

# Classifying Tourists' Photos and Exploring Tourism Destination Image Using a Deep Learning Model

Jiyeon Kim, Youngok Kang \*, Nahye Cho, Soyeon Park

Ewha Womans University, pbmejy@gmail.com ; ykang@ewha.ac.kr; cho.nahye@gmail.com; soy8147@gmail.com;

\* Corresponding author

**Keywords:** Convolutional Neural Network(CNN), Tourism Destination Image(TDI), geo-tagged photo, transfer learning, Inception v3 model, spatial analysis of tourism image

## Abstract:

Recently, people tend to express their own activities, opinions, etc. through SNS (Social Network Service) and exchange them with others. In particular, in the tourism field, people not only post text or photo in tourism destination during or after travel, but also get tourism information through SNS. These SNS data are evaluated as representing the key tourism elements and the image that tourists feel about in tourism destination. As competition in the global tourism market intensifies, the TDI (Tourism Destination Image) has become a crucial concept. Previously, projected image generated by a DMO (Destination Management Organization) has led TDI, but recently, UGC (User Generated Contents) generated by tourists is evaluated to play a major role in generating TDI (Gilbert & Hancock, 2006). Among SNS data, LBSN (Location Based Social Network) data has been widely utilized for various analyses in the tourism field. In particular, Flickr has become a valuable data for various studies because it is not only a representative SNS for sharing photos, but also includes location and time information in photo metadata. Using Flickr data, studies such as analysis of ROA(Region Of Attraction), analysis of tourists movement pattern, and analysis of city images and emotions have been conducted. However, these existing studies using Flickr data have limitations in that they only use metadata and textual data embedded in photos. Recently, as computer vision and image processing technologies have rapidly advanced in the AI (Artificial Intelligence) field, techniques for analyzing visual contents in photo are also increasingly evolving. In particular, as pre-trained models based on deep learning techniques have been released, the application of these models to other areas through transfer learning continues to surge. In the tourism field, a few studies are also being initiated to classify tourist photos based on pre-trained models and analyze TDI. However, when pre-trained models are applied in the specific region, they do not properly classify the photos by reflecting the unique scenery or characteristics of tourism destinations (Kim et al., 2020). With the purpose of overcoming these limitations of previous researches, this study aims to apply computer vision and image processing techniques to analyze the TDI with a large volume of Flickr data. The objectives of this study are three folds; (1) develop tourists' photo classification by analyzing the characteristics of photos posted by tourists, (2) construct a deep learning model to accurately classify photos according to the photo classification developed, (3) analyze the TDI for Korea by combining visual contents with location and user information of photos.

In this study, we first crawled 168,216 Flickr photos posted by tourists within the range of Korea from 2013 to 2018. Then we generated a tourists' photo classification by analyzing the characteristics of photos posted by tourists and referring to the tourism classification of the Tourism Organization. Through this photo classification refining process, the tourists' photos are classified into 75 scenes, and 'difficult to classify' is a scene for photo refining. And 74 scenes are grouped into 12 categories to facilitate future interpretation. In this category, Korea specific elements such as unique Korean palaces, street food, traditional markets, hanbok experience, traditional performances, etc. are included. And then we developed a deep learning model by continuously retraining the Inception-v3 model, which is one of well-known pre-trained CNN architectures. CNN is one of deep neural networks, which is the essential technology leading the state-of-the-art in computer vision for a variety of tasks. Although a number of models have been released so far, Inception-V3 has been still one of the most accurate models in its field for image classification, achieving top-5 error accuracy of 3.58% and top-1 error accuracy of 17.3%. We decided to use the Inception-V3 model for photo classification through transfer learning. In our work, the original network architecture of Inception-V3 is maintained and the pre-trained weights by ImageNet are used to initialize the network. The model was modified so that it could classify photos into 75 scenes in the last softmax layer (Figure 1). The final model shows high accuracy of 85.77% for the top 1 and 95.69% for the top 5 when it is applied to the test data. Finally, the final model has applied the entire dataset to analyze TDI for Korea.

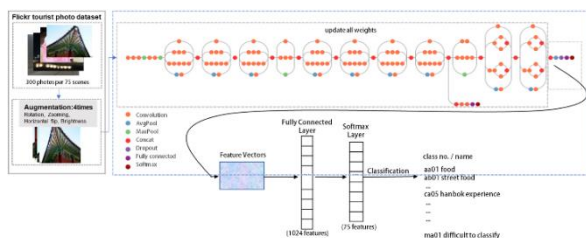


Figure 1. Architecture of fine-tuned Inception-V3 model

Through the analysis of 168,216 Flickr photos, we analyzed the tourism destination image of Seoul. Seoul is a representative attraction in Korea, where 2,264 tourists (69.5%) of 3,259 tourists visited and 80,553 photos (47.9% of the 168,216 photos) were posted. The frequency of photos posted in Seoul by scene and by category are as follows: 'selfies & people', 'food', 'palace', 'conference', and 'building' by scene and 'Urban scenery', 'Korean traditional architecture', 'Food & Beverage', 'Shopping', and 'Activities' by category. When comparing photo frequencies in Seoul with those in Korea, it showed that 'seasonal landscape' decreased while 'conference' increased by scene, and 'Natural Landscape' reduced remarkably while 'Urban Scenery' and 'Korean Traditional Architecture' increased a little by category. Figure 2 shows a map constructed with convex hulls by connecting the outermost points of the points composing the cluster. As a result of cluster analysis, a total of 12 ROAs were derived and attractions by each ROA are investigated. Figure 2 shows the correlation between 12 ROAs and photo categories as a matrix heatmap. There were many photos in the urban scenery category in most regions. However, it was confirmed that the attractive factors of each ROA differ as follows: In Jongno ROA, the proportion of 'Korean traditional architecture' photo is high; In Yongsan, 'cultural & relics' photo; In Lotte World, 'Activities' photo; In Hongdae, 'food & beverage' and 'shopping' photo; In Noryangjin, 'shopping' photo. As well as we tried to derive the type of tourist with photos posted by tourists and found that tourist preference differ by continent (Figure 3).

This study contributes in both technical and practical aspects. In terms of technical aspect, this study is significant in that it developed a tourist photo classification suitable for local characteristics and showed the process of retraining a deep learning model to effectively classify tourism photos. In addition, the result analysis of ROA and TDI in tourist destination can provide an idea of what images DMO should offer to tourists and how to promote the tourist destination by understanding a microscopic characteristics of regions.

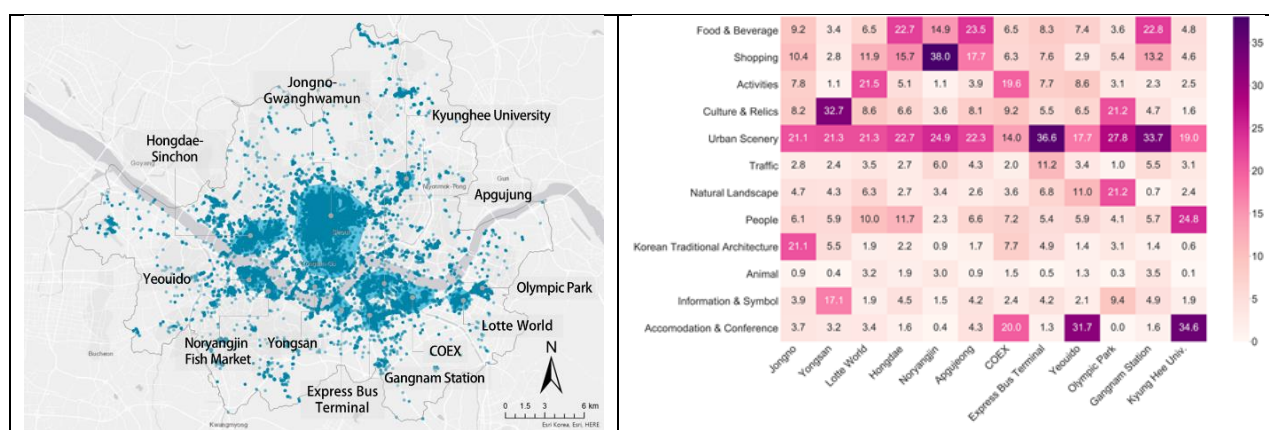


Figure 2. ROAs in Seoul(left) and Matrix heatmap between ROA and photo category(right)

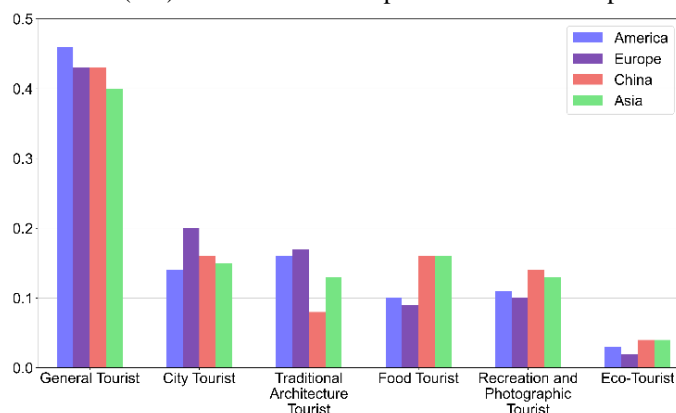


Figure 3. Percentage of tourist type by continent

**Funding Information :** This research was supported by the Technology Advancement Research Program funded by the Ministry of Land, Infrastructure and Transport of the Korean government (Grant No. 20CTAP-C151886-02).

## References

- Gilbert, D., & Hancock, C. (2006). New York City and the transatlantic imagination: French and English tourism and the spectacle of the modern metropolis, 1893-1939. *Journal of Urban History*, 33(1), 77-107. <https://doi.org/10.1177/0096144206290385>.
- Kim, D., Kang, Y., Park, Y., Kim, N., & Lee, J. (2020). Understanding tourists' urban images with geotagged photos using convolutional neural networks. *Spatial Information Research*, 28(2), 241-255. <https://doi.org/10.1007/s41324-019-00285-x>