Good morning every one! Today I want to share you about my program, “Life Long Learning for Handwritten Digits Recognition”.

With the rapid development of deep learning, the accuracy of handwritten digit recognition was pushed very high, even almost reaching 100%. It is difficult to make a breakthrough in recognizing categories of handwritten digits, but there are still many topics related to handwritten digit recognition that can be studied.

Therefore, I proposed a new dataset in which 4 volunteers write digits with 7 virtual different pens using a software in iPad and Apple Pencil. And I also provide a csv file containing 3 kinds of annotations: class label, pen type and writer name.

For data collection, every writer needs to write one digit using 7 different kinds of pen for 3 times, so the dataset has in total 840 images, and all the images has a size of 410\*410. There are some samples showing the diversity of my dataset. Every row is written by one person, and we can see the difference between different types of pens and different writer’s habits. Even for the same writer, sometimes the same digit is written differently. What is more interesting is the last column, although the same parameters are controlled when using the software, different people have different writing habits and stress, so the effects of writing with the same pen are also different. I think this dataset can help us to do more interesting tasks rather than simple classification.

Based on this dataset, I implement life long learning and curriculum learning in practice.

Life long learning wants to use a unified network structure to train separately on different tasks, and the network can also solve all the tasks. Curriculum Learning focuses on the relationship between each task to find the proper learning order, because the order of tasks will affect the performance of the model.

Here I divide the dataset according to different writers, and obtain 4 different sub-tasks. Then use the elastic weight consolidation method and sufficient experiments to check the effectiveness of life long learning.

For image classification, I choose ResNet [2] pretrained on ImageNet as the base network. For the sake of brevity I will not go into depth about resnet. And for life long learning, I use elastic weight consolidation method (EWC [3]). In the loss function, b\_i is a guard, indicating the importance of parameter θi to the previous task. And the basic idea of EWC is that only change the parameters unimportant to previous tasks,

As for the experiment, I define 3 baselines. The first one is Multi-task Learning, which trains ResNet with the whole dataset as upper bound. And then train the resnet continuously using the 4 sub-tasks with and without EWC loss.

I use 3 metrics as in the paper of GEM, Average Accuracy Measures the overall performance after learning all tasks on the test set. Backward Transfer Measures the degree of forgetting on previous task after training on the last task. Forward Transfer Measures how the learned knowledge on previous tasks affects the performance on task Ti before learning task Ti.

There are the experiment settings.

Next is the experiment results, from this picture we can see that the previous task can indeed help the following ones.

And this picture shows the process of life long learning. It’s obvious that Training on previous task can improve the performance on future tasks, and training on later task doesn’t affect performance on previous tasks.

Here is the experiment results on the three metrics. In order to get the accurate results, I repeat the experiment for each setting for 5 times, and get the average values and range.

Life Long Learning: EWC indeed improves ACC for all the tasks. But for BWT and FWT, EWC methods only outperforms the basic one in the last task order, which means that EWC is not so efficient on handling the forgetting problem and knowledge transferring in this dataset.

Curriculum Learning: Train the dataset with meaningful order (Shen-Peng-Cheng-Wang), EWC method can handle the forgetting problem and knowledge transfer- ring to some extent. Besides, this order achieved the best results in average accuracy.