# Homework#4

## Due in Two Weeks (1/8)

**Problem :**

Please complete a CNN model- Lenet-5, which is shown in Fig1. The model has eight layer. Each layer is explained below.

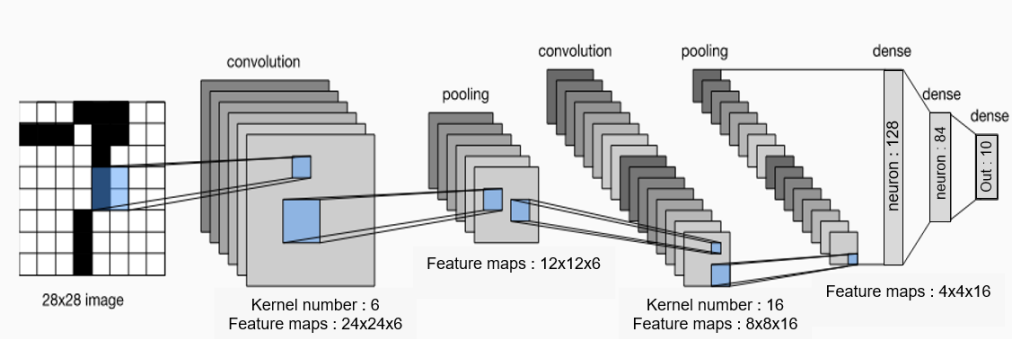


Fig.1 Lenet-5 model architecture

1. Input layer :

The input data are 28x28 images which are from MNIST. We provide 10 input image (number 0 to number 9, each number has one image) that you can test. You need to read the input data by yourself.

2. First convolution layer :

There are six 5x5 kernels that you need to do the convolution operation six times. Hence, there are six output feature maps. The stride is 1 and the activation function is ReLu in this layer.

3. First pooling layer :

The pooling operation is used to compress the feature maps. The first pooling layer adopt max pooling which is to find the maximum value of the input data within the scope of the kernel. The kernel size is 2x2 and the stride is 2. The pooling operation is shown as Fig. 2.

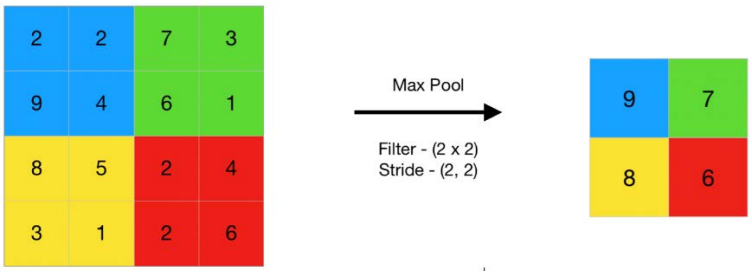


Fig.2 pooling operation

4. Second convolution layer :

In convolution operation, the kernel size is depended on the input data size. Although it is 2 dimension convolution, the kernel size will expand to the same depth as the input data. Besides, the depth of output data is depended on the number of the kernel. For example, In this layer, the input data size is 12x12x6, because the depth of feature maps is 6, the kernel size is 5x5x6. In addition, the number of kernel is 16, so the number of output feature is 16. The convolution operation is shown in the Fig. 3 and Fig. 4. The stride is 1 and the activation function is ReLu in this layer.

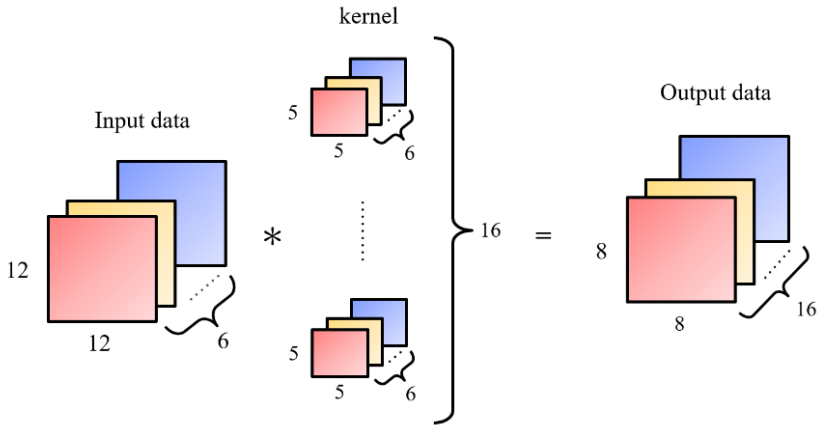


Fig.3 Second convolution layer

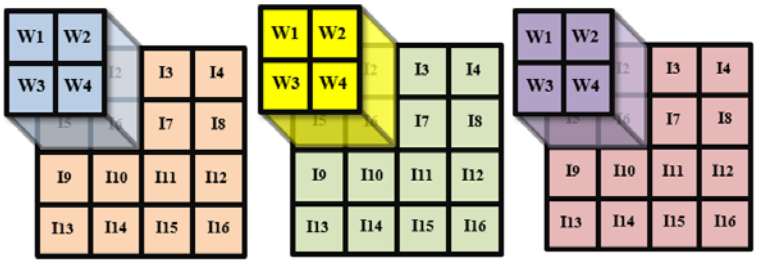


Fig.4 Depth convolution layer operation

5. Second pooling layer :

The operation of this layer is the same as the first pooling layer.

6. First fully connected layer :

Before doing the first fully connected layer operation, it is necessary to flatten the output feature maps from previous layer. The order of input data is arranged from the first feature. The order of each feature map is from left to right, from top to down, as shown in Fig. 5. The operation of fully connected layer is the same as homework2. There are 128 neurons in this layer and the activation function is ReLu.

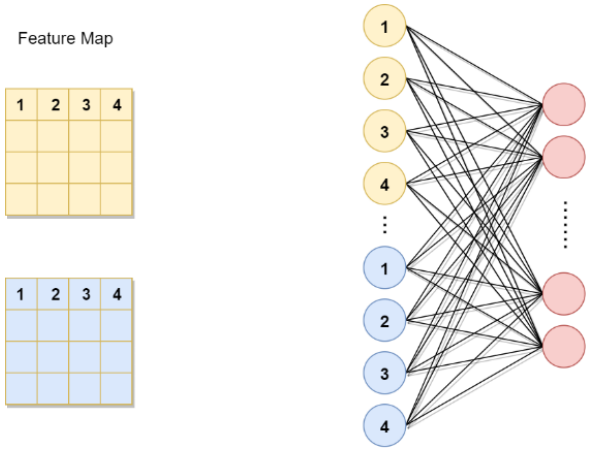


Fig.5 Flatten

7. Second fully connected layer :

There are 84 neurons in this layer and the activation function is ReLu. The operation of this layer is the same as the first fully connected layer.

8. Output layer :

There are 10 neurons in this layer but it has no activation function. The operation of this layer is the same as the first fully connected layer. The value of these neurons is the output data.

The system block is shown as Fig. 5. The ROM is used to store the weight of CNN model. The order of weigh is arranged from the first kernel. The order of each kernel is first horizontal, then vertical and finally depth, as shown in Fig. 6. The order of fully connected layer weigh is arranged from the first neuron, as shown in Fig. 7. We provide a RAM that you can store the temporary data. The RAM has 16KB entry. The RAM can read/write data. When the ram_wr becomes 0, you can write data to RAM. When the ram_wr becomes to 1, you can read data from RAM. The Monitor is used to print the output value. When the output_valid signal change to 1, the result is valid. If the output_valid signal becomes 1 to 10 times, the simulation will be end.

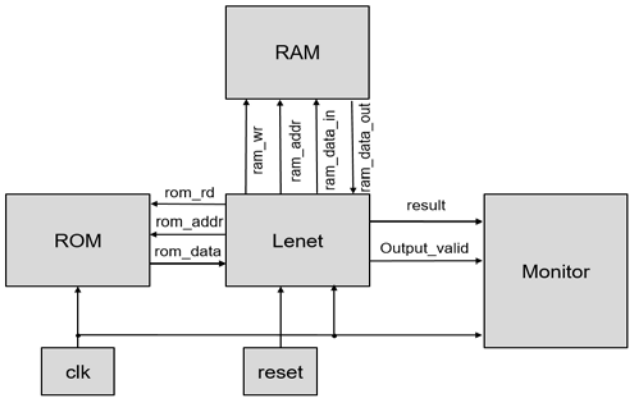


Fig.6 System block

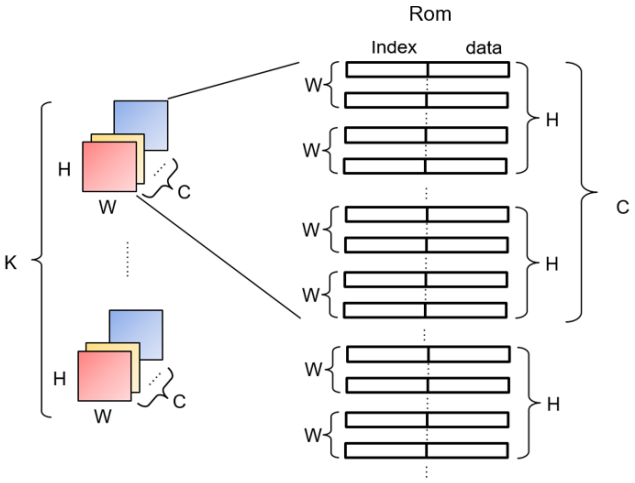


Fig.7 Kernel weight in ROM

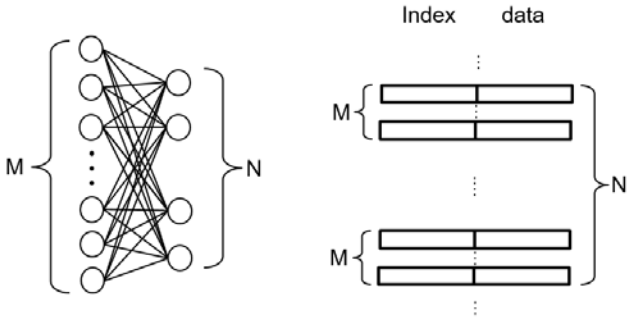


Fig.8 Fully connected layer weight in ROM

Finally, the run_mode.h file is used to determine which data type declaration you want to and which set of input data you want. You can set the value of MODE and NUM.

Requirement:

1. Complete two version of the homework4 (Floating point and fixed point)
2. You only can modify the Lenet block.
3. Please use the Platform Architect (PA) to simulate the system and show the simulation result.
4. Report (Word file Include simulation result of terminal and PA and the design concept. You can explain the advantage of your design for bonus)

Please upload the source code, PA project and report file to your server account (Path: /Desktop/(Student ID #)_(HW4)/)