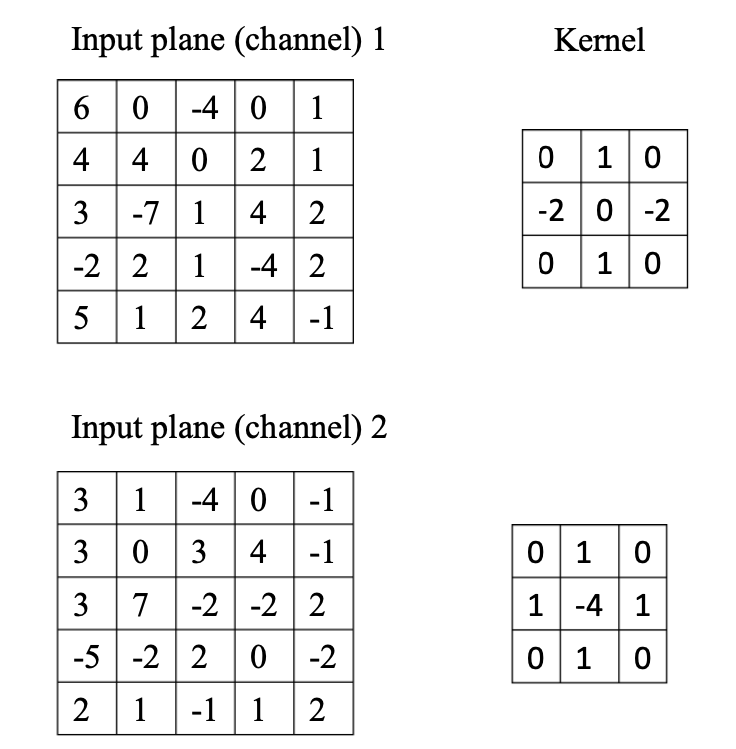
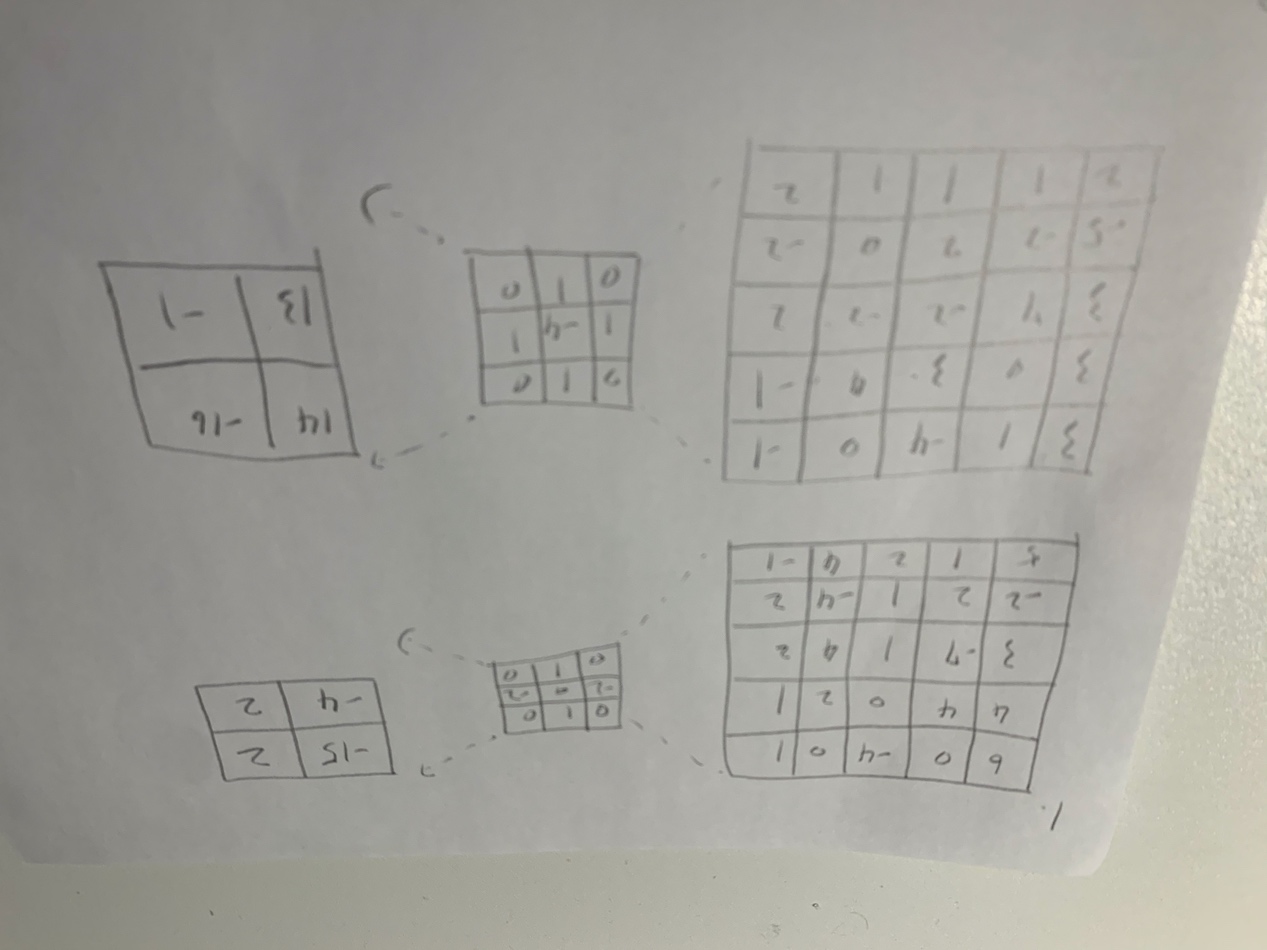
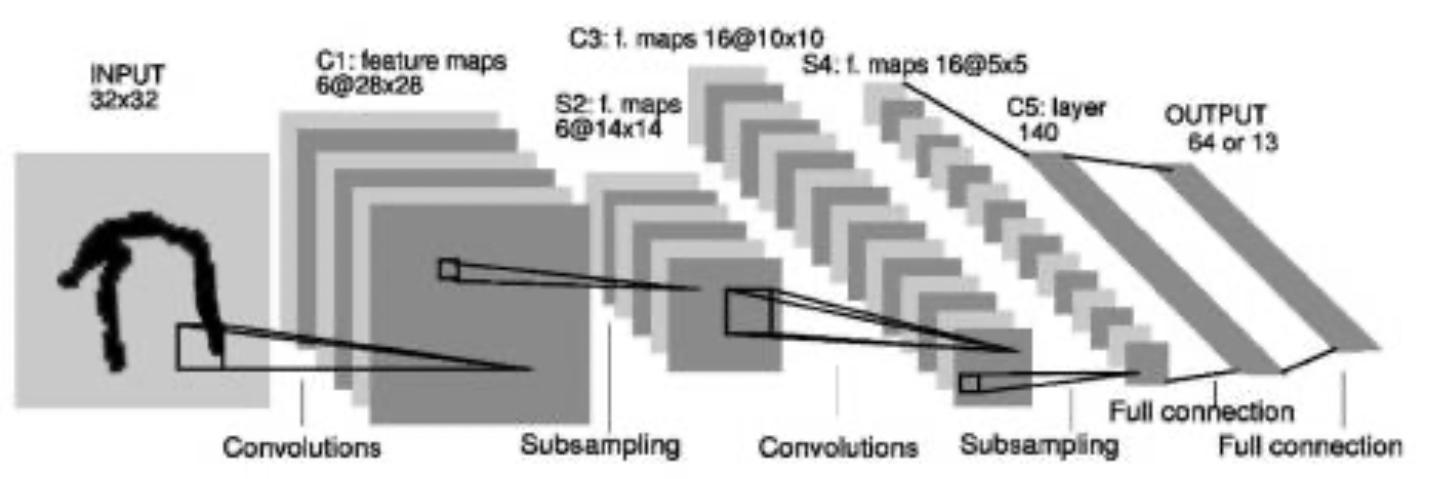
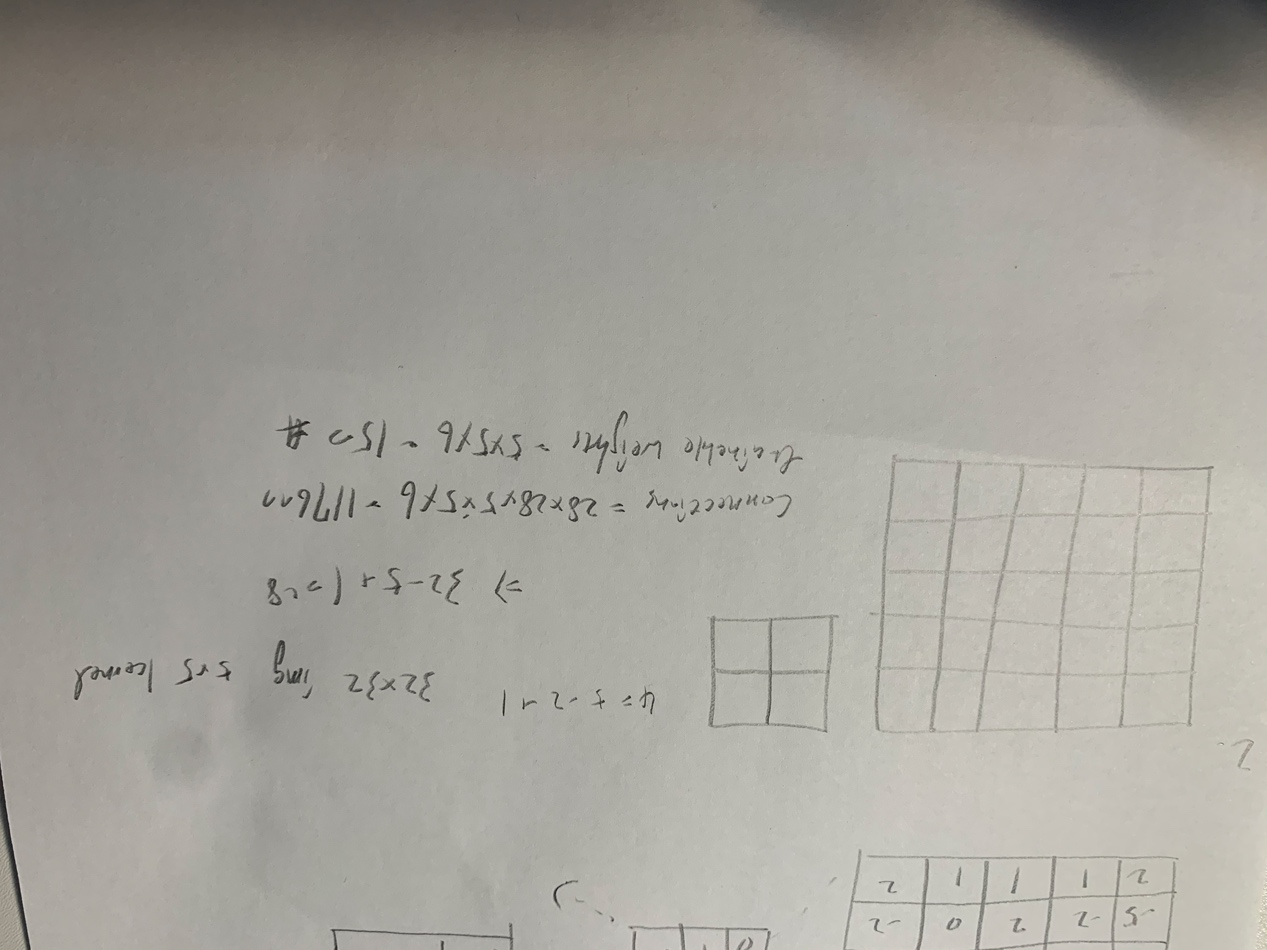
HW 5 姓名：胡喻翔 學號：108598047

1. Assuming that the following is a part of convolution neural networks. Compute the resultant values if it has two input planes (channels), stride of two, no zero-padding, and using the ReLU activation function. Remember that the output from this part of network has only ONE value.





1. The following shows the LeNet-5 architecture. If we use the weight-sharing approach for the convolutional layers, compute the number of connections and trainable weights of the network. To compute the results, you need the following parameters for the convolutional layers: kernel size = 5×5, stride = 1, and no zero-padding. To simplify the computations, ignore the bias weights and let the output units be 10.

1. We mentioned bias and variance dilemma before. If somehow you determine that your fully-connected neural network has the high-bias problem, what would you do? If, on the other hand, your CNN has the high-variance problem, what would you do then?   
   Ans:   
   We can try to add some parameters to increase the complexity of the model; add more training data or improve the representation of the variance of parameters in the population; dropout a certain percentage of the neurons in every training iterations; or we can decrease the regularization of the parameter, by decreasing the performance ratio, we can reduce the bias.
2. One widely used technique to train very deep CNN is called ResNet. Explain how it works.

Ans: The most important concept of ResNet is “skip connection”. Skip connections mitigate the problem of vanishing gradient by allowing this alternate shortcut path for gradient to flow through; Skip connections also allow the model to learn an identity function which ensures that the higher layer will perform at least as good as the lower layer, and not worse. Ideally, if the model of a neural network is the best, we can, then easily optimize the residual mapping to 0, and we’ll got identity mapping remaining only. As the result, no matter how many layers we got, the network will keep at the best status.

1. Find Δ𝑤1 = 𝜂 𝜕𝐽 of the following CNN using backprop. The activation 𝜕𝑤1 function from 𝑞1 to h1 and 𝑞2 to h2 is ReLU, the outputs 𝑦1 and 𝑦2 are softmax ouput, and the cost function is 𝐽 = − log 𝑦1. Let 𝑤1 to 𝑤6 be 1.0, 𝑥1 =0.5, 𝑥2 =1.0, 𝑥3 =-0.5, and 𝜂=0.1.

