CLASSIFYING SIGHTSEEING PHOTOS OF THE HIGASHIYAMA AREA OF KYOTO CITY USING SEMANTIC SEGMENTATION

Taichi Kawakatsu*[1], Yusuke Kimura[2], Junichi Susaki[1], Tetsuharu Oba[1]

[1] Kyoto University

[2] Osaka Institute of Technology

Tourism industry is important to local economies. In order to realize a sustainable tourism that allows tourist to engage in highly satisfying tourism while alleviating excessive congestion in tourist areas, it is necessary to introduce tourism policies that meet the interests of each tourist. In order to understand the interests, data on social media is used instead of conventional questionnaire surveys. In particular, geotagged photos on a photo-sharing service, a photo with geolocation information as a metadata on one of the social media, record subjects of each photographer's interest.

A previous research applied an image recognition to geotagged photos in the Higashiyama area, which is a representative tourist area of Kyoto City, and classified these photos based on the obtained label information on subjects of photos. It brings 1,597 types of labels and their reliability, but it is impossible to obtain information on the percentage of subjects in each photo. Since there is no weight information on such subjects, the result of this classification has a problem that the number of photos whose features cannot be determined is as high as 20% of all photos.

In this research, we focus on a semantic segmentation that can classify each pixel of photos into a class specified in advance. This method enables us to calculate the proportion of each subject in photos and to use it for classification. We aim to clarify the applicability the semantic segmentation to the analysis of sightseeing photos. Targeting the Higashiyama area of Kyoto City, we conduct the followings: 1) We apply two types of models to photos from the photo-sharing service Flickr, the same as those obtained in previous research, and clarified the characteristics of each model and the model suitable for the analysis of sightseeing photos. 2) By clarifying the relationship between the results of the semantic segmentation analysis and that of the existing image recognition analysis, we reclassify photos that were not properly classified by existing classifications in the previous research.

As for 1), we compared two models: pretrained on the CITYSCAPES dataset and the ADE20K dataset. The CITYSCAPES model correctly recognized street photos, but it failed to correctly classify photos such as looking up at the sky, nighttime photos, and indoor photos. The ADE20K model has a large number of classes, thus there were many misclassified classes. However, its effect was small because the ratio of each class in each photo was small. Therefore, the ADE20K model was suitable for segmenting sightseeing photos of the Higashiyama area.

As for 2), we examined the composition ratio of the top class of the ADE20K model for each classification groups by image recognition. The result showed that the "building", "person", "water", "house", and "sculpture" classes can be used to classify photos from their composition ratio. Moreover, based on this feature, we reclassified groups of photos that were not appropriately classified by existing classifications, and were able to reduce the number of photos that belonged to them.

Keywords: semantic segmentation, ADE20K, image recognition, classification of photograph