Convolutional Layer Backpropagation

Maik Schmidt

14. März 2017

1 Single Kernel

$$\begin{bmatrix} f_{0,0} & f_{0,1} & f_{0,2} \\ f_{1,0} & f_{1,1} & f_{1,2} \\ f_{2,0} & f_{2,1} & f_{2,2} \end{bmatrix} * \begin{bmatrix} w_{0,0} & w_{0,1} \\ w_{1,0} & w_{1,1} \end{bmatrix} + \begin{bmatrix} b \end{bmatrix} = \begin{bmatrix} y_{0,0} & y_{0,1} \\ y_{1,0} & y_{1,1} \end{bmatrix}$$

Forward:

$$y^l_{i,j} = \sum_{i'=0}^{\substack{\text{kernel}\\\text{height}\\-1}} \sum_{j'=0}^{\substack{\text{kernel}\\\text{width}\\-1}} w^l_{i',j'} \cdot f^{l-1}_{i+i',j+j'} + b^l$$

$$f^l_{i,j} = h(y^l_{i,j})$$

Backward:

$$\delta^l_{i,j} = \frac{\partial h(y^l_{i,j})}{\partial y^l_{i,j}} \cdot \sum_{i'=0}^{\substack{\text{kernel}\\\text{height}\\-1}} \sum_{j'=0}^{\substack{\text{kernel}\\\text{width}\\-1}} w^{l+1}_{i',j'} \cdot \delta^{l+1}_{i-i',j-j'}$$

with *zero padding*:

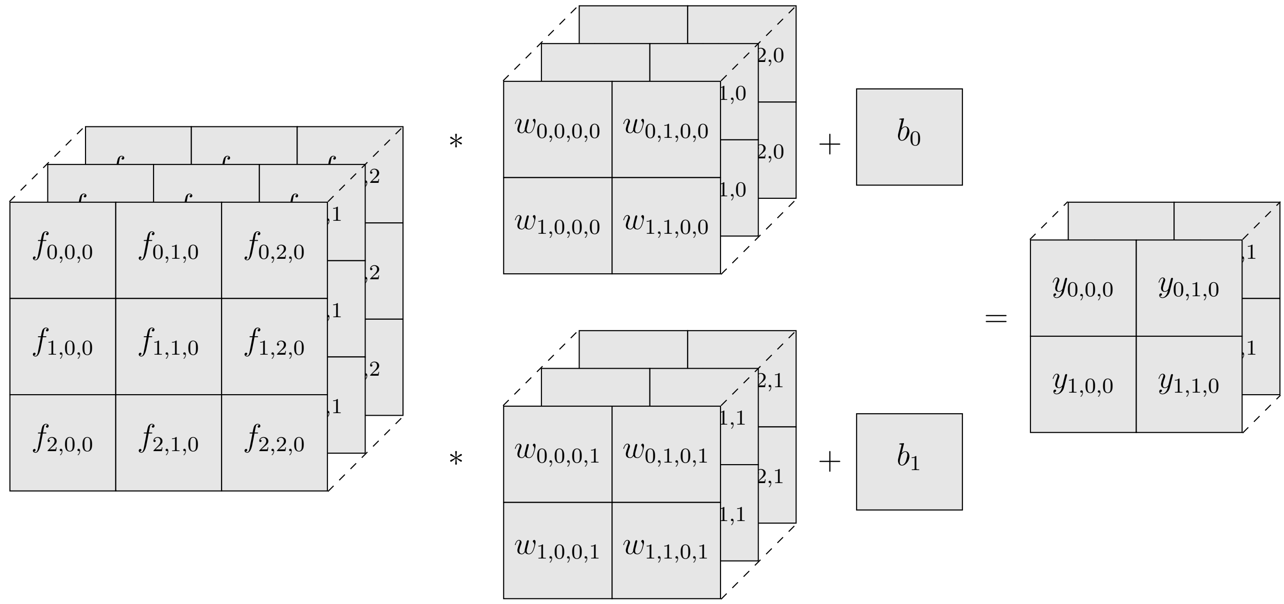
$$\delta^{l+1}_{x,y} := 0 \quad \text{if } x \notin [0, \text{output height}] \text{ or } y \notin [0, \text{output width}]$$

Gradient step:

$$w^l_{i,j} = w^l_{i,j} - \alpha \cdot \sum_{i'=0}^{\substack{\text{output}\\\text{height}\\-1}} \sum_{j'=0}^{\substack{\text{output}\\\text{width}\\-1}} \delta^l_{i',j'} \cdot f^{l-1}_{i+i',j+j'}$$

$$b^l = b^l - \alpha \cdot \sum_{i'=0}^{\substack{\text{output}\\\text{height}\\-1}} \sum_{j'=0}^{\substack{\text{output}\\\text{width}\\-1}} \delta^l_{i',j'}$$

2 Multiple Kernels



Forward:

$$y^l_{i,j,f} = \sum_{i'=0}^{\substack{\text{kernel}\\ \text{height}\\ -1}} \sum_{j'=0}^{\substack{\text{kernel}\\ \text{width}\\ -1}} \sum_{k'=0}^{\substack{\text{kernel}\\ \text{depth}\\ -1}} w^l_{i',j',k',f} \cdot f^{l-1}_{i+i',j+j',k'} + b^l_f$$

$$f^l_{i,j,f} = h(y^l_{i,j,f})$$

Backward:

$$\delta^l_{i,j,f} = \frac{\partial h(y^l_{i,j,f})}{\partial y^l_{i,j,f}} \cdot \sum_{i'=0}^{\substack{\text{kernel}\\ \text{height}\\ -1}} \sum_{j'=0}^{\substack{\text{kernel}\\ \text{width}\\ -1}} \sum_{f'=0}^{\substack{\#\text{ of}\\ \text{kernels}\\ -1}} w^{l+1}_{i',j',f,f'} \cdot \delta^{l+1}_{i-i',j-j',f'}$$

with *zero padding*:

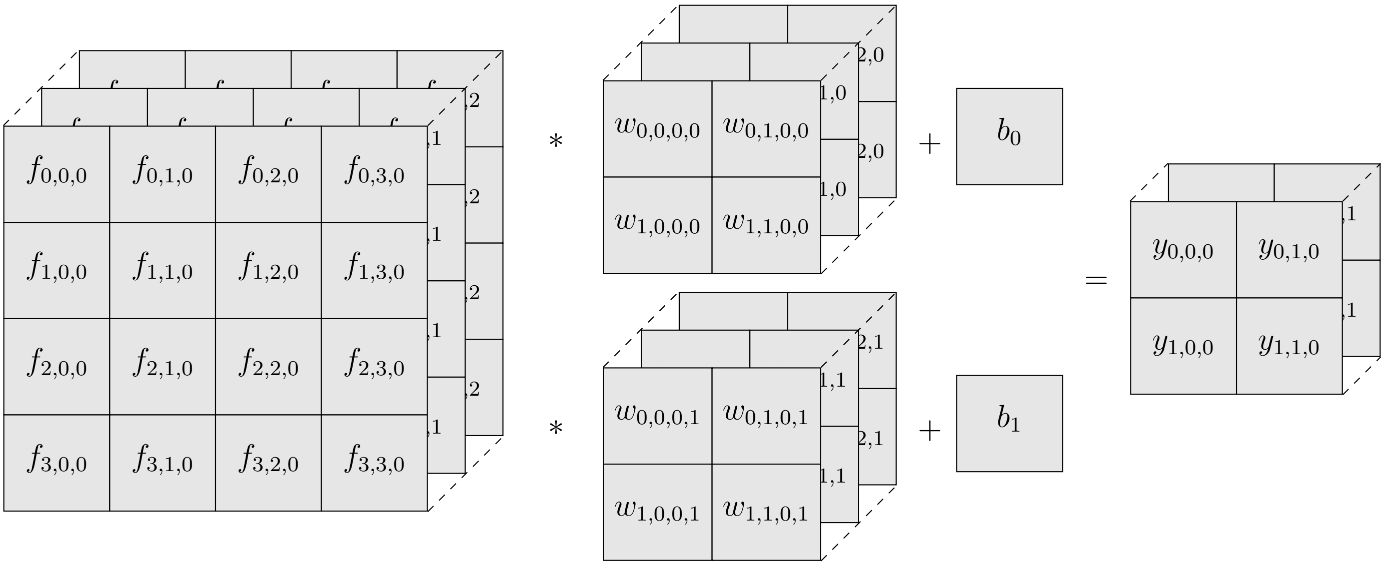
$$\delta^{l+1}_{x,y,z} := 0 \quad \text{if } x \notin [0, \text{output height}] \text{ or } y \notin [0, \text{output width}]$$

Gradient step:

$$w^l_{i,j,k,f} = w^l_{i,j,k,f} - \alpha \cdot \sum_{i'=0}^{\substack{\text{output}\\ \text{height}\\ -1}} \sum_{j'=0}^{\substack{\text{output}\\ \text{width}\\ -1}} \delta^l_{i',j',f} \cdot f^{l-1}_{i+i',j+j',k}$$

$$b^l_f = b^l_f - \alpha \cdot \sum_{i'=0}^{\substack{\text{output}\\ \text{height}\\ -1}} \sum_{j'=0}^{\substack{\text{output}\\ \text{width}\\ -1}} \delta^l_{i',j',f}$$

3 Multiple Kernels with stride



Forward:

$$y^l_{i,j,f} = \sum_{i'=0}^{\substack{\text{kernel}\\\text{height}\\-1}} \sum_{j'=0}^{\substack{\text{kernel}\\\text{width}\\-1}} \sum_{k'=0}^{\substack{\text{kernel}\\\text{depth}\\-1}} w^l_{i',j',k',f} \cdot f^{l-1}_{i\cdot stride+i',j\cdot stride+j',k'} + b^l_f$$

$$f^l_{i,j,f} = h(y^l_{i,j,f})$$

Backward:

$$\delta^l_{i,j,f} = \frac{\partial h(y^l_{i,j,f})}{\partial y^l_{i,j,f}} \cdot \sum_{i'=0}^{\substack{\text{kernel}\\\text{height}\\/stride\\-1}} \sum_{j'=0}^{\substack{\text{kernel}\\\text{width}\\/stride\\-1}} \sum_{f'=0}^{\substack{\#\text{ of}\\\text{kernels}\\-1}} w^{l+1}_{i\%stride+i'\cdot stride,j\%stride+j'\cdot stride,f,f'} \cdot \delta^{l+1}_{i/stride-i',j/stride-j',f'}$$

with *zero padding*:

$$\delta^{l+1}_{x,y,z} := 0 \quad \text{if } x \notin [0, \text{output height}] \text{ or } y \notin [0, \text{output width}]$$

Gradient step:

$$w^l_{i,j,k,f} = w^l_{i,j,k,f} - \alpha \cdot \sum_{i'=0}^{\substack{\text{output}\\\text{height}\\-1}} \sum_{j'=0}^{\substack{\text{output}\\\text{width}\\-1}} \delta^l_{i',j',f} \cdot f^{l-1}_{i+i'\cdot stride,j+j'\cdot stride,k}$$

$$b^l_f = b^l_f - \alpha \cdot \sum_{i'=0}^{\substack{\text{output}\\\text{height}\\-1}} \sum_{j'=0}^{\substack{\text{output}\\\text{width}\\-1}} \delta^l_{i',j',f}$$