COSE474-2023F: Final Project
Landmark Image Classification AI in Seoul

SunKyung Moon

1. Introduction

1.1. Motivation

After the COVID-19 pandemic, the Seoul Metropolitan Government is making large-scale infrastructure investments to revitalize the tourism industry. Several technical attempts are being accompanied, such as installing a ”Seoul-type blue plaque” using smart technology in major tourist facilities to allow tourists to interestingly access the history and culture stories of Seoul. The necessity of an AI classification algorithm that tags which landmark only with the image of each place was found here so that tourists can take major landmarks of Seoul tourism and immediately obtain related information.

1.2. Problem definition

By training the Seoul landmark image data set and the label representing the name of the landmark on the model, we want to create an image classification model that can tag the label for the new landmark image. The data set contains a total of 10 places where tourists visit in Seoul, including Bukchon Hanok Village, Seoul Forest, Lotte World, and Lotte Tower. The label is integer data representing the names of each landmark from 0 to 9. If you put a test image in the model after the training, you should be able to tag the appropriate label for each image.

1.3. Concise description of contriubtion

Among the Transfer Learning models, the ResNet152V2 model was used to approach the problem. The ResNet152V2 model is known to effectively lower errors even though it has a deep neural network structure. In particular, it contributes to effectively classifying and recognizing various landmarks in Seoul by allowing the model to acquire specialized knowledge about specific landmarks necessary for image classification work through transfer learning.

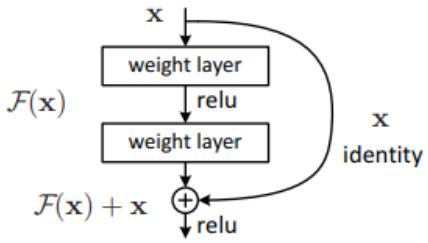

Figure 1. Residual Block

2. Methods

2.1. challenges

In the process of selecting the candidate model, there were many concerns among VGG, ResNet, and CNN classifier. As a result of an in-depth examination of each model parameter and structure, it was concluded that the VGG and ResNet models that increase the accuracy of image classification by increasing the depth of the layer even though it is a CNN architecture base were selected. Since then, considering the limited computer resources, it was determined that ResNet, which reduced complexity through the visible block, was appropriate, which is deeper than VGGG, which has too many parameters.

2.2. Main figure : ResNet

The main concept of the ResNet model is to preserve the information x learned in the previous layer and learn only the physical information of the next layer.

In figure 1, shortcut connection x is simply added, and the structure of learning only F(x) can be confirmed. In other words, the layer layer is grown while reducing complexity by simply adding it to the shortcut connection by explicitly learning only two new conv layers, which are financial information.

$$y = F(x, W_i) + W_s x \tag{1}$$

In the ResNet model, it address the degradation problem by introducing a deep residual learning framework. in most cases, only x is added to the output by identity mapping,

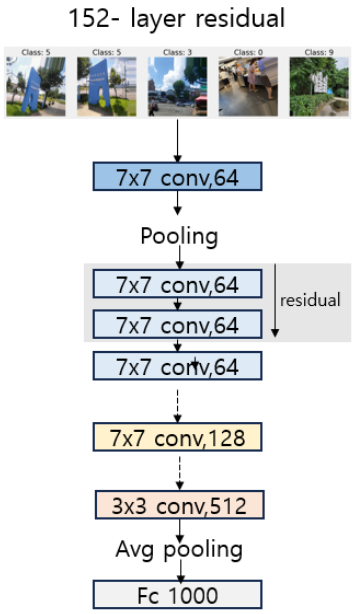

Figure 2. ResNet(input image as Landmarks)

except for the part where the size of the conv layer changes and adjusts x. In this task, ReNet152v2, which has the largest number of layers among ResNet models, was used. The predicted classification model is as figure2.

3. Experiments

3.1. Setting

3.1.1. Data

- train image : (001.PNG 723.PNG)
- test image : (001.PNG 199.PNG)
- train.csv : Data containing correct answer labels by image file name
- test.csv : Data containing only the file name of the image

3.1.2. Data preprocess

The correct answer label is from 0 to 9, and a total of 10 types of landmarks must be classified. The distribution of classes 0 to 9 is close to the uniform distribution as shown in the Figure 3, so a separate process to match the number of classes was not required.

The number of training data was 723, which was small, so augmentation was performed. (See below) The problem of performance deterioration and overfitting can occur when layers are stacked deep in a small image size is raised in the ResNet paper, so the image size was resized to 500by500.

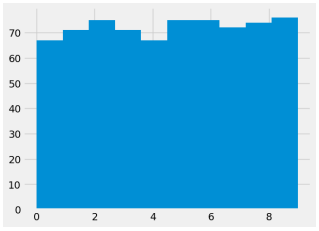

Figure 3. distribution in class(each number means label of Landmark

Figure 4. data augmentaion

3.1.3. Computing resource

Use colab V100 GPU and high capacity RAM

3.1.4. Experimental design/setup

Set to optimizer=Adam(), loss='catechical crosscentropy', metrics=['acc'], and train the ResNet152V2 model with epoch = 20. and batch = 32. After that, we would like to evaluate the performance of the model by comparing it with the DACON leaderboard scores that provide the original data.

4. Results

4.1. Results According to Epoch

Figure 5 is the result of each epoch learned under the conditions of epoch = 20 and batch = 32, set to optimizer = Adam(), loss = 'categorical cross-entropy', metrics = ['acc']. As the epoch increases, the acc (accuracy) score improves, but when the epoch moves from 19 to 20, it is possible to observe the point where the performance decreases further. Therefore, it can be concluded that epoch 19 is most appropriate for training the model in this problem.

4.2. validation, training error plot

For a total of 20 epochs, validation account and training account are as follows. As confirmed in the table (figure), while the epoch increases, the account rises and the loss goes down, but from 20, the loss slightly increases in validation, and the account also slightly decreases. Overfitting, which should be noted in models with deep layers, was not confirmed on the graph.(see Figure 6.)

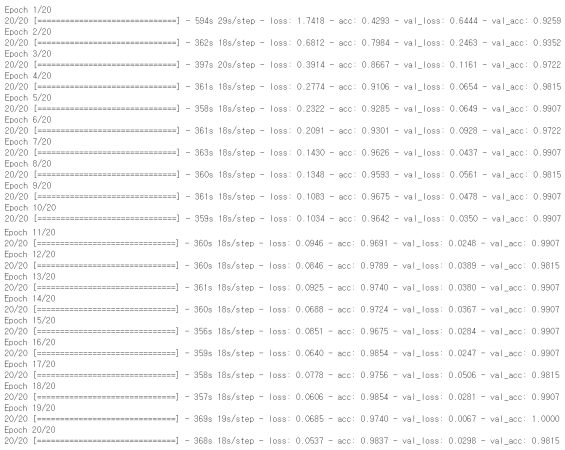

Figure 5. Result According to Epoch

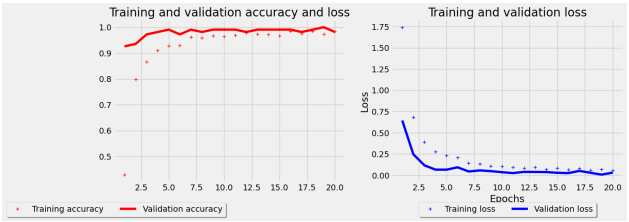

Figure 6. train,validation plot

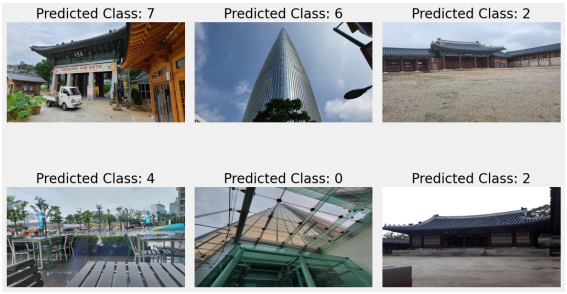

Figure 7. predict results example

method	top-1 err.	top-5 err.
VGG [41] (ILSVRC'14)	-	8.43†
GoogLeNet [44] (ILSVRC'14)	-	7.89
VGG [41] (v5)	24.4	7.1
PReLU-net [13]	21.59	5.71
BN-inception [16]	21.99	5.81
ResNet-34 B	21.84	5.71
ResNet-34 C	21.53	5.60
ResNet-50	20.74	5.25
ResNet-101	19.87	4.60
ResNet-152	**19.38**	**4.49**

Figure 8. sota : ResNet (Error ratesof single-model results on the ImageNet validation set)

4.3. Predict and Scores

The prediction result can be seen in figure 7. Number 6 was the label meaning Lotte Tower, and as can be seen from the result image, it can be seen that it has been classified successfully. As for the DACON score scored by making the predicted label into a csv file, a perfect score of 1 between 0 and 1 was found. Referring to the fact that the top-5 error of the SOTA result of the ResNet paper was 4.49, the number of original data itself is very small, 700 units, and the number of labels to be classified is 10 units. When referring to the validation and training error plot graph, it cannot be judged as overfitting, but it is too early to judge that it is a good performance because the ability of the model to cope with an image other than the correct answer cannot be seen.

4.4. Future direction

Through this project, various landmark images of Seoul have been classified and used for the ongoing smart tourism business. However, the score was perfect because the number of train images was 723, and all 723 were evenly distributed with values corresponding to the correct answer between labels 0 and 9. Although it cannot be said that the direction of problem solving was wrong, it was judged that it was not a practical model in a situation where all the various real-world landmark image data and non-correct data were included.

Therefore, in the future, the task of increasing the number of labels to 30 and extracting more various types of landmark images through Google image scraping remains to be developed into a more practical model. In addition, referring to the ResNet paper, there is a record that the error rate drops to 3.57 when performing sensorble learning. Therefore, in the process of developing into a wide range of landmark classification models in the future, it can be concluded that it is necessary to consider sensorble learning together to deal with vast and diverse landmark image data sets.

References

Kaiming He, Xiangyu Zhang, S. R. J. S. Deep residual learning for image recognition, 2016.

Karen Simonyan, A. Z. Very deep convolutional networks for large-scale image recognition, 2014.

Na, D. Careful deep learning paper reviews and code practice: Deep learning paper review and prac-

tice. 2021. URL https://github.com/ndb796/Deep-Learning-Paper-Review-and-Practice.

Sharma, N., Jain, V., and Mishra, A. An analysis of convolutional neural networks for image classification. *Procedia Computer Science*, 132:377–384, 2018. ISSN 1877-0509. doi: https://doi.org/10.1016/j.procs.2018.05.198. URL https://www.sciencedirect.com/science/article/pii/S1877050918309335. International Conference on Computational Intelligence and Data Science.

Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., Erhan, D., Vanhoucke, V., and Rabinovich, A. Going deeper with convolutions. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 1–9, 2015.

(Sharma et al., 2018) (Na, 2021) (Kaiming He, 2016) (Karen Simonyan, 2014) (Szegedy et al., 2015)