Design of Deep Learning-based Technology for Place Image Collecting

Jin-wookJang

Office of Industry-Academy Cooperation Foundation,Hanyang Women's University,Seoul, Republic of Korea
jinwchng@hywoman.ac.kr

Abstract: This research study designed a location image collecting technology. It provides the exact location information of an image which is not given in the photo to the user. Deep learning technology analysis and collects the images. The purpose of this service system is to provide the exact place name, location and the various information of the place such as nearby recommended attractions when the user upload the image photo to the service system. Suggested system has a deep learning model that has a size of 25.3MB, and the model repeats the learning process 50 times with a total of 15,266 data, performing 93.75% of the final accuracy. In a performance test, the final accuracy of the model is calculated 93.75%. This system can also be linked with various services potentially for further development.

Keywords: Deep Learning, Place Image, CNN(Convolutional Neural Network), Pooling Layer

1. Introduction

To find the location of an image, we normally use location tag saved in the image file. Most photos taken via smartphone do not have their location information. Even though an image has its location information with a local tag, there is potential loss of information when the image file is being uploaded. Eventually, even if there is someone who wants to visit a place in a picture, it is difficult to find the location through the image.

In this research study, a service system using deep learning technology was developed. The system provides location details for input photos without location information.

This service has a possibility to be developed as a service system providing not only location information of images, but also nearby restaurants and recommended attractions.

Enough data is needed to improve the accuracy of the service. Since the accuracy increases as the amount of data increases, we designed the image crawling system that automatically collects data and the deep learning system using images.

2. Relevant Service and Technology

In this research, a system that informs a user of local information based on images would be uploaded by the user. Although this system is similar to Google Image Searching System, it specializes in local image searching unlike Google's. Various recognition techniques, such as CNN (Convolutional Neural Network), Pooling Layer, LeRU(Rectified Linear Unit), and VGGNet(Visual Geometry Group Neural Network) are applied to detect the location information of a place image

- **Google Image Searching System**

Google Image Searching is a service that provides similar images with the primary image and common keyword's referential images when an image's URL is searched, or an image file is uploaded. As this service is available for various purposes, there are numerous customers[1]. Yet it lacks expertise in the fields of multipurpose services. In machine learning, the amount of data is proportional to its accuracy. Even though Google has one of the biggest databases, it cannot be said that the search results from Google are accurate; it is difficult to find related location information from its massive results.

- **CNN(Convolutional Neural Network)**

CNN is a deep learning algorithm that distinguishes an image from others[2]. The Convolutional Neural is able to learn the characteristics of an image by itself, and this ability allows the developed system to learn features of place images and discover the location where an input image is taken at.

- **Pooling Layer**

One of techniques that enhance the efficiency of CNN is Pooling Layer. The layer works as a filter, optimizing an image file. [3] This function reduces the computation in the neural network.

VGGNet(Visual Geometry Group Neural Network)

VGG model uses 3*3 filters instead of large-size filters. The 3*3 filter requires more convolution process than large-size filters do. However, the number of parameters earned by using 3*3 filters is fewer than the number of parameters from large size filters. Since there would be less parameters to be learnt, the model would have faster convergence, and minimized overfitting problem[6].

- **ReLU(Rectified Linear Unit)**

A Neural Network Activation function is what determines its output, accuracy, and efficiency of training model[4, 13]. In this study, Rectified Linear Unit (ReLU) is used as an activation function. ReLU has an advantage that it allows computation process can be run quickly[5, 14]. This is because ReLU function returns zero when inputs approach to zero, or negative.

As shown in Figure 1, when an unnecessary value such as a negative integer is received, ReLU function returns a value of 0 unconditionally. As the result, this function shows a significantly faster processing speed than the other formulas are used as a loss function[11,12].
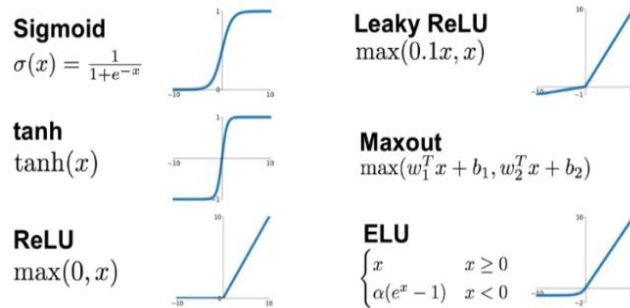

Figure 1. Formulas of Various Loss Function

Therefore, CNN with Max Pooling and ReLU function is considered as the most suitable neural network for two-dimensional data learning. Since the purpose of this study is to train the service system in Jeju images, CNN is chosen to use.

TensorFlow is a machine learning library or engine developed by Google. This has been used in a variety of artificial intelligence fields such as Google search, speech recognition, translation, and AI service. Using TensorFlow, algorithms for image recognition, repetitive neural networks, and neural network learning are easily implemented[3]. The library consists of arithmetic operations, then it is a useful library for this system that needs to process matrix type files[4]. In this study, TensorFlow was used for various calculation processing such as Max Pooling and Loss Function for image learning.There are two primary libraries used in this system implementation. Tensors are composed of matrices, so Numpy library that handles matrices and Matplotlib library which visualizes image data are necessary. Therefore, it is also essential to implement CNN which studies its own images.

3. System Structure

Deep learning-based location information service has a structure like Figure 2. The service has the image collecting part and the deep learning part, which analyzes location information of the given images, and service part providing an actual service with filtered information.
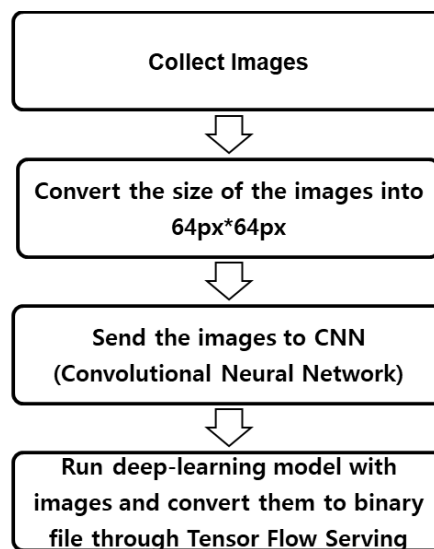

Figure 2. The System Structure

First, to collect the place images as the user's request, write the place names to search and index them using hexadecimal in the Excel file as shown in Figure2, then Image crawling code reads the Excel file and automatically creates subfolders for each index. After that, place images are automatically searched, collected, and sequentially stored in an appropriate subfolder. Second, preprocessing code is used to delete invalid files and convert the training images to 64*64 pixels. Third, the images are converted into NPY format through the Dataset code. In the fourth step, actual learning starts to proceed through the deep learning machine's convolutional layer, max pooling layer, and the ReLU function in the CNN code. After the learning process, the training results are stored in the form of a h5 file, which is a stacked data. Finally, if the images that a user wants to predict are stored in the test_image folder, the prediction is executed based on the learning done in the fourth step. The prediction results are returned in the form of hexadecimal, so the location name of the predicted image can be checked through the classification index table.

4. System Structure

TensorFlow by Google is used in frame working for its deep learning engine, and Python is used as its development language. The deep learning-based local image collecting model on Figure 2 shows the process from the completion of its build to the new model distribution part.

After developing the model, TensorFlow-Serving Part Module was developed as a sub-project of TensorFlow from the perspective of 'How to effectively provide a stored model?', TensorFlow-serving receives an image as an input from the saved model like Figure 3, then returns the converted input file as a HTTP Response format.
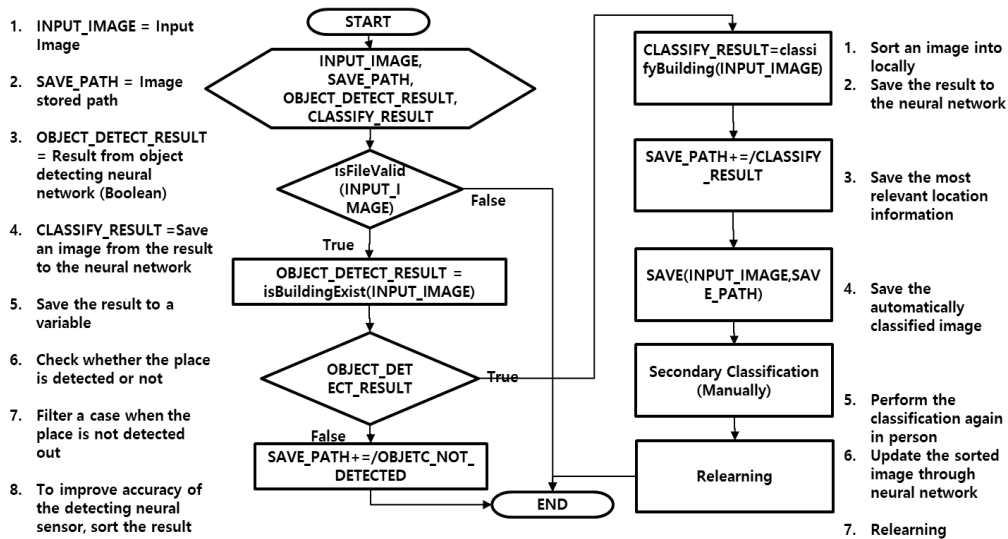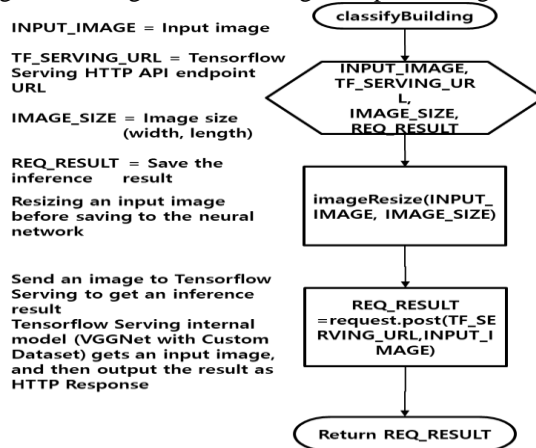
Figure 3. Process of TensorFlow-Serving

Figure 4. Design of Local Image Deep Learning Model

TensorFlow-serving has an advantage that even non-AI (Artificial Intelligence) experts can utilize the system with convenience[7]. The overall structure of the location image's deep learning model looks like Figure 3.

When the system gets an input image, it checks validation of the input file first. If it fails the validity check, it ends the algorithm. Otherwise, the input file would be saved into the object detecting neural network, and the Boolean result would be used to study the input image. Once the algorithm gets true from the previous process, the input image will be classified again at the next step, which is the location sorting neural network. Based on the inference result from the neural network, the system decides a folder where to save the image. After all, the system performs, the second classification, and re-learning on the neural network to improve the accuracy of its data[8]. Local images would be classified as described in Figure 4.

Through the above process, it consolidates the disposal against duplicated images. The structure of interconnections between each layer is formed like Figure 5. When the system is running, the following tensor tube relationship can be seen through the tensor board.
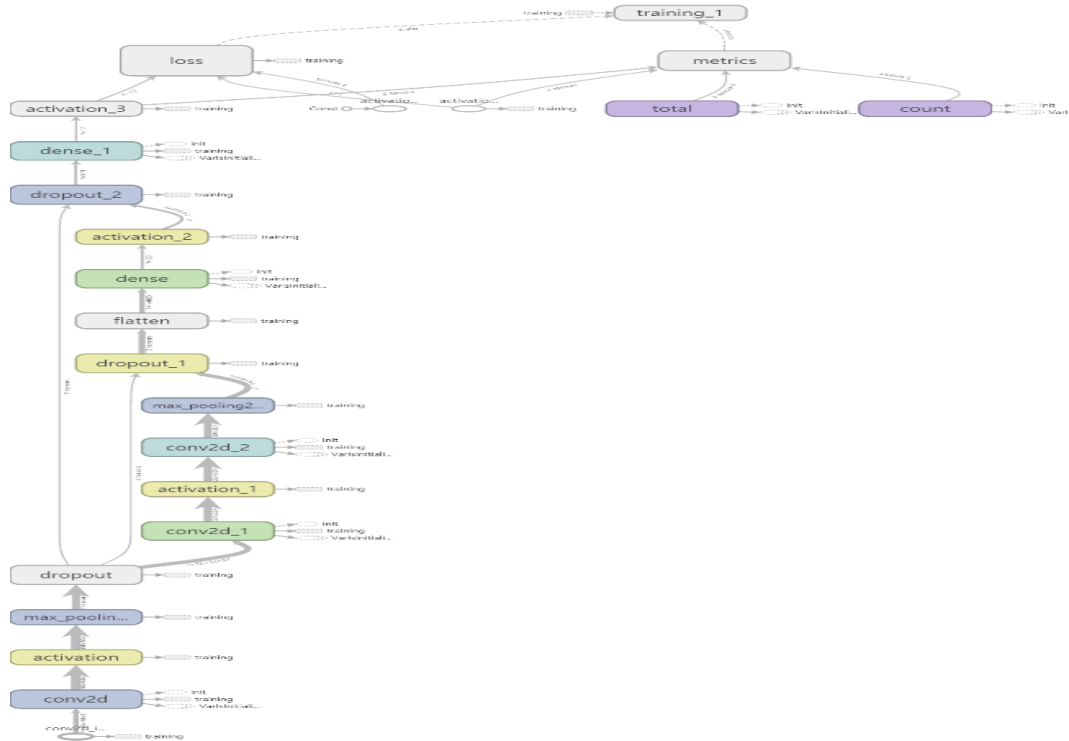

Figure 5. The Interrelationship of each layer

5. Performance Analysis

The accuracy of the model was analyzed to improve the overall system precision and to run a performance test, using a confusion matrix. Table 1 indicates its criteria for classification[9].

Table 1. Confusion Matrix

Headinglevel	Expected Value		
Criteria	Div.	True	False
	True	True Positive(TP)	False Positive(FP)
	False	False Negative(FN)	True Negative(TN)

Accuracy, precision, recall, and F1 score are used to measure the accuracy of deep learning model. The precision from the measurement result is $\frac{TP}{TP+FP} = 0.8461538461538461$,recall is $\frac{TP}{TP+FP} = 0.33460076045627374$, accuracy is $\frac{TP+TN}{TP+FN+FP+TN} = 0.7579214195183777$, and F1 score is calculated by Equation 1: $F1-score = 2 \times \frac{1}{\frac{1}{precision}+\frac{1}{recall}} = 2 \times \frac{precision \times recall}{precision+recall} = 7.66958041958042$.

The size of the deep learning model is 25.3MB, and the deep learning model repeats the learning process 50 times with a total of 15,266 data. Final accuracy of the model is calculated 93.75% as Fig. 6 implies.

Figure 6. The Accuracy Calculation Result

The accuracy of this model increases as the number of repetitions increases, as shown in Figure 7. This model was run repeatedly and as the process was repeated, the accuracy of the model converged to 1. Through the practice, its effectiveness is verified.

Figure 7. Graph of The Deep Learning Model Accuracy
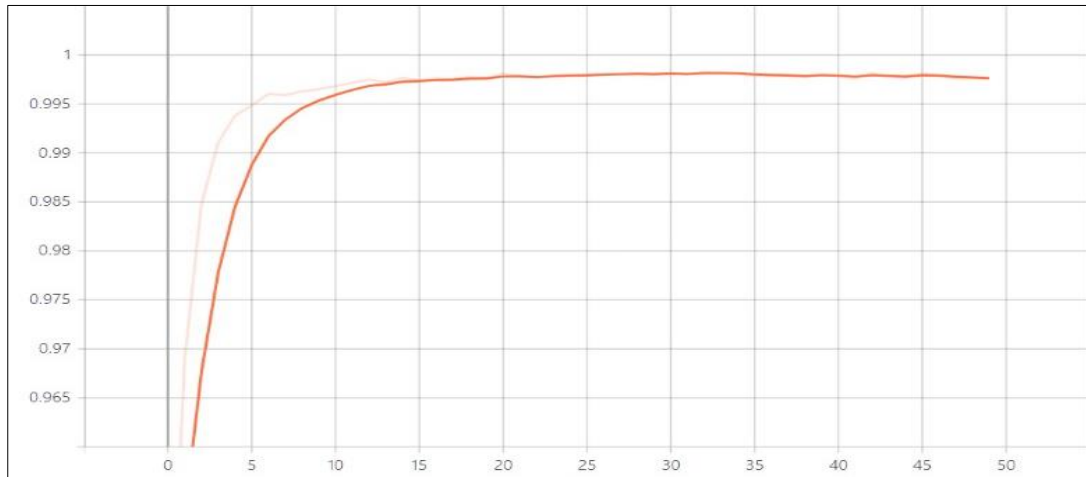
6. Substantiation

To apply this system into practical service, names of places as listed in Table 2 need to be collected and indexed. Following listed 8 places are famous attractions in Seoul, and they are selected to test the performance of the model

Table 2.Index of Place Images

Index	Seoul's Place Images
1	63 Building
2	Seoul City Hall
3	A statue of Admiral Yi Sun-shin
4	A statue of King Sejong
5	Geunjeongjeon Hall in Gyeongbokgung Palace
6	Olympic Park
7	Lotte World Tower
8	Myeongdong Catholic Cathedral
9	The Namsan Tower
10	Seokchon Lake
11	Seodaemun Prison History Museum
12	Seoul Station Old Platform

Table 3 shows the prediction results of each image. If the indexing value of predictions is "1.0", it means success of prediction results. Table 3 shows the prediction results of each image

Table3.Prediction Result Example of Place Image

Index	Seoul's Place Images
	predictions": [0.0, <u>1.0</u>, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0] Label Index 1 indicates that the probability that given image is 'Seoul City Hall' is 100%
	"predictions": [0.0, 0.0, 0.0, 0.0, 0.0, 0.0, <u>1.0</u>, 0.0, 0.0, 0.0, 0.0, 0.0] Label Index 6 indicates that the probability that given image is 'Lotte World Tower' is 100%
	"predictions": [0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 1.0, 0.0, <u>1.0</u>, 0.0, 0.0, 0.0] Label Index 8 indicates that the probability that given image is 'The Namsan Tower' is 100%

7. Conclusion

This research study built the model that collects images without location information and performs deep-learning on them. Through the deep learning model, the system learns and filters input images repeatedly.

Suggested system has a deep learning model that has a size of 25.3MB, and the model repeats the learning process 50 times with a total of 15,266 data, performing 93.75% of the final accuracy. In a performance test, the final accuracy of the model is calculated 93.75%. This system can also be linked with various services potentially for further development.

8. Acknowledgment

References

1. Google Lens. https://lens.google.com/
2. SumitSaha, "A Comprehensive Guide to Convolutional Neural Networks", 2018, https://towardsdatascience.com/a-comprehensive-guide-to-convolutional-neural-networks-the-eli5-way-3bd2b1164a53
3. SavyakHolsa, "CNN | Introduction to Pooling Layer", https://www.geeksforgeeks.org/cnn-introduction-to-pooling-layer/
4. Missing Link, "7 Types of Neural Network Activation Functions: How to Choose?", https://missinglink.ai/guides/neural-network-concepts/7-types-neural-network-activation-functions-right
5. Sagar Sharma, "Activation Functions in Neural Networks", 2017, https://towardsdatascience.com/activation-functions-neural-networks-1cbd9f8d91d6
6. Sik-Ho Tsang, "VGGNet — 1st Runner-Up (Image Classification), Winner (Localization) in ILSVRC 2014",https://medium.com/coinmonks/paper-review-of-vggnet-1st-runner-up-of-ilsvlc-2014-image-classification-d02355543a11, 2018

7. Google Inc,"Tensorflow Serving Documentation. https://www.tensorflow.org/tfx/guide/serving, 2016.

8. Salma Ghoneim. Accuracy, Recall, Precision, F-Score & Specificity, which to optimize on?. Towards Data Science. https://towardsdatascience.com/accuracy-recall-precision-f-score-specificity-which-to-optimize-on-867d3f11124, 2019.

9. H. J´egou, M. Douze, C. Schmid, and P. P´erez,"Aggregating Local Descriptors into a Compact ImageRepresentation," in Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, pp. 3304-3311, 2010.

10. Prabhu, "Understanding of Convolutional Neural Network (CNN) Deep Learning", 2018.

11. Jason Brownlee, "A Gentle Introduction to Pooling Layers for Convolutional Neural Networks", 2019, https://machinelearningmastery.com/pooling-layers-for-convolutional-neuralnetworks/#:~:text=Maximum%20pooling%2C%20or%20max%20pooling,the%20case%20of%20average%20pooling.

12. Jason Brownlee, "A Gentle Introduction to the Rectified Linear Unit (ReLU), 2019, https://machinelearningmastery.com/rectified-linear-activation-function-for-deep-learning-neural-networks/"

13. Konstantinis, A., Rozakis, S., Maria, E. A., & Shu, K. (2018). A definition of bioeconomy through bibliometric networks of the scientific literature. AgBioForum, 21(2), 64-85.

14. Maciejczak, M. (2018). Quality as Value-added Bioeconomy: Analysis of the EU Policies and Empirical Evidence from Polish Agriculture. AgBioForum, 21(2), 86-96.