BP for convolutional layer

Jiashu Xu

October 2020

1 conv layers BP

Let I images of shape (C, H, W), and F filters of shape (FN, C, FH, FW), let Conv2D(I, F) = O (FN, OH, OW) where O is

$$O_{cxy} = \sum_{k=1}^{C} \sum_{i=1}^{FH} \sum_{j=1}^{FW} F_{c,k,i,j} I_{k,x-1+i,y-1+j},$$

for $c = 1, ..., FN$.

We can implement it using im2col, let $\phi(I)$ be im2col(I) such that $\phi(I)$ has shape $(OH \cdot OW, FH \cdot FW \cdot C)$, F_f be reshaped filters $F.reshape(-1, FN)$, thus having shape $(C \cdot FH \cdot FW, FN)$, similarily, defined O_f of shape $(OH \cdot OW, FN)$

Thus $\phi(I) \cdot F_f = O_f$, and $vec(\phi(I) \cdot F_f) = vec(O_f)$

Since $vec(AXB) = (B^T \otimes A)vec(X)$, we have

$$\begin{aligned} vec(O_f) &= vec(\phi(I)F_f I_{FN}) = (I_{FN} \otimes \phi(I))vec(F_f) && (1) \\ &= vec(I_{OH \cdot OW}\phi(I)F_f) = (F_f^T \otimes I_{OH \cdot OW})vec(\phi(I)) && (2) \end{aligned}$$

where $I_{FN}, I_{OH \cdot OW}$ identity matrix of shape same as subscript.

Then

$$\begin{aligned} \frac{\partial L}{\partial vec(F_f)} = (\frac{\partial L}{\partial vec(F_f)^T})^T &= (\frac{\partial L}{\partial vec(O_f)^T} \frac{\partial vec(O_f)}{\partial vec(F_f)^T})^T \\ &= (I_{FN} \otimes \phi(I))^T \frac{\partial L}{\partial vec(O_f)} && \text{(by Eq. (1))} \\ &= (I_{FN} \otimes \phi(I)^T) vec(\frac{\partial L}{\partial O_f}) \\ & && ((A \otimes B)^T = A^T \otimes B^T) \\ &= vec(\phi(I)^T \frac{\partial L}{\partial O_f} I_{FN}) \end{aligned}$$

thus $\frac{\partial L}{\partial F_f} = \phi(I)^T \frac{\partial L}{\partial O_f}$, note same as BP for linear layer.

Also,

$$\begin{aligned}
\frac{\partial L}{\partial vec(\phi(I))^T} &= \frac{\partial L}{\partial vec(O_f)^T}\frac{\partial vec(O_f)}{\partial vec(\phi(I))^T} \\
&= \frac{\partial L}{\partial vec(O_f)^T} \cdot (F_f^T \otimes I_{OH \cdot OW}) && \text{(by Eq. (2))} \\
&= [(F_f \otimes I_{OH \cdot OW}) vec(\frac{\partial L}{\partial O_f})]^T \\
&= vec(I_{OH \cdot OW}\frac{\partial L}{\partial O_f} F_f^T)^T
\end{aligned}$$

thus $\frac{\partial L}{\partial vec(\phi(I))} = vec(\frac{\partial L}{\partial O_f} F_f^T)$, and $\frac{\partial L}{\partial \phi(I)} = \frac{\partial L}{\partial O_f} F_f^T$, note same as BP for linear layer.