National University of Singapore
School of Computing
IT5005 Artificial Intelligence

Convolutional Neural Networks

1. **ConvNets**

   (a) You are given an image $\mathbf{x} \in \mathbb{R}^{4\times4}$ and a filter $\mathbf{W} \in \mathbb{R}^{3\times3}$. **No** extra padding is added.

0.5	0.2	0.1	0.7
0.1	0.6	0.9	0.5
0.0	0.8	0.2	0.7
0.2	0.4	0.0	0.4

0.1	0.2	0.6
0.4	0.3	0.5
0.9	0.8	0.7

Figure 1: (left) The image $\mathbf{x}$. (right) The filter $\mathbf{W}$.

Get the output feature map from this convolution and write down its output dimensions if the kernel moved with a stride of 1×1. No Padding or Pooling operation required. Note that while mathematically, there is a need to flip the kernel matrix, in practice when using CNNs we do not flip the kernel. This saves on the cost of flipping while making little practical difference. For this question, we compute the output feature map with no flipping.

(b) Now, let's say you have access to a very deep, large CNN model. We feed a single image to the network. Each image has $3(C)$ channels (RGB) with a height and width of 224×224 ($H = 224, W = 224$). Here our input is a 3-dimensional tensor ($H\times W\times C$). The first layer of the big CNN is a Convolutional Layer with 96 kernels which has a height and width of 11, and each kernel has the same number of channels as the input channel. The stride is 4×4 and no padding is used. Calculate the output size $H_1 \times W_1 \times C_1$ after the first Convolutional Layer.