Contents lists available at Science-Gate

International Journal of Advanced and Applied Sciences

Journal homepage: http://www.science-gate.com/IJAAS.html

Substantiation of location image classification model using projective template matching and convolutional neural network

Jin-Wook Jang [1, *], Dong-Wook Lee [2]

[1] *Digital Transformation, Agricultural Cooperative University, Goyang City, South Korea*
[2] *Intelligence Mobile Systems, Jacobs University Bremen, Bremen, Germany*

ARTICLE INFO

Article history:
Received 16 December 2021
Received in revised form
27 February 2022
Accepted 4 March 2022

Keywords:
Location image search
Projective template matching
Convolutional neural network
Object detection

ABSTRACT

This study first attempts to observe the action of the CNN and then compares it to test Projective Template Matching and Object Detection as new approaches. In the final model selection, the accuracy of the prediction model and the computational processing time was mainly compared. At last, the combination of the Object Detection model and CNN was selected as a final location classification model with a prediction accuracy of 61%. This final model shows the optimal prediction result by first attempting to detect the common feature regions of the location image and then analyzing the overall feature characteristic. The fact is that CNN is good for training image data with common overall features for classification. This being so, we expect that training several fundamental ROIs can more efficiently train the CNN model than training the pure location images.

1. Introduction

Unlike the image of an object, the characteristics of the place image are not universal. Even in the same location, there is a high probability of photos to be taken in a different view, which lowers the commonality of features in the image. Therefore, the Convolutional Neural Network (CNN) that analyzes and learns common characteristics of image classes is not suitable for use in location image classification. Based on this idea, this study first attempts to observe the action of the CNN and then compares it to test Projective Template Matching and Object Detection as new approaches. This study aims to build an optimal Location Image Classification Model. CNN, Projective Template Matching, and Object Detection model are used. The main programming language is Python, with help of mathematical libraries such as Tensorflow and Numpy. Five attractions from Jeju-island are selected to serve as the study's location image classes. The image data have been collected from searching portals through crawling, saved as a dataset, and been used for the model training and testing.

1.1. Convolutional neural network (CNN)

CNN, a type of deep neural network, is optimized for image analysis. CNN is renowned for its automatic process of generating an image feature extracting filter during data training (Prabhu, 2018). Multiplying the filter to each input image file creates the feature map of the image, which raises the CNN's efficiency in analyzing the image class. In addition, CNN not only uses filters to create image feature maps but also reduces the dimension size of the image through the pooling process. Below Fig. 1 represents the process of how CNN classifies the image.

1.2. Activation function

In this study, the ReLU function is used as an activation function. As shown in Fig. 2, the ReLU function maps negative input values to 0, and positive input values to the linear function y=x. Therefore, the ReLU function does not do any calculation for negative input values, but rather returns 0. This fact lets ReLU take a shorter processing time than other activation functions (Brownlee, 2019a).

1.3. Max pooling

Max Pooling is one of the ways to reduce the size of 2D data. Specifically, this is a method of extracting the value of the most significant pixel from each

[*] Corresponding Author.
Email Address: jjw@nonghyup.ac.kr (J. W. Jang)

https://doi.org/10.21833/ijaas.2022.05.008
Corresponding author's ORCID profile:
https://orcid.org/0000-0002-3605-9157
2313-626X/© 2022 The Authors. Published by IASE.

image patch (Brownlee, 2019b). Pixels with relatively small values are removed and the only pixel with the largest value survives. Fig. 3 compares average pooling and max pooling. It can be seen that the average pooling extracts the average value of pixels in the patch.

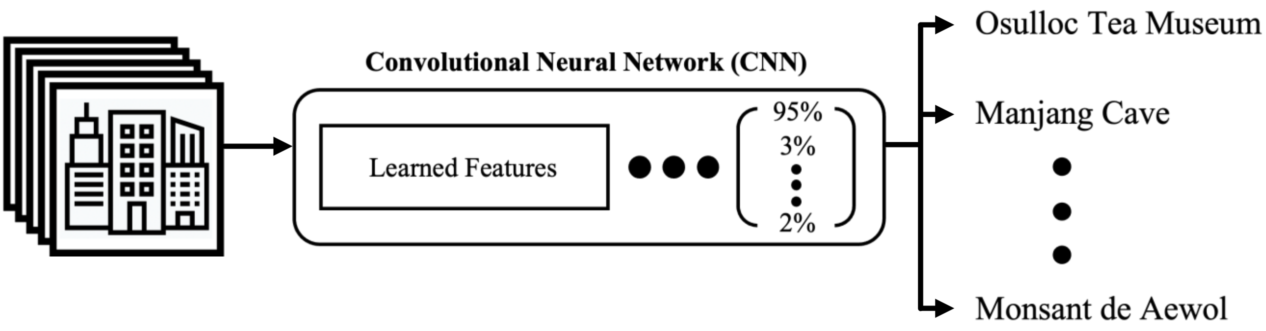


Fig. 1: CNN's image data learning process

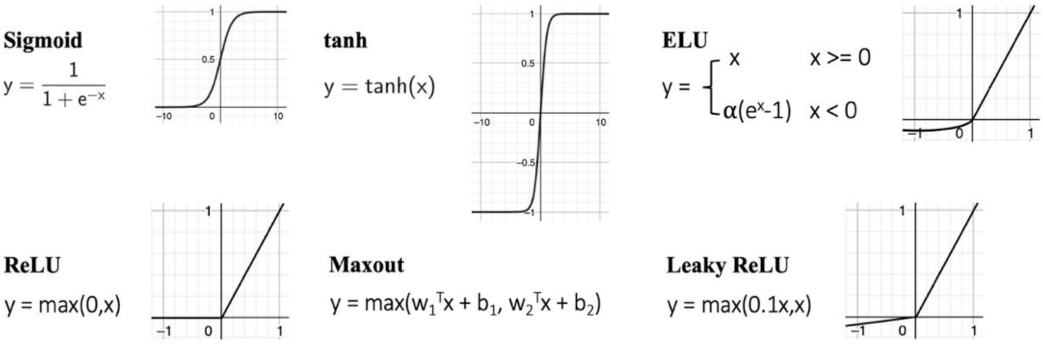
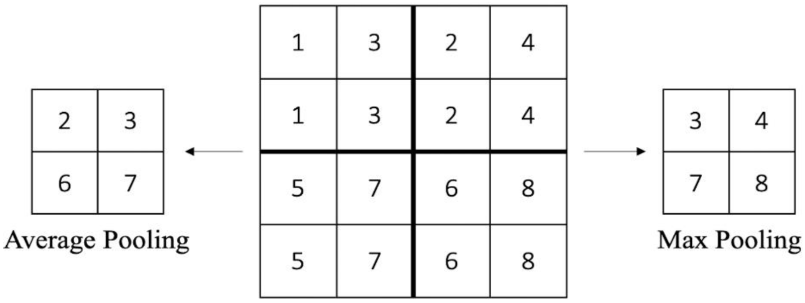


Fig. 2: Activation functions



1	3	2	4
1	3	2	4
5	7	6	8
5	7	6	8

Average Pooling

2	3
6	7

Max Pooling

3	4
7	8

Fig. 3: Max pooling

1.4. Tensorflow

Tensorflow is an open-source machine learning library (or engine) developed by Google, which is utilized in various research fields such as Artificial Intelligence, Deep Learning, and Data Manipulation (Chatterjee, 2020). Tensorflow is useful for processing numerical computations, which applies to this study (Tensorflow, 2020). In this study, Tensorflow is used for computational processes such as max-pooling and activation function.

1.5. Projective template matching

Template Matching is a computer vision technique that detects a specific template in an image. Sliding the window through the image, the algorithm compares the pixel values within the template and the image window. Examples of its usage are Mars exploration sample tube localization, corneal tracking during eye operation, and car license plate recognition (Daftry et al., 2021; Cho et al., 2014; Oh and Park, 2014).

However, since the pixel values are compared one by one in each position, Template Matching is variance to all transformations (translation, scale, rotation, skew, projective). To compensate for this disadvantage, this study intends to use Projective Template Matching.

Projection Template Matching is an algorithm that transforms a given template through a matrix and enables the detection of the template in the image even if the view of the photo differs (Zhang and Akashi, 2016). The following Fig. 4 is a visual representation of how projection transformation works on the template image.

In this study, the 'Robust Projective Template Matching' paper written by Chao Zhang and Takuya Akashj, a genetic algorithm is used for Projection Template Matching (Zhang and Akashi, 2016). This algorithm generates several initial candidate transformation matrices and then finalizes the transformation matrices with the smallest loss rate. The loss rate of each matrix is measured by calculating the Sum of Absolute Differences (SAD). Among them, only a specific percentage of low-loss matrices survive and contribute to the establishment of the next generation. The survival percentage value is set manually by the user. This method continues until only a certain few matrices survive, consequently extracting a Projection Transformation Matrix with the smallest SAD.

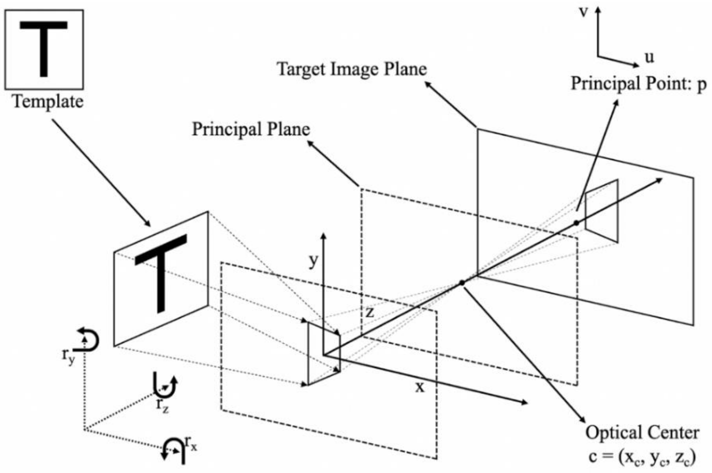


Fig. 4: Visual understanding of projection matrix transformation

2. Requirements for location image classification

This study conducted several tests to build an efficient location image classification model. First, we used CNN, the approach of our previous study, YOLOv5, the model aiming to understand the location image through ROI detection, and the Projective Template Matching, the algorithm without massive data training. Fig. 5 is the structural diagram of the three test approaches.

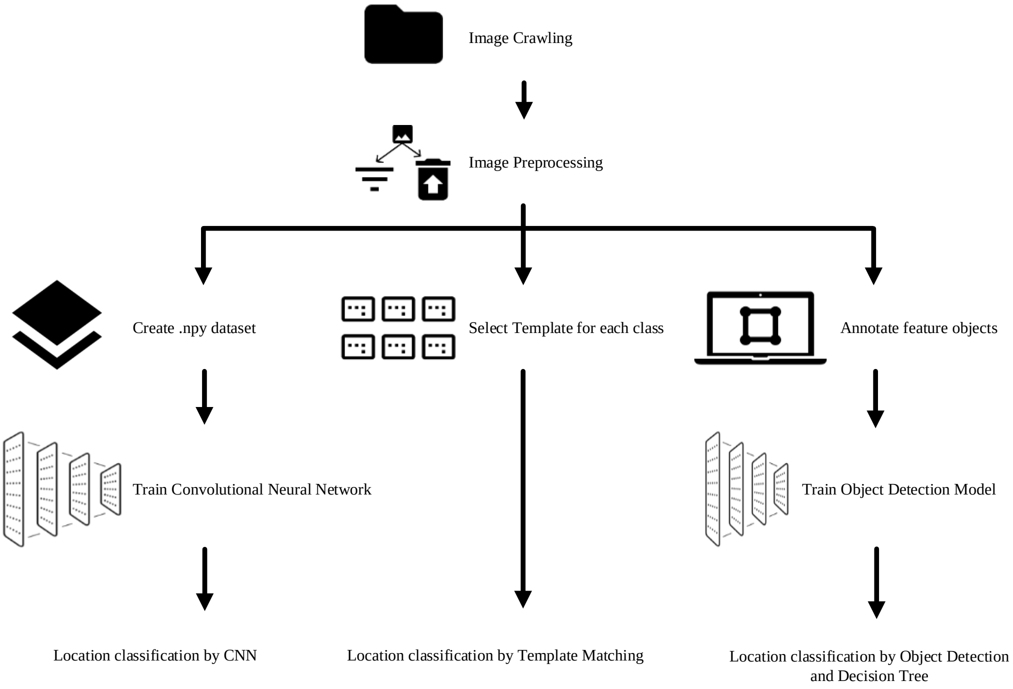


Image Crawling
Image Preprocessing
Create .npy dataset
Select Template for each class
Annotate feature objects
Train Convolutional Neural Network
Train Object Detection Model
Location classification by CNN
Location classification by Template Matching
Location classification by Object Detection and Decision Tree

Fig. 5: The structure of the tests

2.1. Learning convolutional neural network

The first approach of this study is using CNN. After crawling images of five tourist attractions in Jeju Island-which are Osulloc Green Tea Field, Manjang Cave, Monsant Café Aewol, Columnar Joints, and Cheonjiyeon Falls-in the search portals Naver and Google, the images were preprocessed. Images with irrelevant regions, such as an advertisement or an obstacle object, are filtered out. Then, CNN was trained with these preprocessed images. We used VGGNet as the CNN model. The VGGNet was designed and tested by Tensorflow.

The result of the test concluded that the VGGNet model shows a large deviation in classification accuracy depending on the feature of the location image class. CNN is renowned for its optimal performance in analyzing the overall common features of the image class (Prabhu, 2018). However, the location images are hard to have universal common features because of diverse camera views, light variations, etc. In accordance, if all the images of the location class are taken from the same camera view or have a monotonous pattern, the class gets a high rate to be classified.

Class numbers 2 and 3, Monsant Café Aewol, and the Columnar Joints are the location classes that have completely failed to be classified. Inspection of both location classes' images shows that the views of the taken photos in both location classes are diverse.

2.2. Projective template matching

Projection Template Matching is an approach that examines the presence or absence of a template in a test image while transforming the template by the transformation matrix. To test Projection Template Matching through a genetic algorithm, several parameter values should be set by the user. The parameters are shown in Eqs. 1, 2, and 3.

$$\delta = 10000 \tag{1}$$
$$c = 20 \tag{2}$$
$$\lambda = 0.1 \tag{3}$$

δ is the number of first-generation matrix instances, c is the maximum number of matrix instances to be finalized, and λ is the ratio of the number of instances to be transferred from the generation to the next generation. As a result, however, the computation time for finalizing the transformation matrix through the genetic algorithm took the execution time of a linearly increasing function, which is not suitable for a real-time location image classification model. Below Fig. 7 shows the computation time of the algorithm measured in varying numbers of initial matrix instances. Fig. 6 is an image of Osulloc Tea Museum used for Projection Template Matching with the above setting.

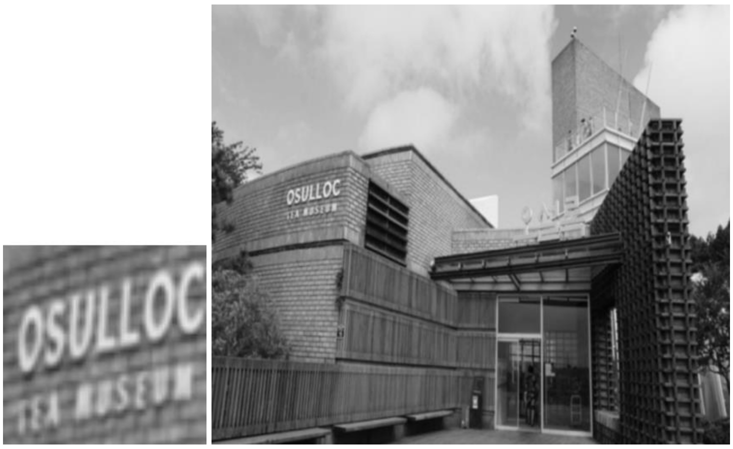


Fig. 6: Template (left) and test image (right) of Osulloc Tea Museum on Jeju Island

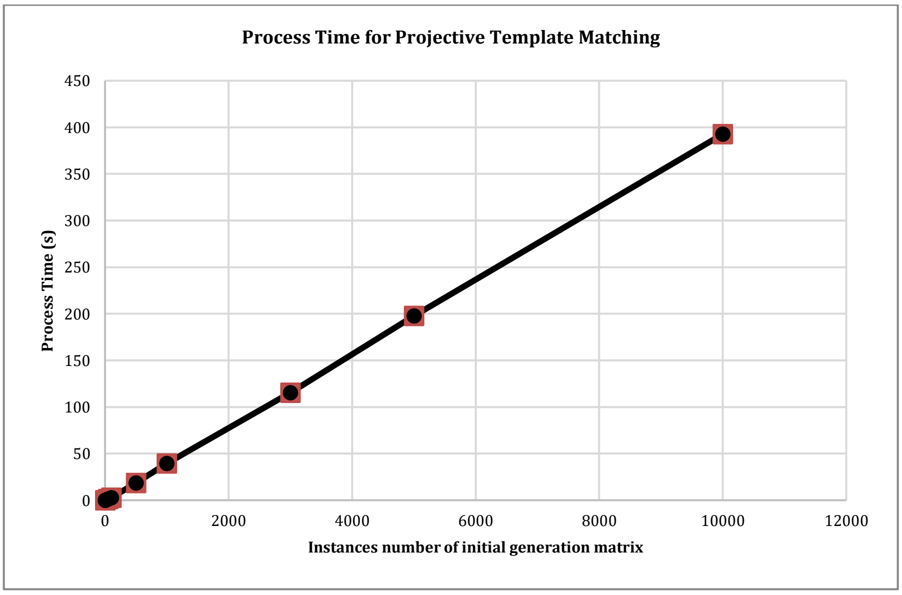


Fig. 7: Projection template matching genetic algorithm execution time

2.3. Object detection with decision tree

This approach is targeting the location classes that contain a common Region of Interest (ROI). Fig. 8 is an example of this test approach in which the water stream of Chonjiyeon Falls is used as the ROI. We used YOLOv5 as an object detection model. YOLOv5 was trained on the ROIs included in five location classes in Jeju-island as templates. For classification, the model tries to find the ROI within the test image by YOLOv5 and then classifies the location image into the corresponding ROI's location class. As a result, even the location classes that were taken from various camera views or did not have a monotonous pattern were well classified when the characteristic ROI is inside.

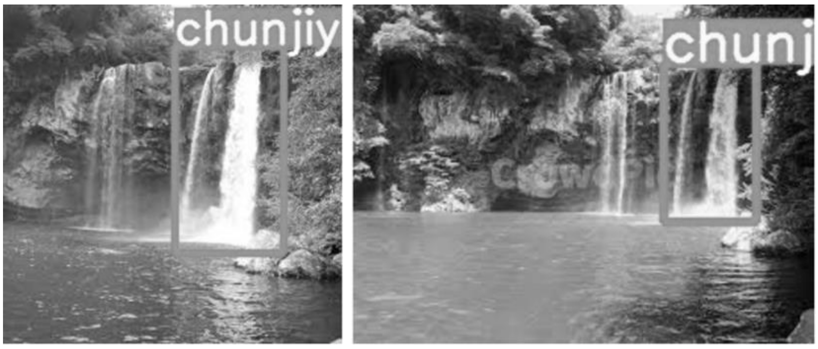


Fig. 8: The template of Cheonjiyeon fall

3. Results and discussion

As a result of the tests, it was checked that the Projection Template Matching using a genetic algorithm is not suitable for real-time location image classification. However, VGGNet and YOLOv5 models showed high classification efficiency in different location classes. If there is a specific ROI in the location image, it is appropriate to use the YOLOv5 model, and if not, it is desirable to use a VGGNet to analyze the overall features of the location image.

Accordingly, we at last integrated the VGGNet and YOLOv5 approaches as a new attempt. Specific ROI is attempted to be detected through YOLOv5 and if not detected, the overall feature of the image is analyzed through the VGGNet. Fig. 9 represents the classification result of all the possible models.

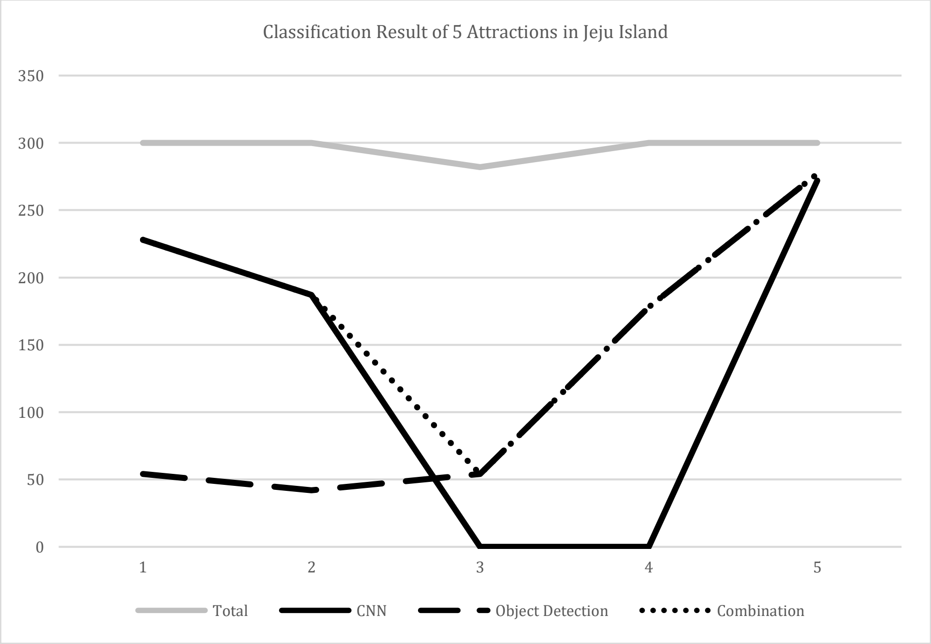


Fig. 9: Comparison of the overall prediction model accuracy

4. Conclusion

CNNs are known to be much optimized for image analysis, and accordingly, they are useful in various research fields such as medical care and architecture. However, the location images are hard to be simply classified by passing the image through CNN. The reason is that the location images can have variant inner features depending on the camera view. Based on this fact, we also tried to detect the ROIs in the location image for further classification. At last, by integrating these two methods-passing through CNN and detecting ROIs-the study found that the model could reach the prediction rate of 61%.

Nevertheless, the prediction rate of 61% still is not enough to be used as a service model that aims to provide correct information to users. Still, the fact is that CNN is good for training image data with common overall features for classification. This being so, we expect that training several fundamental ROIs can more efficiently train the CNN model than training the pure location images. We intend to leave "classification of the location image through learning the distribution of feature objects and patterns" a future research topic.

Acknowledgment

This research was supported by the National Research Foundation of Korea in 2022(No. NRF-2020R1G1A1005872).

Compliance with ethical standards

Conflict of interest

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

References

Brownlee J (2019a). A gentle introduction to the rectified linear unit (ReLU). Machine Learning Mastery, 6. Available online at: https://machinelearningmastery.com/rectified-linear-activation-function-for-deep-learning-neural-networks/

Brownlee J (2019b). A gentle introduction to pooling layers for convolutional neural networks. Machine Learning Mastery, 22. Available online at: https://machinelearningmastery.com/pooling-layers-for-convolutional-neural-networks/

Chatterjee M (2020). What is TensorFlow? The machine learning library explained. Available online at: https://www.mygreatlearning.com/blog/what-is-tensorflow-machine-learning-library-explained/

Cho JH, Ahn CW, and Jun JH (2014). A feature point tracking method by using template matching and buffer. The Journal of the Institute of Internet, Broadcasting and Communication, 14(4): 173-179. https://doi.org/10.7236/JIIBC.2014.14.4.173

Daftry S, Ridge B, Seto W, Pham TH, Ilhardt P, Maggiolino G, and Detry R (2021). Machine vision based sample-tube localization for Mars sample return. In the 2021 IEEE Aerospace Conference (50100), IEEE, Big Sky, USA: 1-12. https://doi.org/10.1109/AERO50100.2021.9438364

Oh S and Park CS (2014). Vehicle license plate recognition system using image binarization and template matching. Journal of the Semiconductor and Display Technology, 13(2): 7-12.

Prabhu R (2018). Understanding of convolutional neural network (CNN)-Deep learning. A Medium Corporation, San Francisco, USA.

Tensorflow (2020). Customization basics: Tensors and operations. Available online at: https://www.tensorflow.org/tutorials/customization/basics

Zhang C and Akashi T (2016). Robust projective template matching. IEICE Transactions on Information and Systems, 99(9): 2341-2350. https://doi.org/10.1587/transinf.2016EDP7038