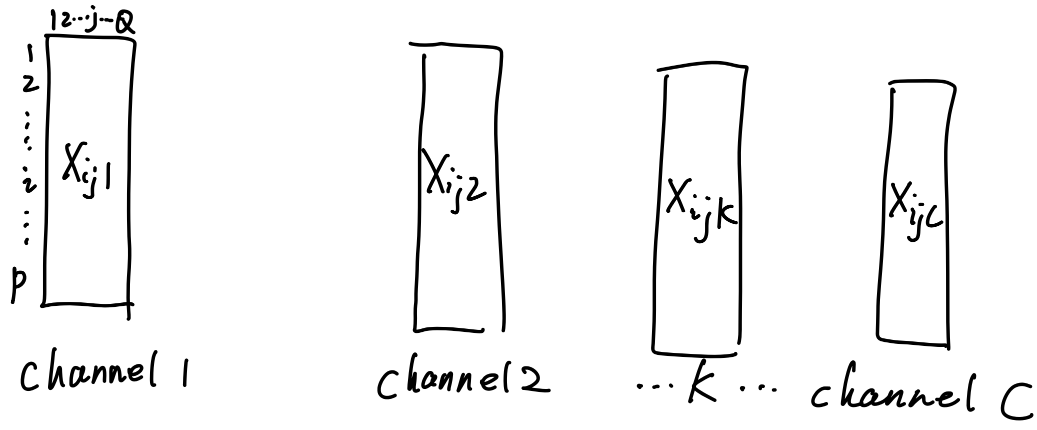
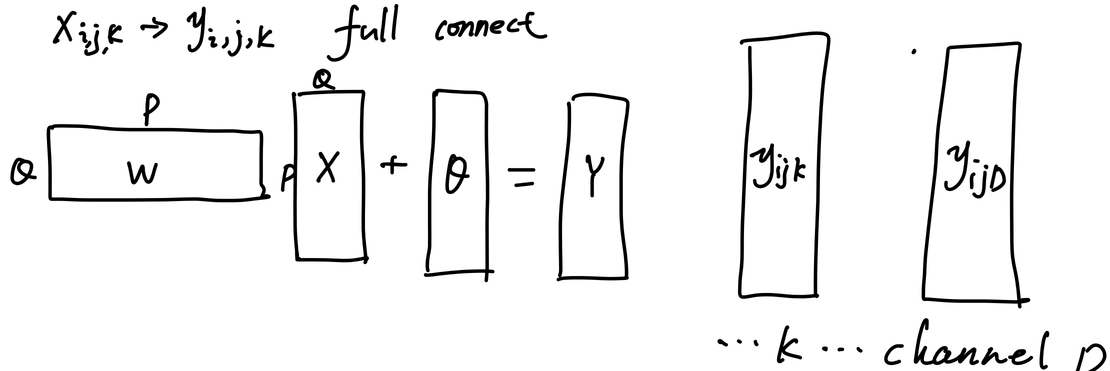
23-S2-Q4

(a) Q: $x_{ijk} \to y_{ijk}$? parameters?

Solution ① understand



$X_{ijk} \to y_{i,j,k}$ full connect



$$\sum \begin{bmatrix} W_{11} & W_{12} & \cdots & W_{1i} & \cdots & W_{1p} \\ W_{21} & W_{22} & \cdots & W_{2i} & \cdots & W_{2p} \\ \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\ W_{j1} & W_{j2} & \cdots & W_{ji} & \cdots & W_{jp} \\ \vdots & \vdots & & \vdots & \ddots & \vdots \\ W_{Q1} & W_{Q2} & & W_{Qi} & \cdots & W_{Qp} \end{bmatrix} \begin{bmatrix} X_{11} & \cdots & X_{1Q} \\ X_{21} & \cdots & X_{2Q} \\ \vdots & \ddots & \vdots \\ X_{i1} & \cdots & X_{iQ} \\ \vdots & \ddots & \vdots \\ X_{p1} & & X_{pQ} \end{bmatrix} + \begin{bmatrix} b_{11} & \cdots & b_{1Q} \\ b_{21} & \cdots & b_{2Q} \\ \vdots & \ddots & \vdots \\ b_{i1} & \cdots & b_{iQ} \\ \vdots & \ddots & \vdots \\ b_{p1} & \cdots & b_{pQ} \end{bmatrix}$$

$$y_{ijk} = \sum_{l=1}^{C} \sum_{j=1}^{Q} \sum_{i=1}^{P} W_{ji} X_{ijl} + b_{ijk} \qquad (1 \le k \le D)$$

parameters $P \times Q \times C \times D + P \times Q \times D$

一张W

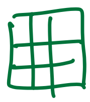
C张W生成1张Y

生成D张Y

1张bias

D张bias

(b) CNN ?



$$y_{i,j,k} = \sum_{u=-1}^{1} \sum_{v=-1}^{1} \sum_{l=1}^{C} w_{u,v,l,k} \, x_{i-u,j-v,l} + b_k$$

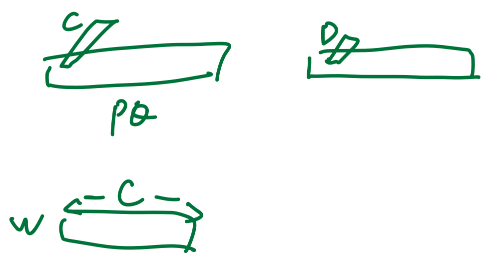
$$3 \times 3 \times C \times D + D = 9 \times C \times D + D$$

(c) $y_{i,j,k} = \sum_{l=1}^{C} w_{l,k} \, x_{i,j,l} + b_k$

$C \times D + D$

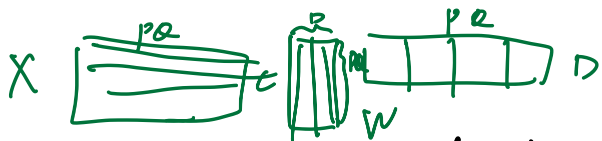
RGB D个特征图 bias

(d) $y_{i,k} = \sum_{l=1}^{C} w_{l,k} \, x_{i,l}$



$C \times D$

(e) $Y = XW$



① Both architectures apply matrix multiplications to process input

② $Y = XW$ shows the CNN perform a linear transformation on the input features at each spatial position, identical across all position.

It is similar to the linear layers used in Transformer, where input are transformed via weight matrices