when the proportion of people are less than 10%, the role of the people is relatively blurred. In the part of statistical analysis, the results from these two models were delivered and synthesized.

Table 1. Categories for visual content analysis

3 Categories	11 Categories	Scenes
Urban Landscape	Building and urban space	bridge, cathedral hall, corridor, European buildings, Islam buildings, overlook, skyscraper, worksite
	Transportation	aircraft, bicycle, cabin, car, helicopter, in car, motorcycle, ship, station, train
Natural landscape	Meteorological phenomena	blue sky, night, overcast, snow, sunset
	Water and mountain	beach, mountain, waterfall, waterside, water surface
	Plant and living beings	bee, butterfly, camel, camera, cat, deer, dinosaur, dog, dragonfly, elephant, fallen, flower, giraffe, green plant, kangaroo, ladybug, leopard, lion, ornamental fish, panda, peacock, penguin, rabbit, rhinoceros, tiger, tortoise
Society and culture	Cultural activity and symbol	dragon dance, fireworks, Fu character, lion dance, red envelope, stage, xi character
	Description and illustration	Map, text
	Entertainment	badminton court, bar, baseball court, billiard room, bowling alley, chess, football court, go, indoor basketball court, mah-jong, Ping-Pong court, playground, swimming pool, tennis court
	Indoor room	air conditioner, bedroom, classroom, dining room, kitchen, library, living room, meeting room, office, washing machine, washroom
	Food and eating place	food, McDonald's, restaurant
	shopping place	mall courtyard, supermarket
	Others	Glasses, high heels, keyboard, little pony, teddy bear, the smurfs, transformer, watch

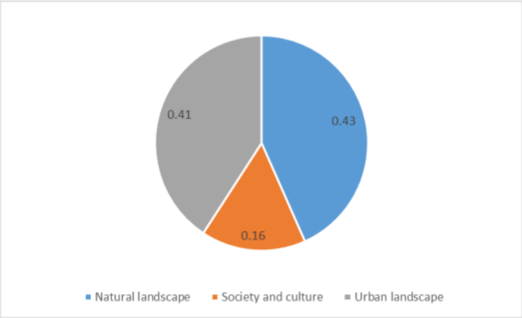
Data Collection and Sample

The dataset of “Yahoo Flickr Creative Commons 100M” (YFCC 100M) is the data resource, which contains 99.2 million photos uploaded by users during the time from 2004 until early 2014. With the help of ArcGIS, all the photos shoot in Beijing were found out according to Beijing administrative boundary. Moreover, by invoking API (Application Programming Interface) data in Flickr, 36497 photos shared only by oversea tourists were taken as the objects for visual content analysis in this study. According to the information of the users' home location, all the photos were uploaded by 1075 tourists from 64 countries/regions and six continents (Asia, Europe, North America, Oceania, South America, and Africa).

Findings

Scene understanding

According to the first level of categories for the photos' scenes, tourists' perception about the natural landscape and urban perception are comparable, both of them overwhelmingly exceeds the perception of the society and culture, which only accounts for 16% of the total (figure 3). Referring to the distributions of 11 categories (figure 4), building, and urban space is the most famous attraction for tourists to perceive, which is followed by the meteorological phenomena, plant and animal, and water and mountain. The differences between the number of food and eating place, description and illustration, cultural activities or symbol, and the transportation perception are not noticeable. The perception of the entertainment is relatively weak, and the shopping place attracts little tourists comparing to the other attractiveness.



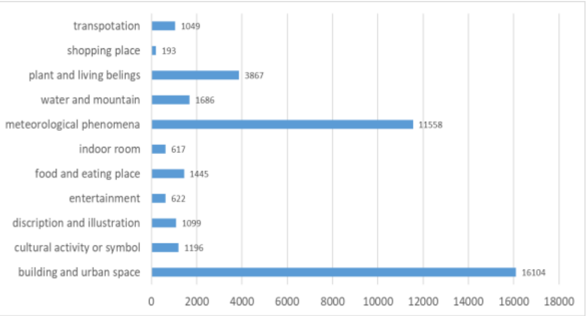
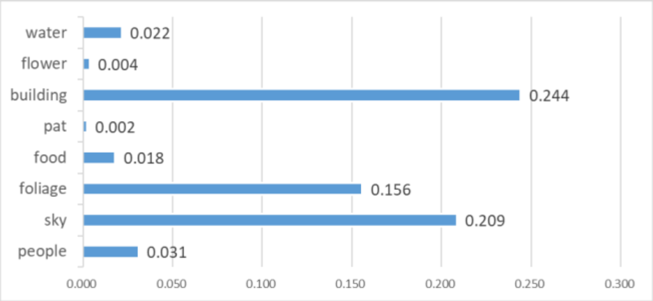


Figure 3. Result of the scene classification according to 3 categories(left)
Figure 4. Result of the scene classification according to 11 categories (right)

Semantic Segmentation

By calculating the average of each semantic element's percentage (figure 5), the statistical results show that the building and sky are the most significant semantic elements, which are followed by the plant. Specifically, the average of people' percentage is 3.1%, and the average of water’s percentage and food's percentage are 2.2% and 1.8%. According to the proportion of people in the photo (figure 6), all the photos were divided into three types. The photos of "no people appeared" account for 61% of the total, which meant tourists prefer a "pure" landscape in the photo. Only 11% of the photos are people dominated while the other 28% of the photos are blurred about the dominated subjects.



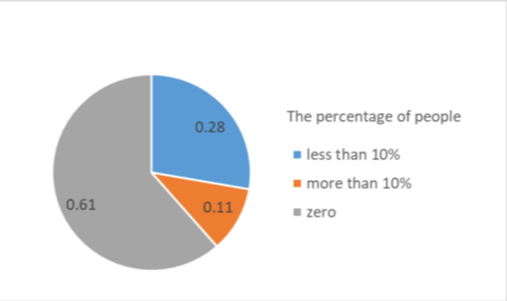


Figure 5. Result of semantic segmentation by eight elements (left)
Figure 6. Result of semantic segmentation according to the percentage of people (right)

Synthesized result

Figure 7. The synthesized results about the presence of tourists' photos

A cross-analysis between the results of the scene recognition and semantic segmentation reveals the following characteristics of tourists' photos (Figure 7). In the scenes of the entertainment and cultural activities or symbol, the photos that taken people as the subject hold the most significant number, which is more or less 30%, which is followed by the scenes of the transportation, indoor room and shopping place. About 10% of the photographs in the scene of the building and urban space could find people. Besides, in every fifteen food-related photos, there is one photo in which people is the subject. In the scenes of the other categories, such as the plant and living beings, mountain and water, the number of photos taking people as the subject is less than 10% of the total. As a summary, people appeared more frequently in the scenes of society and culture than the scenes of urban landscape and natural landscape.

Conclusion

User-generated photography is compelling evidence for exploring tourists' perception of a tourism destination. The result of scene recognition in this study shows that tourists' experiences about the natural and urban landscape are more abundant than the social and cultural aspects. Among them, building and urban space are the most significant attractions to tourists, which is followed by the plant and animal, water, and mountain. While in the social and cultural aspect, food and activities are the most perceived objects by tourists. The results of semantic segmentation show that people are found in 39% of the visitors' photos, and 11% of them are taking people as the subjects. Specifically, people appear more frequently in the scene of the entertainment and cultural activities, and about 10% of the photographs in the scene of the building and urban space are taking people as the subjects.

For the implementation, the above conclusions provide some clues and references for the tourism development and management of Beijing. For example, most of the tourists’ activities in Beijing are sightseeing for the natural and urban landscape. There is a great potential to promote the development of tourism attractions in the cultural and social aspects and furthermore build up a pluralistic system of tourist attractions. Besides, there are more human interactions in the scenes of entertainment, cultural activities, and food. The destination management organizations could consider to enhance tourists' cultural experience and well disseminate the culture exchange in such scenes in Beijing.

There are two main theoretical contributions to this study. Firstly, this study tested the possibility of a new smart way for interpreting the visual content of user-generated photography with two deep learning models-scene recognition and semantic segmentation. It proofed that the new approach has significant advantages of saving time and energy in processing big pictorial data. Secondly, this study explored the applicability of the previous typology theory for photo-based study in the new machine learning approach. According to the performance of the output, in one hand, the deep learning models show weakness in distinguishing the interaction between the tourists and hosts, in another hand, the new classification framework designed based on the outputs provides a basis for a further research of employing deep learning model for the photos' analysis.

Limits and future work

Compared to the conventional way of reading the representation of the photos, the deep learning approach saves a lot of time and energy in analysing the visual content of massive photos. However, the process of designing and running of the model is still time and energy-consuming, various problems and bugs may happen. Taking this study as an example, it took nearly half-year for data processing.

Under the premise of new technology, how these image annotation tools could be well adapted to answer scientific or practical tourism questions is a crucial consideration in the future. In our parallel research, we have already explored several issues. For example, the perception differences of tourists from different places (Zhang et al., 2019), the tourism destination image reflected from tourists' photos of other city different from Beijing (Zhang et al., 2019), and the chances of tourism destination image referring to the information of photos' shooting time (in preparing), etc. With the massive emerging big data and the updated computer vision technology in the future, the possibility of theoretical and practical exploration in the field of tourism is far beyond what have been explored.

Acknowledgment

Thanks to the National Natural Science Foundation of China (No. 51608278) for supporting this research. We are also grateful to Dr. Xin Guo. He has contributed a lot to the establishment of these two deep learning models in this study.

References

Baró, X., Escalera, S., Radeva, P., & Vitrià, J. (2009). Visual content layer for scalable object recognition in urban image databases. Paper presented at the IEEE International Conference on Multimedia & Expo.

Bell, E. (2001). Content analysis of visual images. In C. J. T. Van Leeuwen (Ed.), Handbook of visual analysis (pp. 92-118). London: Sage.

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

Chen, L. C., Papandreou, G., Schroff, F., & Adam, H. (2017). Rethinking Atrous Convolution for Semantic Image Segmentation.

Choi, S., Lehto, X. Y., & Morrison, A. M. (2007). Destination image representation on the web: Content analysis of Macau travel related websites. Tourism Management, 28(1), 118-129.

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

Henderson, J. (2010). Event-Based Analysis of People's Activities and Behavior Using Flickr and Panoramio Geotagged Photo Collections. Paper presented at the Information Visualisation.

Hunter, W. C. (2008). A typology of photographic representations for tourism: Depictions of groomed spaces. Tourism Management, 29(2), 354-365.

Hunter, W. C. (2016). The social construction of tourism online destination image: A comparative semiotic analysis of the visual representation of Seoul. Tourism Management, 54, 221-229.

Jenkins, O. (2003). Photography and travel brochures: The circle of representation. Tourism Geographies, 5(3), 305-328.

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.

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

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

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

Pantic, M., Pentland, A., Nijholt, A., & S. Huang, T. (2007). Human Computing and Machine Understanding of Human Behavior: A Survey, ICMI 2006 and IJCAI 2007 international conference on artifical intelligence for human computing (pp. 47-71). Berlin, Heidelberg: Springer-Verlag.

Stepchenkova, S., & Zhan, F. Z. (2013). Visual destination images of Peru: comparative content analysis of DMO and user-generated photography. Tourism Management, 36(3), 590-601.

Urry, J. (1990). The Tourist Gaze: Leisure and Travel in Contemporary Societies. London: Sage Publications, Thousand Oaks.

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.

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.

You, Q., Luo, J., Jin, H., & Yang, J. (2015). Robust Image Sentiment Analysis Using Progressively Trained and Domain Transferred Deep Networks.

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, K., Chen, D., & Li, C. (2019). How are Tourists Different - Reading Geo-tagged Photos through a Deep Learning Model. Journal of Quality Assurance in Hospitality & Tourism, 1-10.