Rock-Scissors-Paper classifier model

20200045 Kimgeonwoo

(a) Description of the algorithm and the hyperparameters used

As a CNN-based model, the model was composed of convolution layer, batch normalization, max pooling, fc, etc. The role of each layer is as follows.

conv layer : captures the features of the image

batch normalization: increases model stability and accelerates learning speed

max pooling : Downsamples spatial dimensions to reduce the size of feature maps

fc : perform final prediction through extracted features

Hyperparameters: Hyperparameters were set to prevent overfitting in the learning process.

(n_ephoc: 3, lr = 0.01)

The model was completed by adjusting the number of channels in each layer.

(b) Rationale for your choice of algorithm

Our goal can be called image classification. Therefore, the model adopted cnn. When classifying images, we should not look at the whole picture, but at the object we want to classify. To be more precise, features that can distinguish objects should be extracted and images should be classified through that information. When learning the picture as a whole (e.g., MLP), there are too many parameters that can be overfit. In addition, even though they are the same cat, the case located at the upper right and the case located at the lower left can be recognized differently. This is why the image classification model must learn the characteristics of the object, not the location information.

CNN repeatedly performs Convolution and Pooling, which mimic the structure of a human optic nerve, to extract features and perform differential learning to classify an image input through fully connected layers. Therefore, it can be said to be a model suitable for image classification.

Additionally, layers such as batch normalization were added to this, resulting in higher performance.

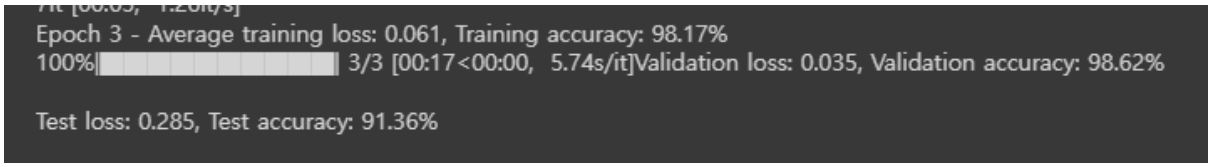
(c) Final performance (train, validation, test) of your code

for dataset1(Archive 1)

train| average loss : 0.061 acc: 98.17%

val | average loss : 0.035 acc: 98.62%

test| average loss : 0.285 acc: 91.36%

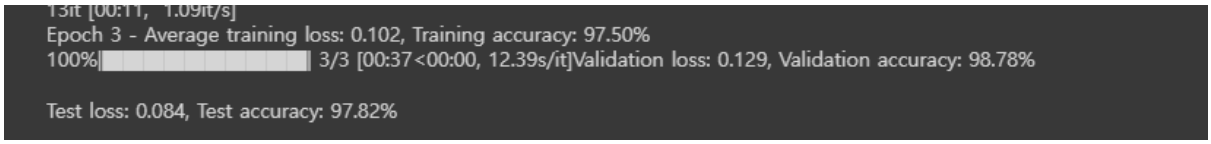


for dataset 2(Archive 2) **+ dataset1**

train| average loss : 0.102 acc: 97.5%

val| average loss : 0.129 acc: 98.78%

test| average loss : 0.084 acc: 97.82%



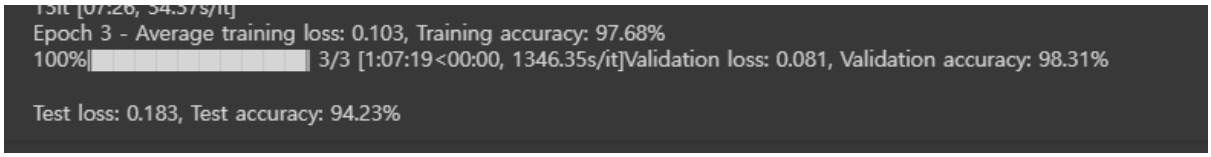
dataset 3(customized) + data set2 + dataset1

Dataset3 is a self-generated data. A total of 2,746, including 937 scissors, 900 rocks, and 909 beams, were used to train data.

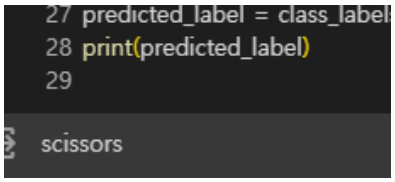
train| average loss : 0.103 acc: 97.68%

val| average loss : 0.081 acc: 98.31%

test| average loss : 0.183 acc: 94.23%



The given test picture was classified well as scissors.



(d) Any other changes to the code you think is worth mentioning

It showed better performance when all data were divided and learned than when all data were learned at once. However, as learning progressed, catastrophic forgetting seemed to occur. Therefore, a strategy was used to learn some of the previously learned datasets together. The entire dataset was mixed and learned.

(e) Citations

https://medium.com/ddiddu-log/%EC%9D%B4%EB%AF%B8%EC%A7%80-%EC%9D%B8%EC%8B%9D%EC%9D%98-%EC%A0%95%EC%9D%98%EC%99%80-%EC%A3%BC%EC%9A%94-%EB%AA%A8%EB%8D%B8-%EB%B9%84%EA%B5%901-%EC%9D%B4%EB%AF%B8%EC%A7%80-%EB%B6%84%EB%A5%98-image-classification-ae7a59bfaf65

https://seoilgun.medium.com/cnn%EC%9D%98-stationarity%EC%99%80-locality-610166700979