# 1 Convolutional Neural Network

## 1.1 Convolution Operation of Matrix

We denote the convolution operation as the following

$$Z = A * W + b$$

Where $A$, $W$, $b$ are matrices with proper dimension. The star $*$ represent the convolution operation of the two matrix $A$ and $W$. We define the matrix $A$ with dimension $I^{in} \times J^{in}$. The matrix $W$ has the dimension $P \times Q$ and we assume $P \leq I^{in}$, $Q \leq J^{in}$. The matrix $Z$ has the dimension $I^{out} \times J^{out}$ . And b is a scalar constant. The convolution operation is defined as

$$z_{ij} = \sum_{p=i}^{i+P} \sum_{q=j}^{j+Q} a_{pq} w_{pq} + b$$

The input dimension and output dimension have the following relationship

$$\begin{aligned} P &= I^{in} - I^{out} + 1 \\ Q &= J^{in} - J^{out} + 1 \end{aligned}$$

Namely,

$$\begin{aligned} I^{out} &= I^{in} - P + 1 \\ J^{out} &= J^{in} - Q + 1 \end{aligned}$$

An intuitive way of thinking the above operation is we put matrix W on top of matrix A and align the top left corner, meaning that $a_{11}$ aligns with $w_{11}$. And for each overlapped element of matrix A and Z, we multiply them and get each product. Finally we sum all the product, add the bias b, then we get $z_{11}$. To get $z_{12}$, we still put W on top of A, but this time we align $a_{12}$ with $w_{11}$. Then we follow the same procedure to calculate the product and summation. We keep going to slide W matrix along the matrix A by one element each time, we can get all the element of matrix Z.

## 1.2 Convolution Operation of Matrix with Stride

In the previous convolution operation, we slide W matrix along the matrix A by one element each time. This means if the first step we place $w_{11}$ on top of $a_{11}$, next time we move W horizontally and place $w_{11}$ on top of $a_{12}$. In the convolution operation with stride, we slide W matrix by several element each time. We call the slide length stride. For example, if $stride = 2$, then after we place $w_{11}$ on the top of $a_{11}$, we place $w_{11}$ on top of $a_{13}$.
The convolution operation is defined as

$$z_{ij} = \sum_{p=i'}^{i'+P} \sum_{q=j'}^{j'+Q} a_{pq} w_{pq} + b$$

Where

$$i^{'} = (i-1) * stride + 1$$
$$j^{'} = (j-1) * stride + 1$$

And the output dimension is

$$I^{out} = \frac{I^{in} - P + 1}{stride}$$
$$J^{out} = \frac{J^{in} - Q + 1}{stride}$$

## 1.3 Convolution Operation of Tensors

We denote the convolution operation as the following(same as above)

$$Z = A * W + b$$

Where $A$, $W$, $b$ are tensors with proper dimension. The star $*$ represent the convolution operation of the two matrix $A$ and $W$. We define the tensor $A$ with dimension $I^{in} \times J^{in} \times K$. The tensor $W$ has the dimension $P \times Q \times K \times N$. The tensor $Z$ has the dimension $I^{out} \times J^{out} \times N$. And $b$ is a vector of size $N$. The convolution operation without stride is defined as

$$z_{ijn} = \sum_{p=i}^{i+P} \sum_{q=j}^{j+Q} \sum_{k=1}^{K} a_{pqk} w_{pqk}^{(n)} + b^{(n)}$$

**Example** Take $A$ as a $2 \times 2 \times 3$ tensor, and $W$ as a $2 \times 2 \times 3 \times 2$ tensor, then

$$\begin{aligned} z_{111} &= a_{111} w_{111}^{(1)} + a_{121} w_{121}^{(1)} + a_{211} w_{211}^{(1)} + a_{221} w_{221}^{(1)} \\ &+ a_{112} w_{112}^{(1)} + a_{122} w_{122}^{(1)} + a_{212} w_{212}^{(1)} + a_{222} w_{222}^{(1)} \\ &+ a_{113} w_{113}^{(1)} + a_{123} w_{123}^{(1)} + a_{213} w_{213}^{(1)} + a_{223} w_{223}^{(1)} \\ &+ b^1 \end{aligned}$$

$$\begin{aligned} z_{112} &= a_{111} w_{111}^{(2)} + a_{121} w_{121}^{(2)} + a_{211} w_{211}^{(2)} + a_{221} w_{221}^{(2)} \\ &+ a_{112} w_{112}^{(2)} + a_{122} w_{122}^{(2)} + a_{212} w_{212}^{(2)} + a_{222} w_{222}^{(2)} \\ &+ a_{113} w_{113}^{(2)} + a_{123} w_{123}^{(2)} + a_{213} w_{213}^{(2)} + a_{223} w_{223}^{(2)} \\ &+ b^1 \end{aligned}$$

The convolution operation with stride is defined as

$$z_{ijn} = \sum_{p=i^{'}}^{i^{'}+P} \sum_{q=j^{'}}^{j+Q} \sum_{k=1}^{K} a_{pqk} w_{pqk}^{(n)} + b^{(n)}$$

Where

$$i^{'} = (i-1)*stride+1$$
$$j^{'} = (j-1)*stride+1$$

## 1.4 Pooling Operation

Pooling operation is used to reduce the dimension of the matrix therefore reduce the number of parameters in the following hidden layers. The max pooling operation is defined as

$$z_{ijk} = max(a_{i^{'}j^{'}k})$$

where

$$i^{'} \in [(i-1)*stride+1, i*stride+1]$$
$$j^{'} \in [(j-1)*stride+1, j*stride+1]$$

The average pooling operation is defined as

$$z_{ijk} = average(a_{i^{'}j^{'}k})$$

And the definition of $a_{i^{'}}$ and $a_{j^{'}}$ is the same as in max pooling.