Original Article

An image search application to streamline the preparation of open image data using deep learning

Masato Kinugasa (Graduate School of Informatics, Nagoya University, kinugasa.masato.t5@s.mail.nagoya-u.ac.jp, Japan)
Ryo Hori (Graduate School of Informatics, Nagoya University, hori.ryo.1001@gmail.com, Japan)
Yunhao Tu (Department of Computer Science, Chubu University, yunhaotu@isc.chubu.ac.jp, Japan)
Mayu Urata (Graduate School of Informatics, Nagoya University, mayu@i.nagoya-u.ac.jp, Japan)
Mamoru Endo (Graduate School of Informatics, Nagoya University, endo@i.nagoya-u.ac.jp, Japan)
Takami Yasuda (Graduate School of Informatics, Nagoya University, yasuda@i.nagoya-u.ac.jp, Japan)

Abstract

When local governments engage in tourism promotion, they can benefit from increasing open image data. However, they face difficulties in searching for appropriate images from a large number of directories and images. This study aims to solve this issue by developing and introducing an application that facilitates the search of potential candidates for open image data. First, to display images in order of attractiveness, we developed a model to assess the attractiveness of images and assigned attractiveness scores to all images of a local government. Second, to support keyword searches with common nouns, descriptions were added to them using an existing generative model. Third, to display a list of similar images, we grouped them based on the cosine similarity of their vector representations. Then, an application was developed that could reference these metadata. This application was introduced into the actual workflows in Hida City. where it received high ratings in a user survey. This application enables public officers to search for attractive images more efficiently and comprehensively. Meanwhile, through this introductory experiment, we have found that some local governments are interested in releasing thousands of open image data. Therefore, it is considered essential to develop an integrated method.

Keywords

deep learning, image classification, image search, open data, local government

1. Introduction

1.1 Background

In recent years, the Cabinet Office and the Cabinet Secretariat have positioned the utilization of open data as the "first step towards regional/municipal DX (Digital Transformation) promotion" [Cabinet Office, 2021]. In the tourism sector, the publication of open data is also expected to enhance public-private collaboration to promoting tourism. For instance, Yahoo Japan Corporation has utilized datasets such as cultural property lists, tourist facility lists, and event lists from open data for the development of a tourist guide map application [Digital Agency, 2023].

Among the various types of open data, open image data holds significant potential. Visually appealing images have the power to capture consumer interest instantly and intuitively. However, such images are frequently subject to intellectual property rights, involving costs related to identifying copyright holders and paying associated fees, which creates a significant barrier to utilization. Therefore, if a local government offers a rich repository of open image data, stakeholders can conduct effective tourism promotion focused on that area more easily, bringing substantial benefits to the local government. Yahoo Japan Corporation has also commented to the Digital Agency that "displaying images is highly effective, and it would be highly advantageous if images could be made available in a more user-friendly manner, including copyright clearance" [Digital Agency, 2023].

When public officers publish open image data, it is necessary to search for appropriate images across numerous directories as a preliminary step. However, this workload, which must be completed whenever tourism content is updated, poses a burden to offices that are facing labor shortages.

To address this issue, we developed an application that allows public officers to easily search for attractive images from a large number of directories.

1.2 Related work

Several studies have examined effective strategies for local governments to utilize visually appealing images in tourism promotion. Hayashi [2020] analyzed tourists' photographs, uncovering that tourists' points of interest vary depending on their nationality and the frequency of their visits to the destination. Similarly, Chen et al. [2023] developed a recommendation system that personalized its results based on tourists' photographs. Furthermore, previous work by Tu et al. [2023] has included methods that support the release of open image data, such as face detection, mosaic processing of faces, and tagging images.

Nevertheless, little research has focused on the practical challenges involved in searching for appropriate images that public officers face when attempting to implement such strategies.

2. Current status and challenges in Hida City

Hida City in Gifu Prefecture, shown in Figure 1, has been focusing on enhancing its economy through the tourism industry. In fact, 812,223 people visited the city's tourist destinations in 2022 [Gifu Prefecture, 2023].

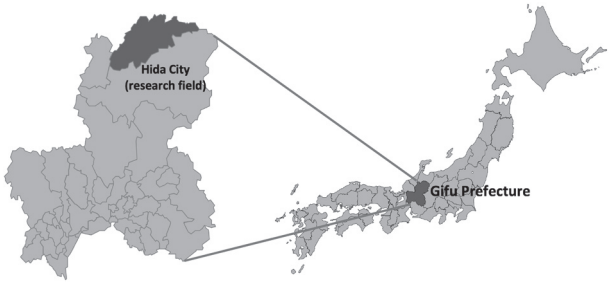


Figure 1: Geographical location of Hida City

The staff of this city have been actively utilizing images for tourism promotion. For instance, when creating the official tourism promotion website "Hida no Tabi," they use various images to share the ambiance of tourist attractions with the viewers [Hida City, n.d.]. As part of these initiatives, they are also highly interested in increasing the amount of open image data.

However, in increasing open image data, they face a challenge: the directories used to store publicity images, which are potential candidates for open image data, have become increasingly unmanageable. The total number of image files has reached 118,855, distributed across 4,826 different directories. These directories have formed a complex hierarchical structure, and images of the same tourist attractions are spread across dozens of different directories.

The challenges that Hida City faces are not limited to these issues; there is also the challenge of time constraints. Currently, these publicity images are stored on an external local storage device. However, the content of the stored images is not being managed organizationally. This situation is now identified as a problem from the perspective of information security. It has been decided that this local storage will be unused in the near future. The internal server that will replace the local storage has severely limited capacity, which means that most of the publicity images that are not released as open image data will be discarded.

This situation causes two issues. First, Hida City staff have to select images to publish as open image data from approximately 120,000 publicity images, but this task requires an unrealistic workload. Second, in practice, they have no choice but to select open image data candidates from only a few hundred images stored in folders they frequently use. As a result, many highly attractive images remain overlooked.

In response to this, the application developed in this study incorporated the following features. Firstly, to reduce work time, attractive images are displayed preferentially. Secondly, to make all images easily searchable, keyword search checks not only file paths of images but also descriptions of the image generated by an existing AI model. Thirdly, similar images are listed with the results to streamline the search process through to the final selection of the most appropriate image.

3. Development of search application functions

3.1 Interviews with Hida City staff

Prior to the development of the application, interviews were conducted with Hida City office staff about their current tasks related to searching for publicity images. The purpose was to define the application requirements and to ensure that the application could be integrated into their existing workflows without causing burden.

These interviews revealed the following characteristics of their current tasks:

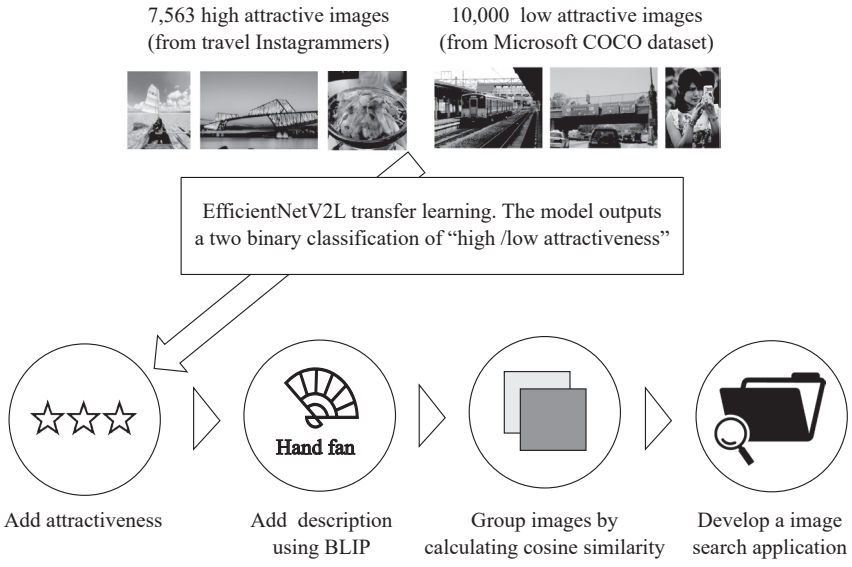


Figure 2: Process flow of system

- Multiple staff might work on the task with their own PCs and some PCs were not connected to the Internet.
- Images in certain directories could not be published as open image data, so they wanted to check the file path on the search results screen, the name of the directories where each image is stored.

Consequently, we designed the application as an executable file (EXE), allowing operation on any PC connected to an external SSD storage of the publicity images. Furthermore, it was decided that the search result screen would show the file path for each image.

3.2 Development of attractiveness assessing model

The development procedure for each function to be introduced in the application is shown in Figure 2.

First, we developed a binary classification model capable of assessing the attractiveness of an image by transfer learning on an image classification model, EfficientNetV2L [Tan and Le, 2021].

As highly attractive training data, 7,563 images were collected from five travel-related Instagram influencers' accounts using the Instagram Graph API [Instagram n.d.a; n.d.b; n.d.c; n.d.d; n.d.e], to achieve a research objective of identifying images deemed attractive for posting on social media platforms and websites.

As less attractive training data, 10,000 images were randomly selected from the publicly available COCO (Common Objects in Context) dataset provided by Microsoft [Lin et al., 2014]. This dataset comprises general images that do not intentionally emphasize specific attractive elements. Therefore, it serves as a suitable contrast to the Instagram images, which are taken with the intent of attracting user interest. Examples of training data for each label are shown in Figures 3 and 4.

This model achieved an accuracy of 99.3 % on a validation dataset, randomly sampled from the training data without being used for training. Therefore, as a final evaluation, the model was tested on a dataset of 300 images created from images of Hida City. As a result, the model achieved an accuracy of 80.7 % on this test dataset.

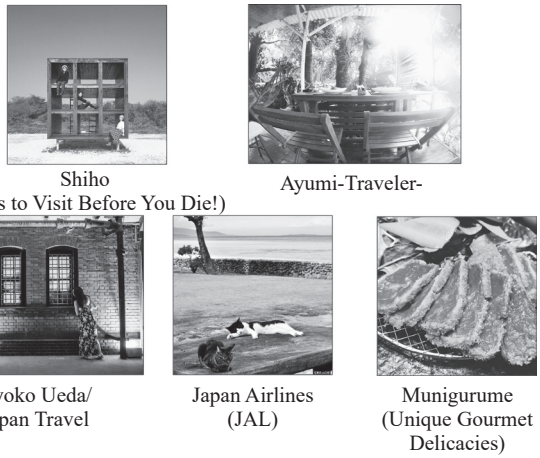


Figure 3: Highly attractive training images

Figure 4: Less attractive training images

Figure 5 shows the test dataset and predictions. Examining the test results in detail, out of 150 images deemed to be highly attractive, the model predicted 10 images were less attractive. These incorrect predictions occurred mainly in images that were either dark or had monotonous colors. Furthermore, out of the other set of 150 images deemed to be less attractive, the model incorrectly predicted 48 images were highly attractive. These false predictions were mostly associated with images featuring people or natural scenes.

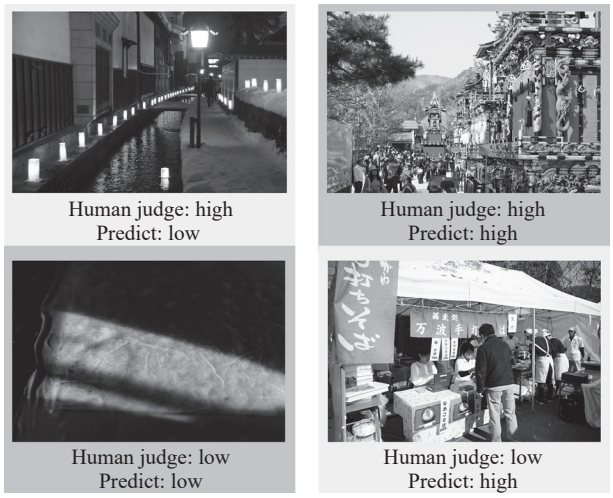


Figure 5: Test dataset and predictions

We assigned an attractiveness score to every publicity image of Hida City by this model, using the probability of "high attractiveness." These attractiveness scores are used for five-star ratings.

3.3 Adding descriptions to images

After that, a description was added to every publicity image using BLIP, a generative model pretrained with a large dataset of image-text pairs [Li et al., 2022]. The descriptions generated by BLIP were originally output in English and then translated into Japanese for use. The generated descriptions expressed the features of the images. For instance, in Figure 6, the left image is described with features like 'stone lantern' and 'yellow

Photo of red gate and stone lantern in Japanese garden with yellow leaves on the ground.

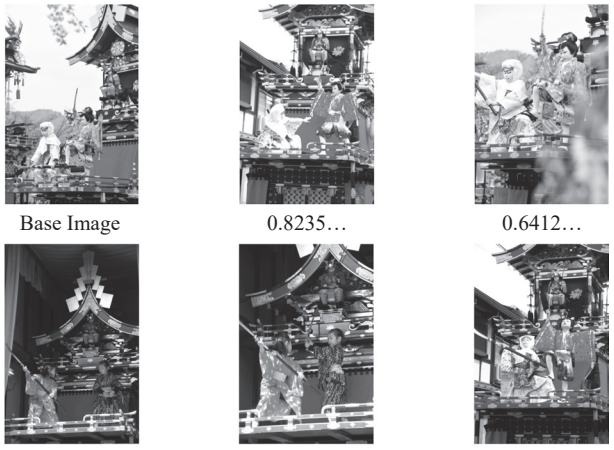
Photo of a plate with a fried egg, salad, and a bowl of soup.

Figure 6: Result with image descriptions

leaves,' while the right image is characterized by 'fried egg' and 'salad.'

3.4 Grouping similar images

Subsequently, EfficientNetV2L was used to extract features from each image. Each image was represented as a vector with 1280 elements. Then similar images were grouped by calculating the cosine similarity between these vectors. The specific steps are as follows: (1) The most attractive ungrouped image was selected as a base image. (2) For this base image, the cosine similarity was calculated with every ungrouped image individually. (3) The same group identification number was assigned to the base image and to images with a cosine similarity to the base image of at least 0.8. (4) These steps were repeated until every image was assigned a group identification number. Examples of the result of calculating cosine similarity are shown in Figure 7. All images were divided into 35,732 similar image groups; the mean number of images per group was 3.33, the median was 1, and the standard deviation was 14.38.



Base Image 0.8235… 0.6412…

0.7542… 0.7056… 0.7679…

Figure 7: Cosine similarity between the base image and other similar images

3.5 Developing the application

Finally, in order to integrate all the information related to the images, we saved the attractiveness score, description, and group identification number for each image in a single file, linking them to each file path of the image. An application was developed that allowed access to this information and reference each file.

4. Designing user experience

Figure 8 shows the first line of keyword search results for 'Goshozakura' and 'night.' With a few simple scrolls on the application screen, staff members can view hundreds of images at a glance without having to figure out complex folder structures.

This application's search function checks whether the keywords entered by a user are included in the file path or description of each image. It also supports AND searches with multiple keywords separated by spaces. Because of the descriptions and file paths, users can combine proper nouns frequently found in file paths with common nouns found in descriptions as search keywords. The application performs a synonym search using Sudachi synonym dictionary [Takaoka et al., 2020] for each keyword, and considers a keyword to be a match if any of the synonyms match the keyword as well as the

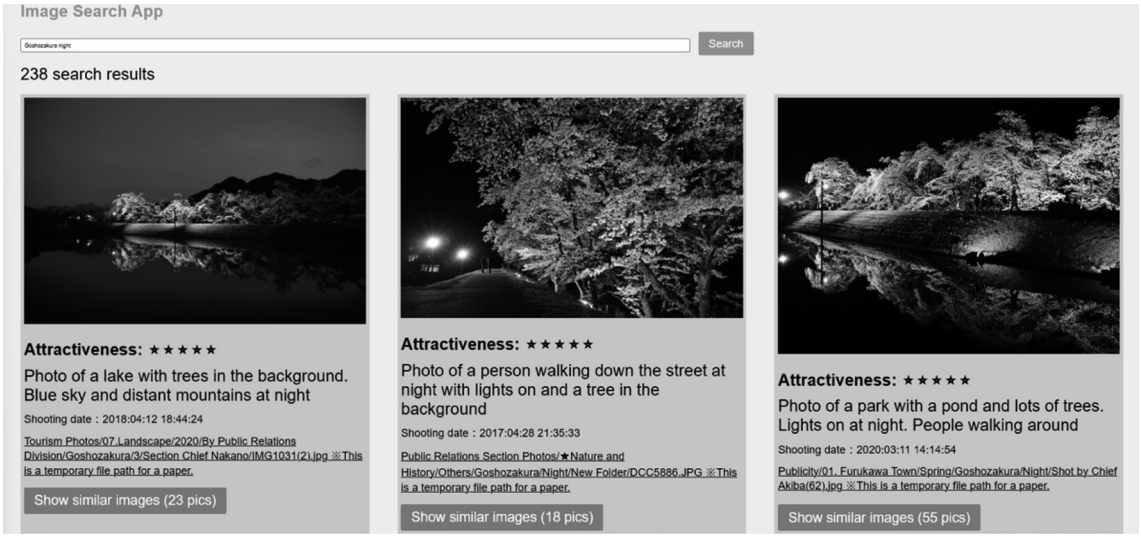


Figure 8: Examples of search result for the two words of "Goshozakura" and "night"

keyword itself.

The searched images are displayed in order of attractiveness scores. Attractiveness scores are rounded to a five-star rating when displayed.

Users can click on the file path in the search results to open the directory containing the corresponding image. This feature makes it easy for users to find the image they need without having to understand the complex directory structures.

Additionally, users can click the "Show similar images" button on the search results screen to display a list of images that belong to the same group as the selected image. This feature allows users to choose a scene from the search results and then carefully select the most appropriate image from among similar ones. As a side note, the keyword search results display all images that meet the search keywords, regardless of whether images with the same group identification number are already displayed. Therefore, images will not become undiscoverable due to grouping assignments.

5. Evaluation and discussion

5.1 Introductory experiment

The application developed in this study was introduced into the actual workflows of Hida City staff on October 31, 2023 (Figure 9). It received positive feedback, with staff noting that it "saved time" and "revealed previously unknown directories and images."

Figure 9: Hida City staff using the application

Furthermore, in January 2024, we conducted a survey regarding the user experience of the application with eight staff from Hida City. In this survey, as shown in Figures 10, 11 and 12, when they were asked, "Has the process of searching publicity images become faster?," the average rating was 4.5 out of 5. When they were asked, "Is there now a wider range of options for searching for publicity images?," the average rating was 4.25 out of 5. When they were asked, "Is the application easy to navigate?," the average rating was 4.875 out of 5.

Thus, the application has been demonstrated to be a valuable tool in addressing the challenge of unorganized publicity images in Hida City. Specifically, it was confirmed that all expected functions of the application worked effectively. First,

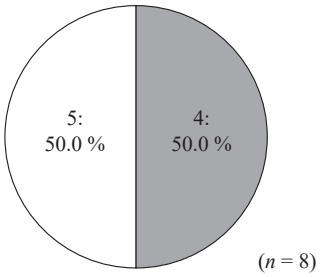


Figure 10: The result of the five-grade user survey of "Has the process of searching publicity images become quicker?"

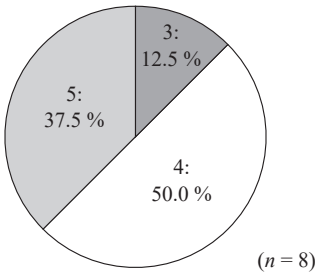


Figure 11: The result of the five-grade user survey of "Is there now a wider range of options for searching for publicity images?"

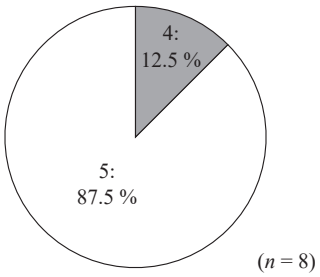


Figure 12: The result of the five-grade user survey of "Is the application easy to navigate?"

prioritizing the attractive images in the display order reduced work time. Second, checking search keywords against both the generated descriptions and the file paths of the images broadened the scope of the search. Third, listing similar images streamlined the task of carefully selecting the most suitable images. Furthermore, the staff were able to use these functions without difficulty. This was because the application operating environment and user experience were in line with the local government's existing workflows. The knowledge we gained through interviews allowed us to achieve that comfortable usability.

5.2 Limitation

On the other hand, some staff expressed the opinion that "about one out of every 100 stored images is worth releasing as open image data. In the future, we would like to publish a much larger number of images as the open image data." Considering that there are approximately 120,000 publicity images stored in Hida City, it can be estimated that they were willing

to increase the open image data to about 1,200 or more images.

However, if staff intend to release large-scale open image data containing thousands of images, the approach of selecting images one by one becomes nearly impossible, even with the aid of this image search application.

This search application is not enough to support large-scale releases of open image data by the local government; therefore, we believe that it is necessary to develop a three-step method (Figure 13). In the first step, many candidates for open image data are automatically extracted based on some criteria such as attractiveness scores and group identification numbers. In the second step, all tourist attractions in the local government are listed, every extracted image is classified as a proper tourist attraction, and checked if all tourist attractions have a sufficient number of candidate images. In the third step, the images of tourist attractions that are not included in the candidates or have insufficient numbers are manually picked up using the image search application. Through this integrated method, the staff can efficiently select thousands of image candidates for open data with minimal operational burden.

Furthermore, it is difficult to quantify how much time spent on a task has been reduced by the introduction of an application. This is because, with or without the application, there is a very large variation in the time required each time depending on factors such as the type of images and the number of images used at one time.

Moreover, there is much room for improvement in search performance: the descriptions generated by BLIP do not include (1) traditional Japanese objects such as "autumn leaves" and "floats" and (2) expressions of cultural background. BLIP also cannot properly generate descriptions longer than about 20 words. In practice, we assigned one to three keywords to each of 100 images, for a total of 196 keywords, with each keyword representing terms that might be used to retrieve the image. We then examined how frequently these keywords, or their synonyms were found within the corresponding image file paths or descriptions. As a result, only 101 out of the 196 keywords (51.5 %) matched. Although the introductory experiment was conducted in Hida City with the application with BLIP descriptions, we would like to improve search performance in the future. For example, when the same experiment as above was conducted with descriptions created by GPT-4o-mini (GPT), the match rate increased to 128 out of 196 keywords (65.3 %). This is because GPT is able to create longer sentences, and also because GPT is able to successfully reflect in the descriptions those words such as "autumn leaves," "*Dashi* (A float used in festivals and believed to carry gods)," and "rice planting," which BLIP is not so good at.

6. Conclusion

This study aimed to address the unorganized state of potential candidates for open image data which has been identified as a bottleneck in increasing open image data.

We developed an image search application to solve this issue. First, to display images in order of attractiveness, a model was developed to assess the attractiveness of images and assigned attractiveness scores to all images of a local government. Second, to support keyword searches that combine proper nouns and common nouns, descriptions were added to all images using an existing generative model, and the search function checks both file paths and descriptions. Third, to display a list of similar images, we grouped them based on the cosine similarity of their vector representations. We then developed an application that could reference these metadata. This application displays the search result images in order of attractiveness scores, support for AND searches that combine proper nouns and common nouns, and allows users to select the most appropriate image among similar ones.

The application was introduced into Hida City's actual workflows last year. Verbal feedback and a survey from Hida City staff indicate highly positive receptions. These outcomes suggest the application enhances the efficiency and comprehensiveness of searching publicity images by public officers.

On the other hand, through this introductory experiment, it has been revealed that some local governments are willing to release thousands of open image data. The method of searching and selecting images one by one is not adequate to achieve this large-scale release. Therefore, it is necessary to develop a three-step method including extraction, classification, and supplementation. In the future, we intend to develop and implement this integrated method.

Acknowledgements

We would like to express our gratitude to the Hida City staff

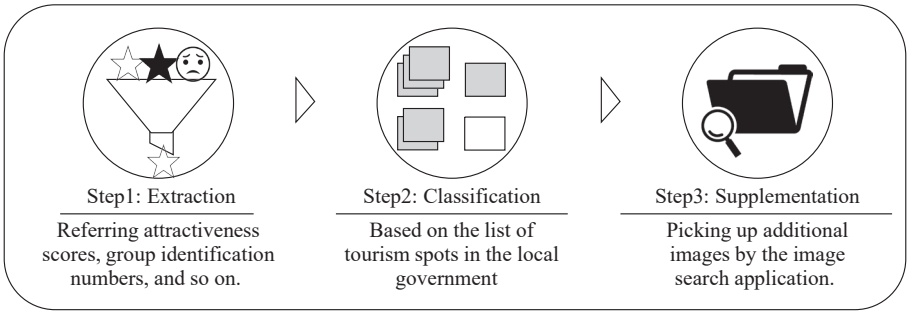


Figure 13: Conception of integrated method

and NEC Solution Innovators, Ltd. for their cooperation in this research.

References

Cabinet Office (2021). *Guidebook for constructing cycle of open data utilization for local development* (Retrieved September 5, 2024 from https://www.chisou.go.jp/sousei/about/mirai/policy/policy4.html). (in Japanese)

Chen, L., Lyu, D., Yu, S., and Chen, G. (2023). Multi-level visual similarity based personalized tourist attraction recommendation using geo-tagged photos. *ACM Transactions on Knowledge Discovery from Data*, Vol. 17, No. 7, 1-18.

Digital Agency (2023). *Examples of applications that are expected to utilize the municipal standard open data set* (Retrieved September 5, 2024 from https://www.digital.go.jp/resources/open_data/municipal-standard-data-set-test). (in Japanese)

Gifu Prefecture (2023). *Survey on the number of tourists visiting Gifu Prefecture in 2022* (Retrieved September 5, 2024 from https://www.pref.gifu.lg.jp/page/329982.html). (in Japanese)

Hayashi, K. (2020). Rediscovering regional attractions through tourist photo survey methods. *Journal of Transportation Science*, Vol. 50, No. 1, 18-23. (in Japanese)

Hida City (n.d.). *Hida City official tourism site "Hida no tabi"* (Retrieved September 5, 2024 from https://www.hida-kankou.jp). (in Japanese)

Instagram (n.d.a). *Shiho (Places to visit before you die!)* (Retrieved June 21, 2023 from https://www.instagram.com/shiho_zekkei).

Instagram (n.d.b). *Ayumi-traveler-* (Retrieved June 21, 2023 from https://www.instagram.com/oooooayumiooooooo/).

Instagram (n.d.c). *Kyoko Ueda/Japan travel* (Retrieved June 21, 2023 from https://www.instagram.com/kyoko1903/).

Instagram (n.d.d). *Japan Airlines (JAL)* (Retrieved June 21, 2023 from https://www.instagram.com/japanairlines_jal/).

Instagram (n.d.e). *Munigurume (Unique gourmet delicacies)* (Retrieved June 21, 2023 from https://www.instagram.com/muni_gurume_japan/).

Li, J., Li, D., Xiong, C., and Hoi, S. (2022). Blip: Bootstrapping language-image pre-training for unified vision-language understanding and generation. *International Conference on Machine Learning*, PMLR, 12888-12900.

Lin, T., Y., Maire, M., Belongie, S., Hays, J., Perona, P., Ramanan, D., Zitnick, C., L., and Dollar, P. (2014). Microsoft coco: Common objects in context. *Computer Vision–ECCV 2014: 13th European Conference*, Part V 13, 740-755.

Takaoka, K., Okabe, Y., Kawahara, N., Sakamoto, M., and Uchida, Y. (2020). Creating a synonym dictionary with refined synonym relations. *Proceedings of the Twenty-sixth Annual Meeting of the Association for Natural Language Processing, Association for Natural Language Processing*, 840-842. (in Japanese)

Tan, M. and Le, Q. (2021). Efficientnetv2: Smaller models and faster training. *International Conference on Machine Learning*, PMLR, 10096-10106.

Tu, Y., Kinugasa, M., Urata, M., Endo, M., and Yasuda, T. (2023). An image processing and utilization system using deep learning for tourism promotion. *2023 IEEE Asia-Pacific Conference on Computer Science and Data Engineering*, IEEE Computer Society, 1-6.

Received: September 30, 2024
Revised: November 5, 2024
Accepted: November 7, 2024
Published: November 30, 2024

https://doi.org/10.37020/jgtr.9.2_111