EE7403 2023-2024 S2

4. (a) $y_{i,j,k} = \sum_{n=1}^{C} \sum_{m=1}^{Q} \sum_{l=1}^{P} w_{l,m,n,k} \cdot x_{l,m,n} + b_k$.

for each channel, there are $(PQC+1)$ learnable parameters

(including C $w_{k,n}$ and 1 b_k)

there's D output channels, and $P \times Q$ pixels hence:

Total number: $PQD(PQC+1) = PQDPQC + PQD$

(b) $y_{i,j,k} = \sum_{n=1}^{C} \sum_{m=-1}^{1} \sum_{l=-1}^{1} w_{l,m,n,k} \cdot x_{i-l,j-m,n} + b_k$

each channel $(3 \times 3 \times C + 1) = 9C + 1$

Total: $D(9C+1) = 9CD + D$

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

each channel $(1 \times 1 \times C + 1) = C + D$

Total: $D(C+1) = CD + D$

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

each channel: C

Total: DC.

(e) Set $W = [w_1, w_2, \dots w_C]^T \in \mathbb{R}^{C \times D}$

$Y = XW$

In Transformer, all tokens share the same learnable parameter, the learnable parameters go through all pixels, which is quite similar to CNN.