CENG4480 Homework 1

Due: 11:59PM, Oct. 21, 2025

All solutions should be submitted to the blackboard in the format of **PDF**.

Q1 (10%) Suppose we have a feature map of size $H \times W$, a convolution kernel with size $m \times n$, stride s. The channel number of the feature map is 1, and there's no zero padding.

1. Suppose $H = W = 5, m = n = 3, s = 1$, if we use vanilla img2col function to map this feature map to a lowered matrix, what will be the size of the lowered matrix?
2. Suppose $H = W = 5, m = n = 3, s = 1$, if we use memory-efficient convolution (MEC) to map this feature map to a lowered matrix, what will be the size of the lowered matrix? How much memory saving can we obtain?

Q2 (25%) Img2col is a key operation that transforms convolution operations into General Matrix Multiply (GEMM) operations. By rearranging image data into columns, it allows convolutions to be computed as matrix multiplications. $X \in \mathbb{R}^{H \times W \times C_i}$ is the original input feature map, $W \in \mathbb{R}^{K \times K \times C_i \times C_o}$. Write down the following two implementations of img2col in C++ language style. We set $H = W = 5$, $C_i = 3$, $C_o = 6$, $K = 3$, stride = 1, padding = 0.

(a) The convolution will be $X' \times W'$ after transformation, where X' is the lowered feature map and W' is the lowered weight matrix.

(b) The convolution will be $W' \times X'$ after transformation, where X' is the lowered feature map and W' is the lowered weight matrix.

Q3 (30%) Convolution is the most important operation in CNN. As shown in Figure 1, the input activation tensor is $\mathcal{X} \in \mathbb{R}^{H \times W \times C}$. Weight tensor is $\mathcal{W} \in \mathbb{R}^{R \times S \times C \times K}$. The output activation tensor is $\mathcal{Y} \in \mathbb{R}^{P \times Q \times K}$. Here we set $H = W = 5$, $C = 6$, $R = S = 3$, $K = 4$. Besides, the stride number is 1 and the padding number is 0.

(a) What are the height and width of the output feature map (P and Q)?

(b) Write down direct convolution by C++ language style.

Q4 (35%) The input feature map is shown in Figure 2, where P1, P2 and P3 has pixel value of 1 in 3 channels. For the convolution kernel settings, the kernel size is 3, number of output channel is 2, stride is 1 and no padding. Submanifold sparse convolution is used to speed up the computation.

(a) Build the input table H_{in}.

(b) Generate the output hash table H_{out}.

(c) Construct the offset map and the rule book.

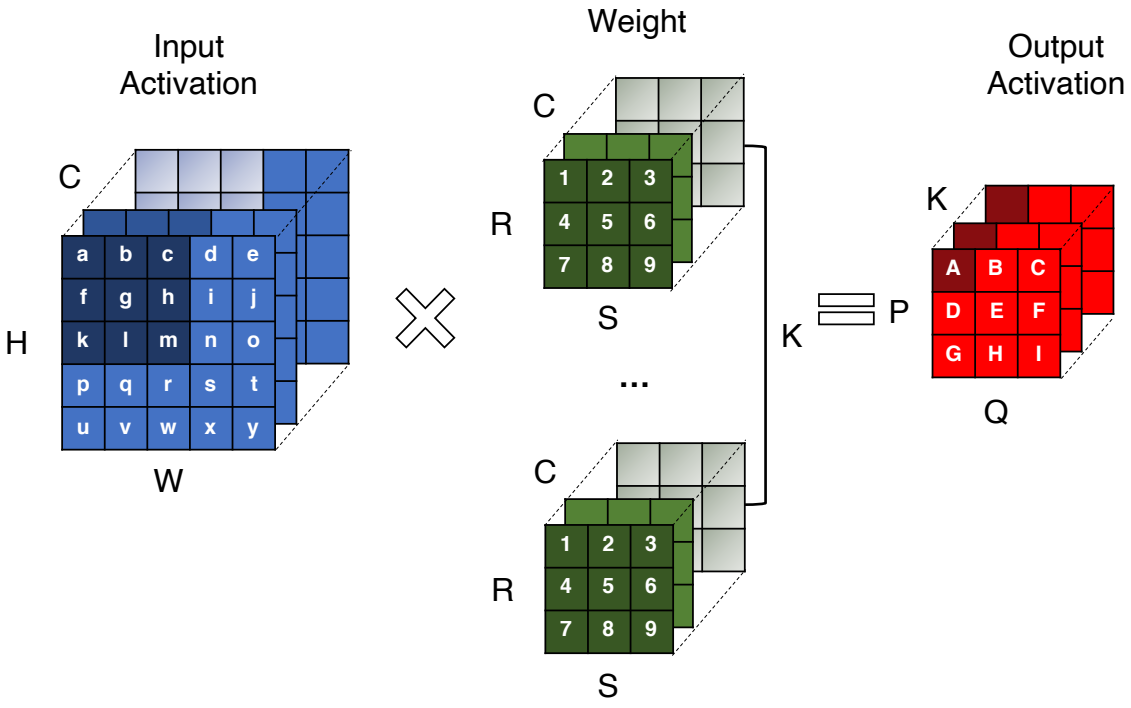

Figure 1: Convolution.

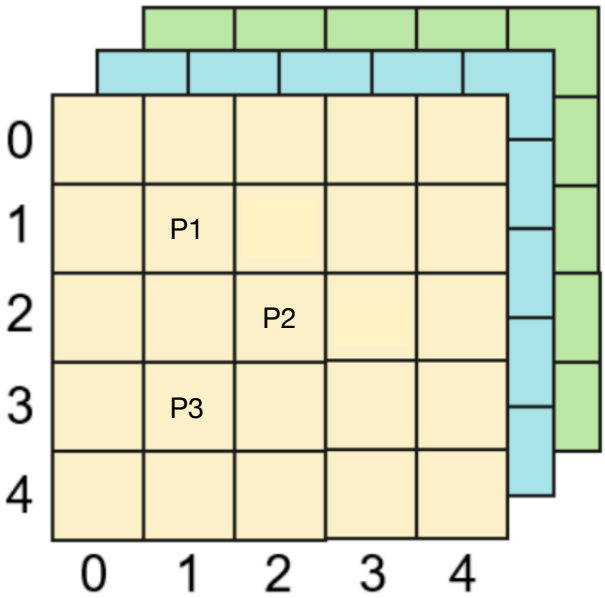

Figure 2: Q4 Input feature map