For convolutional layers, recall we had

\[
\text{vec}(S_{m,i}) = (\phi(\text{pad}(Z_{m,i}))^T \otimes I_{d_{m+1}}) \text{vec}(W^m) + \\
(1_{a^m_{\text{conv}}} b^m_{\text{conv}} \otimes I_{d_{m+1}}) b^m
\]
Thus

\[
\frac{\partial \xi_i}{\partial \text{vec}(W^m)^T} = \frac{\partial \xi_i}{\partial \text{vec}(S^{m,i})^T} \frac{\partial \text{vec}(S^{m,i})}{\partial \text{vec}(W^m)^T}
\]

\[
= \frac{\partial \xi_i}{\partial \text{vec}(S^{m,i})^T} \left( \phi(\text{pad}(Z^{m,i}))^T \otimes I_{d_m+1} \right)
\]

\[
= \text{vec} \left( \frac{\partial \xi_i}{\partial S^{m,i}} \phi(\text{pad}(Z^{m,i}))^T \right)^T
\]

(1)
where (1) is from

\[ \text{vec}(AB)^T = \text{vec}(B)^T (\mathcal{I} \otimes A^T) \]  \hspace{1cm} (2)

\[ = \text{vec}(A)^T (B \otimes \mathcal{I}) \]  \hspace{1cm} (3)

- We applied chain rule here
- Note that we define

\[
\frac{\partial y}{\partial (x)}^T = \begin{bmatrix}
\frac{\partial y_1}{\partial x_1} & \cdots & \frac{\partial y_1}{\partial x_{|x|}} \\
\vdots & \ddots & \vdots \\
\frac{\partial y_{|y|}}{\partial x_1} & \cdots & \frac{\partial y_{|y|}}{\partial x_{|x|}}
\end{bmatrix},
\]  \hspace{1cm} (4)
where $x$ and $y$ are column vectors, and $|x|$, $|y|$ are their lengths.

Thus if

$$y = Ax$$

then

$$y_1 = A_{11}x_1 + \cdots + A_1|x||x||x|$$

and

$$\frac{\partial y}{\partial (x)^T} = \begin{bmatrix} A_{11} & A_{12} & \cdots \\ A_{21} \\ \vdots \end{bmatrix} = A$$
Similarly

\[
\frac{\partial \xi_i}{\partial (b^m)^T} = \frac{\partial \xi_i}{\partial \text{vec}(S^{m,i})^T} \frac{\partial \text{vec}(S^{m,i})}{\partial (b^m)^T}
\]

\[
= \frac{\partial \xi_i}{\partial \text{vec}(S^{m,i})^T} \left( 1 a_{conv}^m b_{conv}^m \otimes I_{dm+1} \right)
\]

\[
= \text{vec} \left( \frac{\partial \xi_i}{\partial S_{m,i}} 1 a_{conv}^m b_{conv}^m \right)^T,
\]

where (5) is from (3).
To calculate (1), \( \phi(\text{pad}(Z^{m,i})) \) has been available from the forward process of calculating the function value.

In (1) and (5), \( \partial \xi_i / \partial S^{m,i} \) is also needed.

We will show that it can be obtained by a backward process.
What we will do is to assume that

\[ \frac{\partial \xi_i}{\partial Z^{m+1,i}} \]

is available.

Then we show details of calculating

\[ \frac{\partial \xi_i}{\partial S^{m,i}} \]  and  \[ \frac{\partial \xi_i}{\partial Z^{m,i}} \]

for layer \( m \).

Thus a back propagation process...
Calculation of $\frac{\partial \xi_i}{\partial S^{m,i}}$

- We have the following workflow.

  $$Z^{m,i} \leftarrow \text{padding} \leftarrow \text{convolution} \leftarrow \sigma(S^{m,i})$$
  $$\leftarrow \text{pooling} \leftarrow Z^{m+1,i}.$$  \hspace{1cm} (6)

- From chain rule,

  $$\frac{\partial \xi_i}{\partial \text{vec}(S^{m,i})^T} = \frac{\partial \xi_i}{\partial \text{vec}(\sigma(S^{m,i}))^T} \frac{\partial \text{vec}(\sigma(S^{m,i}))}{\partial \text{vec}(S^{m,i})^T}$$
If $\sigma$ is a scalar function, then

$$
\frac{\partial \text{vec}(\sigma(S^{m,i}))}{\partial \text{vec}(S^{m,i})^T}
$$

is a squared diagonal matrix of

$$
|\text{vec}(S^{m,i})| \times |\text{vec}(S^{m,i})|
$$
We further assume that the RELU activation function is used. Recall that we assume

$$\sigma'(x) = \begin{cases} 
1 & \text{if } x > 0 \\
0 & \text{otherwise}
\end{cases}$$

though $\sigma(x)$ is not differentiable at $x = 0$
We can define

\[ I[S^{m,i}](p,q) = \begin{cases} 
1 & \text{if } S_{(p,q)}^{m,i} > 0, \\
0 & \text{otherwise,}
\end{cases} \]

and have

\[
\frac{\partial \xi_i}{\partial \text{vec}(S^{m,i})^T} = \frac{\partial \xi_i}{\partial \text{vec}(\sigma(S^{m,i}))^T} \odot \text{vec}(I[S^{m,i}])^T
\]

where \( \odot \) is Hadamard product (i.e., element-wise products)
Q: can we extend this to other scalar activation functions?

Yes, the general form is

$$
\frac{\partial \xi_i}{\partial \text{vec}(S^m,i)^T} = \frac{\partial \xi_i}{\partial \text{vec}(\sigma(S^m,i))^T} \odot \text{vec}(\sigma'(S^m,i))^T
$$

Next,


Calculation of $\partial \xi_i / \partial S_{m, i}^T$ VII

\[
\frac{\partial \xi_i}{\partial \text{vec}(S_{m, i}^T)} = \frac{\partial \xi_i}{\partial \text{vec}(Z_{m+1, i}^T)} \frac{\partial \text{vec}(Z_{m+1, i}^T)}{\partial \text{vec}(\sigma(S_{m, i}^T))} \frac{\partial \text{vec}(\sigma(S_{m, i}^T))}{\partial \text{vec}(S_{m, i}^T)}
\]

\[
= \left( \frac{\partial \xi_i}{\partial \text{vec}(Z_{m+1, i}^T)} \frac{\partial \text{vec}(Z_{m+1, i}^T)}{\partial \text{vec}(\sigma(S_{m, i}^T))} \right) \odot \text{vec}(I[S_{m, i}])^T
\]

\[
= \left( \frac{\partial \xi_i}{\partial \text{vec}(Z_{m+1, i}^T)} P_{\text{pool}}^{m, i} \right) \odot \text{vec}(I[S_{m, i}])^T
\]

(7)
Calculation of $\frac{\partial \xi_i}{\partial S^{m,i}}$ VIII

- Note that (7) is from

$$Z^{m+1,i} = \text{mat}(P_{\text{pool}}^{m,i} \text{vec}(\sigma(S^{m,i})))_{d^{m+1} \times a^{m+1} b^{m+1}}$$

- If a general scalar activation function is considered, (7) is changed to

$$\frac{\partial \xi_i}{\partial \text{vec}(S^{m,i})^T} = \left(\frac{\partial \xi_i}{\partial \text{vec}(Z^{m+1,i})^T} P_{\text{pool}}^{m,i}\right) \odot \text{vec}(\sigma'(S^{m,i}))^T$$
Calculation of $\frac{\partial \xi_i}{\partial S^{m,i}}$

In the end we calculate $\frac{\partial \xi_i}{\partial Z^{m,i}}$ and pass it to the previous layer.
Calculation of $\frac{\partial \xi_i}{\partial S^m, i}$

$$\frac{\partial \xi_i}{\partial \text{vec}(Z^{m,i})^T} = \frac{\partial \xi_i}{\partial \text{vec}(S^{m,i})^T} \frac{\partial \text{vec}(S^{m,i})}{\partial \text{vec}(\phi(\text{pad}(Z^{m,i})))^T} \frac{\partial \text{vec}(\phi(\text{pad}(Z^{m,i})))}{\partial \text{vec}(\text{pad}(Z^{m,i}))^T}$$

$$= \frac{\partial \xi_i}{\partial \text{vec}(S^{m,i})^T} \left( I_{a_{\text{conv}}^m b_{\text{conv}}^m} \otimes W^m \right) P^m \Phi P^m_{\text{pad}} \tag{8}$$

$$= \text{vec} \left( (W^m)^T \frac{\partial \xi_i}{\partial S^{m,i}} \right)^T P^m \Phi P^m_{\text{pad}} \tag{9}$$
Calculation of $\partial \xi_i / \partial S^{m,i}$

where (8) is from

$$\text{vec}(S^{m,i}) = (\mathcal{I}_{a_{\text{conv}}^m b_{\text{conv}}^m} \otimes W^m) \text{vec}(\phi(\text{pad}(Z^{m,i}))) + (1_{a_{\text{conv}}^m b_{\text{conv}}^m} \otimes \mathcal{I}_{d^{m+1}}) b^m$$

and (9) is from (2).