Convolutional Neural Networks: Back Propagation

Deep Learning

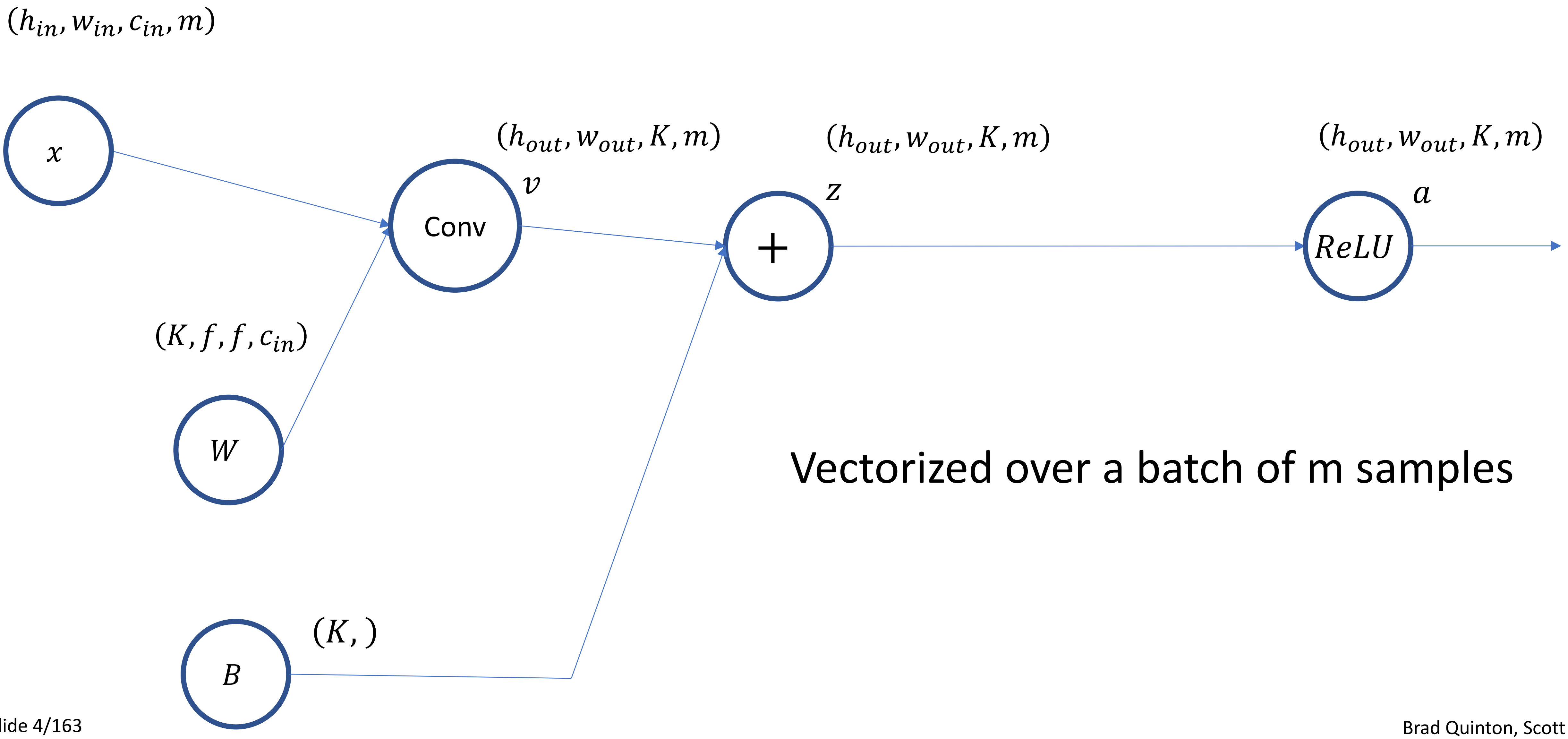
[Brad Quinton](), [Scott Chin]()

Learning Objectives

- More practice with backpropagation
- Understand backprop being applied to convolutional layer
- See some of the “shortcuts” in convolutional layer backprop
- Understand use of numerical gradient

BackPropagation for CNNs

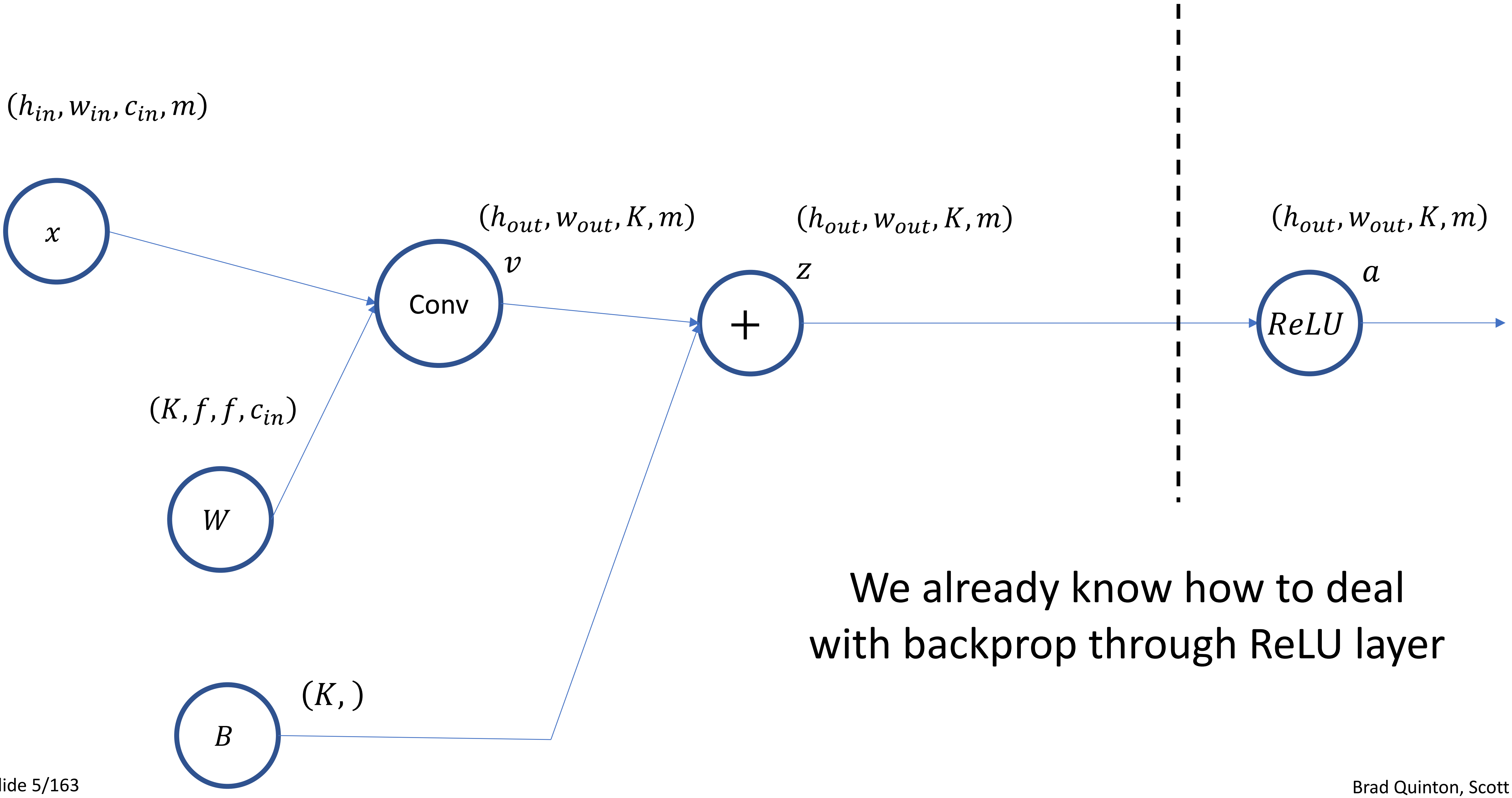
A compute graph for a convolutional layer



Vectorized over a batch of m samples

Brad Quinton, Scott Chin

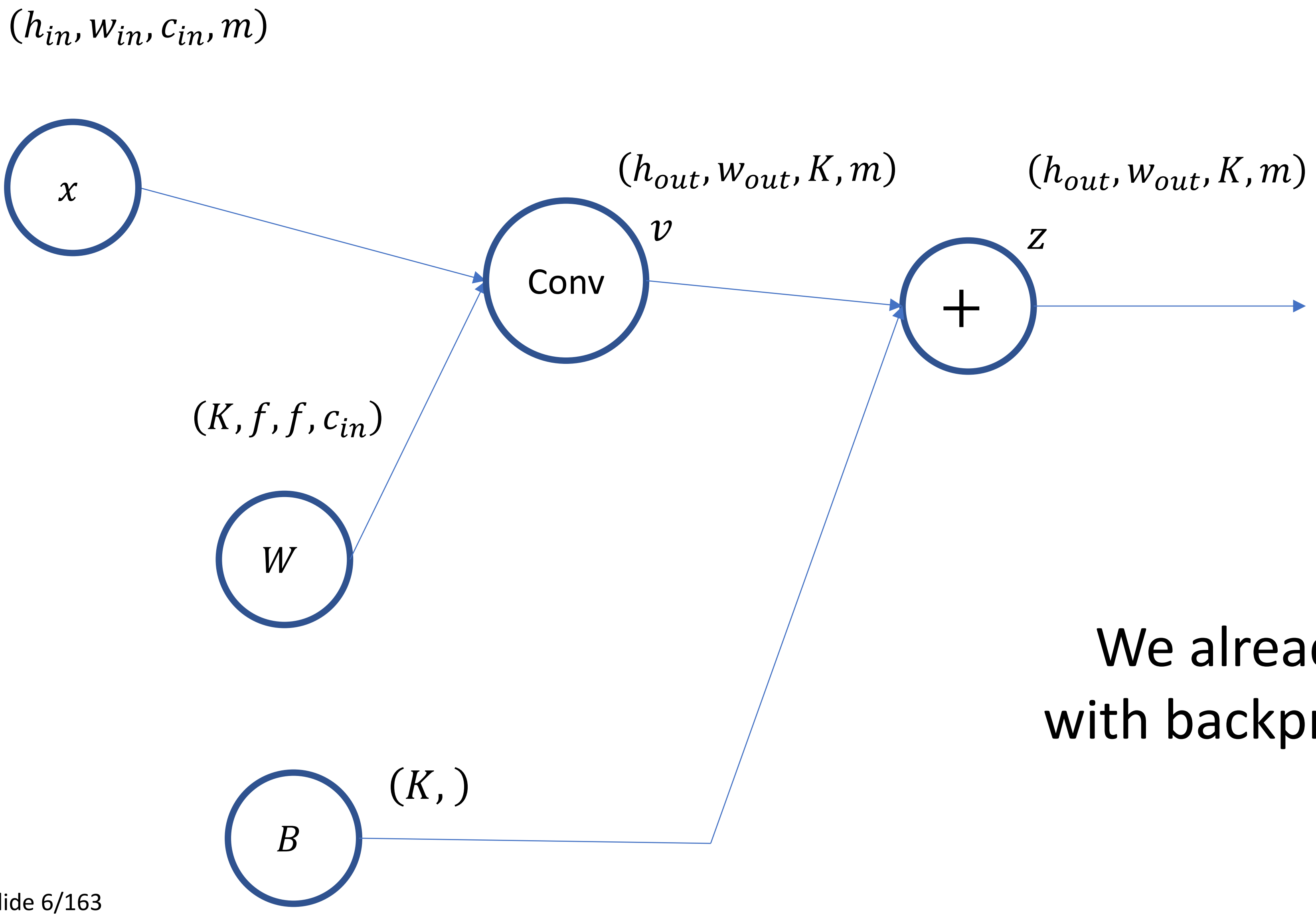
A compute graph for a convolutional layer



We already know how to deal with backprop through ReLU layer

Brad Quinton, Scott Chin

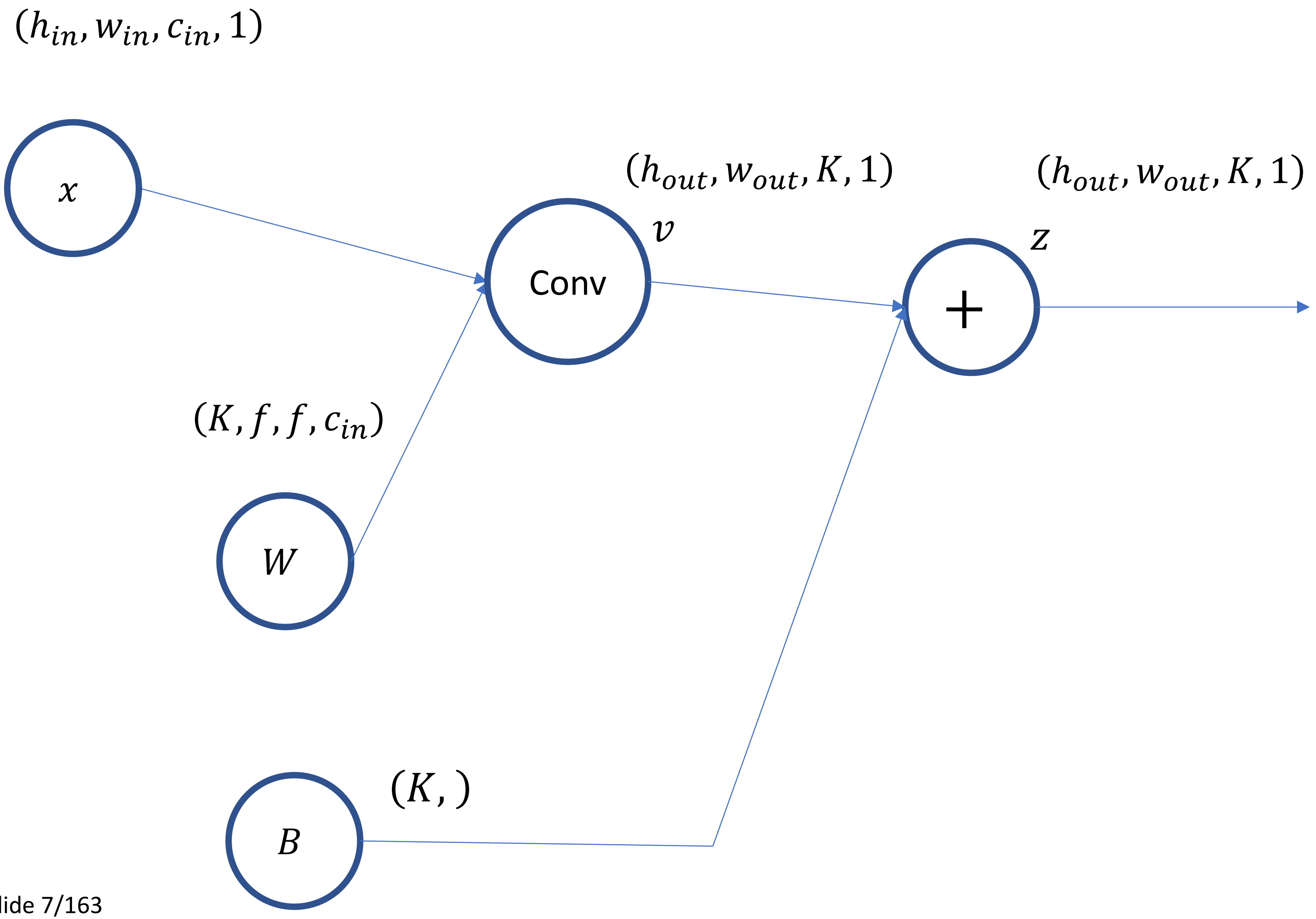
A compute graph for a convolutional layer



We already know how to deal with backprop through ReLU layer

Brad Quinton, Scott Chin

A compute graph for a convolutional layer

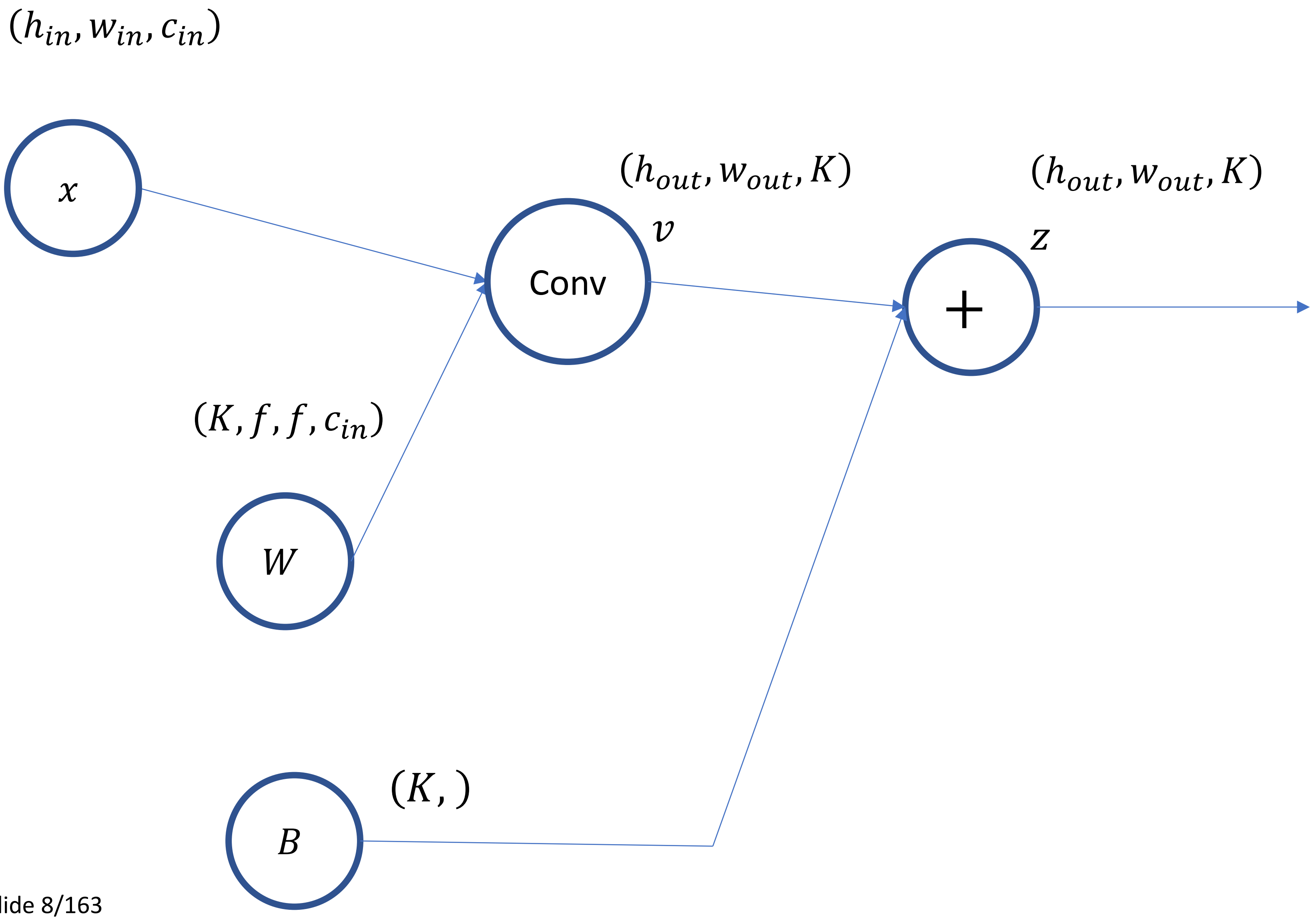


Start by considering:

- One sample (m=1)

Brad Quinton, Scott Chin

A compute graph for a convolutional layer

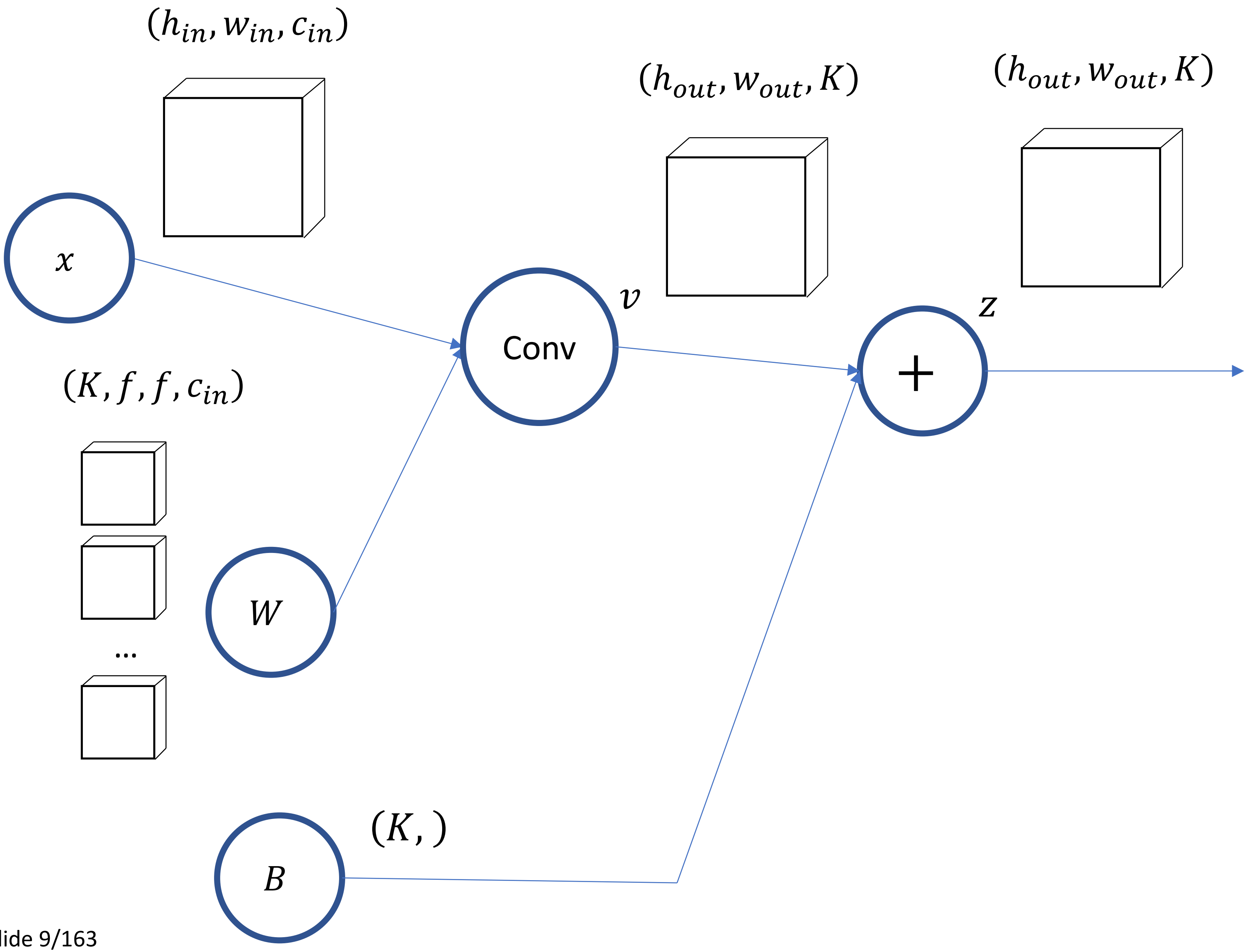


Start by considering:

- One sample (m=1)

Brad Quinton, Scott Chin

A compute graph for a convolutional layer

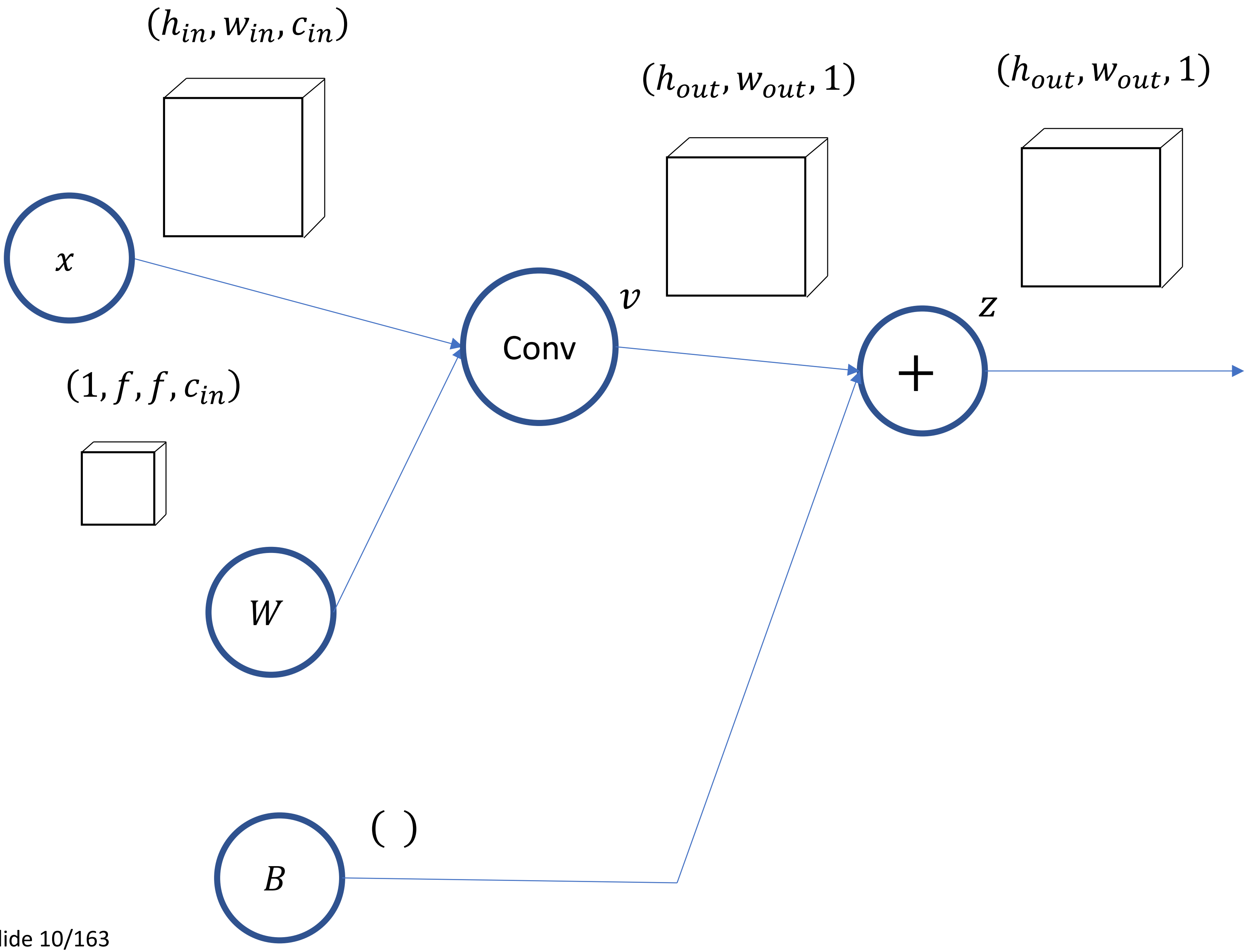


Start by considering:

- One sample (m=1)

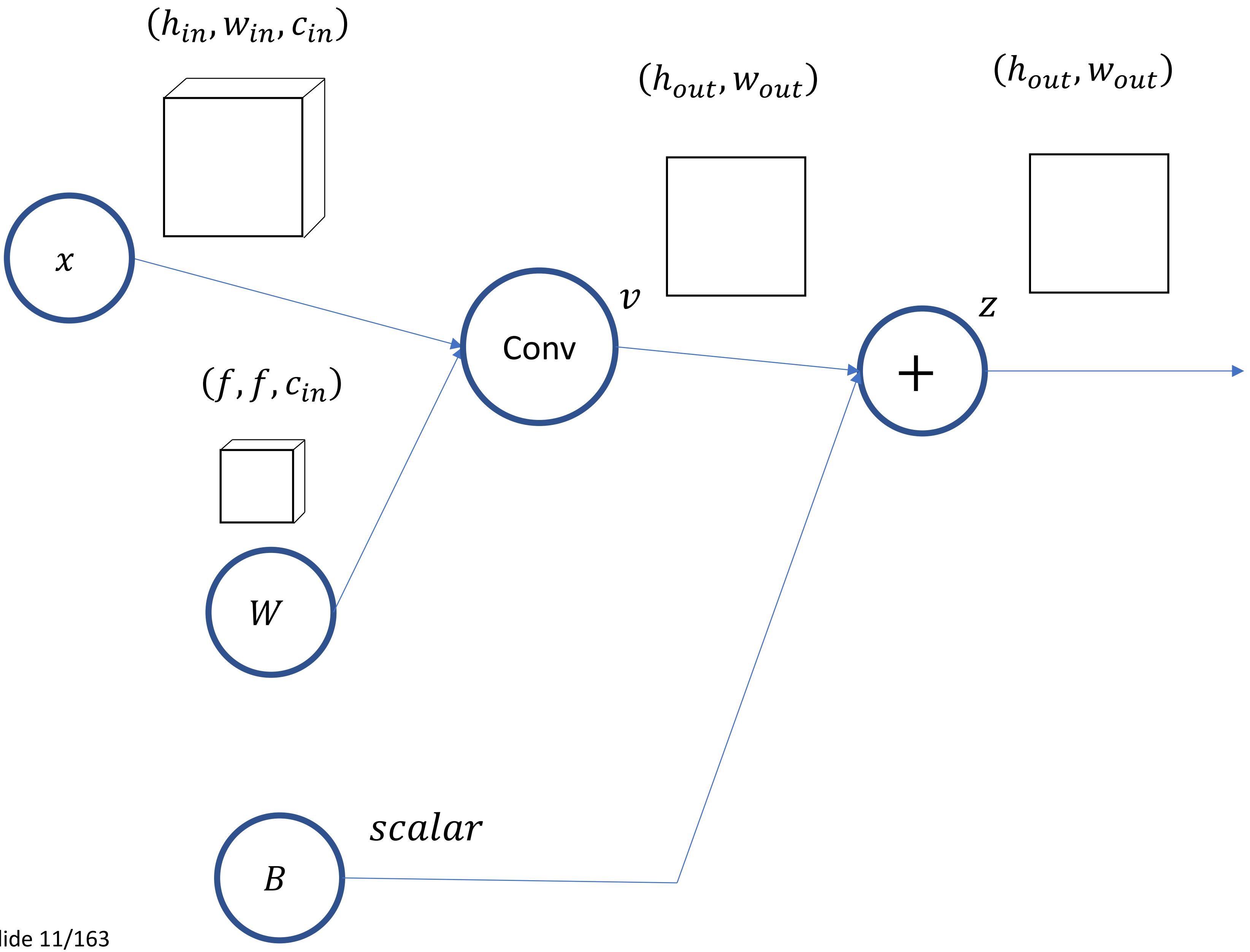
Brad Quinton, Scott Chin

A compute graph for a convolutional layer



Start by considering:
- One sample (m=1)
- One filter (K=1)

A compute graph for a convolutional layer

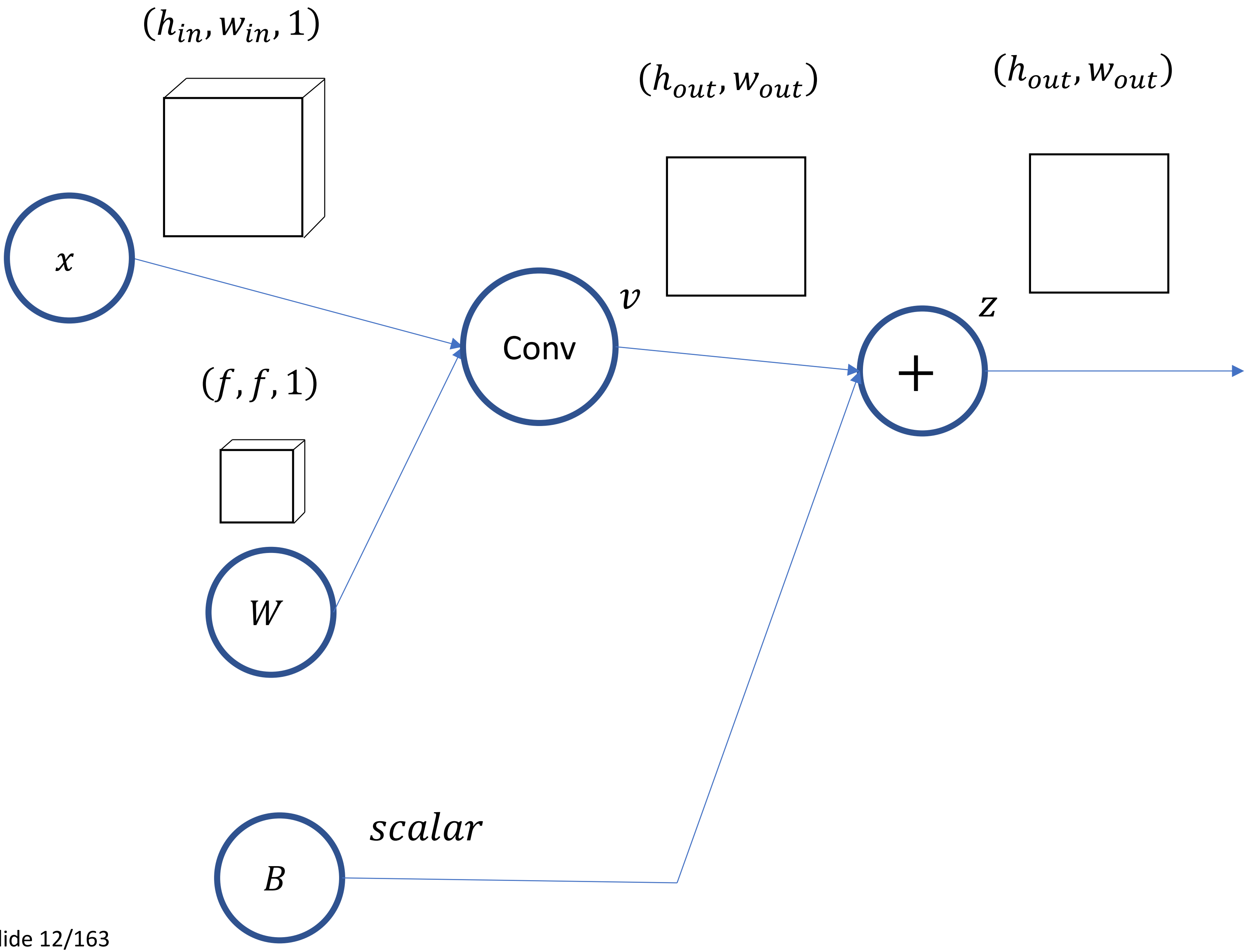


Start by considering:

- One sample (m=1)
- One filter (K=1)

Brad Quinton, Scott Chin

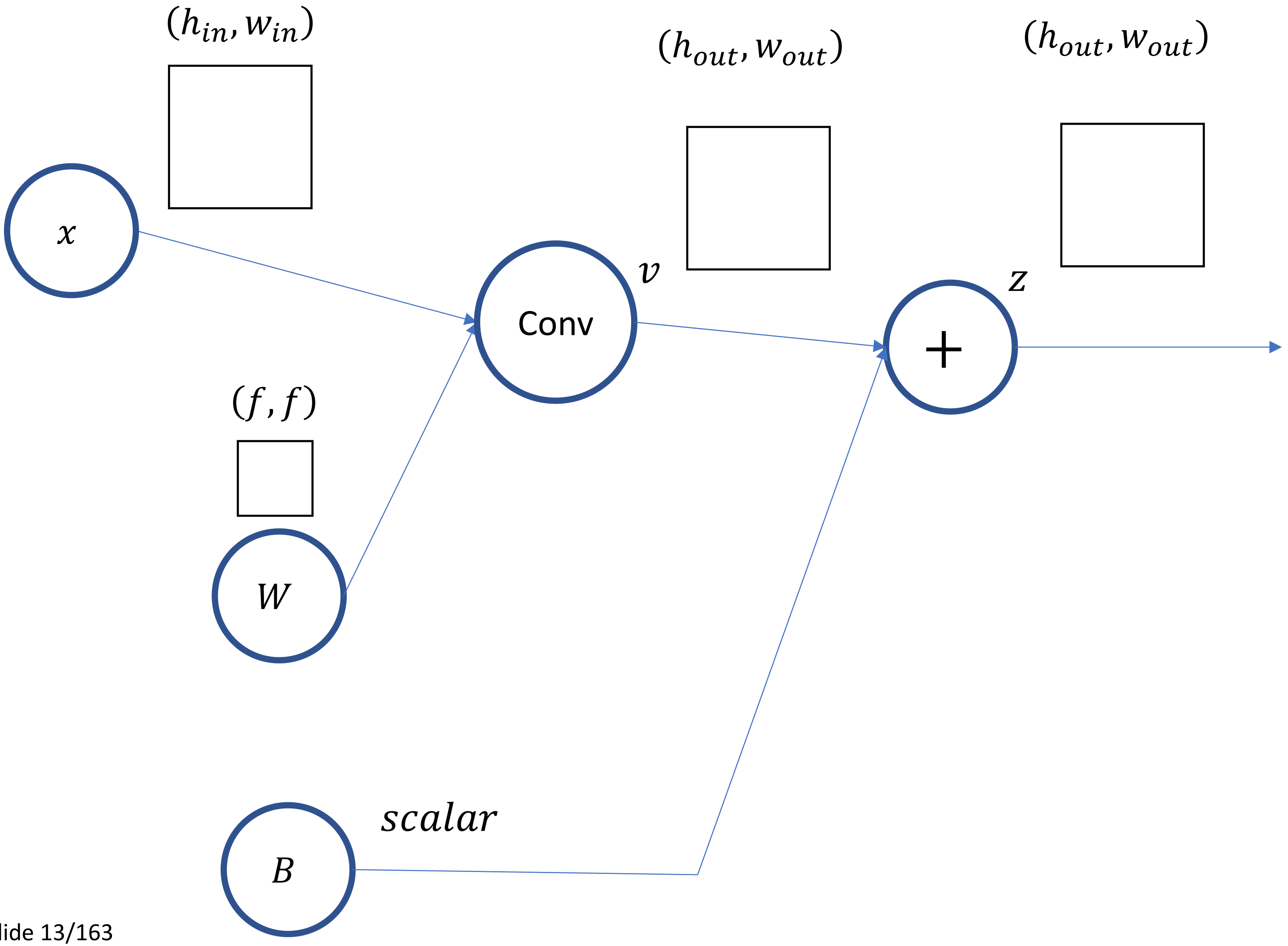
A compute graph for a convolutional layer



Start by considering:

- One sample (m=1)
- One filter (K=1)
- One channel ($c_{in} = 1$)

A compute graph for a convolutional layer

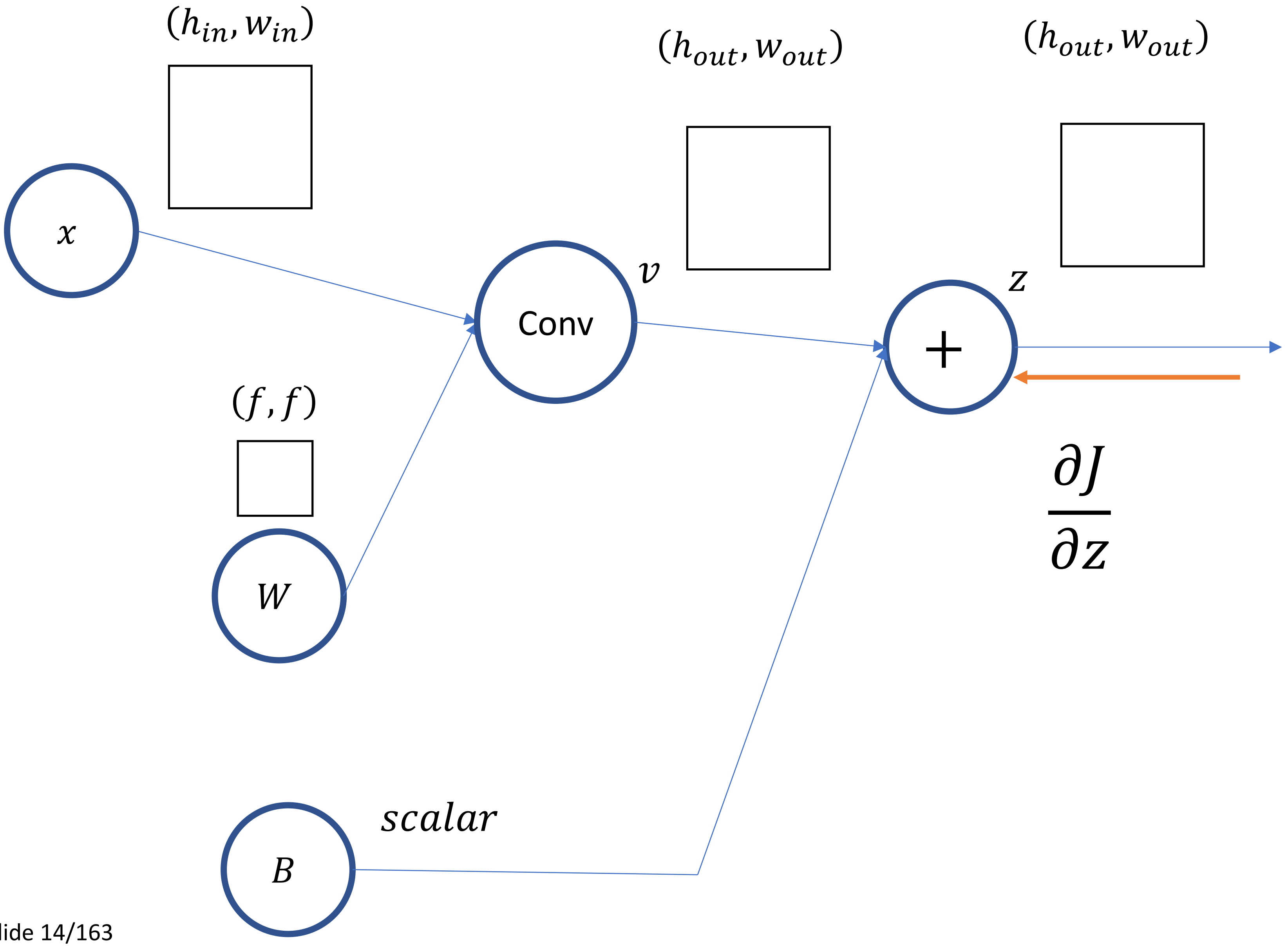


Start by considering:

- One sample (m=1)
- One filter (K=1)
- One channel ($c_{in} = 1$)

Brad Quinton, Scott Chin

A compute graph for a convolutional layer

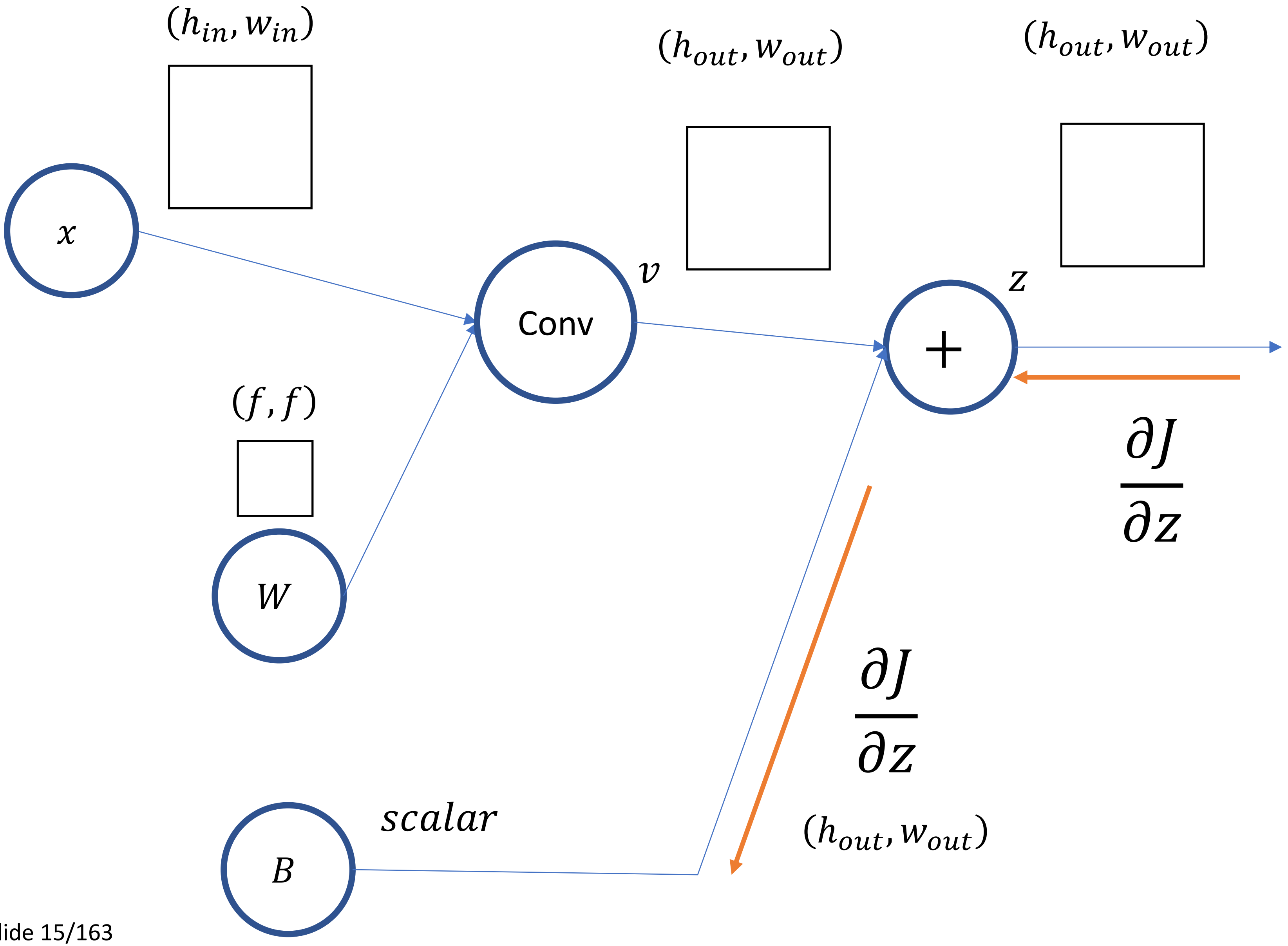


Start by considering:

- One sample (m=1)
- One filter (K=1)
- One channel ($c_{in} = 1$)

Brad Quinton, Scott Chin

A compute graph for a convolutional layer

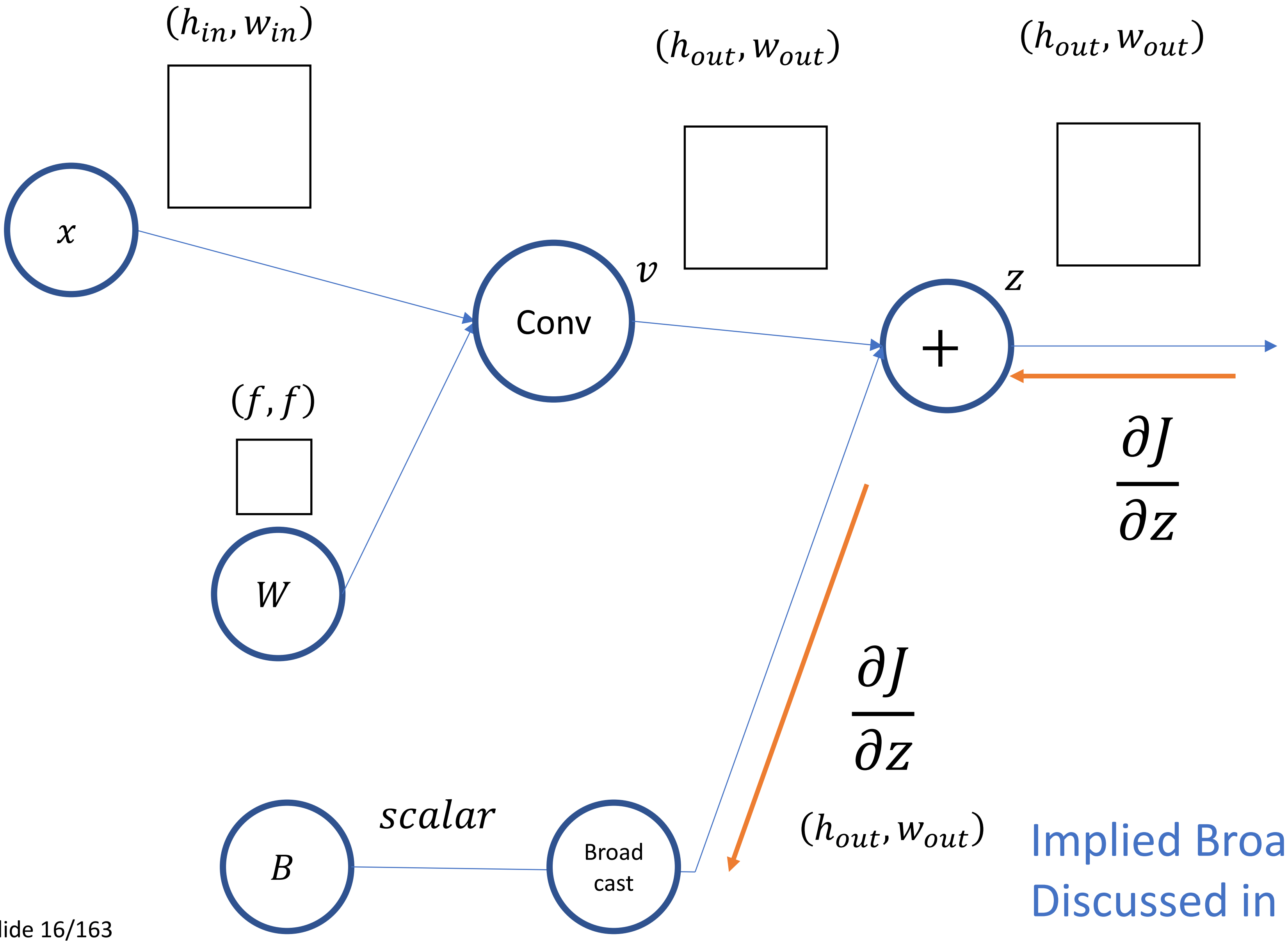


Start by considering:

- One sample (m=1)
- One filter (K=1)
- One channel ($c_{in} = 1$)

Brad Quinton, Scott Chin

A compute graph for a convolutional layer

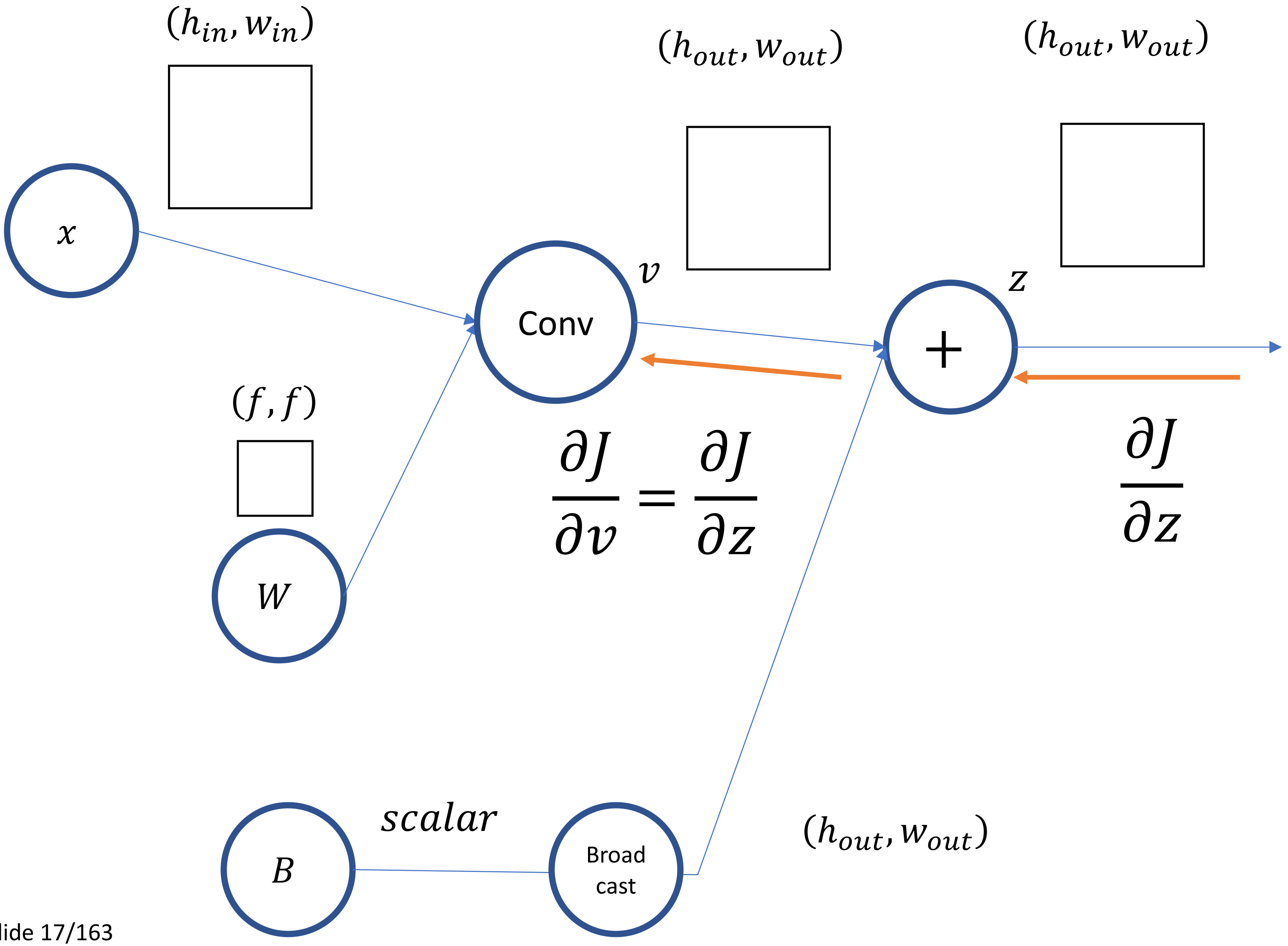


Start by considering:

- One sample (m=1)
- One filter (K=1)
- One channel ($c_{in} = 1$)

Implied Broadcasting here.
Discussed in Lecture 12

A compute graph for a convolutional layer



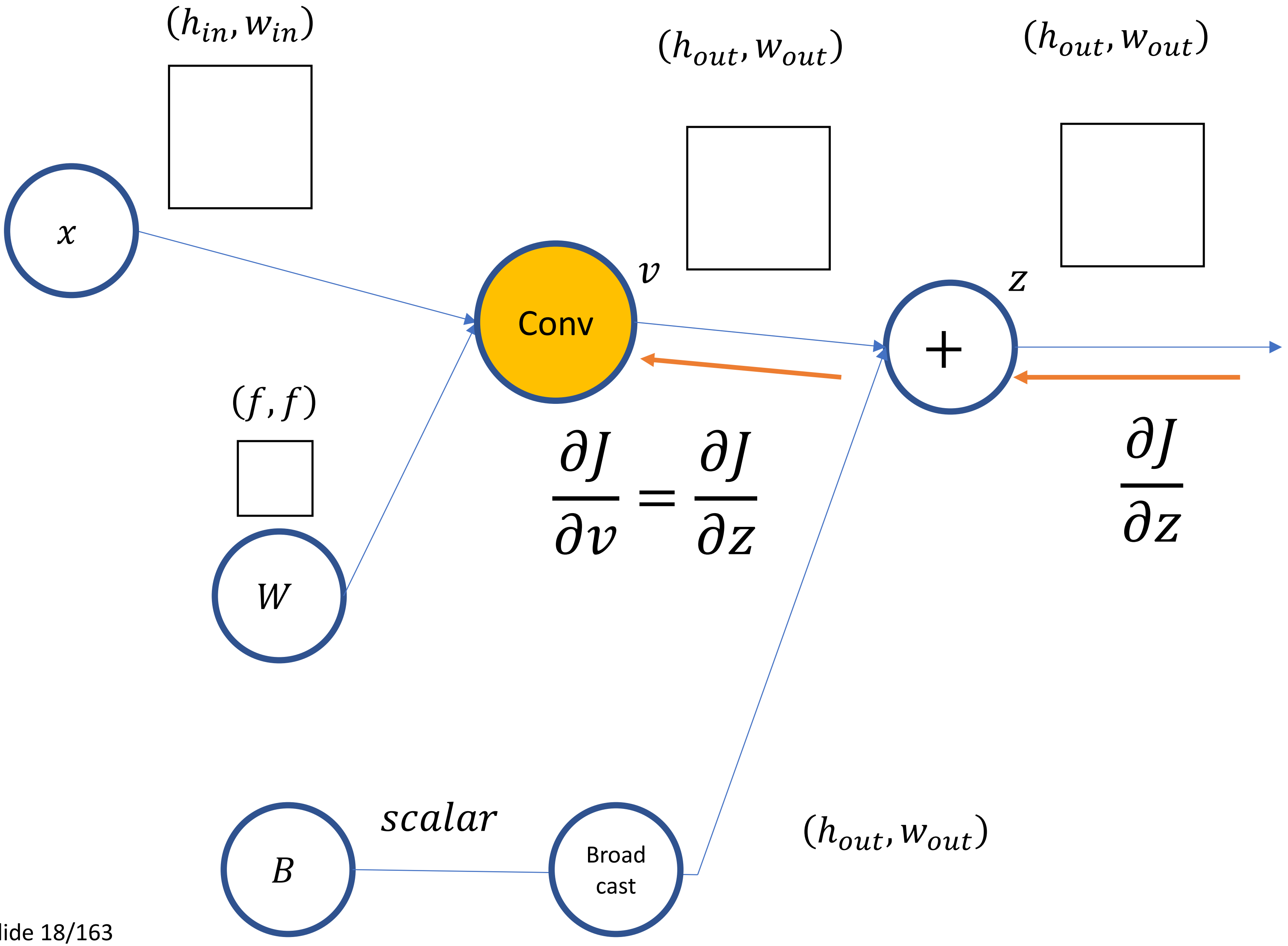
Start by considering:

- One sample (m=1)
- One filter (K=1)
- One channel ($c_{in} = 1$)

Brad Quinton, Scott Chin

A compute graph for a convolutional layer

How to compute backprop through convolution operation?



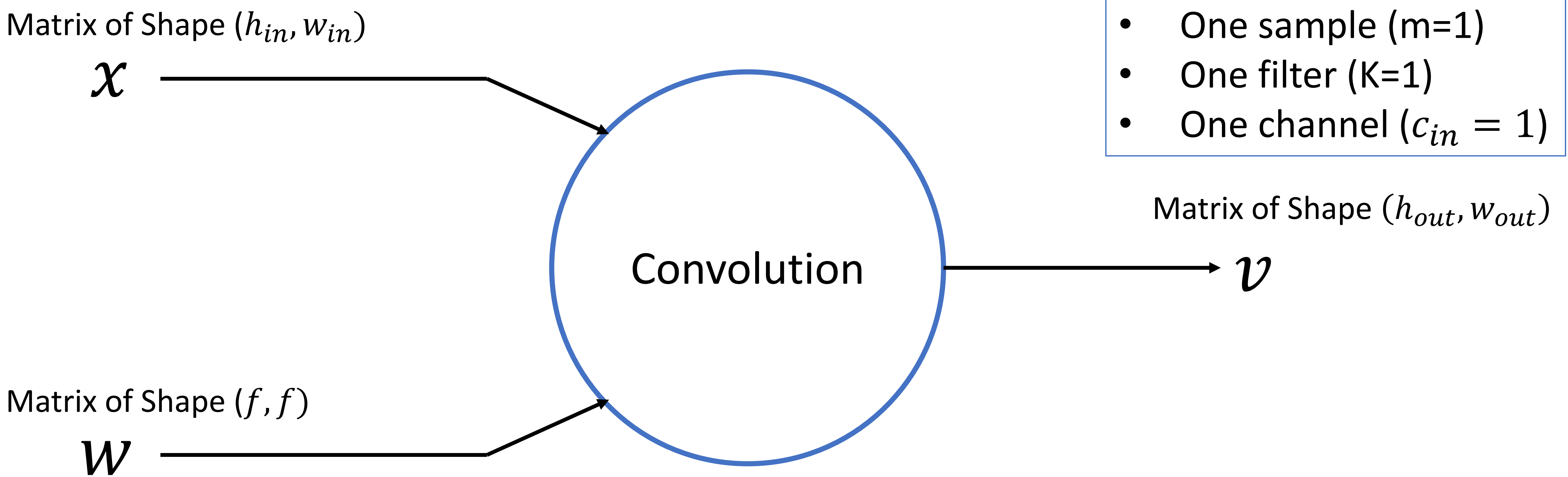
Start by considering:

- One sample (m=1)
- One filter (K=1)
- One channel ($c_{in} = 1$)

Brad Quinton, Scott Chin

Still only considering:

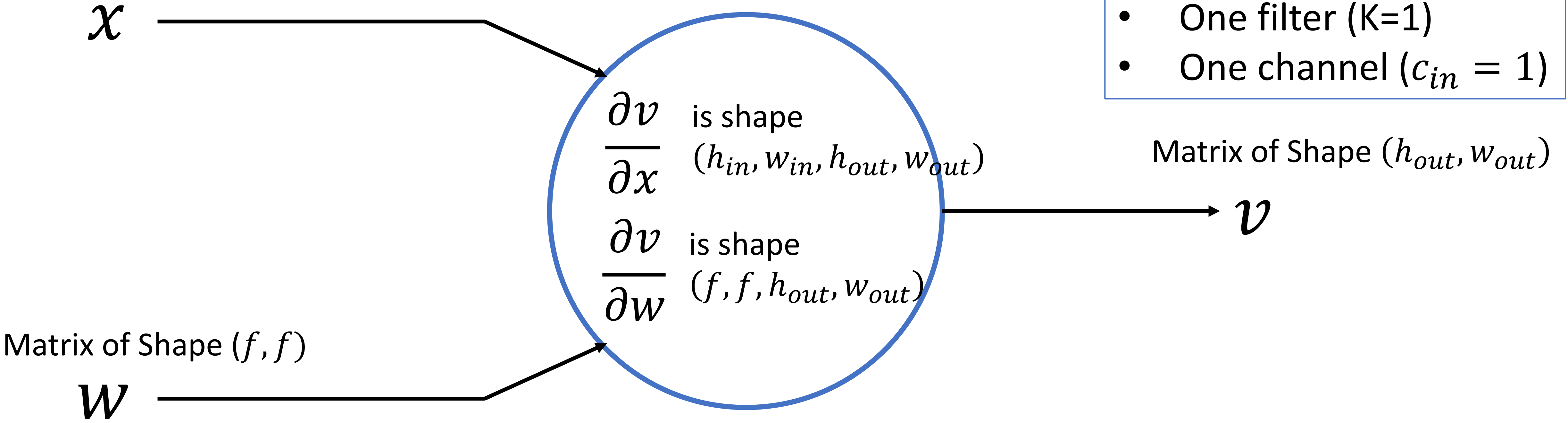
- One sample (m=1)
- One filter (K=1)
- One channel ($c_{in} = 1$)



Local Derivatives are 4D Tensors

Still only considering:

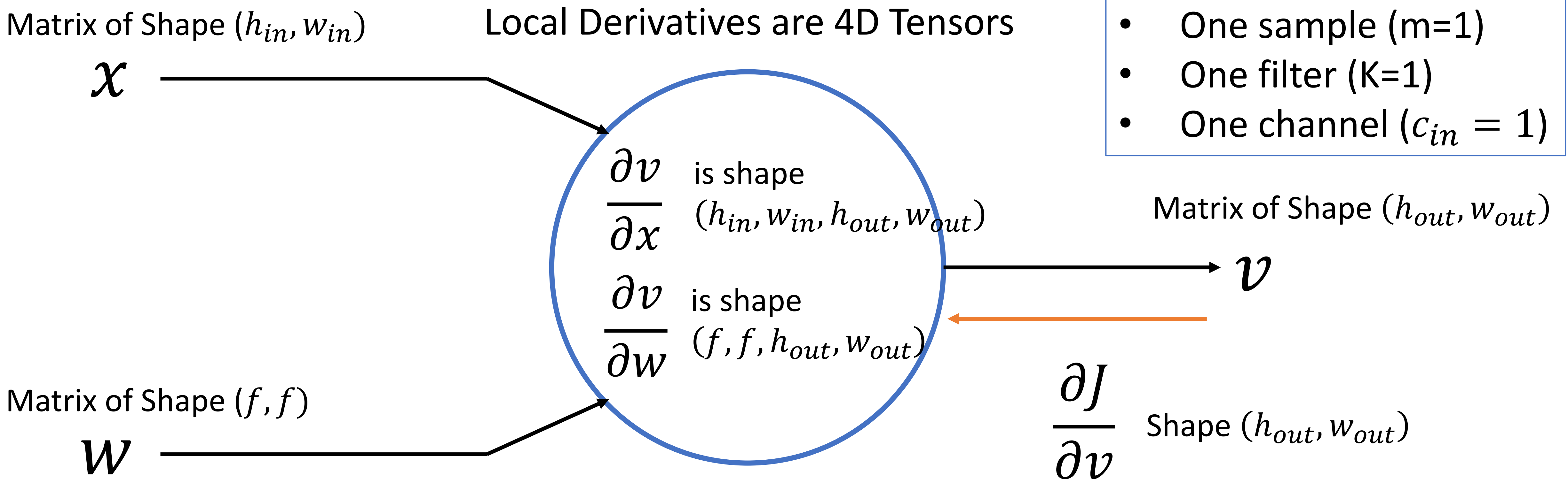
- One sample (m=1)
- One filter (K=1)
- One channel ($c_{in} = 1$)



Local Derivatives are 4D Tensors

Still only considering:

- One sample (m=1)
- One filter (K=1)
- One channel ($c_{in} = 1$)



Upstream gradient is with respect to Cost J (a scalar)

i.e. How does each of the (h_{out}, w_{out}) outputs affect Cost

Local Derivatives are 4D Tensors

Still only considering:
- One sample (m=1)
- One filter (K=1)
- One channel ($c_{in} = 1$)

Matrix of Shape (h_{in}, w_{in})

x

$$\frac{\partial J}{\partial x} = \frac{\partial v}{\partial x}\frac{\partial J}{\partial v}$$

$$(h_{in}, w_{in}, h_{out}, w_{out}) * (h_{out}, w_{out})$$
$$= (h_{in}, w_{in})$$

Matrix of Shape (f, f)

w

$\frac{\partial v}{\partial x}$ is shape $(h_{in}, w_{in}, h_{out}, w_{out})$

$\frac{\partial v}{\partial w}$ is shape (f, f, h_{out}, w_{out})

Matrix of Shape (h_{out}, w_{out})

v

$\frac{\partial J}{\partial v}$ Shape (h_{out}, w_{out})

Upstream gradient is with respect to Cost J (a scalar)

i.e. How does each of the (h_{out}, w_{out}) outputs affect Cost

Chain Rule application is **Tensor-Matrix Multiply**

Local Derivatives are 4D Tensors

Still only considering:
- One sample (m=1)
- One filter (K=1)
- One channel ($c_{in} = 1$)

Matrix of Shape (h_{in}, w_{in})

x

$$\frac{\partial J}{\partial x} = \frac{\partial v}{\partial x}\frac{\partial J}{\partial v}$$

$$(h_{in}, w_{in}, h_{out}, w_{out}) * (h_{out}, w_{out}) = (h_{in}, w_{in})$$

Matrix of Shape (f, f)

w

$$\frac{\partial J}{\partial w} = \frac{\partial v}{\partial w}\frac{\partial J}{\partial v}$$

$$(f, f, h_{out}, w_{out}) * (h_{out}, w_{out}) = (f, f)$$

$\frac{\partial v}{\partial x}$ is shape $(h_{in}, w_{in}, h_{out}, w_{out})$

$\frac{\partial v}{\partial w}$ is shape (f, f, h_{out}, w_{out})

Matrix of Shape (h_{out}, w_{out})

v

$\frac{\partial J}{\partial v}$ Shape (h_{out}, w_{out})

Upstream gradient is with respect to Cost J (a scalar)

i.e. How does each of the (h_{out}, w_{out}) outputs affect Cost

Chain Rule application is **Tensor-Matrix Multiply**

$(h_{in}, w_{in}) = (3,3)$ $\quad$ $(f, f) = (2,2)$ $\quad$ $(h_{out}, w_{out}) = (2,2)$

$$\begin{array}{|c|c|c|}\hline x_{11} & x_{12} & x_{13} \\ \hline x_{21} & x_{22} & x_{23} \\ \hline x_{31} & x_{32} & x_{33} \\ \hline \end{array} \quad \text{conv} \quad \begin{array}{|c|c|}\hline w_{11} & w_{12} \\ \hline w_{21} & w_{22} \\ \hline \end{array} \quad = \quad \begin{array}{|c|c|}\hline v_{11} & v_{12} \\ \hline v_{21} & v_{22} \\ \hline \end{array}$$

$$(h_{in}, w_{in}) = (3,3)$$

$$\begin{bmatrix} x_{11} & x_{12} & x_{13} \\ x_{21} & x_{22} & x_{23} \\ x_{31} & x_{32} & x_{33} \end{bmatrix}$$

conv

$$(f, f) = (2,2)$$

$$\begin{bmatrix} w_{11} & w_{12} \\ w_{21} & w_{22} \end{bmatrix}$$

=

$$(h_{out}, w_{out}) = (2,2)$$

$$\begin{bmatrix} v_{11} & v_{12} \\ v_{21} & v_{22} \end{bmatrix}$$

$$v_{11} = w_{11}x_{11} + w_{12}x_{12} + w_{21}x_{21} + w_{22}x_{22}$$

$$(h_{in}, w_{in}) = (3,3) \qquad (f, f) = (2,2) \qquad (h_{out}, w_{out}) = (2,2)$$

$$\begin{bmatrix} x_{11} & x_{12} & x_{13} \\ x_{21} & x_{22} & x_{23} \\ x_{31} & x_{32} & x_{33} \end{bmatrix} \ \text{conv} \ \begin{bmatrix} w_{11} & w_{12} \\ w_{21} & w_{22} \end{bmatrix} = \begin{bmatrix} v_{11} & v_{12} \\ v_{21} & v_{22} \end{bmatrix}$$

$$v_{11} = w_{11}x_{11} + w_{12}x_{12} + w_{21}x_{21} + w_{22}x_{22}$$

$$v_{12} = w_{11}x_{12} + w_{12}x_{13} + w_{21}x_{22} + w_{22}x_{23}$$

$(h_{in}, w_{in}) = (3,3)$

$$\begin{bmatrix} x_{11} & x_{12} & x_{13} \\ x_{21} & x_{22} & x_{23} \\ x_{31} & x_{32} & x_{33} \end{bmatrix}$$

conv

$(f, f) = (2,2)$

$$\begin{bmatrix} w_{11} & w_{12} \\ w_{21} & w_{22} \end{bmatrix}$$

=

$(h_{out}, w_{out}) = (2,2)$

$$\begin{bmatrix} v_{11} & v_{12} \\ v_{21} & v_{22} \end{bmatrix}$$

$$v_{11} = w_{11}x_{11} + w_{12}x_{12} + w_{21}x_{21} + w_{22}x_{22}$$

$$v_{12} = w_{11}x_{12} + w_{12}x_{13} + w_{21}x_{22} + w_{22}x_{23}$$

$$v_{21} = w_{11}x_{21} + w_{12}x_{22} + w_{21}x_{31} + w_{22}x_{31}$$

$(h_{in}, w_{in}) = (3,3)$ $\quad$ $(f, f) = (2,2)$ $\quad$ $(h_{out}, w_{out}) = (2,2)$

$$\begin{bmatrix} x_{11} & x_{12} & x_{13} \\ x_{21} & x_{22} & x_{23} \\ x_{31} & x_{32} & x_{33} \end{bmatrix} \text{ conv } \begin{bmatrix} w_{11} & w_{12} \\ w_{21} & w_{22} \end{bmatrix} = \begin{bmatrix} v_{11} & v_{12} \\ v_{21} & v_{22} \end{bmatrix}$$

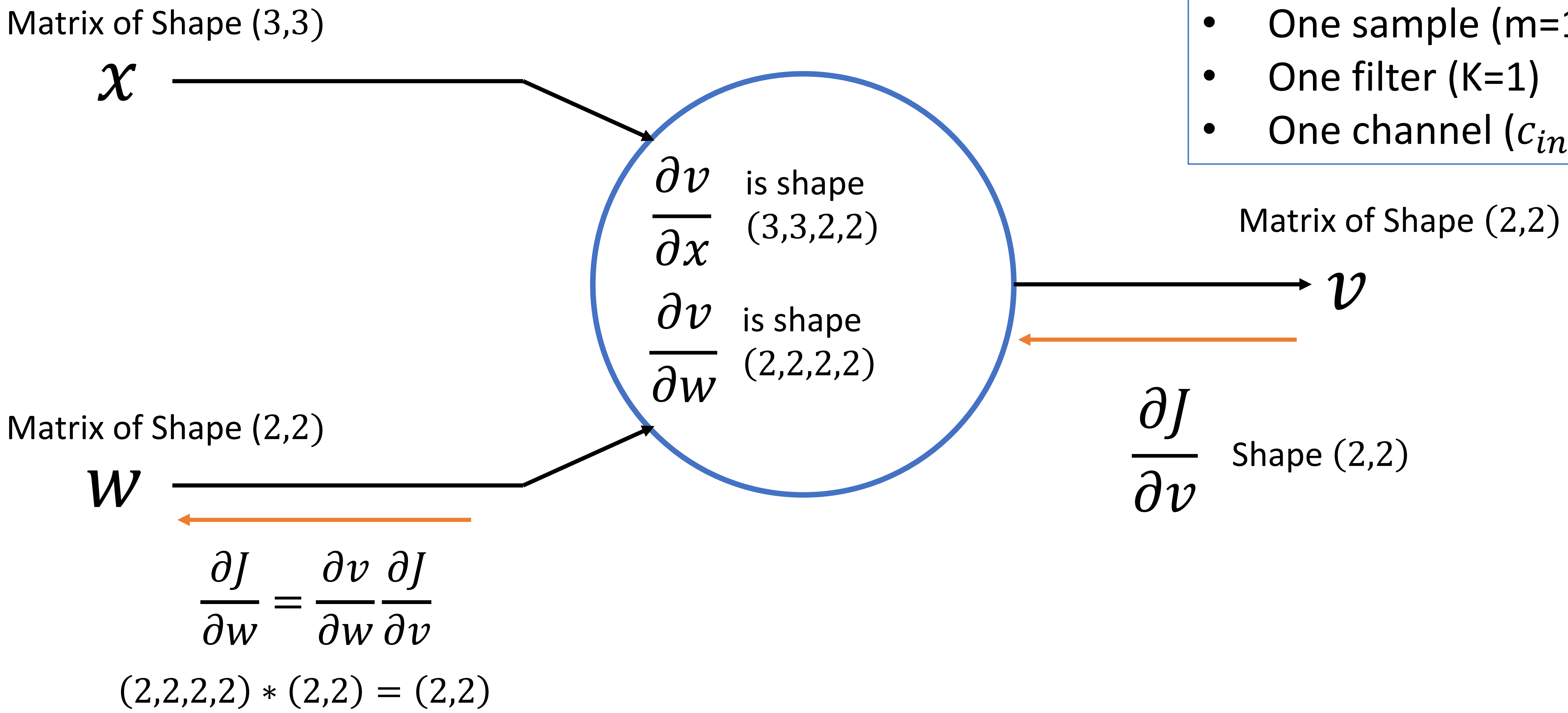
$$v_{11} = w_{11}x_{11} + w_{12}x_{12} + w_{21}x_{21} + w_{22}x_{22}$$

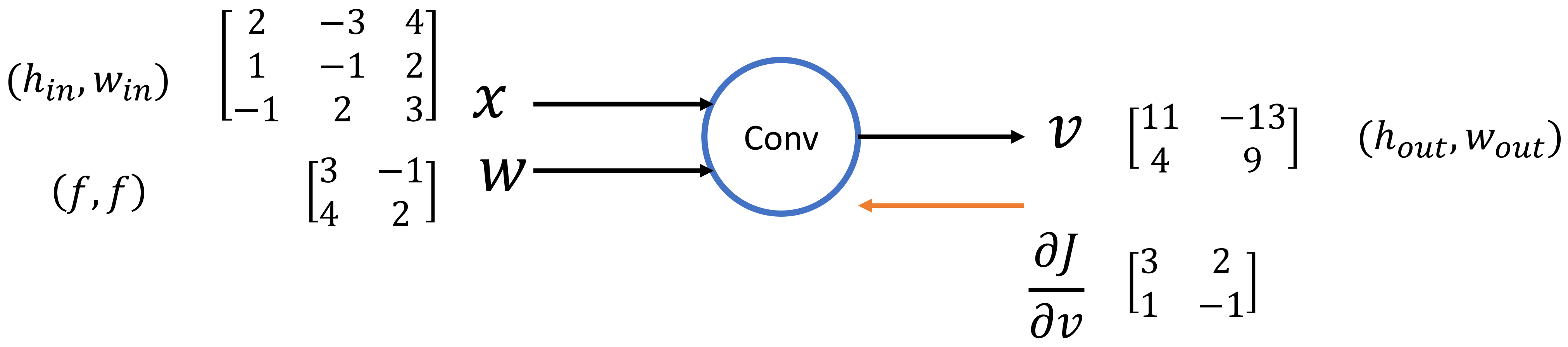
$$v_{12} = w_{11}x_{12} + w_{12}x_{13} + w_{21}x_{22} + w_{22}x_{23}$$

$$v_{21} = w_{11}x_{21} + w_{12}x_{22} + w_{21}x_{31} + w_{22}x_{31}$$

$$v_{22} = w_{11}x_{22} + w_{12}x_{23} + w_{21}x_{32} + w_{22}x_{33}$$

Still only considering:
- One sample (m=1)
- One filter (K=1)
- One channel ($c_{in} = 1$)




(h_{in}, w_{in})
$\begin{bmatrix} 2 & -3 & 4 \\ 1 & -1 & 2 \\ -1 & 2 & 3 \end{bmatrix}$
x
(f, f)
$\begin{bmatrix} 3 & -1 \\ 4 & 2 \end{bmatrix}$
w
Conv
v
$\begin{bmatrix} 11 & -13 \\ 4 & 9 \end{bmatrix}$
(h_{out}, w_{out})
$\frac{\partial J}{\partial v}$
$\begin{bmatrix} 3 & 2 \\ 1 & -1 \end{bmatrix}$

(h_{in}, w_{in}) $\begin{bmatrix} 2 & -3 & 4 \\ 1 & -1 & 2 \\ -1 & 2 & 3 \end{bmatrix}$ x

(f, f) $\begin{bmatrix} 3 & -1 \\ 4 & 2 \end{bmatrix}$ w

Conv

v $\begin{bmatrix} 11 & -13 \\ 4 & 9 \end{bmatrix}$ (h_{out}, w_{out})

$\frac{\partial J}{\partial v}$ $\begin{bmatrix} 3 & 2 \\ 1 & -1 \end{bmatrix}$

$$\frac{\partial J}{\partial w} = \boxed{\frac{\partial v}{\partial w}} \frac{\partial J}{\partial v}$$

Jacobian

(h_{in}, w_{in}) $\begin{bmatrix} 2 & -3 & 4 \\ 1 & -1 & 2 \\ -1 & 2 & 3 \end{bmatrix}$ x

(f, f) $\begin{bmatrix} 3 & -1 \\ 4 & 2 \end{bmatrix}$ w

Conv

v $\begin{bmatrix} 11 & -13 \\ 4 & 9 \end{bmatrix}$ (h_{out}, w_{out})

$\dfrac{\partial J}{\partial v}$ $\begin{bmatrix} 3 & 2 \\ 1 & -1 \end{bmatrix}$

$\dfrac{\partial v}{\partial w}$ is shape $\big((f, f), (h_{out}, w_{out})\big)$

Interpretation: How does each filter weight affect each feature map element

$$\frac{\partial v}{\partial w} = \begin{bmatrix} \dfrac{\partial v}{\partial w_{11}} & \dfrac{\partial v}{\partial w_{12}} \\ \dfrac{\partial v}{\partial w_{21}} & \dfrac{\partial v}{\partial w_{22}} \end{bmatrix}$$

(f, f)

Brad Quinton, Scott Chin

(h_{in}, w_{in}) $\begin{bmatrix} 2 & -3 & 4 \\ 1 & -1 & 2 \\ -1 & 2 & 3 \end{bmatrix}$ x

(f, f) $\begin{bmatrix} 3 & -1 \\ 4 & 2 \end{bmatrix}$ w

Conv

v $\begin{bmatrix} 11 & -13 \\ 4 & 9 \end{bmatrix}$ (h_{out}, w_{out})

$\frac{\partial J}{\partial v}$ $\begin{bmatrix} 3 & 2 \\ 1 & -1 \end{bmatrix}$

$\frac{\partial v}{\partial w}$ is shape $\big((f, f), (h_{out}, w_{out})\big)$

Interpretation: How does each filter weight affect each feature map element

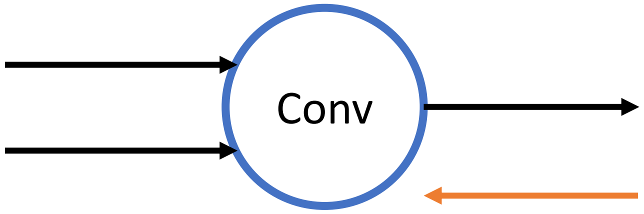
$$\frac{\partial v}{\partial w} = \begin{bmatrix} \boxed{\frac{\partial v}{\partial w_{11}}} & \frac{\partial v}{\partial w_{12}} \\ \frac{\partial v}{\partial w_{21}} & \frac{\partial v}{\partial w_{22}} \end{bmatrix} \qquad \frac{\partial v}{\partial w_{11}} = \begin{bmatrix} \frac{\partial v_{11}}{\partial w_{11}} & \frac{\partial v_{12}}{\partial w_{11}} \\ \frac{\partial v_{21}}{\partial w_{11}} & \frac{\partial v_{22}}{\partial w_{11}} \end{bmatrix}$$

(f, f) $\qquad$ (h_{out}, w_{out})

Interpretation: How does this one weight weight affect each feature map element

(h_{in}, w_{in}) $\begin{bmatrix} 2 & -3 & 4 \\ 1 & -1 & 2 \\ -1 & 2 & 3 \end{bmatrix}$ x

(f, f) $\begin{bmatrix} 3 & -1 \\ 4 & 2 \end{bmatrix}$ w



v $\begin{bmatrix} 11 & -13 \\ 4 & 9 \end{bmatrix}$ (h_{out}, w_{out})

$\frac{\partial J}{\partial v}$ $\begin{bmatrix} 3 & 2 \\ 1 & -1 \end{bmatrix}$

$\frac{\partial v}{\partial w}$ is shape $((f, f), (h_{out}, w_{out}))$

Interpretation: How does each filter weight affect each feature map element

$$\frac{\partial v}{\partial w} = \begin{bmatrix} \boxed{\frac{\partial v}{\partial w_{11}}} & \frac{\partial v}{\partial w_{12}} \\ \frac{\partial v}{\partial w_{21}} & \frac{\partial v}{\partial w_{22}} \end{bmatrix}$$

(f, f)

$$\frac{\partial v}{\partial w_{11}} = \begin{bmatrix} \frac{\partial v_{11}}{\partial w_{11}} & \frac{\partial v_{12}}{\partial w_{11}} \\ \frac{\partial v_{21}}{\partial w_{11}} & \frac{\partial v_{22}}{\partial w_{11}} \end{bmatrix}$$

(h_{out}, w_{out})

$$v_{11} = w_{11}x_{11} + w_{12}x_{12} + w_{21}x_{21} + w_{22}x_{22}$$
$$v_{12} = w_{11}x_{12} + w_{12}x_{13} + w_{21}x_{22} + w_{22}x_{23}$$
$$v_{21} = w_{11}x_{21} + w_{12}x_{22} + w_{21}x_{31} + w_{22}x_{31}$$
$$v_{22} = w_{11}x_{22} + w_{12}x_{23} + w_{21}x_{32} + w_{22}x_{33}$$

Interpretation: How does this one weight weight affect each feature map element

Brad Quinton, Scott Chin

(h_{in}, w_{in}) $\begin{bmatrix} 2 & -3 & 4 \\ 1 & -1 & 2 \\ -1 & 2 & 3 \end{bmatrix}$ x

(f, f) $\begin{bmatrix} 3 & -1 \\ 4 & 2 \end{bmatrix}$ w

Conv

v $\begin{bmatrix} 11 & -13 \\ 4 & 9 \end{bmatrix}$ (h_{out}, w_{out})

$\dfrac{\partial J}{\partial v}$ $\begin{bmatrix} 3 & 2 \\ 1 & -1 \end{bmatrix}$

$\dfrac{\partial v}{\partial w}$ is shape $((f, f), (h_{out}, w_{out}))$

Interpretation: How does each filter weight affect each feature map element

$$\frac{\partial v}{\partial w} = \begin{bmatrix} \boxed{\dfrac{\partial v}{\partial w_{11}}} & \dfrac{\partial v}{\partial w_{12}} \\ \dfrac{\partial v}{\partial w_{21}} & \dfrac{\partial v}{\partial w_{22}} \end{bmatrix}$$

(f, f)

$$\frac{\partial v}{\partial w_{11}} = \begin{bmatrix} \boxed{\dfrac{\partial v_{11}}{\partial w_{11}}} & \dfrac{\partial v_{12}}{\partial w_{11}} \\ \dfrac{\partial v_{21}}{\partial w_{11}} & \dfrac{\partial v_{22}}{\partial w_{11}} \end{bmatrix}$$

(h_{out}, w_{out})

$$\boxed{v_{11} = w_{11}x_{11} + w_{12}x_{12} + w_{21}x_{21} + w_{22}x_{22}}$$

$$v_{12} = w_{11}x_{12} + w_{12}x_{13} + w_{21}x_{22} + w_{22}x_{23}$$

$$v_{21} = w_{11}x_{21} + w_{12}x_{22} + w_{21}x_{31} + w_{22}x_{31}$$

$$v_{22} = w_{11}x_{22} + w_{12}x_{23} + w_{21}x_{32} + w_{22}x_{33}$$

Interpretation: How does this one weight weight affect each feature map element

(h_{in}, w_{in}) $\begin{bmatrix} 2 & -3 & 4 \\ 1 & -1 & 2 \\ -1 & 2 & 3 \end{bmatrix}$ x

(f, f) $\begin{bmatrix} 3 & -1 \\ 4 & 2 \end{bmatrix}$ w

Conv

v $\begin{bmatrix} 11 & -13 \\ 4 & 9 \end{bmatrix}$ (h_{out}, w_{out})

$\frac{\partial J}{\partial v}$ $\begin{bmatrix} 3 & 2 \\ 1 & -1 \end{bmatrix}$

$\frac{\partial v}{\partial w}$ is shape $((f, f), (h_{out}, w_{out}))$

Interpretation: How does each filter weight affect each feature map element

$$\frac{\partial v}{\partial w} = \begin{bmatrix} \boxed{\frac{\partial v}{\partial w_{11}}} & \frac{\partial v}{\partial w_{12}} \\ \frac{\partial v}{\partial w_{21}} & \frac{\partial v}{\partial w_{22}} \end{bmatrix}$$

(f, f)

$$\frac{\partial v}{\partial w_{11}} = \begin{bmatrix} \boxed{x_{11}} & \frac{\partial v_{12}}{\partial w_{11}} \\ \frac{\partial v_{21}}{\partial w_{11}} & \frac{\partial v_{22}}{\partial w_{11}} \end{bmatrix}$$

(h_{out}, w_{out})

$$\boxed{v_{11} = w_{11}x_{11} + w_{12}x_{12} + w_{21}x_{21} + w_{22}x_{22}}$$

$$v_{12} = w_{11}x_{12} + w_{12}x_{13} + w_{21}x_{22} + w_{22}x_{23}$$

$$v_{21} = w_{11}x_{21} + w_{12}x_{22} + w_{21}x_{31} + w_{22}x_{31}$$

$$v_{22} = w_{11}x_{22} + w_{12}x_{23} + w_{21}x_{32} + w_{22}x_{33}$$

Interpretation: How does this one weight weight affect each feature map element

Brad Quinton, Scott Chin

(h_{in}, w_{in}) $\begin{bmatrix} 2 & -3 & 4 \\ 1 & -1 & 2 \\ -1 & 2 & 3 \end{bmatrix}$ x

(f, f) $\begin{bmatrix} 3 & -1 \\ 4 & 2 \end{bmatrix}$ w

Conv

v $\begin{bmatrix} 11 & -13 \\ 4 & 9 \end{bmatrix}$ (h_{out}, w_{out})

$\frac{\partial J}{\partial v}$ $\begin{bmatrix} 3 & 2 \\ 1 & -1 \end{bmatrix}$

$\frac{\partial v}{\partial w}$ is shape $((f, f), (h_{out}, w_{out}))$

Interpretation: How does each filter weight affect each feature map element

$$\frac{\partial v}{\partial w} = \begin{bmatrix} \boxed{\frac{\partial v}{\partial w_{11}}} & \frac{\partial v}{\partial w_{12}} \\ \frac{\partial v}{\partial w_{21}} & \frac{\partial v}{\partial w_{22}} \end{bmatrix}$$

(f, f)

$$\frac{\partial v}{\partial w_{11}} = \begin{bmatrix} x_{11} & \boxed{\frac{\partial v_{12}}{\partial w_{11}}} \\ \frac{\partial v_{21}}{\partial w_{11}} & \frac{\partial v_{22}}{\partial w_{11}} \end{bmatrix}$$

(h_{out}, w_{out})

$$v_{11} = w_{11}x_{11} + w_{12}x_{12} + w_{21}x_{21} + w_{22}x_{22}$$

$$\boxed{v_{12} = w_{11}x_{12} + w_{12}x_{13} + w_{21}x_{22} + w_{22}x_{23}}$$

$$v_{21} = w_{11}x_{21} + w_{12}x_{22} + w_{21}x_{31} + w_{22}x_{31}$$

$$v_{22} = w_{11}x_{22} + w_{12}x_{23} + w_{21}x_{32} + w_{22}x_{33}$$

Interpretation: How does this one weight weight affect each feature map element

Brad Quinton, Scott Chin

(h_{in}, w_{in}) $\begin{bmatrix} 2 & -3 & 4 \\ 1 & -1 & 2 \\ -1 & 2 & 3 \end{bmatrix}$ x

(f, f) $\begin{bmatrix} 3 & -1 \\ 4 & 2 \end{bmatrix}$ w

$x, w \rightarrow$ Conv $\rightarrow v$

v $\begin{bmatrix} 11 & -13 \\ 4 & 9 \end{bmatrix}$ (h_{out}, w_{out})

$\frac{\partial J}{\partial v}$ $\begin{bmatrix} 3 & 2 \\ 1 & -1 \end{bmatrix}$

$\frac{\partial v}{\partial w}$ is shape $((f, f), (h_{out}, w_{out}))$

Interpretation: How does each filter weight affect each feature map element

$$\frac{\partial v}{\partial w} = \begin{bmatrix} \frac{\partial v}{\partial w_{11}} & \frac{\partial v}{\partial w_{12}} \\ \frac{\partial v}{\partial w_{21}} & \frac{\partial v}{\partial w_{22}} \end{bmatrix}$$

(f, f)

$$\frac{\partial v}{\partial w_{11}} = \begin{bmatrix} x_{11} & x_{12} \\ \frac{\partial v_{21}}{\partial w_{11}} & \frac{\partial v_{22}}{\partial w_{11}} \end{bmatrix}$$

(h_{out}, w_{out})

$$v_{11} = w_{11}x_{11} + w_{12}x_{12} + w_{21}x_{21} + w_{22}x_{22}$$

$$v_{12} = w_{11}x_{12} + w_{12}x_{13} + w_{21}x_{22} + w_{22}x_{23}$$

$$v_{21} = w_{11}x_{21} + w_{12}x_{22} + w_{21}x_{31} + w_{22}x_{31}$$

$$v_{22} = w_{11}x_{22} + w_{12}x_{23} + w_{21}x_{32} + w_{22}x_{33}$$

Interpretation: How does this one weight weight affect each feature map element

Brad Quinton, Scott Chin

(h_{in}, w_{in}) $\begin{bmatrix} 2 & -3 & 4 \\ 1 & -1 & 2 \\ -1 & 2 & 3 \end{bmatrix}$ x

(f, f) $\begin{bmatrix} 3 & -1 \\ 4 & 2 \end{bmatrix}$ w

$x, w \xrightarrow{\text{Conv}} v$ $\begin{bmatrix} 11 & -13 \\ 4 & 9 \end{bmatrix}$ (h_{out}, w_{out})

$\frac{\partial J}{\partial v}$ $\begin{bmatrix} 3 & 2 \\ 1 & -1 \end{bmatrix}$

$\frac{\partial v}{\partial w}$ is shape $((f, f), (h_{out}, w_{out}))$

Interpretation: How does each filter weight affect each feature map element

$$\frac{\partial v}{\partial w} = \begin{bmatrix} \boxed{\frac{\partial v}{\partial w_{11}}} & \frac{\partial v}{\partial w_{12}} \\ \frac{\partial v}{\partial w_{21}} & \frac{\partial v}{\partial w_{22}} \end{bmatrix}$$

(f, f)

$$\frac{\partial v}{\partial w_{11}} = \begin{bmatrix} x_{11} & x_{12} \\ \boxed{\frac{\partial v_{21}}{\partial w_{11}}} & \frac{\partial v_{22}}{\partial w_{11}} \end{bmatrix}$$

(h_{out}, w_{out})

$$v_{11} = w_{11}x_{11} + w_{12}x_{12} + w_{21}x_{21} + w_{22}x_{22}$$

$$v_{12} = w_{11}x_{12} + w_{12}x_{13} + w_{21}x_{22} + w_{22}x_{23}$$

$$\boxed{v_{21} = w_{11}x_{21} + w_{12}x_{22} + w_{21}x_{31} + w_{22}x_{31}}$$

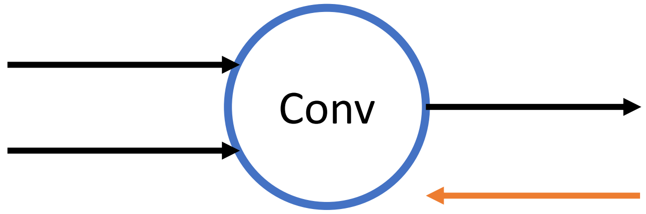
$$v_{22} = w_{11}x_{22} + w_{12}x_{23} + w_{21}x_{32} + w_{22}x_{33}$$

Interpretation: How does this one weight weight affect each feature map element

Brad Quinton, Scott Chin

(h_{in}, w_{in}) $\begin{bmatrix} 2 & -3 & 4 \\ 1 & -1 & 2 \\ -1 & 2 & 3 \end{bmatrix}$ x

(f, f) $\begin{bmatrix} 3 & -1 \\ 4 & 2 \end{bmatrix}$ w



v $\begin{bmatrix} 11 & -13 \\ 4 & 9 \end{bmatrix}$ (h_{out}, w_{out})

$\frac{\partial J}{\partial v}$ $\begin{bmatrix} 3 & 2 \\ 1 & -1 \end{bmatrix}$

$\frac{\partial v}{\partial w}$ is shape $((f, f), (h_{out}, w_{out}))$

Interpretation: How does each filter weight affect each feature map element

$$\frac{\partial v}{\partial w} = \begin{bmatrix} \boxed{\frac{\partial v}{\partial w_{11}}} & \frac{\partial v}{\partial w_{12}} \\ \frac{\partial v}{\partial w_{21}} & \frac{\partial v}{\partial w_{22}} \end{bmatrix}$$

(f, f)

$$\frac{\partial v}{\partial w_{11}} = \begin{bmatrix} x_{11} & x_{12} \\ \boxed{x_{21}} & \frac{\partial v_{22}}{\partial w_{11}} \end{bmatrix}$$

(h_{out}, w_{out})

$$v_{11} = w_{11}x_{11} + w_{12}x_{12} + w_{21}x_{21} + w_{22}x_{22}$$

$$v_{12} = w_{11}x_{12} + w_{12}x_{13} + w_{21}x_{22} + w_{22}x_{23}$$

$$\boxed{v_{21} = w_{11}x_{21} + w_{12}x_{22} + w_{21}x_{31} + w_{22}x_{31}}$$

$$v_{22} = w_{11}x_{22} + w_{12}x_{23} + w_{21}x_{32} + w_{22}x_{33}$$

Interpretation: How does this one weight weight affect each feature map element

Brad Quinton, Scott Chin

(h_{in}, w_{in}) $\begin{bmatrix} 2 & -3 & 4 \\ 1 & -1 & 2 \\ -1 & 2 & 3 \end{bmatrix}$ x

(f, f) $\begin{bmatrix} 3 & -1 \\ 4 & 2 \end{bmatrix}$ w

Conv

v $\begin{bmatrix} 11 & -13 \\ 4 & 9 \end{bmatrix}$ (h_{out}, w_{out})

$\dfrac{\partial J}{\partial v}$ $\begin{bmatrix} 3 & 2 \\ 1 & -1 \end{bmatrix}$

$\dfrac{\partial v}{\partial w}$ is shape $((f, f), (h_{out}, w_{out}))$

Interpretation: How does each filter weight affect each feature map element

$$\frac{\partial v}{\partial w} = \begin{bmatrix} \boxed{\dfrac{\partial v}{\partial w_{11}}} & \dfrac{\partial v}{\partial w_{12}} \\ \dfrac{\partial v}{\partial w_{21}} & \dfrac{\partial v}{\partial w_{22}} \end{bmatrix}$$

(f, f)

$$\frac{\partial v}{\partial w_{11}} = \begin{bmatrix} x_{11} & x_{12} \\ x_{21} & \boxed{\dfrac{\partial v_{22}}{\partial w_{11}}} \end{bmatrix}$$

(h_{out}, w_{out})

$$v_{11} = w_{11}x_{11} + w_{12}x_{12} + w_{21}x_{21} + w_{22}x_{22}$$

$$v_{12} = w_{11}x_{12} + w_{12}x_{13} + w_{21}x_{22} + w_{22}x_{23}$$

$$v_{21} = w_{11}x_{21} + w_{12}x_{22} + w_{21}x_{31} + w_{22}x_{31}$$

$$\boxed{v_{22} = w_{11}x_{22} + w_{12}x_{23} + w_{21}x_{32} + w_{22}x_{33}}$$

Interpretation: How does this one weight weight affect each feature map element

(h_{in}, w_{in}) $\begin{bmatrix} 2 & -3 & 4 \\ 1 & -1 & 2 \\ -1 & 2 & 3 \end{bmatrix}$ x

(f, f) $\begin{bmatrix} 3 & -1 \\ 4 & 2 \end{bmatrix}$ w

Conv

v $\begin{bmatrix} 11 & -13 \\ 4 & 9 \end{bmatrix}$ (h_{out}, w_{out})

$\frac{\partial J}{\partial v}$ $\begin{bmatrix} 3 & 2 \\ 1 & -1 \end{bmatrix}$

$\frac{\partial v}{\partial w}$ is shape $((f, f), (h_{out}, w_{out}))$

Interpretation: How does each filter weight affect each feature map element

$$\frac{\partial v}{\partial w} = \begin{bmatrix} \boxed{\frac{\partial v}{\partial w_{11}}} & \frac{\partial v}{\partial w_{12}} \\ \frac{\partial v}{\partial w_{21}} & \frac{\partial v}{\partial w_{22}} \end{bmatrix}$$

(f, f)

$$\frac{\partial v}{\partial w_{11}} = \begin{bmatrix} x_{11} & x_{12} \\ x_{21} & \boxed{x_{22}} \end{bmatrix}$$

(h_{out}, w_{out})

$$v_{11} = w_{11}x_{11} + w_{12}x_{12} + w_{21}x_{21} + w_{22}x_{22}$$

$$v_{12} = w_{11}x_{12} + w_{12}x_{13} + w_{21}x_{22} + w_{22}x_{23}$$

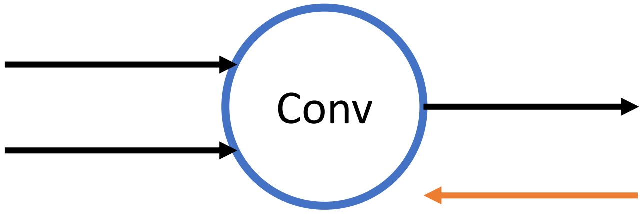
$$v_{21} = w_{11}x_{21} + w_{12}x_{22} + w_{21}x_{31} + w_{22}x_{31}$$

$$\boxed{v_{22} = w_{11}x_{22} + w_{12}x_{23} + w_{21}x_{32} + w_{22}x_{33}}$$

Interpretation: How does this one weight weight affect each feature map element

(h_{in}, w_{in}) $\begin{bmatrix} 2 & -3 & 4 \\ 1 & -1 & 2 \\ -1 & 2 & 3 \end{bmatrix}$ x

(f, f) $\begin{bmatrix} 3 & -1 \\ 4 & 2 \end{bmatrix}$ w



v $\begin{bmatrix} 11 & -13 \\ 4 & 9 \end{bmatrix}$ (h_{out}, w_{out})

$\frac{\partial J}{\partial v}$ $\begin{bmatrix} 3 & 2 \\ 1 & -1 \end{bmatrix}$

$\frac{\partial v}{\partial w}$ is shape $((f, f), (h_{out}, w_{out}))$

Interpretation: How does each filter weight affect each feature map element

$$\frac{\partial v}{\partial w} = \begin{bmatrix} \boxed{\frac{\partial v}{\partial w_{11}}} & \frac{\partial v}{\partial w_{12}} \\ \frac{\partial v}{\partial w_{21}} & \frac{\partial v}{\partial w_{22}} \end{bmatrix}$$

(f, f)

$$\frac{\partial v}{\partial w_{11}} = \begin{bmatrix} x_{11} & x_{12} \\ x_{21} & x_{22} \end{bmatrix}$$

(h_{out}, w_{out})

$$v_{11} = w_{11}x_{11} + w_{12}x_{12} + w_{21}x_{21} + w_{22}x_{22}$$

$$v_{12} = w_{11}x_{12} + w_{12}x_{13} + w_{21}x_{22} + w_{22}x_{23}$$

$$v_{21} = w_{11}x_{21} + w_{12}x_{22} + w_{21}x_{31} + w_{22}x_{31}$$

$$v_{22} = w_{11}x_{22} + w_{12}x_{23} + w_{21}x_{32} + w_{22}x_{33}$$

Interpretation: How does this one weight weight affect each feature map element

Brad Quinton, Scott Chin

(h_{in}, w_{in}) $\begin{bmatrix} 2 & -3 & 4 \\ 1 & -1 & 2 \\ -1 & 2 & 3 \end{bmatrix}$ x → Conv → v $\begin{bmatrix} 11 & -13 \\ 4 & 9 \end{bmatrix}$ (h_{out}, w_{out})

(f, f) $\begin{bmatrix} 3 & -1 \\ 4 & 2 \end{bmatrix}$ w → Conv

$\frac{\partial J}{\partial v}$ $\begin{bmatrix} 3 & 2 \\ 1 & -1 \end{bmatrix}$

$\frac{\partial v}{\partial w}$ is shape $((f, f), (h_{out}, w_{out}))$

Interpretation: How does each filter weight affect each feature map element

$$\frac{\partial v}{\partial w} = \begin{bmatrix} \boxed{\frac{\partial v}{\partial w_{11}}} & \frac{\partial v}{\partial w_{12}} \\ \frac{\partial v}{\partial w_{21}} & \frac{\partial v}{\partial w_{22}} \end{bmatrix}$$

(f, f)

$$\frac{\partial v}{\partial w_{11}} = \begin{bmatrix} 2 & -3 \\ 1 & -1 \end{bmatrix}$$

(h_{out}, w_{out})

$$v_{11} = w_{11}x_{11} + w_{12}x_{12} + w_{21}x_{21} + w_{22}x_{22}$$

$$v_{12} = w_{11}x_{12} + w_{12}x_{13} + w_{21}x_{22} + w_{22}x_{23}$$

$$v_{21} = w_{11}x_{21} + w_{12}x_{22} + w_{21}x_{31} + w_{22}x_{31}$$

$$v_{22} = w_{11}x_{22} + w_{12}x_{23} + w_{21}x_{32} + w_{22}x_{33}$$

Interpretation: How does this one weight weight affect each feature map element

Brad Quinton, Scott Chin

(h_{in}, w_{in}) $\begin{bmatrix} 2 & -3 & 4 \\ 1 & -1 & 2 \\ -1 & 2 & 3 \end{bmatrix}$ x

(f, f) $\begin{bmatrix} 3 & -1 \\ 4 & 2 \end{bmatrix}$ w

Conv

v $\begin{bmatrix} 11 & -13 \\ 4 & 9 \end{bmatrix}$ (h_{out}, w_{out})

$\frac{\partial J}{\partial v}$ $\begin{bmatrix} 3 & 2 \\ 1 & -1 \end{bmatrix}$

$\frac{\partial v}{\partial w}$ is shape $((f, f), (h_{out}, w_{out}))$

Interpretation: How does each filter weight affect each feature map element

$$\frac{\partial v}{\partial w} = \begin{bmatrix} \frac{\partial v}{\partial w_{11}} & \boxed{\frac{\partial v}{\partial w_{12}}} \\ \frac{\partial v}{\partial w_{21}} & \frac{\partial v}{\partial w_{22}} \end{bmatrix}$$

(f, f)

$$\frac{\partial v}{\partial w_{11}} = \begin{bmatrix} 2 & -3 \\ 1 & -1 \end{bmatrix}$$

$$\frac{\partial v}{\partial w_{12}} = \begin{bmatrix} \frac{\partial v_{11}}{\partial w_{12}} & \frac{\partial v_{12}}{\partial w_{12}} \\ \frac{\partial v_{21}}{\partial w_{12}} & \frac{\partial v_{22}}{\partial w_{12}} \end{bmatrix}$$

(h_{out}, w_{out})

$$v_{11} = w_{11}x_{11} + w_{12}x_{12} + w_{21}x_{21} + w_{22}x_{22}$$

$$v_{12} = w_{11}x_{12} + w_{12}x_{13} + w_{21}x_{22} + w_{22}x_{23}$$

$$v_{21} = w_{11}x_{21} + w_{12}x_{22} + w_{21}x_{31} + w_{22}x_{31}$$

$$v_{22} = w_{11}x_{22} + w_{12}x_{23} + w_{21}x_{32} + w_{22}x_{33}$$

Interpretation: How does this one weight weight affect each feature map element

Brad Quinton, Scott Chin

(h_{in}, w_{in}) $\begin{bmatrix} 2 & -3 & 4 \\ 1 & -1 & 2 \\ -1 & 2 & 3 \end{bmatrix}$ x

(f, f) $\begin{bmatrix} 3 & -1 \\ 4 & 2 \end{bmatrix}$ w

$x, w \longrightarrow$ Conv $\longrightarrow v$ $\begin{bmatrix} 11 & -13 \\ 4 & 9 \end{bmatrix}$ (h_{out}, w_{out})

$\dfrac{\partial J}{\partial v}$ $\begin{bmatrix} 3 & 2 \\ 1 & -1 \end{bmatrix}$

$\dfrac{\partial v}{\partial w}$ is shape $((f, f), (h_{out}, w_{out}))$

Interpretation: How does each filter weight affect each feature map element

$$\frac{\partial v}{\partial w} = \begin{bmatrix} \frac{\partial v}{\partial w_{11}} & \boxed{\frac{\partial v}{\partial w_{12}}} \\ \frac{\partial v}{\partial w_{21}} & \frac{\partial v}{\partial w_{22}} \end{bmatrix}$$

(f, f)

$$\frac{\partial v}{\partial w_{11}} = \begin{bmatrix} 2 & -3 \\ 1 & -1 \end{bmatrix}$$

$$\frac{\partial v}{\partial w_{12}} = \begin{bmatrix} \frac{\partial v_{11}}{\partial w_{12}} & \frac{\partial v_{12}}{\partial w_{12}} \\ \frac{\partial v_{21}}{\partial w_{12}} & \frac{\partial v_{22}}{\partial w_{12}} \end{bmatrix}$$

(h_{out}, w_{out})

$$v_{11} = w_{11}x_{11} + w_{12}x_{12} + w_{21}x_{21} + w_{22}x_{22}$$
$$v_{12} = w_{11}x_{12} + w_{12}x_{13} + w_{21}x_{22} + w_{22}x_{23}$$
$$v_{21} = w_{11}x_{21} + w_{12}x_{22} + w_{21}x_{31} + w_{22}x_{31}$$
$$v_{22} = w_{11}x_{22} + w_{12}x_{23} + w_{21}x_{32} + w_{22}x_{33}$$

Interpretation: How does this one weight weight affect each feature map element

Brad Quinton, Scott Chin

(h_{in}, w_{in}) $\begin{bmatrix} 2 & -3 & 4 \\ 1 & -1 & 2 \\ -1 & 2 & 3 \end{bmatrix}$ x

(f, f) $\begin{bmatrix} 3 & -1 \\ 4 & 2 \end{bmatrix}$ w

Conv

v $\begin{bmatrix} 11 & -13 \\ 4 & 9 \end{bmatrix}$ (h_{out}, w_{out})

$\frac{\partial J}{\partial v}$ $\begin{bmatrix} 3 & 2 \\ 1 & -1 \end{bmatrix}$

$\frac{\partial v}{\partial w}$ is shape $((f, f), (h_{out}, w_{out}))$

Interpretation: How does each filter weight affect each feature map element

$$\frac{\partial v}{\partial w} = \begin{bmatrix} \frac{\partial v}{\partial w_{11}} & \boxed{\frac{\partial v}{\partial w_{12}}} \\ \frac{\partial v}{\partial w_{21}} & \frac{\partial v}{\partial w_{22}} \end{bmatrix}$$

(f, f)

$$\frac{\partial v}{\partial w_{11}} = \begin{bmatrix} 2 & -3 \\ 1 & -1 \end{bmatrix}$$

$$\frac{\partial v}{\partial w_{12}} = \begin{bmatrix} x_{12} & x_{13} \\ x_{22} & x_{23} \end{bmatrix}$$

(h_{out}, w_{out})

$$v_{11} = w_{11}x_{11} + \boxed{w_{12}x_{12}} + w_{21}x_{21} + w_{22}x_{22}$$
$$v_{12} = w_{11}x_{12} + \boxed{w_{12}x_{13}} + w_{21}x_{22} + w_{22}x_{23}$$
$$v_{21} = w_{11}x_{21} + \boxed{w_{12}x_{22}} + w_{21}x_{31} + w_{22}x_{31}$$
$$v_{22} = w_{11}x_{22} + \boxed{w_{12}x_{23}} + w_{21}x_{32} + w_{22}x_{33}$$

Interpretation: How does this one weight weight affect each feature map element

Brad Quinton, Scott Chin

(h_{in}, w_{in}) $\begin{bmatrix} 2 & -3 & 4 \\ 1 & -1 & 2 \\ -1 & 2 & 3 \end{bmatrix}$ x $\longrightarrow$ Conv $\longrightarrow$ v $\begin{bmatrix} 11 & -13 \\ 4 & 9 \end{bmatrix}$ (h_{out}, w_{out})

(f, f) $\begin{bmatrix} 3 & -1 \\ 4 & 2 \end{bmatrix}$ w $\longrightarrow$ Conv

$\dfrac{\partial J}{\partial v}$ $\begin{bmatrix} 3 & 2 \\ 1 & -1 \end{bmatrix}$

$\dfrac{\partial v}{\partial w}$ is shape $((f, f), (h_{out}, w_{out}))$

Interpretation: How does each filter weight affect each feature map element

$$\frac{\partial v}{\partial w} = \begin{bmatrix} \frac{\partial v}{\partial w_{11}} & \boxed{\frac{\partial v}{\partial w_{12}}} \\ \frac{\partial v}{\partial w_{21}} & \frac{\partial v}{\partial w_{22}} \end{bmatrix}$$

(f, f)

$$\frac{\partial v}{\partial w_{11}} = \begin{bmatrix} 2 & -3 \\ 1 & -1 \end{bmatrix}$$

$$\frac{\partial v}{\partial w_{12}} = \begin{bmatrix} -3 & 4 \\ -1 & 2 \end{bmatrix}$$

(h_{out}, w_{out})

$$v_{11} = w_{11}x_{11} + \boxed{w_{12}x_{12}} + w_{21}x_{21} + w_{22}x_{22}$$
$$v_{12} = w_{11}x_{12} + \boxed{w_{12}x_{13}} + w_{21}x_{22} + w_{22}x_{23}$$
$$v_{21} = w_{11}x_{21} + \boxed{w_{12}x_{22}} + w_{21}x_{31} + w_{22}x_{31}$$
$$v_{22} = w_{11}x_{22} + \boxed{w_{12}x_{23}} + w_{21}x_{32} + w_{22}x_{33}$$

Interpretation: How does this one weight weight affect each feature map element

Brad Quinton, Scott Chin

(h_{in}, w_{in}) $\begin{bmatrix} 2 & -3 & 4 \\ 1 & -1 & 2 \\ -1 & 2 & 3 \end{bmatrix}$ x

(f, f) $\begin{bmatrix} 3 & -1 \\ 4 & 2 \end{bmatrix}$ w

Conv

v $\begin{bmatrix} 11 & -13 \\ 4 & 9 \end{bmatrix}$ (h_{out}, w_{out})

$\frac{\partial J}{\partial v}$ $\begin{bmatrix} 3 & 2 \\ 1 & -1 \end{bmatrix}$

$\frac{\partial v}{\partial w}$ is shape $\big((f, f), (h_{out}, w_{out})\big)$

Interpretation: How does each filter weight affect each feature map element

$$\frac{\partial v}{\partial w} = \begin{bmatrix} \frac{\partial v}{\partial w_{11}} & \frac{\partial v}{\partial w_{12}} \\ \boxed{\frac{\partial v}{\partial w_{21}}} & \frac{\partial v}{\partial w_{22}} \end{bmatrix}$$

(f, f)

$$\frac{\partial v}{\partial w_{11}} = \begin{bmatrix} 2 & -3 \\ 1 & -1 \end{bmatrix}$$

$$\frac{\partial v}{\partial w_{12}} = \begin{bmatrix} -3 & 4 \\ -1 & 2 \end{bmatrix}$$

$$\frac{\partial v}{\partial w_{21}} = \begin{bmatrix} x_{21} & x_{22} \\ x_{31} & x_{32} \end{bmatrix}$$

(h_{out}, w_{out})

$$v_{11} = w_{11}x_{11} + w_{12}x_{12} + \boxed{w_{21}x_{21}} + w_{22}x_{22}$$

$$v_{12} = w_{11}x_{12} + w_{12}x_{13} + \boxed{w_{21}x_{22}} + w_{22}x_{23}$$

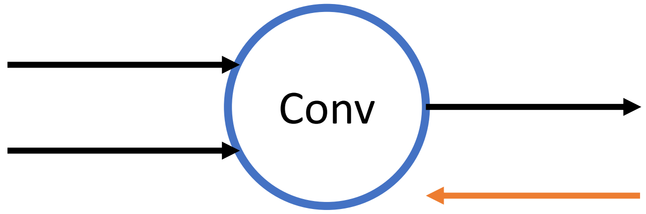
$$v_{21} = w_{11}x_{21} + w_{12}x_{22} + \boxed{w_{21}x_{31}} + w_{22}x_{31}$$

$$v_{22} = w_{11}x_{22} + w_{12}x_{23} + \boxed{w_{21}x_{32}} + w_{22}x_{33}$$

Interpretation: How does this one weight weight affect each feature map element

(h_{in}, w_{in}) $\begin{bmatrix} 2 & -3 & 4 \\ 1 & -1 & 2 \\ -1 & 2 & 3 \end{bmatrix}$ x

(f, f) $\begin{bmatrix} 3 & -1 \\ 4 & 2 \end{bmatrix}$ w



v $\begin{bmatrix} 11 & -13 \\ 4 & 9 \end{bmatrix}$ (h_{out}, w_{out})

$\frac{\partial J}{\partial v}$ $\begin{bmatrix} 3 & 2 \\ 1 & -1 \end{bmatrix}$

$\frac{\partial v}{\partial w}$ is shape $\left((f, f), (h_{out}, w_{out})\right)$

Interpretation: How does each filter weight affect each feature map element

$$\frac{\partial v}{\partial w} = \begin{bmatrix} \frac{\partial v}{\partial w_{11}} & \frac{\partial v}{\partial w_{12}} \\ \boxed{\frac{\partial v}{\partial w_{21}}} & \frac{\partial v}{\partial w_{22}} \end{bmatrix}$$

(f, f)

$$\frac{\partial v}{\partial w_{11}} = \begin{bmatrix} 2 & -3 \\ 1 & -1 \end{bmatrix}$$

$$\frac{\partial v}{\partial w_{12}} = \begin{bmatrix} -3 & 4 \\ -1 & 2 \end{bmatrix}$$

$$\frac{\partial v}{\partial w_{21}} = \begin{bmatrix} 1 & -1 \\ -1 & 2 \end{bmatrix}$$

(h_{out}, w_{out})

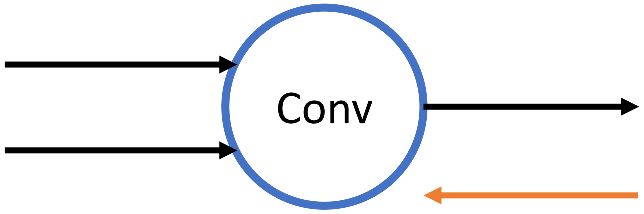
$$v_{11} = w_{11}x_{11} + w_{12}x_{12} + w_{21}x_{21} + w_{22}x_{22}$$
$$v_{12} = w_{11}x_{12} + w_{12}x_{13} + w_{21}x_{22} + w_{22}x_{23}$$
$$v_{21} = w_{11}x_{21} + w_{12}x_{22} + w_{21}x_{31} + w_{22}x_{31}$$
$$v_{22} = w_{11}x_{22} + w_{12}x_{23} + w_{21}x_{32} + w_{22}x_{33}$$

Interpretation: How does this one weight weight affect each feature map element

Brad Quinton, Scott Chin

(h_{in}, w_{in}) $\begin{bmatrix} 2 & -3 & 4 \\ 1 & -1 & 2 \\ -1 & 2 & 3 \end{bmatrix}$ x

(f, f) $\begin{bmatrix} 3 & -1 \\ 4 & 2 \end{bmatrix}$ w



v $\begin{bmatrix} 11 & -13 \\ 4 & 9 \end{bmatrix}$ (h_{out}, w_{out})

$\frac{\partial J}{\partial v}$ $\begin{bmatrix} 3 & 2 \\ 1 & -1 \end{bmatrix}$

$\frac{\partial v}{\partial w}$ is shape $((f, f), (h_{out}, w_{out}))$

Interpretation: How does each filter weight affect each feature map element

$$\frac{\partial v}{\partial w} = \begin{bmatrix} \frac{\partial v}{\partial w_{11}} & \frac{\partial v}{\partial w_{12}} \\ \frac{\partial v}{\partial w_{21}} & \boxed{\frac{\partial v}{\partial w_{22}}} \end{bmatrix}$$

(f, f)

$$\frac{\partial v}{\partial w_{11}} = \begin{bmatrix} 2 & -3 \\ 1 & -1 \end{bmatrix}$$

$$\frac{\partial v}{\partial w_{12}} = \begin{bmatrix} -3 & 4 \\ -1 & 2 \end{bmatrix}$$

$$\frac{\partial v}{\partial w_{21}} = \begin{bmatrix} 1 & -1 \\ -1 & 2 \end{bmatrix}$$

$$\frac{\partial v}{\partial w_{22}} = \begin{bmatrix} x_{22} & x_{23} \\ x_{32} & x_{33} \end{bmatrix}$$

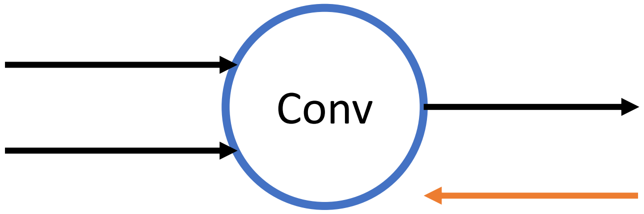
(h_{out}, w_{out})

$$v_{11} = w_{11}x_{11} + w_{12}x_{12} + w_{21}x_{21} + w_{22}x_{22}$$
$$v_{12} = w_{11}x_{12} + w_{12}x_{13} + w_{21}x_{22} + w_{22}x_{23}$$
$$v_{21} = w_{11}x_{21} + w_{12}x_{22} + w_{21}x_{31} + w_{22}x_{31}$$
$$v_{22} = w_{11}x_{22} + w_{12}x_{23} + w_{21}x_{32} + w_{22}x_{33}$$

Brad Quinton, Scott Chin

(h_{in}, w_{in}) $\begin{bmatrix} 2 & -3 & 4 \\ 1 & -1 & 2 \\ -1 & 2 & 3 \end{bmatrix}$ x

(f, f) $\begin{bmatrix} 3 & -1 \\ 4 & 2 \end{bmatrix}$ w



v $\begin{bmatrix} 11 & -13 \\ 4 & 9 \end{bmatrix}$ (h_{out}, w_{out})

$\frac{\partial J}{\partial v}$ $\begin{bmatrix} 3 & 2 \\ 1 & -1 \end{bmatrix}$

$\frac{\partial v}{\partial w}$ is shape $((f, f), (h_{out}, w_{out}))$

Interpretation: How does each filter weight affect each feature map element

$$\frac{\partial v}{\partial w} = \begin{bmatrix} \frac{\partial v}{\partial w_{11}} & \frac{\partial v}{\partial w_{12}} \\ \frac{\partial v}{\partial w_{21}} & \boxed{\frac{\partial v}{\partial w_{22}}} \end{bmatrix}$$

(f, f)

$$\frac{\partial v}{\partial w_{11}} = \begin{bmatrix} 2 & -3 \\ 1 & -1 \end{bmatrix}$$

$$\frac{\partial v}{\partial w_{12}} = \begin{bmatrix} -3 & 4 \\ -1 & 2 \end{bmatrix}$$

$$\frac{\partial v}{\partial w_{21}} = \begin{bmatrix} 1 & -1 \\ -1 & 2 \end{bmatrix}$$

$$\frac{\partial v}{\partial w_{22}} = \begin{bmatrix} -1 & 2 \\ 2 & 3 \end{bmatrix}$$

(h_{out}, w_{out})

$$v_{11} = w_{11}x_{11} + w_{12}x_{12} + w_{21}x_{21} + w_{22}x_{22}$$
$$v_{12} = w_{11}x_{12} + w_{12}x_{13} + w_{21}x_{22} + w_{22}x_{23}$$
$$v_{21} = w_{11}x_{21} + w_{12}x_{22} + w_{21}x_{31} + w_{22}x_{31}$$
$$v_{22} = w_{11}x_{22} + w_{12}x_{23} + w_{21}x_{32} + w_{22}x_{33}$$

Brad Quinton, Scott Chin

(h_{in}, w_{in}) $\begin{bmatrix} 2 & -3 & 4 \\ 1 & -1 & 2 \\ -1 & 2 & 3 \end{bmatrix}$ x

(f, f) $\begin{bmatrix} 3 & -1 \\ 4 & 2 \end{bmatrix}$ w

Conv

v $\begin{bmatrix} 11 & -13 \\ 4 & 9 \end{bmatrix}$ (h_{out}, w_{out})

$\frac{\partial J}{\partial v}$ $\begin{bmatrix} 3 & 2 \\ 1 & -1 \end{bmatrix}$

$\frac{\partial v}{\partial w}$ is shape $\left((f, f), (h_{out}, w_{out})\right)$

Interpretation: How does each filter weight affect each feature map element

$$\frac{\partial v}{\partial w} = \begin{bmatrix} \frac{\partial v}{\partial w_{11}} & \frac{\partial v}{\partial w_{12}} \\ \frac{\partial v}{\partial w_{21}} & \frac{\partial v}{\partial w_{22}} \end{bmatrix}$$

(f, f)

$$\frac{\partial v}{\partial w_{11}} = \begin{bmatrix} 2 & -3 \\ 1 & -1 \end{bmatrix}$$

$$\frac{\partial v}{\partial w_{12}} = \begin{bmatrix} -3 & 4 \\ -1 & 2 \end{bmatrix}$$

$$\frac{\partial v}{\partial w_{21}} = \begin{bmatrix} 1 & -1 \\ -1 & 2 \end{bmatrix}$$

$$\frac{\partial v}{\partial w_{22}} = \begin{bmatrix} -1 & 2 \\ 2 & 3 \end{bmatrix}$$

(h_{out}, w_{out})

Brad Quinton, Scott Chin

(h_{in}, w_{in}) $\begin{bmatrix} 2 & -3 & 4 \\ 1 & -1 & 2 \\ -1 & 2 & 3 \end{bmatrix}$ x

(f, f) $\begin{bmatrix} 3 & -1 \\ 4 & 2 \end{bmatrix}$ w

Conv

v $\begin{bmatrix} 11 & -13 \\ 4 & 9 \end{bmatrix}$ (h_{out}, w_{out})

$\frac{\partial J}{\partial v}$ $\begin{bmatrix} 3 & 2 \\ 1 & -1 \end{bmatrix}$

$\frac{\partial v}{\partial w}$ is shape $\big((f, f), (h_{out}, w_{out})\big)$

Interpretation: How does each filter weight affect each feature map element

$$\frac{\partial v}{\partial w} = \begin{bmatrix} \frac{\partial v}{\partial w_{11}} & \frac{\partial v}{\partial w_{12}} \\ \frac{\partial v}{\partial w_{21}} & \frac{\partial v}{\partial w_{22}} \end{bmatrix}$$

(f, f)

$$\frac{\partial v}{\partial w_{11}} = \begin{bmatrix} 2 & -3 \\ 1 & -1 \end{bmatrix}$$

$$\frac{\partial v}{\partial w_{12}} = \begin{bmatrix} -3 & 4 \\ -1 & 2 \end{bmatrix}$$

$$\frac{\partial v}{\partial w_{21}} = \begin{bmatrix} 1 & -1 \\ -1 & 2 \end{bmatrix}$$

$$\frac{\partial v}{\partial w_{22}} = \begin{bmatrix} -1 & 2 \\ 2 & 3 \end{bmatrix}$$

(h_{out}, w_{out})

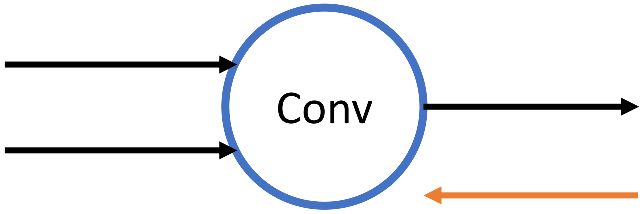
Can write the Jacobian this way

$$\frac{\partial v}{\partial w} = \begin{bmatrix} \begin{bmatrix} 2 & -3 \\ 1 & -1 \end{bmatrix} & \begin{bmatrix} -3 & 4 \\ -1 & 2 \end{bmatrix} \\ \begin{bmatrix} 1 & -1 \\ -1 & 2 \end{bmatrix} & \begin{bmatrix} -1 & 2 \\ 2 & 3 \end{bmatrix} \end{bmatrix}$$

$\big((f, f), (h_{out}, w_{out})\big)$

(h_{in}, w_{in}) $\begin{bmatrix} 2 & -3 & 4 \\ 1 & -1 & 2 \\ -1 & 2 & 3 \end{bmatrix}$ x

(f, f) $\begin{bmatrix} 3 & -1 \\ 4 & 2 \end{bmatrix}$ w



v $\begin{bmatrix} 11 & -13 \\ 4 & 9 \end{bmatrix}$ (h_{out}, w_{out})

$\frac{\partial J}{\partial v}$ $\begin{bmatrix} 3 & 2 \\ 1 & -1 \end{bmatrix}$

$\frac{\partial v}{\partial w}$ is shape $((f, f), (h_{out}, w_{out}))$

Interpretation: How does each filter weight affect each feature map element

$$\frac{\partial v}{\partial w} = \begin{bmatrix} \begin{bmatrix} 2 & -3 \\ 1 & -1 \end{bmatrix} & \begin{bmatrix} -3 & 4 \\ -1 & 2 \end{bmatrix} \\ \begin{bmatrix} 1 & -1 \\ -1 & 2 \end{bmatrix} & \begin{bmatrix} -1 & 2 \\ 2 & 3 \end{bmatrix} \end{bmatrix}$$

$((f, f), (h_{out}, w_{out}))$

Now we have the full Jacobian

(h_{in}, w_{in}) $\begin{bmatrix} 2 & -3 & 4 \\ 1 & -1 & 2 \\ -1 & 2 & 3 \end{bmatrix}$ x

(f, f) $\begin{bmatrix} 3 & -1 \\ 4 & 2 \end{bmatrix}$ w

Conv

v $\begin{bmatrix} 11 & -13 \\ 4 & 9 \end{bmatrix}$ (h_{out}, w_{out})

$\frac{\partial J}{\partial v}$ $\begin{bmatrix} 3 & 2 \\ 1 & -1 \end{bmatrix}$

$\frac{\partial v}{\partial w}$ is shape $((f, f), (h_{out}, w_{out}))$

Interpretation: How does each filter weight affect each feature map element

$$\frac{\partial v}{\partial w} = \begin{bmatrix} \begin{bmatrix} 2 & -3 \\ 1 & -1 \end{bmatrix} & \begin{bmatrix} -3 & 4 \\ -1 & 2 \end{bmatrix} \\ \begin{bmatrix} 1 & -1 \\ -1 & 2 \end{bmatrix} & \begin{bmatrix} -1 & 2 \\ 2 & 3 \end{bmatrix} \end{bmatrix}$$

$((f, f), (h_{out}, w_{out}))$

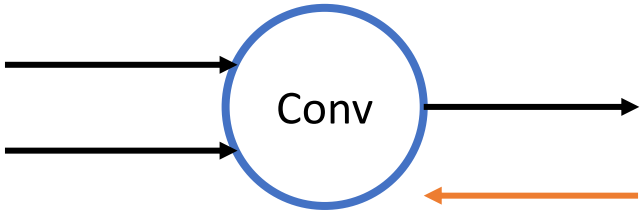
As expected, Jacobian can get big.
E.g. Second Conv layer of AlexNet
(5,5,256,27,27) ~4.7million numbers

Now we have the full Jacobian

Brad Quinton, Scott Chin

(h_{in}, w_{in}) $\begin{bmatrix} 2 & -3 & 4 \\ 1 & -1 & 2 \\ -1 & 2 & 3 \end{bmatrix}$ x

(f, f) $\begin{bmatrix} 3 & -1 \\ 4 & 2 \end{bmatrix}$ w



v $\begin{bmatrix} 11 & -13 \\ 4 & 9 \end{bmatrix}$ (h_{out}, w_{out})

$\frac{\partial J}{\partial v}$ $\begin{bmatrix} 3 & 2 \\ 1 & -1 \end{bmatrix}$

$\frac{\partial v}{\partial w}$ is shape $((f,f),(h_{out},w_{out}))$

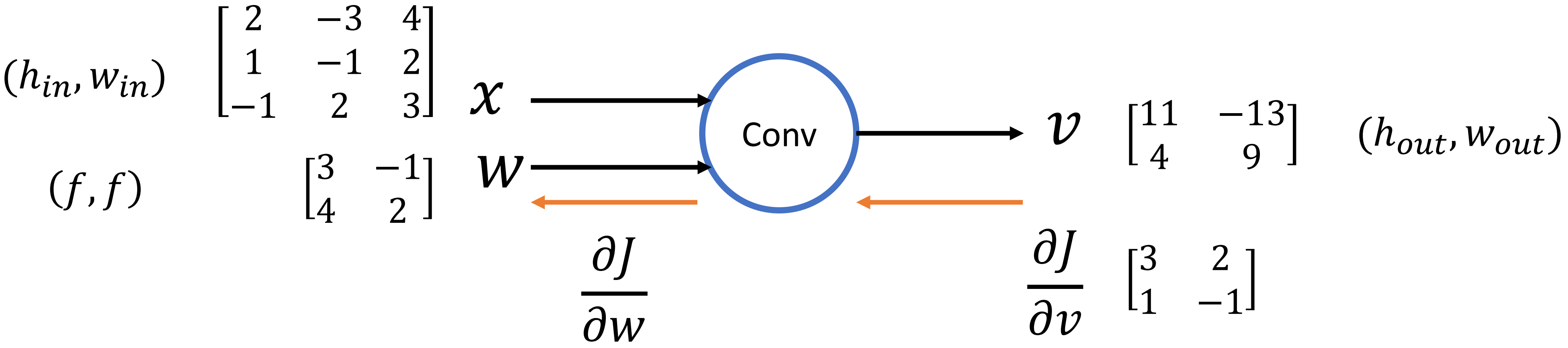
Interpretation: How does each filter weight affect each feature map element

$$\frac{\partial v}{\partial w} = \begin{bmatrix} \begin{bmatrix} 2 & -3 \\ 1 & -1 \end{bmatrix} & \begin{bmatrix} -3 & 4 \\ -1 & 2 \end{bmatrix} \\ \begin{bmatrix} 1 & -1 \\ -1 & 2 \end{bmatrix} & \begin{bmatrix} -1 & 2 \\ 2 & 3 \end{bmatrix} \end{bmatrix}$$

$((f,f),(h_{out},w_{out}))$

- For fully-connected layer, Jacobian had a lot of 0's. (Each neuron has own set of weights. Therefore they do not affect the output of other neurons)
- For conv layer every weight affects every output. Parameter sharing)

Now we have the full Jacobian



Apply chain rule to compute weight gradients

$$\frac{\partial J}{\partial w} = \frac{\partial v}{\partial w}\frac{\partial J}{\partial v} = \begin{bmatrix} \begin{bmatrix} 2 & -3 \\ 1 & -1 \end{bmatrix} & \begin{bmatrix} -3 & 4 \\ -1 & 2 \end{bmatrix} \\ \begin{bmatrix} 1 & -1 \\ -1 & 2 \end{bmatrix} & \begin{bmatrix} -1 & 2 \\ 2 & 3 \end{bmatrix} \end{bmatrix} \begin{bmatrix} 3 & 2 \\ 1 & -1 \end{bmatrix} = \begin{bmatrix} ? & ? \\ ? & ? \end{bmatrix}$$

$((f, f), (h_{out}, w_{out}))$ (h_{out}, w_{out}) (f, f)

Brad Quinton, Scott Chin

(h_{in}, w_{in}) $\begin{bmatrix} 2 & -3 & 4 \\ 1 & -1 & 2 \\ -1 & 2 & 3 \end{bmatrix}$ x

(f, f) $\begin{bmatrix} 3 & -1 \\ 4 & 2 \end{bmatrix}$ w

Conv

v $\begin{bmatrix} 11 & -13 \\ 4 & 9 \end{bmatrix}$ (h_{out}, w_{out})

$\frac{\partial J}{\partial w}$

$\frac{\partial J}{\partial v}$ $\begin{bmatrix} 3 & 2 \\ 1 & -1 \end{bmatrix}$

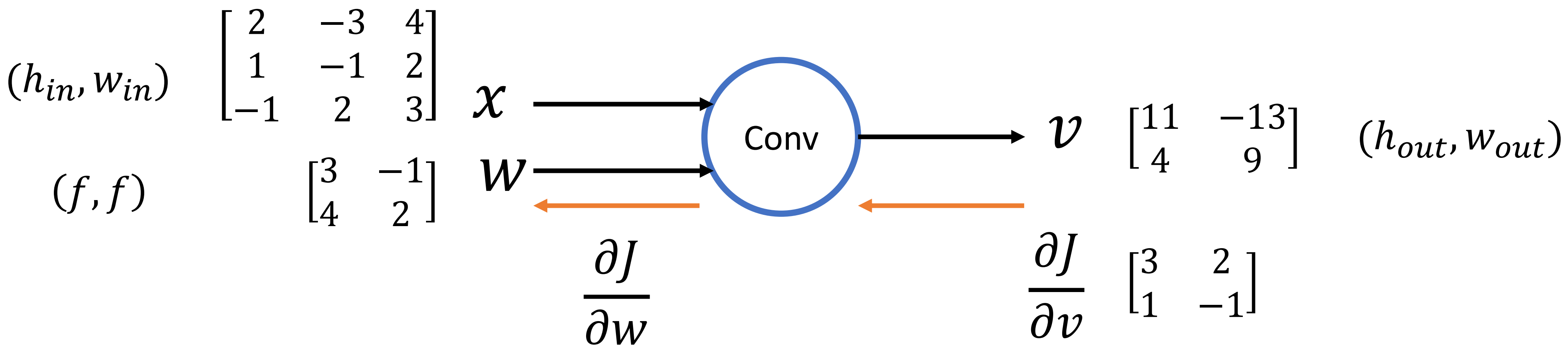
Apply chain rule to compute weight gradients

$$\frac{\partial J}{\partial w} = \frac{\partial v}{\partial w}\frac{\partial J}{\partial v} = \begin{bmatrix} \begin{bmatrix} 2 & -3 \\ 1 & -1 \end{bmatrix} & \begin{bmatrix} -3 & 4 \\ -1 & 2 \end{bmatrix} \\ \begin{bmatrix} 1 & -1 \\ -1 & 2 \end{bmatrix} & \begin{bmatrix} -1 & 2 \\ 2 & 3 \end{bmatrix} \end{bmatrix} \begin{bmatrix} 3 & -2 \\ 1 & -1 \end{bmatrix} = \begin{bmatrix} \boxed{12} & ? \\ ? & ? \end{bmatrix}$$

$((f, f), (h_{out}, w_{out}))$ (h_{out}, w_{out}) (f, f)

$$\frac{\partial J}{\partial w_{11}} = 2 * 3 + (-3) * (-2) + 1 * 1 + (-1)(-1) = 12$$

Recall: Computing an Inner product
(i.e. sum of elementwise multiply)



Apply chain rule to compute weight gradients

$$\frac{\partial J}{\partial w} = \frac{\partial v}{\partial w}\frac{\partial J}{\partial v} = \underbrace{\begin{bmatrix} \begin{bmatrix} 2 & -3 \\ 1 & -1 \end{bmatrix} & \begin{bmatrix} -3 & 4 \\ -1 & 2 \end{bmatrix} \\ \begin{bmatrix} 1 & -1 \\ -1 & 2 \end{bmatrix} & \begin{bmatrix} -1 & 2 \\ 2 & 3 \end{bmatrix} \end{bmatrix}}_{((f,f),(h_{out}, w_{out}))} \underbrace{\begin{bmatrix} 3 & -2 \\ 1 & -1 \end{bmatrix}}_{(h_{out}, w_{out})} = \underbrace{\begin{bmatrix} 12 & \boxed{-20} \\ ? & ? \end{bmatrix}}_{(f,f)}$$

$$\frac{\partial J}{\partial w_{12}} = -3 * 3 + 4 * (-2) + (-1) * 1 + (2)(-1) = -20$$

Recall: Computing an Inner product
(i.e. sum of elementwise multiply)

Brad Quinton, Scott Chin

(h_{in}, w_{in}) $\begin{bmatrix} 2 & -3 & 4 \\ 1 & -1 & 2 \\ -1 & 2 & 3 \end{bmatrix}$ x

(f, f) $\begin{bmatrix} 3 & -1 \\ 4 & 2 \end{bmatrix}$ w

Conv

v $\begin{bmatrix} 11 & -13 \\ 4 & 9 \end{bmatrix}$ (h_{out}, w_{out})

$\frac{\partial J}{\partial w}$

$\frac{\partial J}{\partial v}$ $\begin{bmatrix} 3 & 2 \\ 1 & -1 \end{bmatrix}$

Apply chain rule to compute weight gradients

$$\frac{\partial J}{\partial w} = \frac{\partial v}{\partial w}\frac{\partial J}{\partial v} = \begin{bmatrix} \begin{bmatrix} 2 & -3 \\ 1 & -1 \end{bmatrix} & \begin{bmatrix} -3 & 4 \\ -1 & 2 \end{bmatrix} \\ \begin{bmatrix} 1 & -1 \\ -1 & 2 \end{bmatrix} & \begin{bmatrix} -1 & 2 \\ 2 & 3 \end{bmatrix} \end{bmatrix} \begin{bmatrix} 3 & -2 \\ 1 & -1 \end{bmatrix} = \begin{bmatrix} 12 & -20 \\ \boxed{2} & ? \end{bmatrix}$$

$((f, f), (h_{out}, w_{out}))$ (h_{out}, w_{out}) (f, f)

$$\frac{\partial J}{\partial w_{21}} = 1 * 3 + (-1) * (-2) + (-1) * 1 + (2)(-1) = 2$$

Recall: Computing an Inner product
(i.e. sum of elementwise multiply)

(h_{in}, w_{in}) $\begin{bmatrix} 2 & -3 & 4 \\ 1 & -1 & 2 \\ -1 & 2 & 3 \end{bmatrix}$ $x \rightarrow$ Conv $\rightarrow v$ $\begin{bmatrix} 11 & -13 \\ 4 & 9 \end{bmatrix}$ (h_{out}, w_{out})

(f, f) $\begin{bmatrix} 3 & -1 \\ 4 & 2 \end{bmatrix}$ $w \rightarrow$ Conv

$\frac{\partial J}{\partial w} \leftarrow$ Conv $\leftarrow \frac{\partial J}{\partial v}$ $\begin{bmatrix} 3 & 2 \\ 1 & -1 \end{bmatrix}$

Apply chain rule to compute weight gradients

$$\frac{\partial J}{\partial w} = \frac{\partial v}{\partial w}\frac{\partial J}{\partial v} = \begin{bmatrix} \begin{bmatrix} 2 & -3 \\ 1 & -1 \end{bmatrix} & \begin{bmatrix} -3 & 4 \\ -1 & 2 \end{bmatrix} \\ \begin{bmatrix} 1 & -1 \\ -1 & 2 \end{bmatrix} & \begin{bmatrix} -1 & 2 \\ 2 & 3 \end{bmatrix} \end{bmatrix} \begin{bmatrix} 3 & -2 \\ 1 & -1 \end{bmatrix} = \begin{bmatrix} 12 & -20 \\ 2 & \boxed{-8} \end{bmatrix}$$

$((f, f), (h_{out}, w_{out}))$ (h_{out}, w_{out}) (f, f)

$$\frac{\partial J}{\partial w_{22}} = -1 * 3 + 2 * (-2) + 2 * 1 + 3 * (-1) = -8$$

Recall: Computing an Inner product
(i.e. sum of elementwise multiply)

Brad Quinton, Scott Chin

(h_{in}, w_{in}) $\begin{bmatrix} 2 & -3 & 4 \\ 1 & -1 & 2 \\ -1 & 2 & 3 \end{bmatrix}$ x

(f, f) $\begin{bmatrix} 3 & -1 \\ 4 & 2 \end{bmatrix}$ w

Conv

v $\begin{bmatrix} 11 & -13 \\ 4 & 9 \end{bmatrix}$ (h_{out}, w_{out})

$\frac{\partial J}{\partial w}$

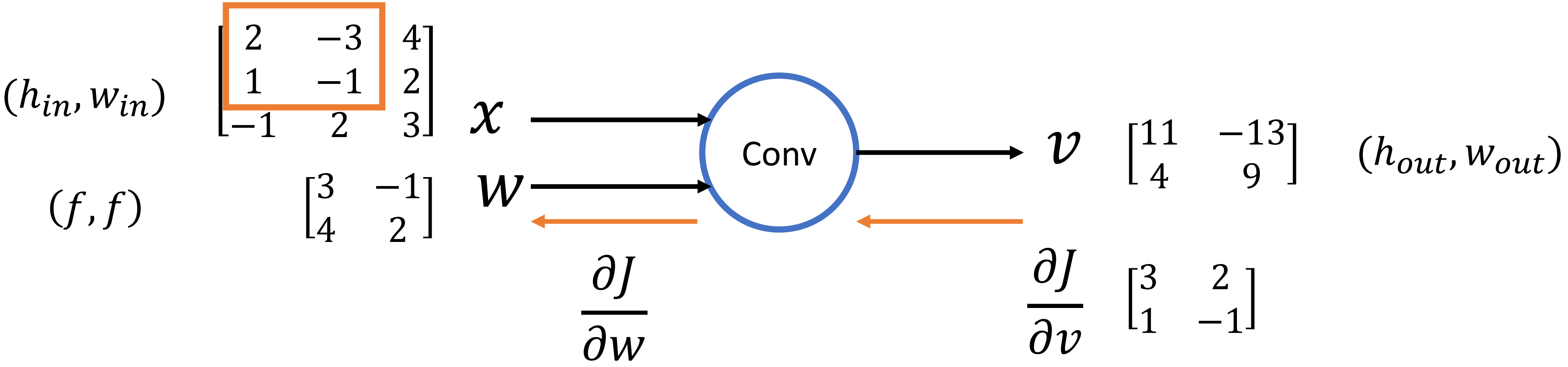
$\frac{\partial J}{\partial v}$ $\begin{bmatrix} 3 & 2 \\ 1 & -1 \end{bmatrix}$

Apply chain rule to compute weight gradients

$$\frac{\partial J}{\partial w} = \frac{\partial v}{\partial w}\frac{\partial J}{\partial v} = \underbrace{\begin{bmatrix} \begin{bmatrix} 2 & -3 \\ 1 & -1 \end{bmatrix} & \begin{bmatrix} -3 & 4 \\ -1 & 2 \end{bmatrix} \\ \begin{bmatrix} 1 & -1 \\ -1 & 2 \end{bmatrix} & \begin{bmatrix} -1 & 2 \\ 2 & 3 \end{bmatrix} \end{bmatrix}}_{((f,f),(h_{out}, w_{out}))} \underbrace{\begin{bmatrix} 3 & -2 \\ 1 & -1 \end{bmatrix}}_{(h_{out}, w_{out})} = \underbrace{\begin{bmatrix} 12 & -20 \\ 2 & -8 \end{bmatrix}}_{(f,f)}$$

Notice any patterns to this Jacobian?

Brad Quinton, Scott Chin



Apply chain rule to compute weight gradients

$$\frac{\partial J}{\partial w} = \frac{\partial v}{\partial w}\frac{\partial J}{\partial v} = \underbrace{\begin{bmatrix} \boxed{\begin{bmatrix} 2 & -3 \\ 1 & -1 \end{bmatrix}} & \begin{bmatrix} -3 & 4 \\ -1 & 2 \end{bmatrix} \\ \begin{bmatrix} 1 & -1 \\ -1 & 2 \end{bmatrix} & \begin{bmatrix} -1 & 2 \\ 2 & 3 \end{bmatrix} \end{bmatrix}}_{((f,f),(h_{out},w_{out}))} \underbrace{\begin{bmatrix} 3 & -2 \\ 1 & -1 \end{bmatrix}}_{(h_{out},w_{out})} = \underbrace{\begin{bmatrix} 12 & -20 \\ 2 & -8 \end{bmatrix}}_{(f,f)}$$

Notice any patterns to this Jacobian?

Brad Quinton, Scott Chin

(h_{in}, w_{in}) $\begin{bmatrix} 2 & -3 & 4 \\ 1 & -1 & 2 \\ -1 & 2 & 3 \end{bmatrix}$ $x \longrightarrow$ Conv $\longrightarrow v$ $\begin{bmatrix} 11 & -13 \\ 4 & 9 \end{bmatrix}$ (h_{out}, w_{out})

(f, f) $\begin{bmatrix} 3 & -1 \\ 4 & 2 \end{bmatrix}$ $w \longrightarrow$ Conv

$\dfrac{\partial J}{\partial w} \longleftarrow$ Conv $\longleftarrow \dfrac{\partial J}{\partial v}$ $\begin{bmatrix} 3 & 2 \\ 1 & -1 \end{bmatrix}$

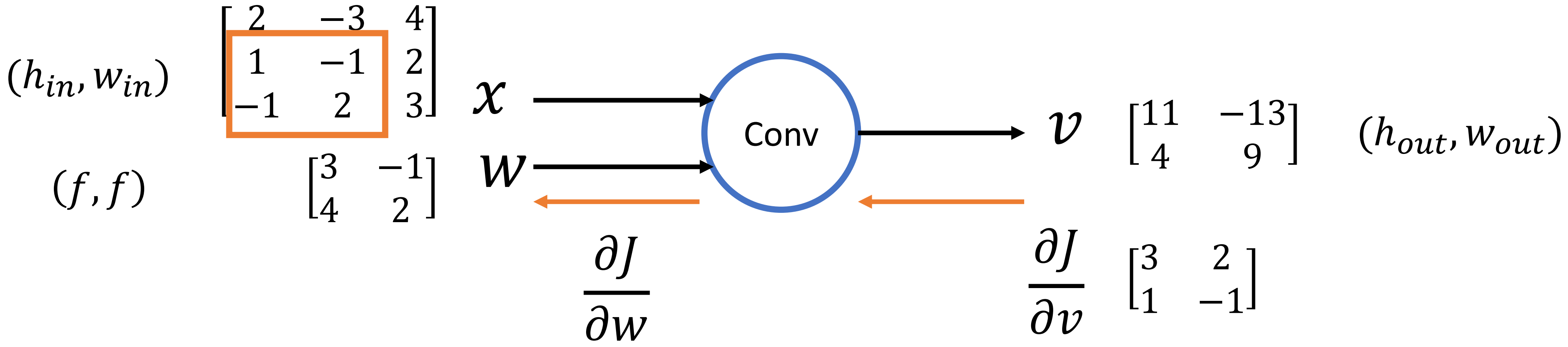
Apply chain rule to compute weight gradients

$$\frac{\partial J}{\partial w} = \frac{\partial v}{\partial w}\frac{\partial J}{\partial v} = \begin{bmatrix} \begin{bmatrix} 2 & -3 \\ 1 & -1 \end{bmatrix} & \boxed{\begin{bmatrix} -3 & 4 \\ -1 & 2 \end{bmatrix}} \\ \begin{bmatrix} 1 & -1 \\ -1 & 2 \end{bmatrix} & \begin{bmatrix} -1 & 2 \\ 2 & 3 \end{bmatrix} \end{bmatrix} \begin{bmatrix} 3 & -2 \\ 1 & -1 \end{bmatrix} = \begin{bmatrix} 12 & -20 \\ 2 & -8 \end{bmatrix}$$

$$\big((f, f), (h_{out}, w_{out})\big) \quad (h_{out}, w_{out}) \quad (f, f)$$

Notice any patterns to this Jacobian?

Brad Quinton, Scott Chin



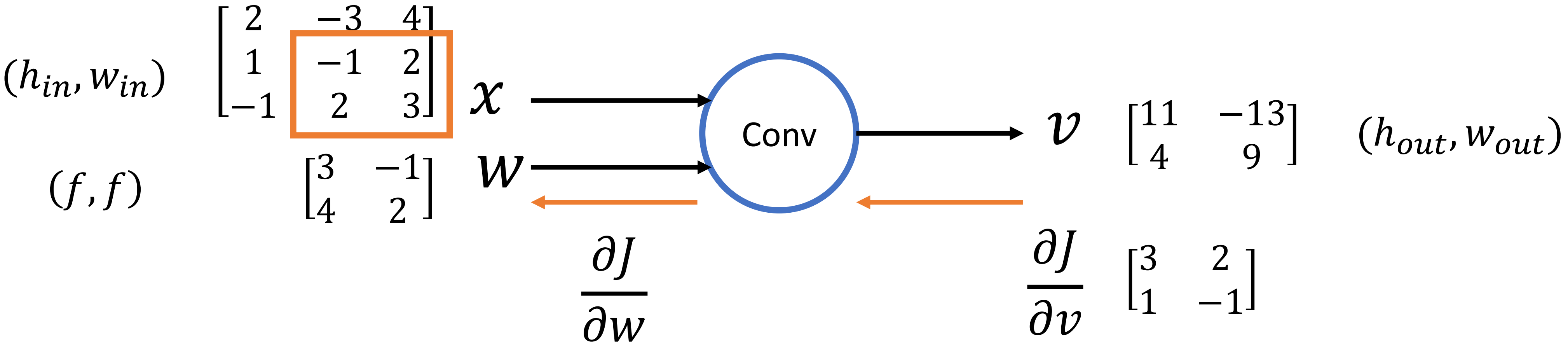
Apply chain rule to compute weight gradients

$$\frac{\partial J}{\partial w} = \frac{\partial v}{\partial w}\frac{\partial J}{\partial v} = \begin{bmatrix} \begin{bmatrix} 2 & -3 \\ 1 & -1 \end{bmatrix} & \begin{bmatrix} -3 & 4 \\ -1 & 2 \end{bmatrix} \\ \boxed{\begin{bmatrix} 1 & -1 \\ -1 & 2 \end{bmatrix}} & \begin{bmatrix} -1 & 2 \\ 2 & 3 \end{bmatrix} \end{bmatrix} \begin{bmatrix} 3 & -2 \\ 1 & -1 \end{bmatrix} = \begin{bmatrix} 12 & -20 \\ 2 & -8 \end{bmatrix}$$

$\big((f, f), (h_{out}, w_{out})\big)$ (h_{out}, w_{out}) (f, f)

Notice any patterns to this Jacobian?

Brad Quinton, Scott Chin



Apply chain rule to compute weight gradients

$$\frac{\partial J}{\partial w} = \frac{\partial v}{\partial w}\frac{\partial J}{\partial v} = \begin{bmatrix} \begin{bmatrix} 2 & -3 \\ 1 & -1 \end{bmatrix} & \begin{bmatrix} -3 & 4 \\ -1 & 2 \end{bmatrix} \\ \begin{bmatrix} 1 & -1 \\ -1 & 2 \end{bmatrix} & \boxed{\begin{bmatrix} -1 & 2 \\ 2 & 3 \end{bmatrix}} \end{bmatrix} \begin{bmatrix} 3 & -2 \\ 1 & -1 \end{bmatrix} = \begin{bmatrix} 12 & -20 \\ 2 & -8 \end{bmatrix}$$

$((f, f), (h_{out}, w_{out}))$ $\quad (h_{out}, w_{out})$ $\quad (f, f)$

Notice any patterns to this Jacobian?

Brad Quinton, Scott Chin

(h_{in}, w_{in}) $\begin{bmatrix}2 & -3 & 4\\ 1 & -1 & 2\\ -1 & 2 & 3\end{bmatrix}$ x

(f, f) $\begin{bmatrix}3 & -1\\ 4 & 2\end{bmatrix}$ w

Conv

v $\begin{bmatrix}11 & -13\\ 4 & 9\end{bmatrix}$ (h_{out}, w_{out})

$\frac{\partial J}{\partial w}$

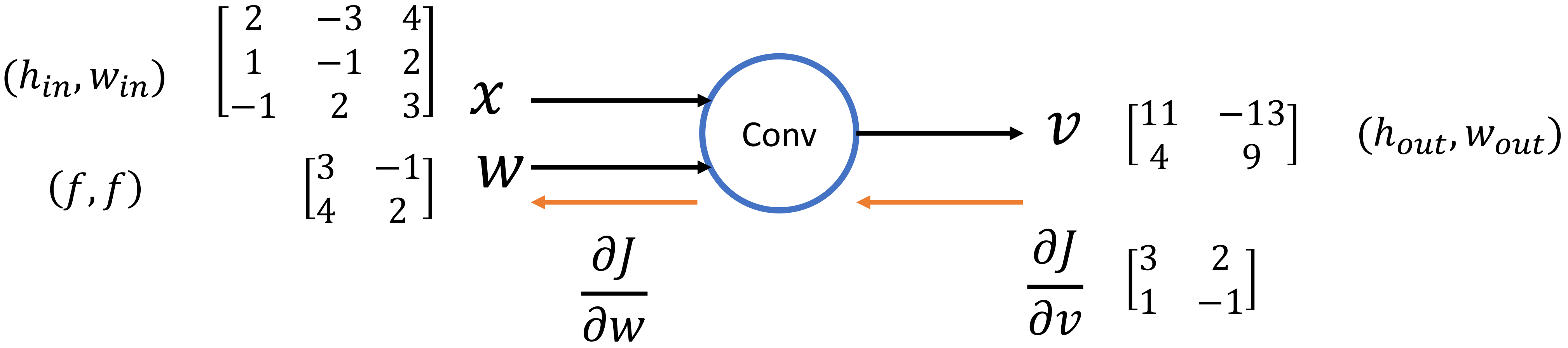
$\frac{\partial J}{\partial v}$ $\begin{bmatrix}3 & 2\\ 1 & -1\end{bmatrix}$

Apply chain rule to compute weight gradients

$$\frac{\partial J}{\partial w} = \frac{\partial v}{\partial w}\frac{\partial J}{\partial v} = \begin{bmatrix}\begin{bmatrix}2 & -3\\ 1 & -1\end{bmatrix} & \begin{bmatrix}-3 & 4\\ -1 & 2\end{bmatrix}\\ \begin{bmatrix}1 & -1\\ -1 & 2\end{bmatrix} & \boxed{\begin{bmatrix}-1 & 2\\ 2 & 3\end{bmatrix}}\end{bmatrix}\begin{bmatrix}3 & -2\\ 1 & -1\end{bmatrix} = \begin{bmatrix}12 & -20\\ 2 & -8\end{bmatrix}$$

$((f, f), (h_{out}, w_{out}))$ (h_{out}, w_{out}) (f, f)

Each Jacobian slice is a sliding window over the input x



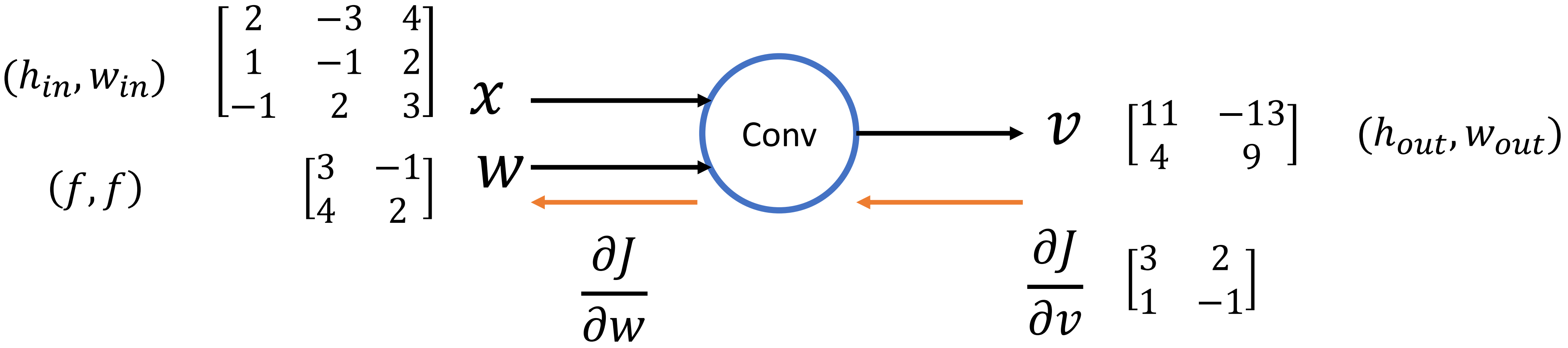
Apply chain rule to compute weight gradients

$$\frac{\partial J}{\partial w} = \frac{\partial v}{\partial w}\frac{\partial J}{\partial v} = \begin{bmatrix} \begin{bmatrix} 2 & -3 \\ 1 & -1 \end{bmatrix} & \begin{bmatrix} -3 & 4 \\ -1 & 2 \end{bmatrix} \\ \begin{bmatrix} 1 & -1 \\ -1 & 2 \end{bmatrix} & \begin{bmatrix} -1 & 2 \\ 2 & 3 \end{bmatrix} \end{bmatrix} \begin{bmatrix} 3 & -2 \\ 1 & -1 \end{bmatrix} = \begin{bmatrix} 12 & -20 \\ 2 & -8 \end{bmatrix}$$

$((f, f), (h_{out}, w_{out}))$ $\quad (h_{out}, w_{out})$ $\quad (f, f)$

We then perform elementwise multiply and sum of each jacobian slice with upstream gradient. What is this operation?

Brad Quinton, Scott Chin



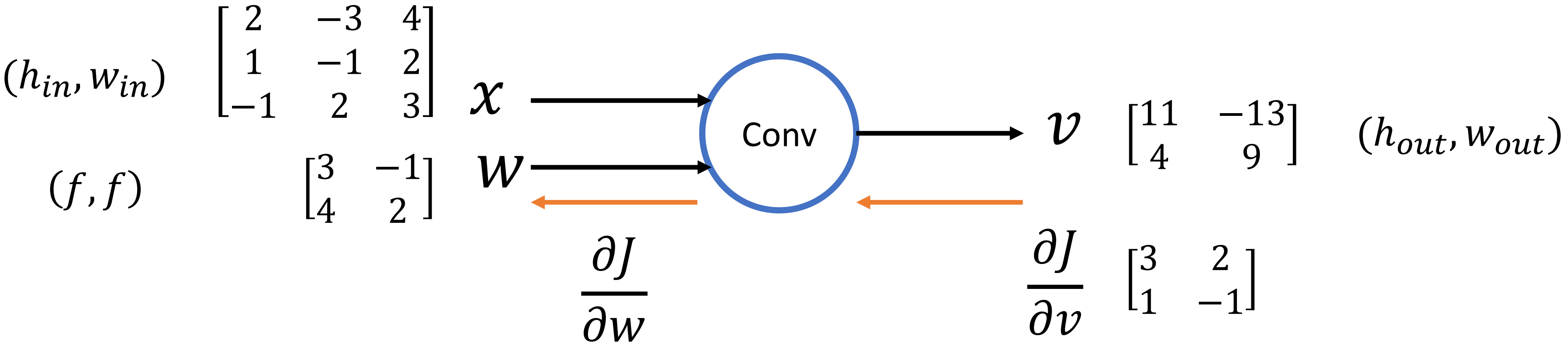
Apply chain rule to compute weight gradients

$$\frac{\partial J}{\partial w} = \frac{\partial v}{\partial w}\frac{\partial J}{\partial v} = \underbrace{\begin{bmatrix} \begin{bmatrix} 2 & -3 \\ 1 & -1 \end{bmatrix} & \begin{bmatrix} -3 & 4 \\ -1 & 2 \end{bmatrix} \\ \begin{bmatrix} 1 & -1 \\ -1 & 2 \end{bmatrix} & \begin{bmatrix} -1 & 2 \\ 2 & 3 \end{bmatrix} \end{bmatrix}}_{((f,f),(h_{out},w_{out}))} \underbrace{\begin{bmatrix} 3 & -2 \\ 1 & -1 \end{bmatrix}}_{(h_{out},w_{out})} = \underbrace{\begin{bmatrix} 12 & -20 \\ 2 & -8 \end{bmatrix}}_{(f,f)}$$

It’s a convolution operation!

$$\frac{\partial J}{\partial w} = x \text{ conv} \frac{\partial J}{\partial v}$$

Brad Quinton, Scott Chin



Instead of using Jacobian

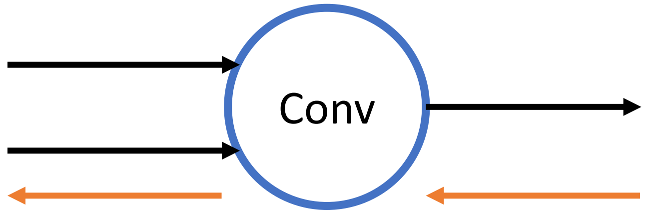
$$\frac{\partial J}{\partial w} = x \text{ conv} \frac{\partial J}{\partial v} = \begin{bmatrix} ? & ? \\ ? & ? \end{bmatrix}$$

$$\frac{\partial J}{\partial w} = \frac{\partial v}{\partial w}\frac{\partial J}{\partial v} = \begin{bmatrix} \begin{bmatrix} 2 & -3 \\ 1 & -1 \end{bmatrix} & \begin{bmatrix} -3 & 4 \\ -1 & 2 \end{bmatrix} \\ \begin{bmatrix} 1 & -1 \\ -1 & 2 \end{bmatrix} & \begin{bmatrix} -1 & 2 \\ 2 & 3 \end{bmatrix} \end{bmatrix} \begin{bmatrix} 3 & -2 \\ 1 & -1 \end{bmatrix} = \begin{bmatrix} 12 & -20 \\ 2 & -8 \end{bmatrix}$$

Brad Quinton, Scott Chin

(h_{in}, w_{in}) $\begin{bmatrix} \boxed{\begin{matrix} 2 & -3 \\ 1 & -1 \end{matrix}} & \begin{matrix} 4 \\ 2 \end{matrix} \\ \begin{matrix} -1 & 2 \end{matrix} & 3 \end{bmatrix}$ x

(f, f) $\begin{bmatrix} 3 & -1 \\ 4 & 2 \end{bmatrix}$ w



$\dfrac{\partial J}{\partial w}$

v $\begin{bmatrix} 11 & -13 \\ 4 & 9 \end{bmatrix}$ (h_{out}, w_{out})

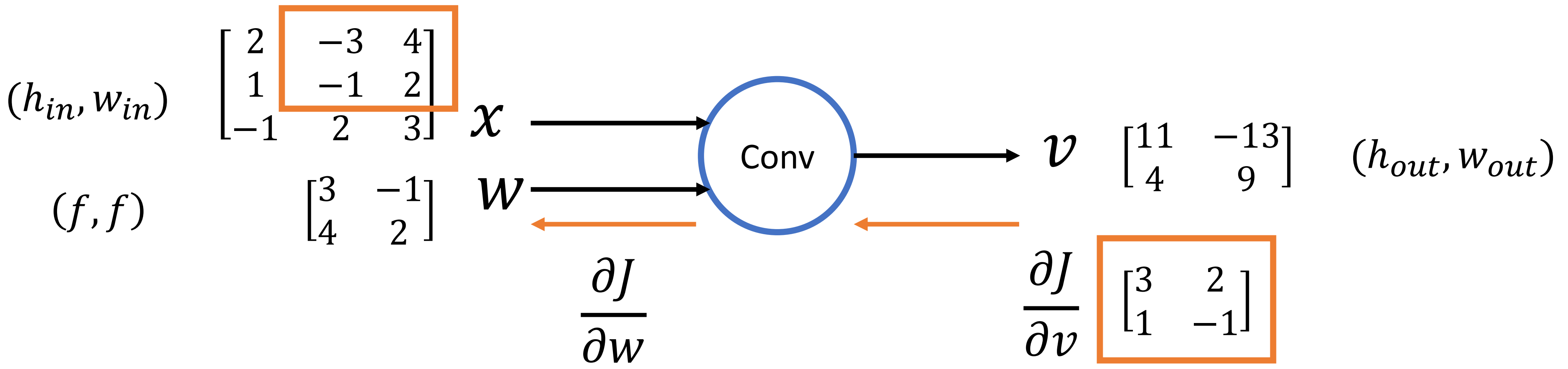
$\dfrac{\partial J}{\partial v}$ $\boxed{\begin{bmatrix} 3 & 2 \\ 1 & -1 \end{bmatrix}}$

Instead of using Jacobian

$$\frac{\partial J}{\partial w} = x \text{ conv} \frac{\partial J}{\partial v} = \begin{bmatrix} \boxed{12} & ? \\ ? & ? \end{bmatrix}$$

$$\frac{\partial J}{\partial w} = \frac{\partial v}{\partial w}\frac{\partial J}{\partial v} = \begin{bmatrix} \begin{bmatrix} 2 & -3 \\ 1 & -1 \end{bmatrix} & \begin{bmatrix} -3 & 4 \\ -1 & 2 \end{bmatrix} \\ \begin{bmatrix} 1 & -1 \\ -1 & 2 \end{bmatrix} & \begin{bmatrix} -1 & 2 \\ 2 & 3 \end{bmatrix} \end{bmatrix} \begin{bmatrix} 3 & -2 \\ 1 & -1 \end{bmatrix} = \begin{bmatrix} 12 & -20 \\ 2 & -8 \end{bmatrix}$$

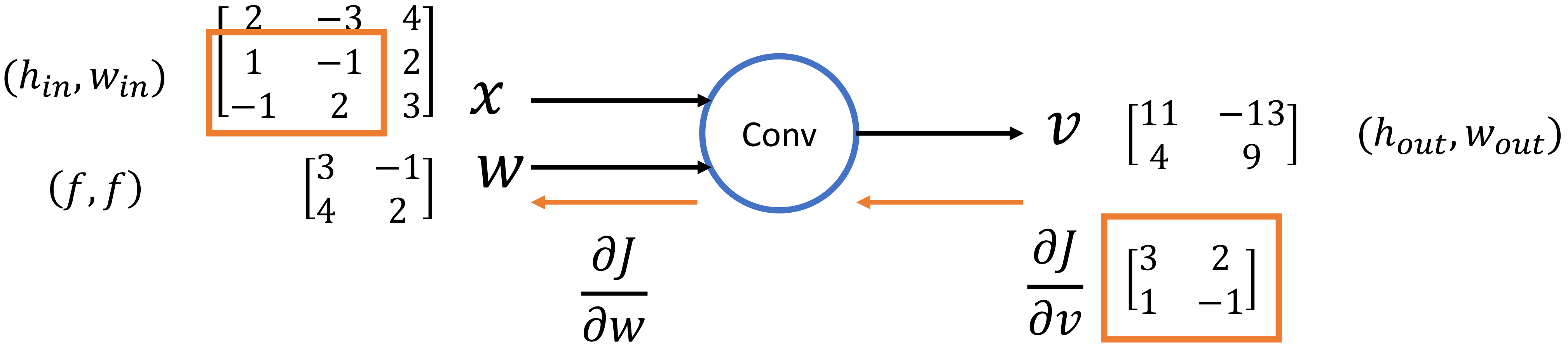
Brad Quinton, Scott Chin



Instead of using Jacobian

$$\frac{\partial J}{\partial w} = x \text{ conv} \frac{\partial J}{\partial v} = \begin{bmatrix} 12 & \boxed{-20} \\ ? & ? \end{bmatrix}$$

$$\frac{\partial J}{\partial w} = \frac{\partial v}{\partial w} \frac{\partial J}{\partial v} = \begin{bmatrix} \begin{bmatrix} 2 & -3 \\ 1 & -1 \end{bmatrix} & \begin{bmatrix} -3 & 4 \\ -1 & 2 \end{bmatrix} \\ \begin{bmatrix} 1 & -1 \\ -1 & 2 \end{bmatrix} & \begin{bmatrix} -1 & 2 \\ 2 & 3 \end{bmatrix} \end{bmatrix} \begin{bmatrix} 3 & -2 \\ 1 & -1 \end{bmatrix} = \begin{bmatrix} 12 & -20 \\ 2 & -8 \end{bmatrix}$$



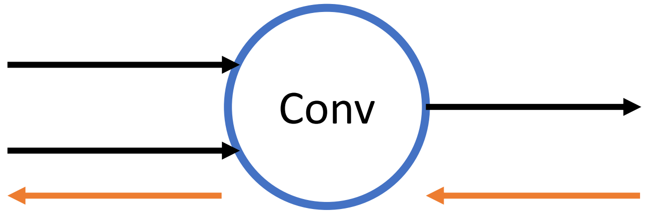
Instead of using Jacobian

$$\frac{\partial J}{\partial w} = x \text{ conv} \frac{\partial J}{\partial v} = \begin{bmatrix} 12 & -20 \\ \boxed{2} & ? \end{bmatrix}$$

$$\frac{\partial J}{\partial w} = \frac{\partial v}{\partial w}\frac{\partial J}{\partial v} = \begin{bmatrix} \begin{bmatrix} 2 & -3 \\ 1 & -1 \end{bmatrix} & \begin{bmatrix} -3 & 4 \\ -1 & 2 \end{bmatrix} \\ \begin{bmatrix} 1 & -1 \\ -1 & 2 \end{bmatrix} & \begin{bmatrix} -1 & 2 \\ 2 & 3 \end{bmatrix} \end{bmatrix} \begin{bmatrix} 3 & -2 \\ 1 & -1 \end{bmatrix} = \begin{bmatrix} 12 & -20 \\ 2 & -8 \end{bmatrix}$$

(h_{in}, w_{in}) $\begin{bmatrix} 2 & -3 & 4 \\ 1 & -1 & 2 \\ -1 & 2 & 3 \end{bmatrix}$ x

(f, f) $\begin{bmatrix} 3 & -1 \\ 4 & 2 \end{bmatrix}$ w



v $\begin{bmatrix} 11 & -13 \\ 4 & 9 \end{bmatrix}$ (h_{out}, w_{out})

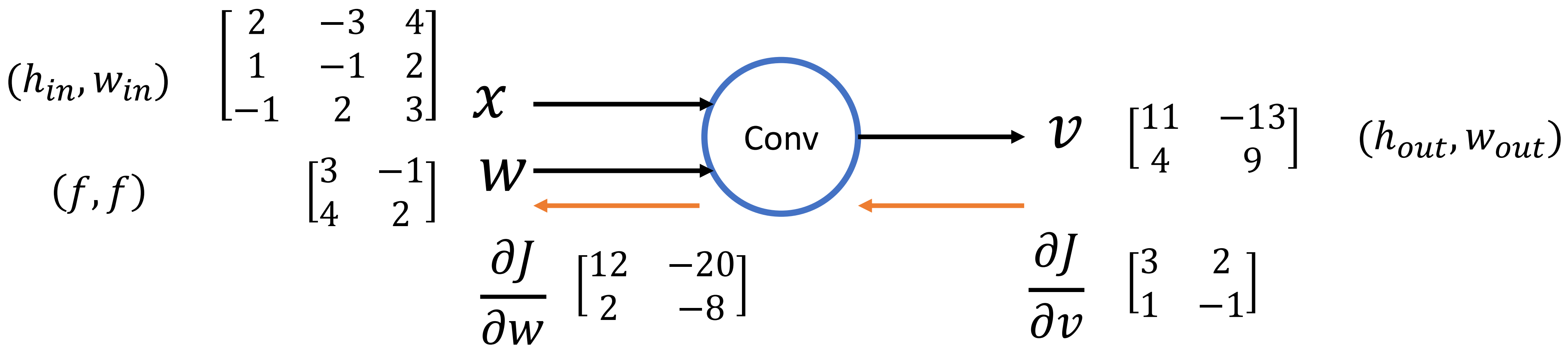
$\frac{\partial J}{\partial w}$

$\frac{\partial J}{\partial v}$ $\boxed{\begin{bmatrix} 3 & 2 \\ 1 & -1 \end{bmatrix}}$

Instead of using Jacobian

$$\frac{\partial J}{\partial w} = x \text{ conv} \frac{\partial J}{\partial v} = \begin{bmatrix} 12 & -20 \\ 2 & \boxed{-8} \end{bmatrix}$$

$$\frac{\partial J}{\partial w} = \frac{\partial v}{\partial w} \frac{\partial J}{\partial v} = \begin{bmatrix} \begin{bmatrix} 2 & -3 \\ 1 & -1 \end{bmatrix} & \begin{bmatrix} -3 & 4 \\ -1 & 2 \end{bmatrix} \\ \begin{bmatrix} 1 & -1 \\ -1 & 2 \end{bmatrix} & \begin{bmatrix} -1 & 2 \\ 2 & 3 \end{bmatrix} \end{bmatrix} \begin{bmatrix} 3 & -2 \\ 1 & -1 \end{bmatrix} = \begin{bmatrix} 12 & -20 \\ 2 & -8 \end{bmatrix}$$



$$\boxed{\frac{\partial J}{\partial w} = x \text{ conv} \frac{\partial J}{\partial v}}$$

- No Jacobian require at all!
- This convolution yields same result as doing the full 4D-Tensor Jacobian and upstream gradient matrix multiply!
- You will implement this in Assignment 4

Still only considering:

- One sample (m=1)
- One filter (K=1)
- One channel ($c_{in} = 1$)

Matrix of Shape (3,3)

x

$$\frac{\partial J}{\partial x} = \frac{\partial v}{\partial x}\frac{\partial J}{\partial v}$$

$$(3,3,2,2) * (2,2) = (3,3,\ )$$

$\frac{\partial v}{\partial x}$ is shape $(3,3,2,2)$

$\frac{\partial v}{\partial w}$ is shape $(2,2,2,2)$

Matrix of Shape (2,2)

v

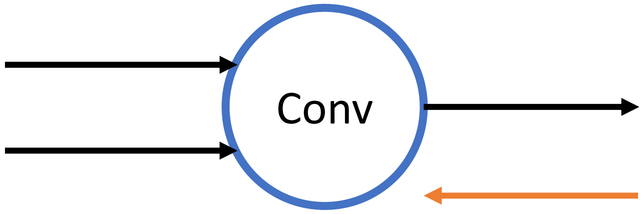
$\frac{\partial J}{\partial v}$ Shape $(2,2)$

Matrix of Shape (2,2)

w

(h_{in}, w_{in}) $\begin{bmatrix}2 & -3 & 4\\ 1 & -1 & 2\\ -1 & 2 & 3\end{bmatrix}$ x

(f, f) $\begin{bmatrix}3 & -1\\ 4 & 2\end{bmatrix}$ w



v $\begin{bmatrix}11 & -13\\ 4 & 9\end{bmatrix}$ (h_{out}, w_{out})

$\frac{\partial J}{\partial v}$ $\begin{bmatrix}3 & 2\\ 1 & -1\end{bmatrix}$

$$\frac{\partial J}{\partial x} = \boxed{\frac{\partial v}{\partial x}}\frac{\partial J}{\partial v}$$

Jacobian

(h_{in}, w_{in}) $\begin{bmatrix} 2 & -3 & 4 \\ 1 & -1 & 2 \\ -1 & 2 & 3 \end{bmatrix}$ x

(f, f) $\begin{bmatrix} 3 & -1 \\ 4 & 2 \end{bmatrix}$ w

$$x, w \longrightarrow \text{Conv} \longrightarrow v$$

v $\begin{bmatrix} 11 & -13 \\ 4 & 9 \end{bmatrix}$ (h_{out}, w_{out})

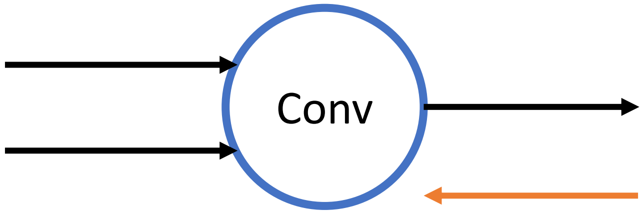
$\frac{\partial J}{\partial v}$ $\begin{bmatrix} 3 & 2 \\ 1 & -1 \end{bmatrix}$

$\frac{\partial v}{\partial x}$ is shape $((h_{in}, w_{in}) \ , (h_{out}, w_{out}))$

Interpretation: How does each input value affect each feature map element

(h_{in}, w_{in}) $\begin{bmatrix} 2 & -3 & 4 \\ 1 & -1 & 2 \\ -1 & 2 & 3 \end{bmatrix}$ x

(f, f) $\begin{bmatrix} 3 & -1 \\ 4 & 2 \end{bmatrix}$ w



v $\begin{bmatrix} 11 & -13 \\ 4 & 9 \end{bmatrix}$ (h_{out}, w_{out})

$\frac{\partial J}{\partial v}$ $\begin{bmatrix} 3 & 2 \\ 1 & -1 \end{bmatrix}$

$\frac{\partial v}{\partial x}$ is shape $((h_{in}, w_{in}) \; , (h_{out}, w_{out}))$

Interpretation: How does each input value affect each feature map element

$$\frac{\partial v}{\partial x} = \begin{bmatrix} \frac{\partial v}{\partial x_{11}} & \frac{\partial v}{\partial x_{12}} & \frac{\partial v}{\partial x_{13}} \\ \frac{\partial v}{\partial x_{21}} & \frac{\partial v}{\partial x_{22}} & \frac{\partial v}{\partial x_{23}} \\ \frac{\partial v}{\partial x_{31}} & \frac{\partial v}{\partial x_{32}} & \frac{\partial v}{\partial x_{33}} \end{bmatrix}$$

(h_{in}, w_{in})

Brad Quinton, Scott Chin

(h_{in}, w_{in}) $\begin{bmatrix} 2 & -3 & 4 \\ 1 & -1 & 2 \\ -1 & 2 & 3 \end{bmatrix}$ x

(f, f) $\begin{bmatrix} 3 & -1 \\ 4 & 2 \end{bmatrix}$ w

Conv

v $\begin{bmatrix} 11 & -13 \\ 4 & 9 \end{bmatrix}$ (h_{out}, w_{out})

$\frac{\partial J}{\partial v}$ $\begin{bmatrix} 3 & 2 \\ 1 & -1 \end{bmatrix}$

$\frac{\partial v}{\partial x}$ is shape $((h_{in}, w_{in}) \ , (h_{out}, w_{out}))$

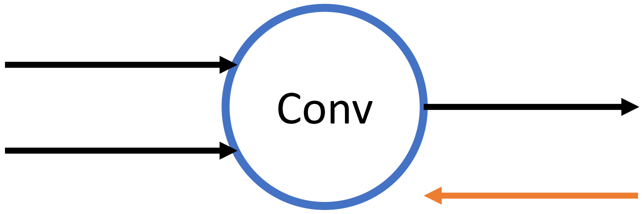
$$\frac{\partial v}{\partial x} = \begin{bmatrix} \boxed{\frac{\partial v}{\partial x_{11}}} & \frac{\partial v}{\partial x_{12}} & \frac{\partial v}{\partial x_{13}} \\ \frac{\partial v}{\partial x_{21}} & \frac{\partial v}{\partial x_{22}} & \frac{\partial v}{\partial x_{23}} \\ \frac{\partial v}{\partial x_{31}} & \frac{\partial v}{\partial x_{32}} & \frac{\partial v}{\partial x_{33}} \end{bmatrix} \qquad \frac{\partial v}{\partial x_{11}} = \begin{bmatrix} \frac{\partial v_{11}}{\partial x_{11}} & \frac{\partial v_{12}}{\partial x_{11}} \\ \frac{\partial v_{21}}{\partial x_{11}} & \frac{\partial v_{22}}{\partial x_{11}} \end{bmatrix}$$

(h_{in}, w_{in}) (h_{out}, w_{out})

Interpretation: How does this one in put value affect each feature map element

(h_{in}, w_{in}) $\begin{bmatrix} 2 & -3 & 4 \\ 1 & -1 & 2 \\ -1 & 2 & 3 \end{bmatrix}$ x

(f, f) $\begin{bmatrix} 3 & -1 \\ 4 & 2 \end{bmatrix}$ w



v $\begin{bmatrix} 11 & -13 \\ 4 & 9 \end{bmatrix}$ (h_{out}, w_{out})

$\frac{\partial J}{\partial v}$ $\begin{bmatrix} 3 & 2 \\ 1 & -1 \end{bmatrix}$

$\frac{\partial v}{\partial x}$ is shape $((h_{in}, w_{in}), (h_{out}, w_{out}))$

$$\frac{\partial v}{\partial x} = \begin{bmatrix} \boxed{\frac{\partial v}{\partial x_{11}}} & \frac{\partial v}{\partial x_{12}} & \frac{\partial v}{\partial x_{13}} \\ \frac{\partial v}{\partial x_{21}} & \frac{\partial v}{\partial x_{22}} & \frac{\partial v}{\partial x_{23}} \\ \frac{\partial v}{\partial x_{31}} & \frac{\partial v}{\partial x_{32}} & \frac{\partial v}{\partial x_{33}} \end{bmatrix} \qquad \frac{\partial v}{\partial x_{11}} = \begin{bmatrix} \frac{\partial v_{11}}{\partial x_{11}} & \frac{\partial v_{12}}{\partial x_{11}} \\ \frac{\partial v_{21}}{\partial x_{11}} & \frac{\partial v_{22}}{\partial x_{11}} \end{bmatrix}$$

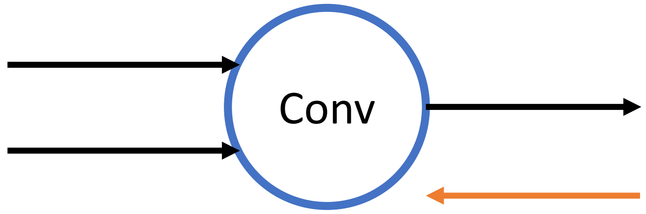
(h_{in}, w_{in}) (h_{out}, w_{out})

$$v_{11} = w_{11}x_{11} + w_{12}x_{12} + w_{21}x_{21} + w_{22}x_{22}$$
$$v_{12} = w_{11}x_{12} + w_{12}x_{13} + w_{21}x_{22} + w_{22}x_{23}$$
$$v_{21} = w_{11}x_{21} + w_{12}x_{22} + w_{21}x_{31} + w_{22}x_{31}$$
$$v_{22} = w_{11}x_{22} + w_{12}x_{23} + w_{21}x_{32} + w_{22}x_{33}$$

Brad Quinton, Scott Chin

(h_{in}, w_{in}) $\begin{bmatrix} 2 & -3 & 4 \\ 1 & -1 & 2 \\ -1 & 2 & 3 \end{bmatrix}$ x

(f, f) $\begin{bmatrix} 3 & -1 \\ 4 & 2 \end{bmatrix}$ w



v $\begin{bmatrix} 11 & -13 \\ 4 & 9 \end{bmatrix}$ (h_{out}, w_{out})

$\frac{\partial J}{\partial v}$ $\begin{bmatrix} 3 & 2 \\ 1 & -1 \end{bmatrix}$

$\frac{\partial v}{\partial x}$ is shape $((h_{in}, w_{in}), (h_{out}, w_{out}))$

$$\frac{\partial v}{\partial x} = \begin{bmatrix} \boxed{\frac{\partial v}{\partial x_{11}}} & \frac{\partial v}{\partial x_{12}} & \frac{\partial v}{\partial x_{13}} \\ \frac{\partial v}{\partial x_{21}} & \frac{\partial v}{\partial x_{22}} & \frac{\partial v}{\partial x_{23}} \\ \frac{\partial v}{\partial x_{31}} & \frac{\partial v}{\partial x_{32}} & \frac{\partial v}{\partial x_{33}} \end{bmatrix}$$

(h_{in}, w_{in})

$$\frac{\partial v}{\partial x_{11}} = \begin{bmatrix} w_{11} & 0 \\ 0 & 0 \end{bmatrix}$$

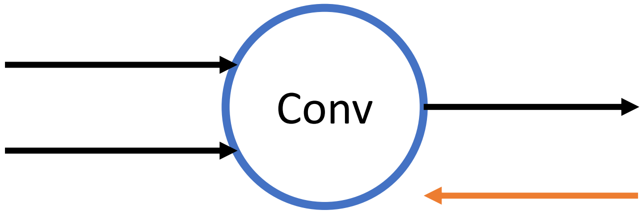
(h_{out}, w_{out})

$$v_{11} = \boxed{w_{11}x_{11}} + w_{12}x_{12} + w_{21}x_{21} + w_{22}x_{22}$$
$$v_{12} = w_{11}x_{12} + w_{12}x_{13} + w_{21}x_{22} + w_{22}x_{23}$$
$$v_{21} = w_{11}x_{21} + w_{12}x_{22} + w_{21}x_{31} + w_{22}x_{31}$$
$$v_{22} = w_{11}x_{22} + w_{12}x_{23} + w_{21}x_{32} + w_{22}x_{33}$$

Brad Quinton, Scott Chin

(h_{in}, w_{in}) $\begin{bmatrix} 2 & -3 & 4 \\ 1 & -1 & 2 \\ -1 & 2 & 3 \end{bmatrix}$ x

(f, f) $\begin{bmatrix} 3 & -1 \\ 4 & 2 \end{bmatrix}$ w



v $\begin{bmatrix} 11 & -13 \\ 4 & 9 \end{bmatrix}$ (h_{out}, w_{out})

$\dfrac{\partial J}{\partial v}$ $\begin{bmatrix} 3 & 2 \\ 1 & -1 \end{bmatrix}$

$\dfrac{\partial v}{\partial x}$ is shape $((h_{in}, w_{in})\ , (h_{out}, w_{out}))$

$$\frac{\partial v}{\partial x} = \begin{bmatrix} \frac{\partial v}{\partial x_{11}} & \boxed{\frac{\partial v}{\partial x_{12}}} & \frac{\partial v}{\partial x_{13}} \\ \frac{\partial v}{\partial x_{21}} & \frac{\partial v}{\partial x_{22}} & \frac{\partial v}{\partial x_{23}} \\ \frac{\partial v}{\partial x_{31}} & \frac{\partial v}{\partial x_{32}} & \frac{\partial v}{\partial x_{33}} \end{bmatrix}$$

(h_{in}, w_{in})

$$\frac{\partial v}{\partial x_{12}} = \begin{bmatrix} w_{12} & w_{11} \\ 0 & 0 \end{bmatrix}$$

(h_{out}, w_{out})

$$v_{11} = w_{11}x_{11} + \boxed{w_{12}x_{12}} + w_{21}x_{21} + w_{22}x_{22}$$

$$v_{12} = \boxed{w_{11}x_{12}} + w_{12}x_{13} + w_{21}x_{22} + w_{22}x_{23}$$

$$v_{21} = w_{11}x_{21} + w_{12}x_{22} + w_{21}x_{31} + w_{22}x_{31}$$

$$v_{22} = w_{11}x_{22} + w_{12}x_{23} + w_{21}x_{32} + w_{22}x_{33}$$

Brad Quinton, Scott Chin

(h_{in}, w_{in}) $\begin{bmatrix} 2 & -3 & 4 \\ 1 & -1 & 2 \\ -1 & 2 & 3 \end{bmatrix}$ x

(f, f) $\begin{bmatrix} 3 & -1 \\ 4 & 2 \end{bmatrix}$ w

Conv

v $\begin{bmatrix} 11 & -13 \\ 4 & 9 \end{bmatrix}$ (h_{out}, w_{out})

$\frac{\partial J}{\partial v}$ $\begin{bmatrix} 3 & 2 \\ 1 & -1 \end{bmatrix}$

$\frac{\partial v}{\partial x}$ is shape $((h_{in}, w_{in}), (h_{out}, w_{out}))$

$$\frac{\partial v}{\partial x} = \begin{bmatrix} \frac{\partial v}{\partial x_{11}} & \frac{\partial v}{\partial x_{12}} & \boxed{\frac{\partial v}{\partial x_{13}}} \\ \frac{\partial v}{\partial x_{21}} & \frac{\partial v}{\partial x_{22}} & \frac{\partial v}{\partial x_{23}} \\ \frac{\partial v}{\partial x_{31}} & \frac{\partial v}{\partial x_{32}} & \frac{\partial v}{\partial x_{33}} \end{bmatrix}$$

(h_{in}, w_{in})

$$\frac{\partial v}{\partial x_{13}} = \begin{bmatrix} 0 & w_{12} \\ 0 & 0 \end{bmatrix}$$

(h_{out}, w_{out})

$$v_{11} = w_{11}x_{11} + w_{12}x_{12} + w_{21}x_{21} + w_{22}x_{22}$$
$$v_{12} = w_{11}x_{12} + \boxed{w_{12}x_{13}} + w_{21}x_{22} + w_{22}x_{23}$$
$$v_{21} = w_{11}x_{21} + w_{12}x_{22} + w_{21}x_{31} + w_{22}x_{31}$$
$$v_{22} = w_{11}x_{22} + w_{12}x_{23} + w_{21}x_{32} + w_{22}x_{33}$$

Brad Quinton, Scott Chin

(h_{in}, w_{in}) $\begin{bmatrix} 2 & -3 & 4 \\ 1 & -1 & 2 \\ -1 & 2 & 3 \end{bmatrix}$ x

(f, f) $\begin{bmatrix} 3 & -1 \\ 4 & 2 \end{bmatrix}$ w

Conv

v $\begin{bmatrix} 11 & -13 \\ 4 & 9 \end{bmatrix}$ (h_{out}, w_{out})

$\frac{\partial J}{\partial v}$ $\begin{bmatrix} 3 & 2 \\ 1 & -1 \end{bmatrix}$

$\frac{\partial v}{\partial x}$ is shape $((h_{in}, w_{in}) \ , (h_{out}, w_{out}))$

$$\frac{\partial v}{\partial x} = \begin{bmatrix} \frac{\partial v}{\partial x_{11}} & \frac{\partial v}{\partial x_{12}} & \frac{\partial v}{\partial x_{13}} \\ \boxed{\frac{\partial v}{\partial x_{21}}} & \frac{\partial v}{\partial x_{22}} & \frac{\partial v}{\partial x_{23}} \\ \frac{\partial v}{\partial x_{31}} & \frac{\partial v}{\partial x_{32}} & \frac{\partial v}{\partial x_{33}} \end{bmatrix}$$

(h_{in}, w_{in})

$$\frac{\partial v}{\partial x_{21}} = \begin{bmatrix} w_{21} & 0 \\ w_{11} & 0 \end{bmatrix}$$

(h_{out}, w_{out})

$$v_{11} = w_{11}x_{11} + w_{12}x_{12} + \boxed{w_{21}x_{21}} + w_{22}x_{22}$$

$$v_{12} = w_{11}x_{12} + w_{12}x_{13} + w_{21}x_{22} + w_{22}x_{23}$$

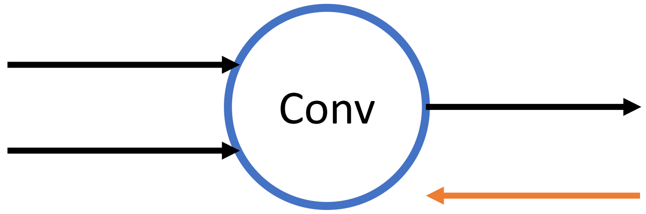
$$v_{21} = \boxed{w_{11}x_{21}} + w_{12}x_{22} + w_{21}x_{31} + w_{22}x_{31}$$

$$v_{22} = w_{11}x_{22} + w_{12}x_{23} + w_{21}x_{32} + w_{22}x_{33}$$

Brad Quinton, Scott Chin

(h_{in}, w_{in}) $\begin{bmatrix} 2 & -3 & 4 \\ 1 & -1 & 2 \\ -1 & 2 & 3 \end{bmatrix}$ x

(f, f) $\begin{bmatrix} 3 & -1 \\ 4 & 2 \end{bmatrix}$ w



v $\begin{bmatrix} 11 & -13 \\ 4 & 9 \end{bmatrix}$ (h_{out}, w_{out})

$\frac{\partial J}{\partial v}$ $\begin{bmatrix} 3 & 2 \\ 1 & -1 \end{bmatrix}$

$\frac{\partial v}{\partial x}$ is shape $((h_{in}, w_{in}) \ , (h_{out}, w_{out}))$

$$\frac{\partial v}{\partial x} = \begin{bmatrix} \frac{\partial v}{\partial x_{11}} & \frac{\partial v}{\partial x_{12}} & \frac{\partial v}{\partial x_{13}} \\ \frac{\partial v}{\partial x_{21}} & \boxed{\frac{\partial v}{\partial x_{22}}} & \frac{\partial v}{\partial x_{23}} \\ \frac{\partial v}{\partial x_{31}} & \frac{\partial v}{\partial x_{32}} & \frac{\partial v}{\partial x_{33}} \end{bmatrix}$$

$$\frac{\partial v}{\partial x_{22}} = \begin{bmatrix} w_{22} & w_{21} \\ w_{12} & w_{11} \end{bmatrix}$$

(h_{in}, w_{in}) (h_{out}, w_{out})

$$v_{11} = w_{11}x_{11} + w_{12}x_{12} + w_{21}x_{21} + \boxed{w_{22}x_{22}}$$
$$v_{12} = w_{11}x_{12} + w_{12}x_{13} + \boxed{w_{21}x_{22}} + w_{22}x_{23}$$
$$v_{21} = w_{11}x_{21} + \boxed{w_{12}x_{22}} + w_{21}x_{31} + w_{22}x_{31}$$
$$v_{22} = \boxed{w_{11}x_{22}} + w_{12}x_{23} + w_{21}x_{32} + w_{22}x_{33}$$

Brad Quinton, Scott Chin

(h_{in}, w_{in}) $\begin{bmatrix} 2 & -3 & 4 \\ 1 & -1 & 2 \\ -1 & 2 & 3 \end{bmatrix}$ x

(f, f) $\begin{bmatrix} 3 & -1 \\ 4 & 2 \end{bmatrix}$ w

Conv

v $\begin{bmatrix} 11 & -13 \\ 4 & 9 \end{bmatrix}$ (h_{out}, w_{out})

$\frac{\partial J}{\partial v}$ $\begin{bmatrix} 3 & 2 \\ 1 & -1 \end{bmatrix}$

$\frac{\partial v}{\partial x}$ is shape $((h_{in}, w_{in})\ , (h_{out}, w_{out}))$

$$\frac{\partial v}{\partial x} = \begin{bmatrix} \frac{\partial v}{\partial x_{11}} & \frac{\partial v}{\partial x_{12}} & \frac{\partial v}{\partial x_{13}} \\ \frac{\partial v}{\partial x_{21}} & \frac{\partial v}{\partial x_{22}} & \frac{\partial v}{\partial x_{23}} \\ \frac{\partial v}{\partial x_{31}} & \frac{\partial v}{\partial x_{32}} & \frac{\partial v}{\partial x_{33}} \end{bmatrix} = \begin{bmatrix} \begin{bmatrix} w_{11} & 0 \\ 0 & 0 \end{bmatrix} & \begin{bmatrix} w_{12} & w_{11} \\ 0 & 0 \end{bmatrix} & \begin{bmatrix} 0 & w_{12} \\ 0 & 0 \end{bmatrix} \\ \begin{bmatrix} w_{21} & 0 \\ w_{11} & 0 \end{bmatrix} & \begin{bmatrix} w_{22} & w_{21} \\ w_{12} & w_{11} \end{bmatrix} & \begin{bmatrix} 0 & w_{22} \\ 0 & w_{12} \end{bmatrix} \\ \begin{bmatrix} 0 & 0 \\ w_{21} & 0 \end{bmatrix} & \begin{bmatrix} 0 & 0 \\ w_{22} & w_{21} \end{bmatrix} & \begin{bmatrix} 0 & 0 \\ 0 & w_{22} \end{bmatrix} \end{bmatrix}$$

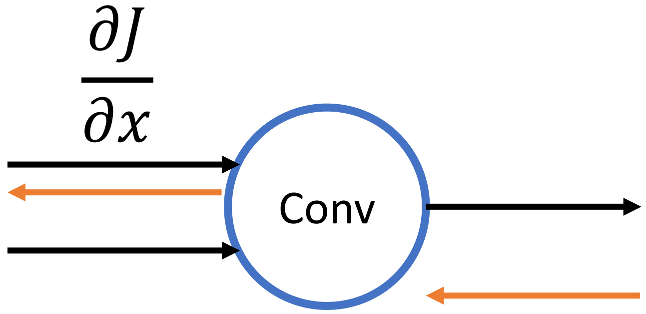
$((h_{in}, w_{in})\ , (h_{out}, w_{out}))$

This is the full Jacobian

Brad Quinton, Scott Chin

(h_{in}, w_{in}) $\begin{bmatrix} 2 & -3 & 4 \\ 1 & -1 & 2 \\ -1 & 2 & 3 \end{bmatrix}$ x

(f, f) $\begin{bmatrix} 3 & -1 \\ 4 & 2 \end{bmatrix}$ w



v $\begin{bmatrix} 11 & -13 \\ 4 & 9 \end{bmatrix}$ (h_{out}, w_{out})

$\frac{\partial J}{\partial v}$ $\begin{bmatrix} 3 & 2 \\ 1 & -1 \end{bmatrix}$

Apply chain rule to compute input gradients

$$\frac{\partial J}{\partial x} = \frac{\partial v}{\partial x}\frac{\partial J}{\partial v} = \begin{bmatrix} \begin{bmatrix} w_{11} & 0 \\ 0 & 0 \end{bmatrix} & \begin{bmatrix} w_{12} & w_{11} \\ 0 & 0 \end{bmatrix} & \begin{bmatrix} 0 & w_{12} \\ 0 & 0 \end{bmatrix} \\ \begin{bmatrix} w_{21} & 0 \\ w_{11} & 0 \end{bmatrix} & \begin{bmatrix} w_{22} & w_{21} \\ w_{12} & w_{11} \end{bmatrix} & \begin{bmatrix} 0 & w_{22} \\ 0 & w_{12} \end{bmatrix} \\ \begin{bmatrix} 0 & 0 \\ w_{21} & 0 \end{bmatrix} & \begin{bmatrix} 0 & 0 \\ w_{22} & w_{21} \end{bmatrix} & \begin{bmatrix} 0 & 0 \\ 0 & w_{22} \end{bmatrix} \end{bmatrix} \begin{bmatrix} \frac{\partial J}{\partial v_{11}} & \frac{\partial J}{\partial v_{12}} \\ \frac{\partial J}{\partial v_{21}} & \frac{\partial J}{\partial v_{22}} \end{bmatrix} = \begin{bmatrix} ? & ? & ? \\ ? & ? & ? \\ ? & ? & ? \end{bmatrix}$$

$(h_{in}, w_{in}, h_{out}, w_{out})$ $\quad$ (h_{out}, w_{out}) $\quad$ (h_{in}, w_{in})

Brad Quinton, Scott Chin

Apply chain rule to compute input gradients

$$\frac{\partial J}{\partial x} = \frac{\partial v}{\partial x}\frac{\partial J}{\partial v} = \begin{bmatrix} \begin{bmatrix} w_{11} & 0 \\ 0 & 0 \end{bmatrix} & \begin{bmatrix} w_{12} & w_{11} \\ 0 & 0 \end{bmatrix} & \begin{bmatrix} 0 & w_{12} \\ 0 & 0 \end{bmatrix} \\ \begin{bmatrix} w_{21} & 0 \\ w_{11} & 0 \end{bmatrix} & \begin{bmatrix} w_{22} & w_{21} \\ w_{12} & w_{11} \end{bmatrix} & \begin{bmatrix} 0 & w_{22} \\ 0 & w_{12} \end{bmatrix} \\ \begin{bmatrix} 0 & 0 \\ w_{21} & 0 \end{bmatrix} & \begin{bmatrix} 0 & 0 \\ w_{22} & w_{21} \end{bmatrix} & \begin{bmatrix} 0 & 0 \\ 0 & w_{22} \end{bmatrix} \end{bmatrix} \begin{bmatrix} \frac{\partial J}{\partial v_{11}} & \frac{\partial J}{\partial v_{12}} \\ \frac{\partial J}{\partial v_{21}} & \frac{\partial J}{\partial v_{22}} \end{bmatrix} =$$

$$\begin{bmatrix} w_{11}\frac{\partial J}{\partial v_{11}} & w_{12}\frac{\partial J}{\partial v_{11}} + w_{11}\frac{\partial J}{\partial v_{12}} & w_{12}\frac{\partial J}{\partial v_{12}} \\ w_{21}\frac{\partial J}{\partial v_{11}} + w_{11}\frac{\partial J}{\partial v_{21}} & w_{22}\frac{\partial J}{\partial v_{11}} + w_{21}\frac{\partial J}{\partial v_{12}} + w_{12}\frac{\partial J}{\partial v_{21}} + w_{11}\frac{\partial J}{\partial v_{22}} & w_{22}\frac{\partial J}{\partial v_{12}} + w_{12}\frac{\partial J}{\partial v_{22}} \\ w_{21}\frac{\partial J}{\partial v_{21}} & w_{22}\frac{\partial J}{\partial v_{21}} + w_{21}\frac{\partial J}{\partial v_{22}} & w_{22}\frac{\partial J}{\partial v_{22}} \end{bmatrix}$$

Brad Quinton, Scott Chin

Apply chain rule to compute input gradients

$$\frac{\partial J}{\partial x} = \frac{\partial v}{\partial x}\frac{\partial J}{\partial v} = \begin{bmatrix} \begin{bmatrix} w_{11} & 0 \\ 0 & 0 \end{bmatrix} & \begin{bmatrix} w_{12} & w_{11} \\ 0 & 0 \end{bmatrix} & \begin{bmatrix} 0 & w_{12} \\ 0 & 0 \end{bmatrix} \\ \begin{bmatrix} w_{21} & 0 \\ w_{11} & 0 \end{bmatrix} & \begin{bmatrix} w_{22} & w_{21} \\ w_{12} & w_{11} \end{bmatrix} & \begin{bmatrix} 0 & w_{22} \\ 0 & w_{12} \end{bmatrix} \\ \begin{bmatrix} 0 & 0 \\ w_{21} & 0 \end{bmatrix} & \begin{bmatrix} 0 & 0 \\ w_{22} & w_{21} \end{bmatrix} & \begin{bmatrix} 0 & 0 \\ 0 & w_{22} \end{bmatrix} \end{bmatrix} \begin{bmatrix} \frac{\partial J}{\partial v_{11}} & \frac{\partial J}{\partial v_{12}} \\ \frac{\partial J}{\partial v_{21}} & \frac{\partial J}{\partial v_{22}} \end{bmatrix} =$$

What is the pattern here?

$$\begin{bmatrix} w_{11}\frac{\partial J}{\partial v_{11}} & w_{12}\frac{\partial J}{\partial v_{11}} + w_{11}\frac{\partial J}{\partial v_{12}} & w_{12}\frac{\partial J}{\partial v_{12}} \\ w_{21}\frac{\partial J}{\partial v_{11}} + w_{11}\frac{\partial J}{\partial v_{21}} & w_{22}\frac{\partial J}{\partial v_{11}} + w_{21}\frac{\partial J}{\partial v_{12}} + w_{12}\frac{\partial J}{\partial v_{21}} + w_{11}\frac{\partial J}{\partial v_{22}} & w_{22}\frac{\partial J}{\partial v_{12}} + w_{12}\frac{\partial J}{\partial v_{22}} \\ w_{21}\frac{\partial J}{\partial v_{21}} & w_{22}\frac{\partial J}{\partial v_{21}} + w_{21}\frac{\partial J}{\partial v_{22}} & w_{22}\frac{\partial J}{\partial v_{22}} \end{bmatrix}$$

Brad Quinton, Scott Chin

Apply chain rule to compute input gradients

$$\frac{\partial J}{\partial x} = \frac{\partial v}{\partial x}\frac{\partial J}{\partial v} = \begin{bmatrix} \begin{bmatrix} w_{11} & 0 \\ 0 & 0 \end{bmatrix} & \begin{bmatrix} w_{12} & w_{11} \\ 0 & 0 \end{bmatrix} & \begin{bmatrix} 0 & w_{12} \\ 0 & 0 \end{bmatrix} \\ \begin{bmatrix} w_{21} & 0 \\ w_{11} & 0 \end{bmatrix} & \begin{bmatrix} w_{22} & w_{21} \\ w_{12} & w_{11} \end{bmatrix} & \begin{bmatrix} 0 & w_{22} \\ 0 & w_{12} \end{bmatrix} \\ \begin{bmatrix} 0 & 0 \\ w_{21} & 0 \end{bmatrix} & \begin{bmatrix} 0 & 0 \\ w_{22} & w_{21} \end{bmatrix} & \begin{bmatrix} 0 & 0 \\ 0 & w_{22} \end{bmatrix} \end{bmatrix} \begin{bmatrix} \frac{\partial J}{\partial v_{11}} & \frac{\partial J}{\partial v_{12}} \\ \frac{\partial J}{\partial v_{21}} & \frac{\partial J}{\partial v_{22}} \end{bmatrix} =$$

What is the pattern here? It's another convolution!

$$\begin{bmatrix} w_{11}\frac{\partial J}{\partial v_{11}} & w_{12}\frac{\partial J}{\partial v_{11}} + w_{11}\frac{\partial J}{\partial v_{12}} & w_{12}\frac{\partial J}{\partial v_{12}} \\ w_{21}\frac{\partial J}{\partial v_{11}} + w_{11}\frac{\partial J}{\partial v_{21}} & w_{22}\frac{\partial J}{\partial v_{11}} + w_{21}\frac{\partial J}{\partial v_{12}} + w_{12}\frac{\partial J}{\partial v_{21}} + w_{11}\frac{\partial J}{\partial v_{22}} & w_{22}\frac{\partial J}{\partial v_{12}} + w_{12}\frac{\partial J}{\partial v_{22}} \\ w_{21}\frac{\partial J}{\partial v_{21}} & w_{22}\frac{\partial J}{\partial v_{21}} + w_{21}\frac{\partial J}{\partial v_{22}} & w_{22}\frac{\partial J}{\partial v_{22}} \end{bmatrix}$$

Can get the same result with the following convolution

$\frac{\partial J}{\partial v_{11}}$	$\frac{\partial J}{\partial v_{12}}$
$\frac{\partial J}{\partial v_{21}}$	$\frac{\partial J}{\partial v_{22}}$

$$\begin{bmatrix} w_{11}\frac{\partial J}{\partial v_{11}} & w_{12}\frac{\partial J}{\partial v_{11}} + w_{11}\frac{\partial J}{\partial v_{12}} & w_{12}\frac{\partial J}{\partial v_{12}} \\ w_{21}\frac{\partial J}{\partial v_{11}} + w_{11}\frac{\partial J}{\partial v_{21}} & w_{22}\frac{\partial J}{\partial v_{11}} + w_{21}\frac{\partial J}{\partial v_{12}} + w_{12}\frac{\partial J}{\partial v_{21}} + w_{11}\frac{\partial J}{\partial v_{22}} & w_{22}\frac{\partial J}{\partial v_{12}} + w_{12}\frac{\partial J}{\partial v_{22}} \\ w_{21}\frac{\partial J}{\partial v_{21}} & w_{22}\frac{\partial J}{\partial v_{21}} + w_{21}\frac{\partial J}{\partial v_{22}} & w_{22}\frac{\partial J}{\partial v_{22}} \end{bmatrix}$$

Brad Quinton, Scott Chin

Can get the same result with the following convolution

0	0	0	0
0	$\frac{\partial J}{\partial v_{11}}$	$\frac{\partial J}{\partial v_{12}}$	0
0	$\frac{\partial J}{\partial v_{21}}$	$\frac{\partial J}{\partial v_{22}}$	0
0	0	0	0

$$\begin{bmatrix} w_{11}\frac{\partial J}{\partial v_{11}} & w_{12}\frac{\partial J}{\partial v_{11}} + w_{11}\frac{\partial J}{\partial v_{12}} & w_{12}\frac{\partial J}{\partial v_{12}} \\ w_{21}\frac{\partial J}{\partial v_{11}} + w_{11}\frac{\partial J}{\partial v_{21}} & w_{22}\frac{\partial J}{\partial v_{11}} + w_{21}\frac{\partial J}{\partial v_{12}} + w_{12}\frac{\partial J}{\partial v_{21}} + w_{11}\frac{\partial J}{\partial v_{22}} & w_{22}\frac{\partial J}{\partial v_{12}} + w_{12}\frac{\partial J}{\partial v_{22}} \\ w_{21}\frac{\partial J}{\partial v_{21}} & w_{22}\frac{\partial J}{\partial v_{21}} + w_{21}\frac{\partial J}{\partial v_{22}} & w_{22}\frac{\partial J}{\partial v_{22}} \end{bmatrix}$$

Brad Quinton, Scott Chin

Can get the same result with the following convolution

$$\begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & \frac{\partial J}{\partial v_{11}} & \frac{\partial J}{\partial v_{12}} & 0 \\ 0 & \frac{\partial J}{\partial v_{21}} & \frac{\partial J}{\partial v_{22}} & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} \text{ conv } \begin{bmatrix} w_{11} & w_{12} \\ w_{21} & w_{22} \end{bmatrix}$$

$$\begin{bmatrix} w_{11}\frac{\partial J}{\partial v_{11}} & w_{12}\frac{\partial J}{\partial v_{11}} + w_{11}\frac{\partial J}{\partial v_{12}} & w_{12}\frac{\partial J}{\partial v_{12}} \\ w_{21}\frac{\partial J}{\partial v_{11}} + w_{11}\frac{\partial J}{\partial v_{21}} & w_{22}\frac{\partial J}{\partial v_{11}} + w_{21}\frac{\partial J}{\partial v_{12}} + w_{12}\frac{\partial J}{\partial v_{21}} + w_{11}\frac{\partial J}{\partial v_{22}} & w_{22}\frac{\partial J}{\partial v_{12}} + w_{12}\frac{\partial J}{\partial v_{22}} \\ w_{21}\frac{\partial J}{\partial v_{21}} & w_{22}\frac{\partial J}{\partial v_{21}} + w_{21}\frac{\partial J}{\partial v_{22}} & w_{22}\frac{\partial J}{\partial v_{22}} \end{bmatrix}$$

Brad Quinton, Scott Chin

Can get the same result with the following convolution

$$
\begin{array}{|c|c|c|c|}
\hline
0 & 0 & 0 & 0 \\
\hline
0 & \frac{\partial J}{\partial v_{11}} & \frac{\partial J}{\partial v_{12}} & 0 \\
\hline
0 & \frac{\partial J}{\partial v_{21}} & \frac{\partial J}{\partial v_{22}} & 0 \\
\hline
0 & 0 & 0 & 0 \\
\hline
\end{array}
\quad \text{conv} \quad
\begin{array}{|c|c|}
\hline
w_{21} & w_{11} \\
\hline
w_{22} & w_{12} \\
\hline
\end{array}
$$

90 Degrees

$$
\begin{bmatrix}
w_{11}\frac{\partial J}{\partial v_{11}} & w_{12}\frac{\partial J}{\partial v_{11}} + w_{11}\frac{\partial J}{\partial v_{12}} & w_{12}\frac{\partial J}{\partial v_{12}} \\
w_{21}\frac{\partial J}{\partial v_{11}} + w_{11}\frac{\partial J}{\partial v_{21}} & w_{22}\frac{\partial J}{\partial v_{11}} + w_{21}\frac{\partial J}{\partial v_{12}} + w_{12}\frac{\partial J}{\partial v_{21}} + w_{11}\frac{\partial J}{\partial v_{22}} & w_{22}\frac{\partial J}{\partial v_{12}} + w_{12}\frac{\partial J}{\partial v_{22}} \\
w_{21}\frac{\partial J}{\partial v_{21}} & w_{22}\frac{\partial J}{\partial v_{21}} + w_{21}\frac{\partial J}{\partial v_{22}} & w_{22}\frac{\partial J}{\partial v_{22}}
\end{bmatrix}
$$

Brad Quinton, Scott Chin

Can get the same result with the following convolution

$$
\begin{bmatrix}
0 & 0 & 0 & 0 \\
0 & \frac{\partial J}{\partial v_{11}} & \frac{\partial J}{\partial v_{12}} & 0 \\
0 & \frac{\partial J}{\partial v_{21}} & \frac{\partial J}{\partial v_{22}} & 0 \\
0 & 0 & 0 & 0
\end{bmatrix}
\text{ conv }
\begin{bmatrix}
w_{22} & w_{21} \\
w_{12} & w_{11}
\end{bmatrix}
$$

90 Degrees

$$
\begin{bmatrix}
w_{11}\frac{\partial J}{\partial v_{11}} & w_{12}\frac{\partial J}{\partial v_{11}} + w_{11}\frac{\partial J}{\partial v_{12}} & w_{12}\frac{\partial J}{\partial v_{12}} \\
w_{21}\frac{\partial J}{\partial v_{11}} + w_{11}\frac{\partial J}{\partial v_{21}} & w_{22}\frac{\partial J}{\partial v_{11}} + w_{21}\frac{\partial J}{\partial v_{12}} + w_{12}\frac{\partial J}{\partial v_{21}} + w_{11}\frac{\partial J}{\partial v_{22}} & w_{22}\frac{\partial J}{\partial v_{12}} + w_{12}\frac{\partial J}{\partial v_{22}} \\
w_{21}\frac{\partial J}{\partial v_{21}} & w_{22}\frac{\partial J}{\partial v_{21}} + w_{21}\frac{\partial J}{\partial v_{22}} & w_{22}\frac{\partial J}{\partial v_{22}}
\end{bmatrix}
$$

Brad Quinton, Scott Chin

Can get the same result with the following convolution

$$\begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & \frac{\partial J}{\partial v_{11}} & \frac{\partial J}{\partial v_{12}} & 0 \\ 0 & \frac{\partial J}{\partial v_{21}} & \frac{\partial J}{\partial v_{22}} & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} \text{ conv } \begin{bmatrix} w_{22} & w_{21} \\ w_{12} & w_{11} \end{bmatrix} =$$

$$\begin{bmatrix} w_{11}\frac{\partial J}{\partial v_{11}} & w_{12}\frac{\partial J}{\partial v_{11}} + w_{11}\frac{\partial J}{\partial v_{12}} & w_{12}\frac{\partial J}{\partial v_{12}} \\ w_{21}\frac{\partial J}{\partial v_{11}} + w_{11}\frac{\partial J}{\partial v_{21}} & w_{22}\frac{\partial J}{\partial v_{11}} + w_{21}\frac{\partial J}{\partial v_{12}} + w_{12}\frac{\partial J}{\partial v_{21}} + w_{11}\frac{\partial J}{\partial v_{22}} & w_{22}\frac{\partial J}{\partial v_{12}} + w_{12}\frac{\partial J}{\partial v_{22}} \\ w_{21}\frac{\partial J}{\partial v_{21}} & w_{22}\frac{\partial J}{\partial v_{21}} + w_{21}\frac{\partial J}{\partial v_{22}} & w_{22}\frac{\partial J}{\partial v_{22}} \end{bmatrix}$$

Brad Quinton, Scott Chin

Can get the same result with the following convolution

$$\begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & \frac{\partial J}{\partial v_{11}} & \frac{\partial J}{\partial v_{12}} & 0 \\ 0 & \frac{\partial J}{\partial v_{21}} & \frac{\partial J}{\partial v_{22}} & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} \text{ conv } \begin{bmatrix} w_{22} & w_{21} \\ w_{12} & w_{11} \end{bmatrix} =$$

$$\begin{bmatrix} \boxed{w_{11}\frac{\partial J}{\partial v_{11}}} & w_{12}\frac{\partial J}{\partial v_{11}} + w_{11}\frac{\partial J}{\partial v_{12}} & w_{12}\frac{\partial J}{\partial v_{12}} \\ w_{21}\frac{\partial J}{\partial v_{11}} + w_{11}\frac{\partial J}{\partial v_{21}} & w_{22}\frac{\partial J}{\partial v_{11}} + w_{21}\frac{\partial J}{\partial v_{12}} + w_{12}\frac{\partial J}{\partial v_{21}} + w_{11}\frac{\partial J}{\partial v_{22}} & w_{22}\frac{\partial J}{\partial v_{12}} + w_{12}\frac{\partial J}{\partial v_{22}} \\ w_{21}\frac{\partial J}{\partial v_{21}} & w_{22}\frac{\partial J}{\partial v_{21}} + w_{21}\frac{\partial J}{\partial v_{22}} & w_{22}\frac{\partial J}{\partial v_{22}} \end{bmatrix}$$

Brad Quinton, Scott Chin

Can get the same result with the following convolution

$$\begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & \frac{\partial J}{\partial v_{11}} & \frac{\partial J}{\partial v_{12}} & 0 \\ 0 & \frac{\partial J}{\partial v_{21}} & \frac{\partial J}{\partial v_{22}} & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} \text{ conv } \begin{bmatrix} w_{22} & w_{21} \\ w_{12} & w_{11} \end{bmatrix} =$$

$$\begin{bmatrix} w_{11}\frac{\partial J}{\partial v_{11}} & \boxed{w_{12}\frac{\partial J}{\partial v_{11}} + w_{11}\frac{\partial J}{\partial v_{12}}} & w_{12}\frac{\partial J}{\partial v_{12}} \\ w_{21}\frac{\partial J}{\partial v_{11}} + w_{11}\frac{\partial J}{\partial v_{21}} & w_{22}\frac{\partial J}{\partial v_{11}} + w_{21}\frac{\partial J}{\partial v_{12}} + w_{12}\frac{\partial J}{\partial v_{21}} + w_{11}\frac{\partial J}{\partial v_{22}} & w_{22}\frac{\partial J}{\partial v_{12}} + w_{12}\frac{\partial J}{\partial v_{22}} \\ w_{21}\frac{\partial J}{\partial v_{21}} & w_{22}\frac{\partial J}{\partial v_{21}} + w_{21}\frac{\partial J}{\partial v_{22}} & w_{22}\frac{\partial J}{\partial v_{22}} \end{bmatrix}$$

Can get the same result with the following convolution

$$\begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & \frac{\partial J}{\partial v_{11}} & \frac{\partial J}{\partial v_{12}} & 0 \\ 0 & \frac{\partial J}{\partial v_{21}} & \frac{\partial J}{\partial v_{22}} & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} \text{ conv } \begin{bmatrix} w_{22} & w_{21} \\ w_{12} & w_{11} \end{bmatrix} =$$

$$\begin{bmatrix} w_{11}\frac{\partial J}{\partial v_{11}} & w_{12}\frac{\partial J}{\partial v_{11}} + w_{11}\frac{\partial J}{\partial v_{12}} & \boxed{w_{12}\frac{\partial J}{\partial v_{12}}} \\ w_{21}\frac{\partial J}{\partial v_{11}} + w_{11}\frac{\partial J}{\partial v_{21}} & w_{22}\frac{\partial J}{\partial v_{11}} + w_{21}\frac{\partial J}{\partial v_{12}} + w_{12}\frac{\partial J}{\partial v_{21}} + w_{11}\frac{\partial J}{\partial v_{22}} & w_{22}\frac{\partial J}{\partial v_{12}} + w_{12}\frac{\partial J}{\partial v_{22}} \\ w_{21}\frac{\partial J}{\partial v_{21}} & w_{22}\frac{\partial J}{\partial v_{21}} + w_{21}\frac{\partial J}{\partial v_{22}} & w_{22}\frac{\partial J}{\partial v_{22}} \end{bmatrix}$$

Brad Quinton, Scott Chin

Can get the same result with the following convolution

$$
\begin{bmatrix}
0 & 0 & 0 & 0 \\
0 & \frac{\partial J}{\partial v_{11}} & \frac{\partial J}{\partial v_{12}} & 0 \\
0 & \frac{\partial J}{\partial v_{21}} & \frac{\partial J}{\partial v_{22}} & 0 \\
0 & 0 & 0 & 0
\end{bmatrix}
\text{ conv }
\begin{bmatrix}
w_{22} & w_{21} \\
w_{12} & w_{11}
\end{bmatrix}
=
$$

$$
\begin{bmatrix}
w_{11}\frac{\partial J}{\partial v_{11}} & w_{12}\frac{\partial J}{\partial v_{11}} + w_{11}\frac{\partial J}{\partial v_{12}} & w_{12}\frac{\partial J}{\partial v_{12}} \\
\boxed{w_{21}\frac{\partial J}{\partial v_{11}} + w_{11}\frac{\partial J}{\partial v_{21}}} & w_{22}\frac{\partial J}{\partial v_{11}} + w_{21}\frac{\partial J}{\partial v_{12}} + w_{12}\frac{\partial J}{\partial v_{21}} + w_{11}\frac{\partial J}{\partial v_{22}} & w_{22}\frac{\partial J}{\partial v_{12}} + w_{12}\frac{\partial J}{\partial v_{22}} \\
w_{21}\frac{\partial J}{\partial v_{21}} & w_{22}\frac{\partial J}{\partial v_{21}} + w_{21}\frac{\partial J}{\partial v_{22}} & w_{22}\frac{\partial J}{\partial v_{22}}
\end{bmatrix}
$$

Brad Quinton, Scott Chin

Can get the same result with the following convolution

$$
\begin{bmatrix}
0 & 0 & 0 & 0 \\
0 & \frac{\partial J}{\partial v_{11}} & \frac{\partial J}{\partial v_{12}} & 0 \\
0 & \frac{\partial J}{\partial v_{21}} & \frac{\partial J}{\partial v_{22}} & 0 \\
0 & 0 & 0 & 0
\end{bmatrix}
\ \text{conv}\
\begin{bmatrix}
w_{22} & w_{21} \\
w_{12} & w_{11}
\end{bmatrix}
=
$$

$$
\begin{bmatrix}
w_{11}\frac{\partial J}{\partial v_{11}} & w_{12}\frac{\partial J}{\partial v_{11}} + w_{11}\frac{\partial J}{\partial v_{12}} & w_{12}\frac{\partial J}{\partial v_{12}} \\
w_{21}\frac{\partial J}{\partial v_{11}} + w_{11}\frac{\partial J}{\partial v_{21}} & \boxed{w_{22}\frac{\partial J}{\partial v_{11}} + w_{21}\frac{\partial J}{\partial v_{12}} + w_{12}\frac{\partial J}{\partial v_{21}} + w_{11}\frac{\partial J}{\partial v_{22}}} & w_{22}\frac{\partial J}{\partial v_{12}} + w_{12}\frac{\partial J}{\partial v_{22}} \\
w_{21}\frac{\partial J}{\partial v_{21}} & w_{22}\frac{\partial J}{\partial v_{21}} + w_{21}\frac{\partial J}{\partial v_{22}} & w_{22}\frac{\partial J}{\partial v_{22}}
\end{bmatrix}
$$

Brad Quinton, Scott Chin

Can get the same result with the following convolution

$$\begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & \frac{\partial J}{\partial v_{11}} & \frac{\partial J}{\partial v_{12}} & 0 \\ 0 & \frac{\partial J}{\partial v_{21}} & \frac{\partial J}{\partial v_{22}} & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} \text{ conv } \begin{bmatrix} w_{22} & w_{21} \\ w_{12} & w_{11} \end{bmatrix} =$$

$$\begin{bmatrix} w_{11}\frac{\partial J}{\partial v_{11}} & w_{12}\frac{\partial J}{\partial v_{11}} + w_{11}\frac{\partial J}{\partial v_{12}} & w_{12}\frac{\partial J}{\partial v_{12}} \\ w_{21}\frac{\partial J}{\partial v_{11}} + w_{11}\frac{\partial J}{\partial v_{21}} & w_{22}\frac{\partial J}{\partial v_{11}} + w_{21}\frac{\partial J}{\partial v_{12}} + w_{12}\frac{\partial J}{\partial v_{21}} + w_{11}\frac{\partial J}{\partial v_{22}} & \boxed{w_{22}\frac{\partial J}{\partial v_{12}} + w_{12}\frac{\partial J}{\partial v_{22}}} \\ w_{21}\frac{\partial J}{\partial v_{21}} & w_{22}\frac{\partial J}{\partial v_{21}} + w_{21}\frac{\partial J}{\partial v_{22}} & w_{22}\frac{\partial J}{\partial v_{22}} \end{bmatrix}$$

Brad Quinton, Scott Chin

Can get the same result with the following convolution

$$\begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & \frac{\partial J}{\partial v_{11}} & \frac{\partial J}{\partial v_{12}} & 0 \\ 0 & \frac{\partial J}{\partial v_{21}} & \frac{\partial J}{\partial v_{22}} & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} \text{ conv } \begin{bmatrix} w_{22} & w_{21} \\ w_{12} & w_{11} \end{bmatrix} =$$

$$\begin{bmatrix} w_{11}\frac{\partial J}{\partial v_{11}} & w_{12}\frac{\partial J}{\partial v_{11}} + w_{11}\frac{\partial J}{\partial v_{12}} & w_{12}\frac{\partial J}{\partial v_{12}} \\ w_{21}\frac{\partial J}{\partial v_{11}} + w_{11}\frac{\partial J}{\partial v_{21}} & w_{22}\frac{\partial J}{\partial v_{11}} + w_{21}\frac{\partial J}{\partial v_{12}} + w_{12}\frac{\partial J}{\partial v_{21}} + w_{11}\frac{\partial J}{\partial v_{22}} & w_{22}\frac{\partial J}{\partial v_{12}} + w_{12}\frac{\partial J}{\partial v_{22}} \\ \boxed{w_{21}\frac{\partial J}{\partial v_{21}}} & w_{22}\frac{\partial J}{\partial v_{21}} + w_{21}\frac{\partial J}{\partial v_{22}} & w_{22}\frac{\partial J}{\partial v_{22}} \end{bmatrix}$$

Brad Quinton, Scott Chin

Can get the same result with the following convolution

$$\begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & \frac{\partial J}{\partial v_{11}} & \frac{\partial J}{\partial v_{12}} & 0 \\ 0 & \frac{\partial J}{\partial v_{21}} & \frac{\partial J}{\partial v_{22}} & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} \text{ conv } \begin{bmatrix} w_{22} & w_{21} \\ w_{12} & w_{11} \end{bmatrix} =$$

$$\begin{bmatrix} w_{11}\frac{\partial J}{\partial v_{11}} & w_{12}\frac{\partial J}{\partial v_{11}} + w_{11}\frac{\partial J}{\partial v_{12}} & w_{12}\frac{\partial J}{\partial v_{12}} \\ w_{21}\frac{\partial J}{\partial v_{11}} + w_{11}\frac{\partial J}{\partial v_{21}} & w_{22}\frac{\partial J}{\partial v_{11}} + w_{21}\frac{\partial J}{\partial v_{12}} + w_{12}\frac{\partial J}{\partial v_{21}} + w_{11}\frac{\partial J}{\partial v_{22}} & w_{22}\frac{\partial J}{\partial v_{12}} + w_{12}\frac{\partial J}{\partial v_{22}} \\ w_{21}\frac{\partial J}{\partial v_{21}} & \boxed{w_{22}\frac{\partial J}{\partial v_{21}} + w_{21}\frac{\partial J}{\partial v_{22}}} & w_{22}\frac{\partial J}{\partial v_{22}} \end{bmatrix}$$

Brad Quinton, Scott Chin

Can get the same result with the following convolution

$$
\begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & \frac{\partial J}{\partial v_{11}} & \frac{\partial J}{\partial v_{12}} & 0 \\ 0 & \frac{\partial J}{\partial v_{21}} & \frac{\partial J}{\partial v_{22}} & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} \text{ conv } \begin{bmatrix} w_{22} & w_{21} \\ w_{12} & w_{11} \end{bmatrix} =
$$

$$
\begin{bmatrix} w_{11}\frac{\partial J}{\partial v_{11}} & w_{12}\frac{\partial J}{\partial v_{11}} + w_{11}\frac{\partial J}{\partial v_{12}} & w_{12}\frac{\partial J}{\partial v_{12}} \\ w_{21}\frac{\partial J}{\partial v_{11}} + w_{11}\frac{\partial J}{\partial v_{21}} & w_{22}\frac{\partial J}{\partial v_{11}} + w_{21}\frac{\partial J}{\partial v_{12}} + w_{12}\frac{\partial J}{\partial v_{21}} + w_{11}\frac{\partial J}{\partial v_{22}} & w_{22}\frac{\partial J}{\partial v_{12}} + w_{12}\frac{\partial J}{\partial v_{22}} \\ w_{21}\frac{\partial J}{\partial v_{21}} & w_{22}\frac{\partial J}{\partial v_{21}} + w_{21}\frac{\partial J}{\partial v_{22}} & \boxed{w_{22}\frac{\partial J}{\partial v_{22}}} \end{bmatrix}
$$

Brad Quinton, Scott Chin

Can get the same result with the following convolution

$$\begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & \frac{\partial J}{\partial v_{11}} & \frac{\partial J}{\partial v_{12}} & 0 \\ 0 & \frac{\partial J}{\partial v_{21}} & \frac{\partial J}{\partial v_{22}} & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} \text{ conv } \begin{bmatrix} w_{22} & w_{21} \\ w_{12} & w_{11} \end{bmatrix} = \begin{bmatrix} \frac{\partial J}{\partial x_{11}} & \frac{\partial J}{\partial x_{12}} & \frac{\partial J}{\partial x_{13}} \\ \frac{\partial J}{\partial x_{21}} & \frac{\partial J}{\partial x_{22}} & \frac{\partial J}{\partial x_{23}} \\ \frac{\partial J}{\partial x_{31}} & \frac{\partial J}{\partial x_{32}} & \frac{\partial J}{\partial x_{33}} \end{bmatrix}$$

$$\boxed{\frac{\partial J}{\partial x} = pad\left(\frac{\partial J}{\partial v}\right) \text{ conv rot180}(w)}$$

- No Jacobian require at all!
- This convolution yields same result as doing the full 4D-Tensor Jacobian and upstream gradient matrix multiply!
- You will implement this in Assignment 4

Can get the same result with the following convolution

$$
\begin{array}{|c|c|c|c|}
\hline
0 & 0 & 0 & 0 \\
\hline
0 & \frac{\partial J}{\partial v_{11}} & \frac{\partial J}{\partial v_{12}} & 0 \\
\hline
0 & \frac{\partial J}{\partial v_{21}} & \frac{\partial J}{\partial v_{22}} & 0 \\
\hline
0 & 0 & 0 & 0 \\
\hline
\end{array}
\quad \text{conv} \quad
\begin{array}{|c|c|}
\hline
w_{22} & w_{21} \\
\hline
w_{12} & w_{11} \\
\hline
\end{array}
\quad = \quad
\begin{array}{|c|c|c|}
\hline
\frac{\partial J}{\partial x_{11}} & \frac{\partial J}{\partial x_{12}} & \frac{\partial J}{\partial x_{13}} \\
\hline
\frac{\partial J}{\partial x_{21}} & \frac{\partial J}{\partial x_{22}} & \frac{\partial J}{\partial x_{23}} \\
\hline
\frac{\partial J}{\partial x_{31}} & \frac{\partial J}{\partial x_{32}} & \frac{\partial J}{\partial x_{33}} \\
\hline
\end{array}
$$

$$
\boxed{\frac{\partial J}{\partial x} = pad\left(\frac{\partial J}{\partial v}\right) \ \text{conv rot180}(w)}
$$

How much to pad?

- No Jacobian require at all!
- This convolution yields same result as doing the full 4D-Tensor Jacobian and upstream gradient matrix multiply!
- You will implement this in Assignment 4

Brad Quinton, Scott Chin

From the forward propagation convolution operation:

$$\begin{bmatrix} x_{11} & x_{12} & x_{13} \\ x_{21} & x_{22} & x_{23} \\ x_{31} & x_{32} & x_{33} \end{bmatrix} \text{ conv } \begin{bmatrix} w_{11} & w_{12} \\ w_{21} & w_{22} \end{bmatrix} = \begin{bmatrix} v_{11} & v_{12} \\ v_{21} & v_{22} \end{bmatrix}$$

(h_{in}, w_{in}) $\quad$ (f, f) $\quad$ (h_{out}, w_{out})

$$h_{out} = h_{in} - f + 1 \qquad w_{out} = w_{in} - f + 1$$

Brad Quinton, Scott Chin

From the back propagation convolution operation:

$$\underset{(h'_{out}, w'_{out})}{\begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & \frac{\partial J}{\partial v_{11}} & \frac{\partial J}{\partial v_{12}} & 0 \\ 0 & \frac{\partial J}{\partial v_{21}} & \frac{\partial J}{\partial v_{22}} & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}} \text{ conv } \underset{(f, f)}{\begin{bmatrix} w_{22} & w_{21} \\ w_{12} & w_{11} \end{bmatrix}} = \underset{(h_{in}, w_{in})}{\begin{bmatrix} \frac{\partial J}{\partial x_{11}} & \frac{\partial J}{\partial x_{12}} & \frac{\partial J}{\partial x_{13}} \\ \frac{\partial J}{\partial x_{21}} & \frac{\partial J}{\partial x_{22}} & \frac{\partial J}{\partial x_{23}} \\ \frac{\partial J}{\partial x_{31}} & \frac{\partial J}{\partial x_{32}} & \frac{\partial J}{\partial x_{33}} \end{bmatrix}}$$

$$h_{out} = h_{in} - f + 1 \qquad w_{out} = w_{in} - f + 1$$

$$h'_{out} = h_{out} + 2p_h \qquad w'_{out} = w_{out} + 2p_w$$

Brad Quinton, Scott Chin

From the back propagation convolution operation:

$$\underset{(h'_{out}, w'_{out})}{\begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & \frac{\partial J}{\partial v_{11}} & \frac{\partial J}{\partial v_{12}} & 0 \\ 0 & \frac{\partial J}{\partial v_{21}} & \frac{\partial J}{\partial v_{22}} & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}} \text{ conv } \underset{(f, f)}{\begin{bmatrix} w_{22} & w_{21} \\ w_{12} & w_{11} \end{bmatrix}} = \underset{(h_{in}, w_{in})}{\begin{bmatrix} \frac{\partial J}{\partial x_{11}} & \frac{\partial J}{\partial x_{12}} & \frac{\partial J}{\partial x_{13}} \\ \frac{\partial J}{\partial x_{21}} & \frac{\partial J}{\partial x_{22}} & \frac{\partial J}{\partial x_{23}} \\ \frac{\partial J}{\partial x_{31}} & \frac{\partial J}{\partial x_{32}} & \frac{\partial J}{\partial x_{33}} \end{bmatrix}}$$

$$h_{out} = h_{in} - f + 1$$
$$h'_{out} = h_{out} + 2p_h$$
$$h_{in} = h'_{out} - f + 1$$

$$w_{out} = w_{in} - f + 1$$
$$w'_{out} = w_{out} + 2p_w$$
$$w_{in} = w'_{out} - f + 1$$

Brad Quinton, Scott Chin

From the back propagation convolution operation:

$$\begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & \frac{\partial J}{\partial v_{11}} & \frac{\partial J}{\partial v_{12}} & 0 \\ 0 & \frac{\partial J}{\partial v_{21}} & \frac{\partial J}{\partial v_{22}} & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} \text{ conv } \begin{bmatrix} w_{22} & w_{21} \\ w_{12} & w_{11} \end{bmatrix} = \begin{bmatrix} \frac{\partial J}{\partial x_{11}} & \frac{\partial J}{\partial x_{12}} & \frac{\partial J}{\partial x_{13}} \\ \frac{\partial J}{\partial x_{21}} & \frac{\partial J}{\partial x_{22}} & \frac{\partial J}{\partial x_{23}} \\ \frac{\partial J}{\partial x_{31}} & \frac{\partial J}{\partial x_{32}} & \frac{\partial J}{\partial x_{33}} \end{bmatrix}$$

(h'_{out}, w'_{out}) (f, f) (h_{in}, w_{in})

$$h_{out} = h_{in} - f + 1$$
$$h'_{out} = h_{out} + 2p_h$$
$$h_{in} = h'_{out} - f + 1$$

$$w_{out} = w_{in} - f + 1$$
$$w'_{out} = w_{out} + 2p_w$$
$$w_{in} = w'_{out} - f + 1$$

$$\boxed{p_h = f - 1}$$

$$\boxed{p_w = f - 1}$$

Summary (m=1, K=1, c_in=1)

$$\frac{\partial J}{\partial w} = x \text{ conv} \frac{\partial J}{\partial v}$$

$$\begin{bmatrix} x_{11} & x_{12} & x_{13} \\ x_{21} & x_{22} & x_{23} \\ x_{31} & x_{32} & x_{33} \end{bmatrix} \text{ conv } \begin{bmatrix} \frac{\partial J}{\partial v_{11}} & \frac{\partial J}{\partial v_{12}} \\ \frac{\partial J}{\partial v_{21}} & \frac{\partial J}{\partial v_{22}} \end{bmatrix} = \begin{bmatrix} \frac{\partial J}{\partial w_{11}} & \frac{\partial J}{\partial w_{12}} \\ \frac{\partial J}{\partial w_{21}} & \frac{\partial J}{\partial w_{22}} \end{bmatrix}$$

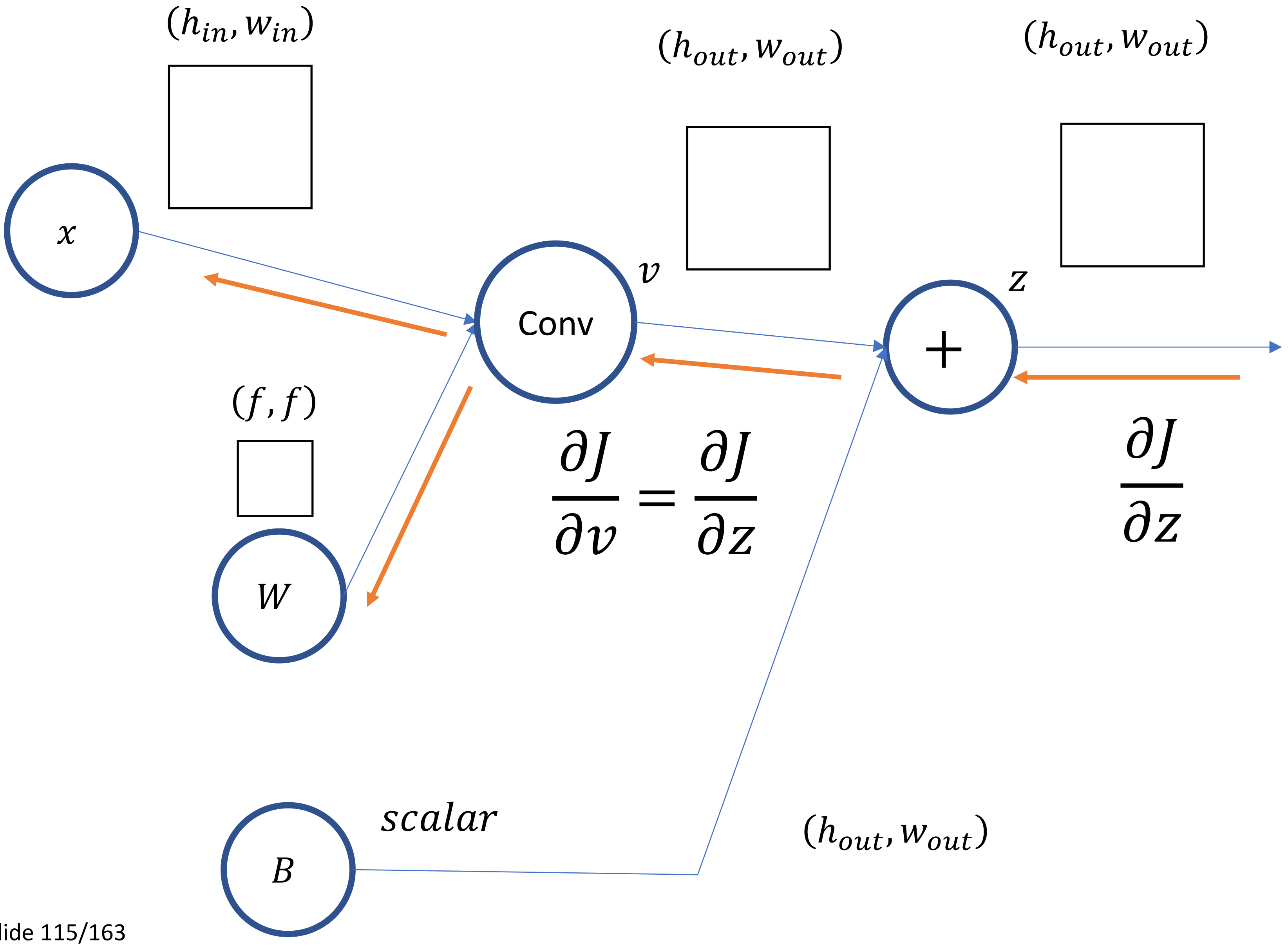
$$\frac{\partial J}{\partial x} = pad\left(\frac{\partial J}{\partial v}\right) \text{ conv rot180}(w)$$

Pad by $p = f - 1$

$$\begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & \frac{\partial J}{\partial v_{11}} & \frac{\partial J}{\partial v_{12}} & 0 \\ 0 & \frac{\partial J}{\partial v_{21}} & \frac{\partial J}{\partial v_{22}} & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} \text{ conv } \begin{bmatrix} w_{22} & w_{21} \\ w_{12} & w_{11} \end{bmatrix} = \begin{bmatrix} \frac{\partial J}{\partial x_{11}} & \frac{\partial J}{\partial x_{12}} & \frac{\partial J}{\partial x_{13}} \\ \frac{\partial J}{\partial x_{21}} & \frac{\partial J}{\partial x_{22}} & \frac{\partial J}{\partial x_{23}} \\ \frac{\partial J}{\partial x_{31}} & \frac{\partial J}{\partial x_{32}} & \frac{\partial J}{\partial x_{33}} \end{bmatrix}$$

Brad Quinton, Scott Chin

A compute graph for a convolutional layer

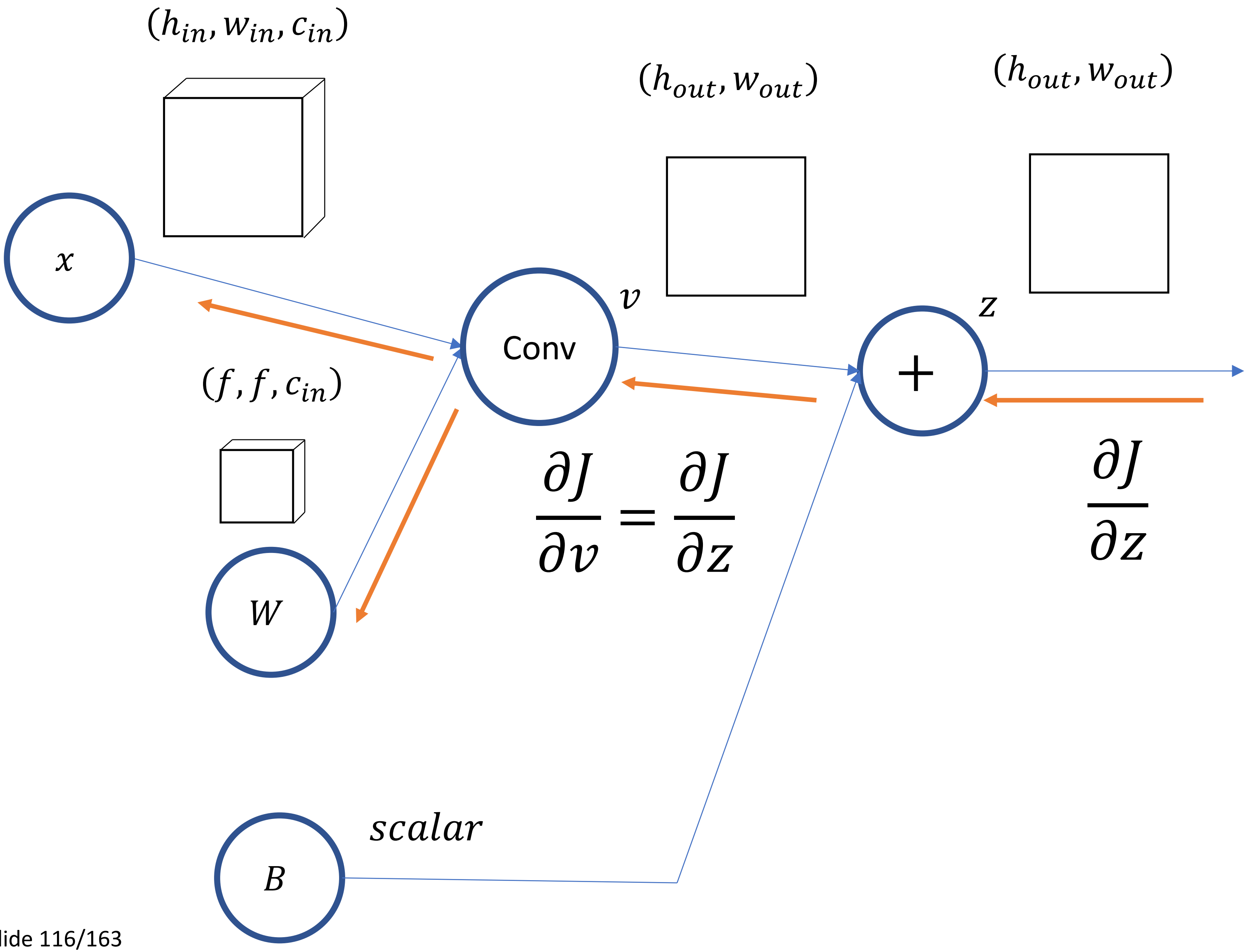


Start by considering:

- One sample (m=1)
- One filter (K=1)
- One channel ($c_{in} = 1$)

Brad Quinton, Scott Chin

A compute graph for a convolutional layer



Start by considering:

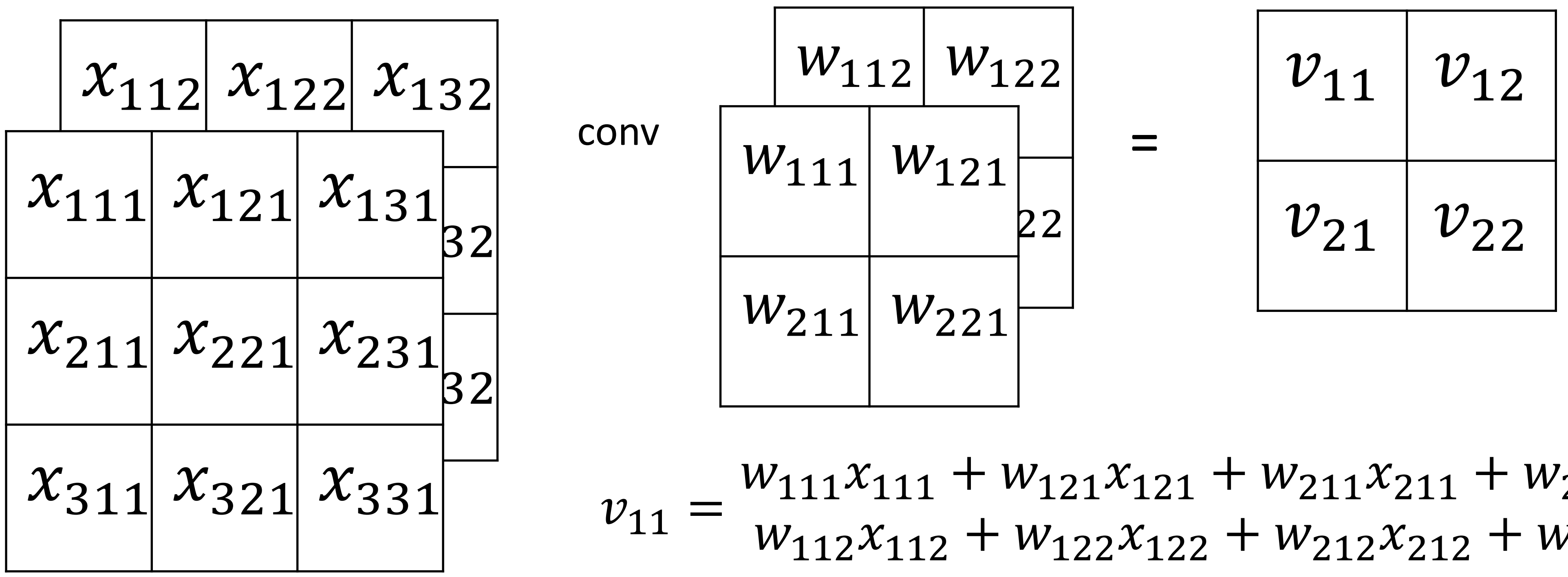
- One sample (m=1)
- One filter (K=1)
- ~~One channel ($c_{in} = 1$)~~

Brad Quinton, Scott Chin

For Weight Gradients $\frac{\partial J}{\partial w}$

- Recall from forward propagation, each channel of weights operate independently on each channel of the input volume.

$(h_{in}, w_{in}) = (3,3,2)$ $\quad (f, f) = (2,2,2)$ $\quad (h_{out}, w_{out}) = (2,2)$



$$v_{11} = \begin{array}{c} w_{111}x_{111} + w_{121}x_{121} + w_{211}x_{211} + w_{221}x_{221} + \\ w_{112}x_{112} + w_{122}x_{122} + w_{212}x_{212} + w_{222}x_{222} \end{array}$$

$$\frac{\partial J}{\partial w} = x \text{ conv} \frac{\partial J}{\partial v}$$

Compute each channel Independently.

For channel 1

$$\begin{bmatrix} x_{111} & x_{121} & x_{131} \\ x_{211} & x_{221} & x_{231} \\ x_{311} & x_{321} & x_{331} \end{bmatrix} \text{ conv } \begin{bmatrix} \frac{\partial J}{\partial v_{11}} & \frac{\partial J}{\partial v_{12}} \\ \frac{\partial J}{\partial v_{21}} & \frac{\partial J}{\partial v_{22}} \end{bmatrix} = \begin{bmatrix} \frac{\partial J}{\partial w_{111}} & \frac{\partial J}{\partial w_{121}} \\ \frac{\partial J}{\partial w_{211}} & \frac{\partial J}{\partial w_{221}} \end{bmatrix}$$

For channel 2

$$\begin{bmatrix} x_{112} & x_{122} & x_{132} \\ x_{212} & x_{222} & x_{232} \\ x_{312} & x_{322} & x_{332} \end{bmatrix} \text{ conv } \begin{bmatrix} \frac{\partial J}{\partial v_{11}} & \frac{\partial J}{\partial v_{12}} \\ \frac{\partial J}{\partial v_{21}} & \frac{\partial J}{\partial v_{22}} \end{bmatrix} = \begin{bmatrix} \frac{\partial J}{\partial w_{112}} & \frac{\partial J}{\partial w_{122}} \\ \frac{\partial J}{\partial w_{212}} & \frac{\partial J}{\partial w_{222}} \end{bmatrix}$$

For Input Volume

$$\frac{\partial J}{\partial x} = pad\left(\frac{\partial J}{\partial v}\right) \text{ conv rot180}(w)$$

For channel 1

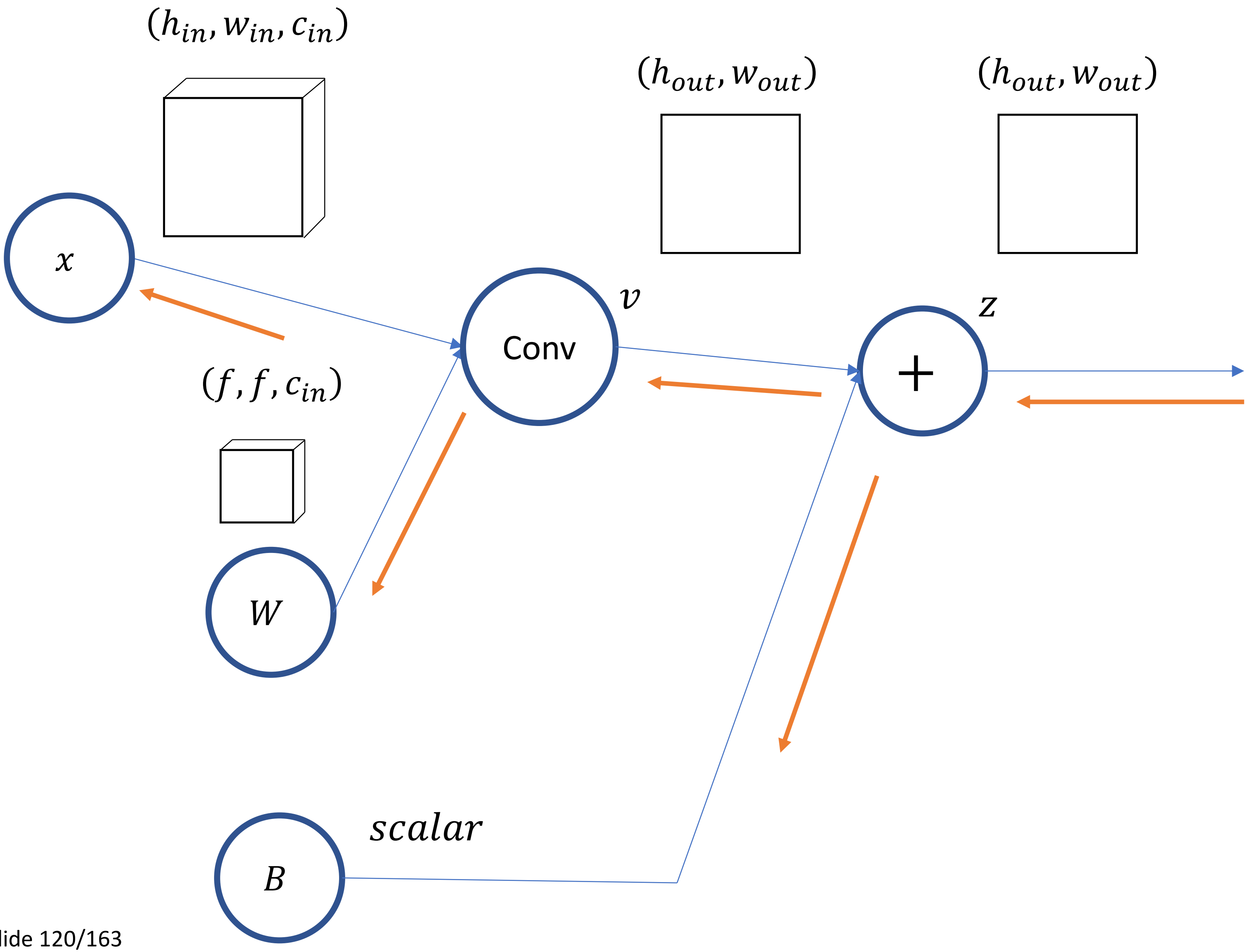
$$\begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & \frac{\partial J}{\partial v_{11}} & \frac{\partial J}{\partial v_{12}} & 0 \\ 0 & \frac{\partial J}{\partial v_{21}} & \frac{\partial J}{\partial v_{22}} & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} \text{ conv } \begin{bmatrix} w_{221} & w_{211} \\ w_{121} & w_{111} \end{bmatrix} = \begin{bmatrix} \frac{\partial J}{\partial x_{111}} & \frac{\partial J}{\partial x_{121}} & \frac{\partial J}{\partial x_{131}} \\ \frac{\partial J}{\partial x_{211}} & \frac{\partial J}{\partial x_{221}} & \frac{\partial J}{\partial x_{231}} \\ \frac{\partial J}{\partial x_{311}} & \frac{\partial J}{\partial x_{321}} & \frac{\partial J}{\partial x_{331}} \end{bmatrix}$$

For channel 2

$$\begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & \frac{\partial J}{\partial v_{11}} & \frac{\partial J}{\partial v_{12}} & 0 \\ 0 & \frac{\partial J}{\partial v_{21}} & \frac{\partial J}{\partial v_{22}} & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} \text{ conv } \begin{bmatrix} w_{222} & w_{212} \\ w_{122} & w_{112} \end{bmatrix} = \begin{bmatrix} \frac{\partial J}{\partial x_{112}} & \frac{\partial J}{\partial x_{122}} & \frac{\partial J}{\partial x_{132}} \\ \frac{\partial J}{\partial x_{212}} & \frac{\partial J}{\partial x_{222}} & \frac{\partial J}{\partial x_{232}} \\ \frac{\partial J}{\partial x_{312}} & \frac{\partial J}{\partial x_{322}} & \frac{\partial J}{\partial x_{332}} \end{bmatrix}$$

Brad Quinton, Scott Chin

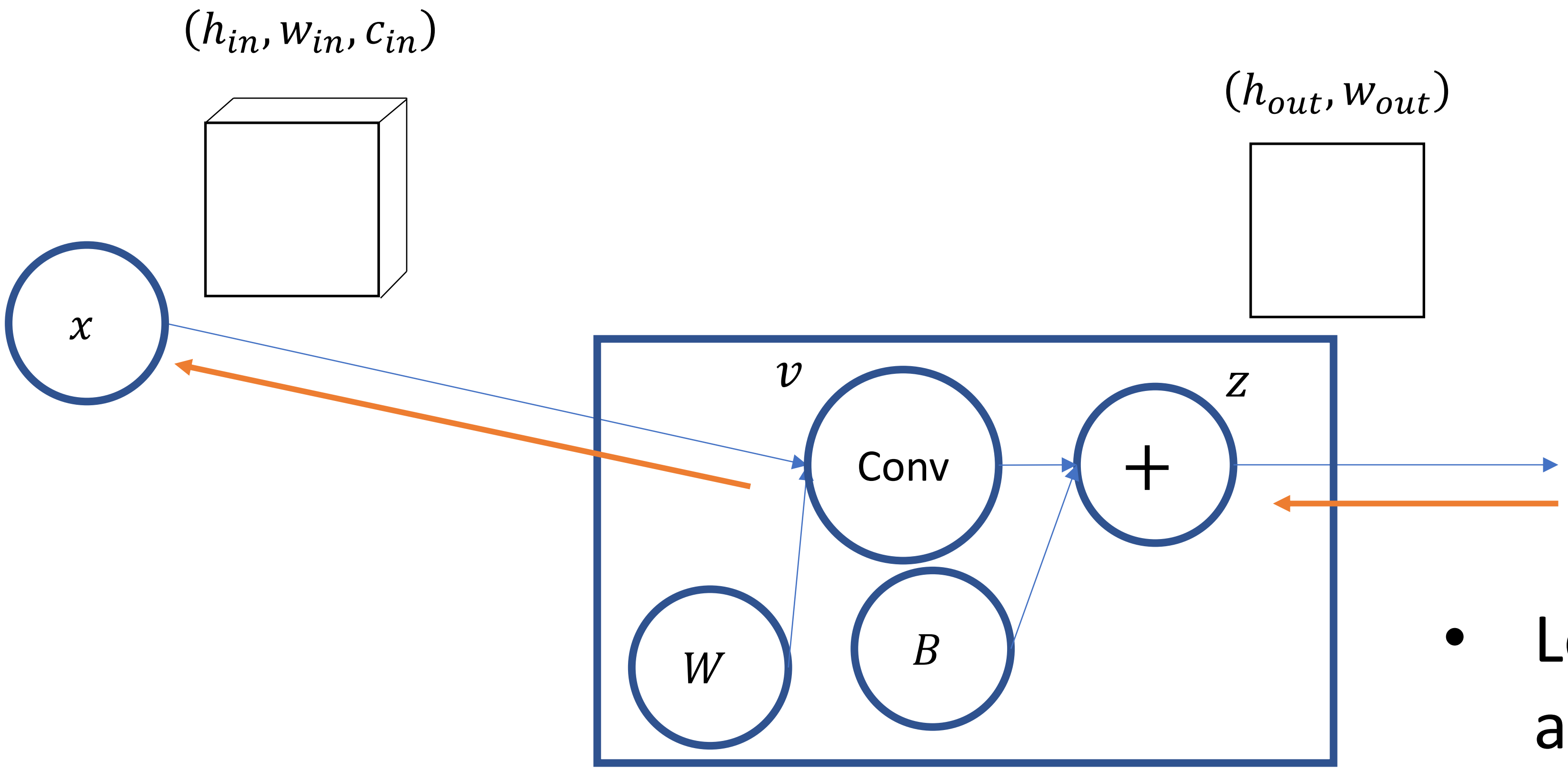
A compute graph for a convolutional layer



We know how to compute all the gradients now for:

- One sample (m=1)
- One filter (K=1)
- ~~One channel~~ $\cancel{(c_{in} = 1)}$

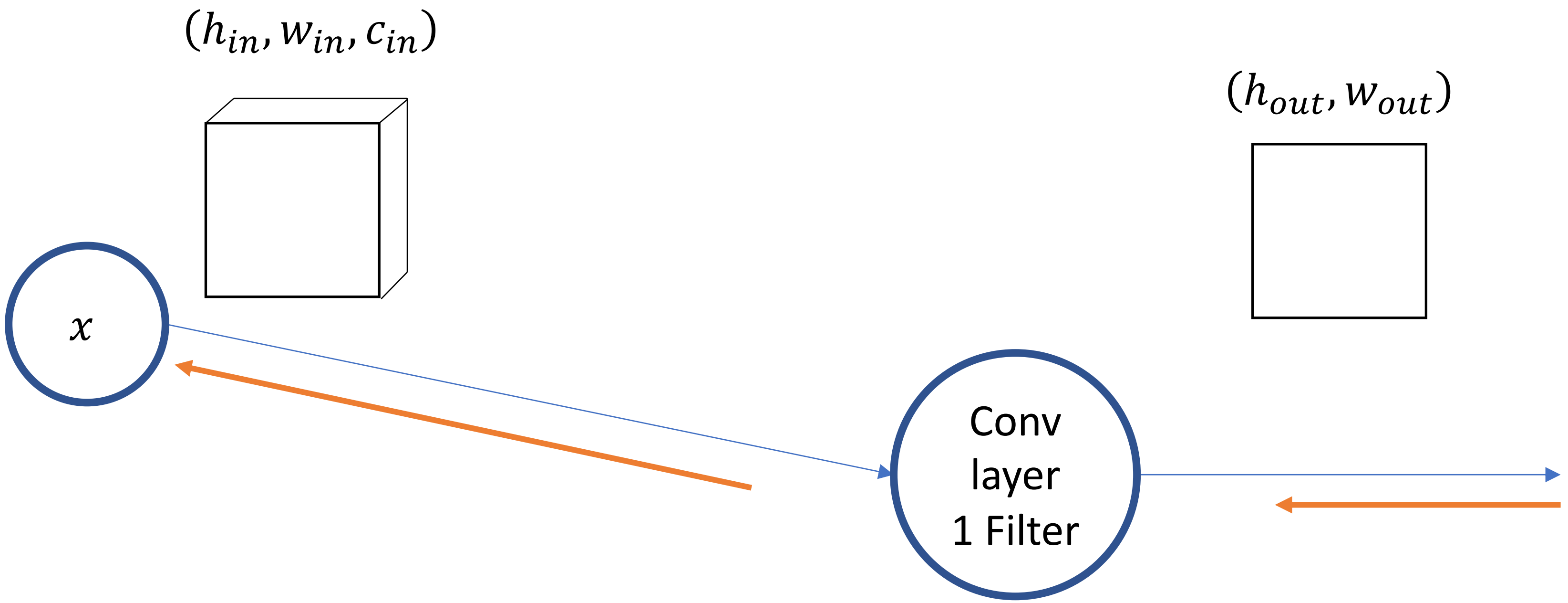
A compute graph for a convolutional layer



- Let's abstract all this part away for now since we know how to do this part

Brad Quinton, Scott Chin

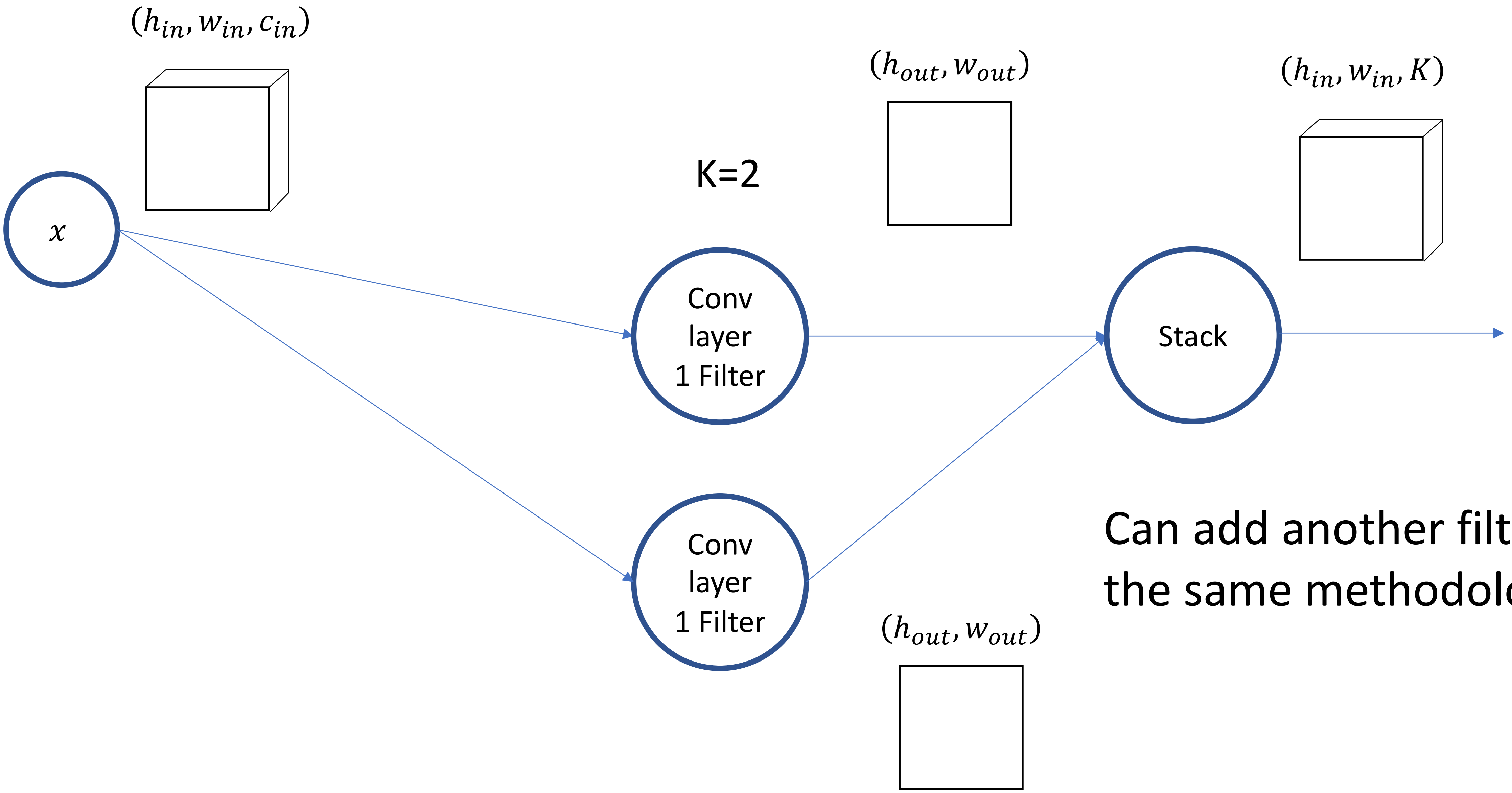
A compute graph for a convolutional layer



- Let’s abstract all this part away for now since we know how to do this part
- Model it as a single compute graph node

Brad Quinton, Scott Chin

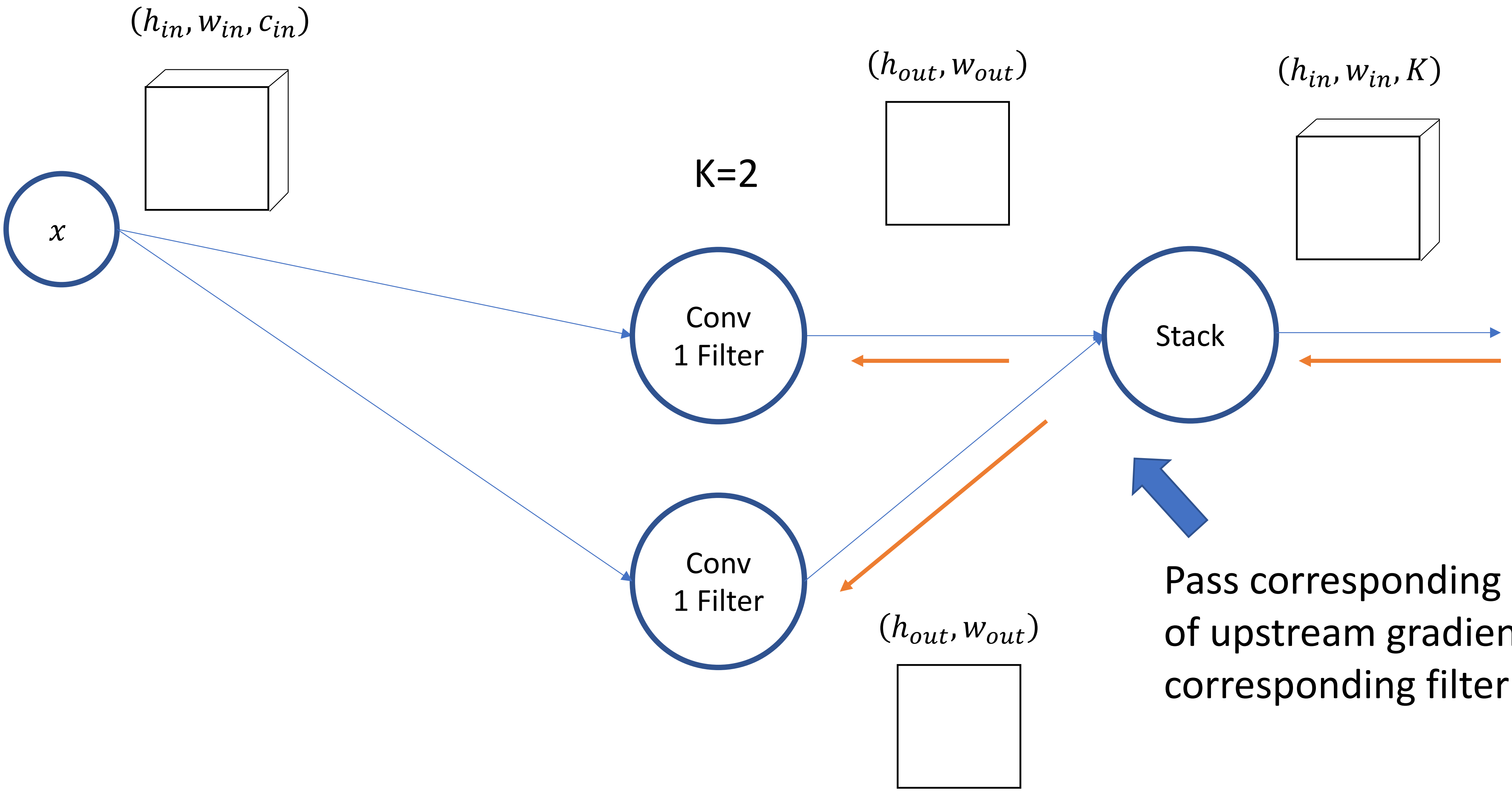
A compute graph for a convolutional layer



Can add another filter using the same methodology

Brad Quinton, Scott Chin

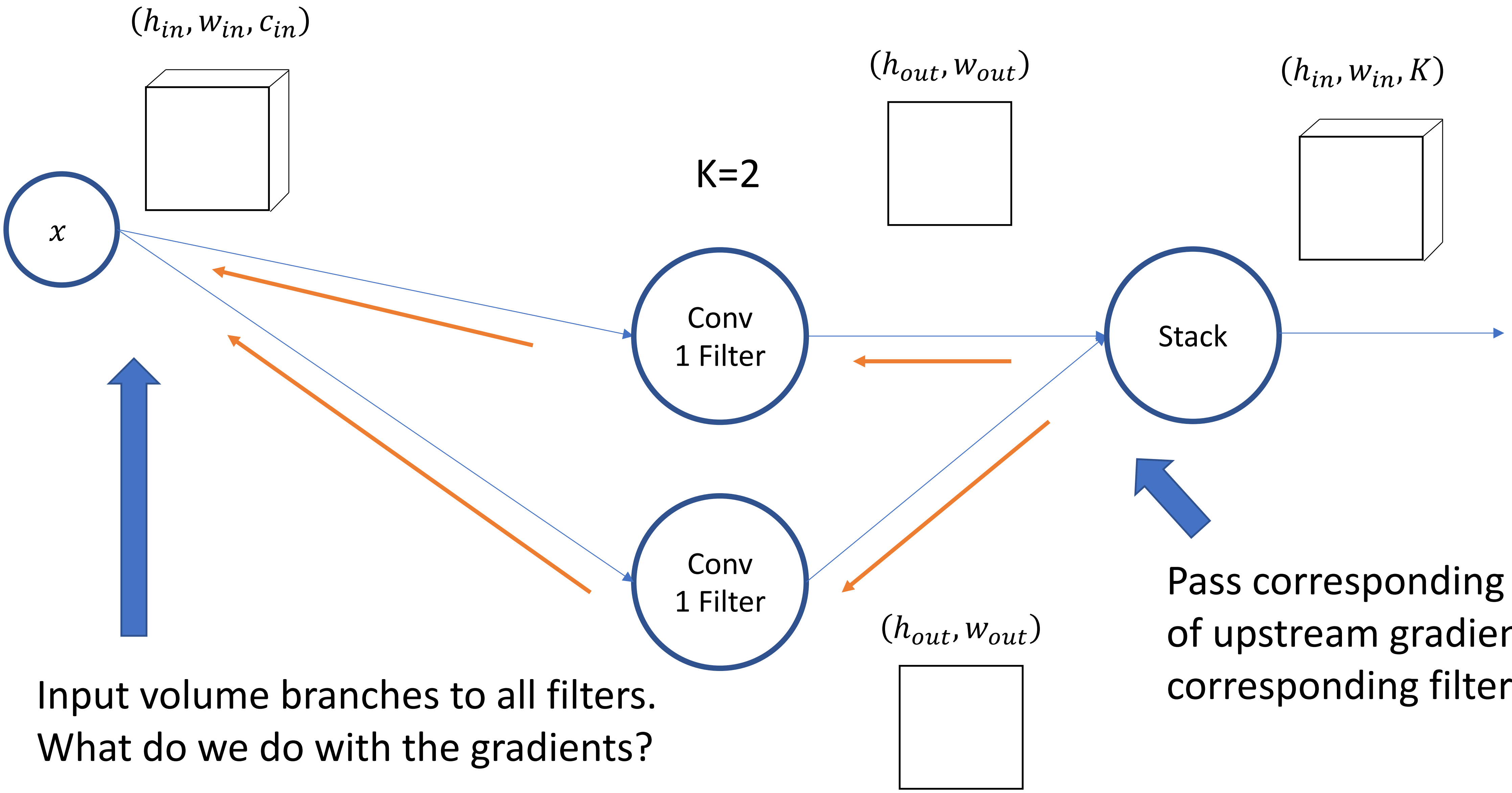
A compute graph for a convolutional layer



Pass corresponding channel of upstream gradients to corresponding filter

Brad Quinton, Scott Chin

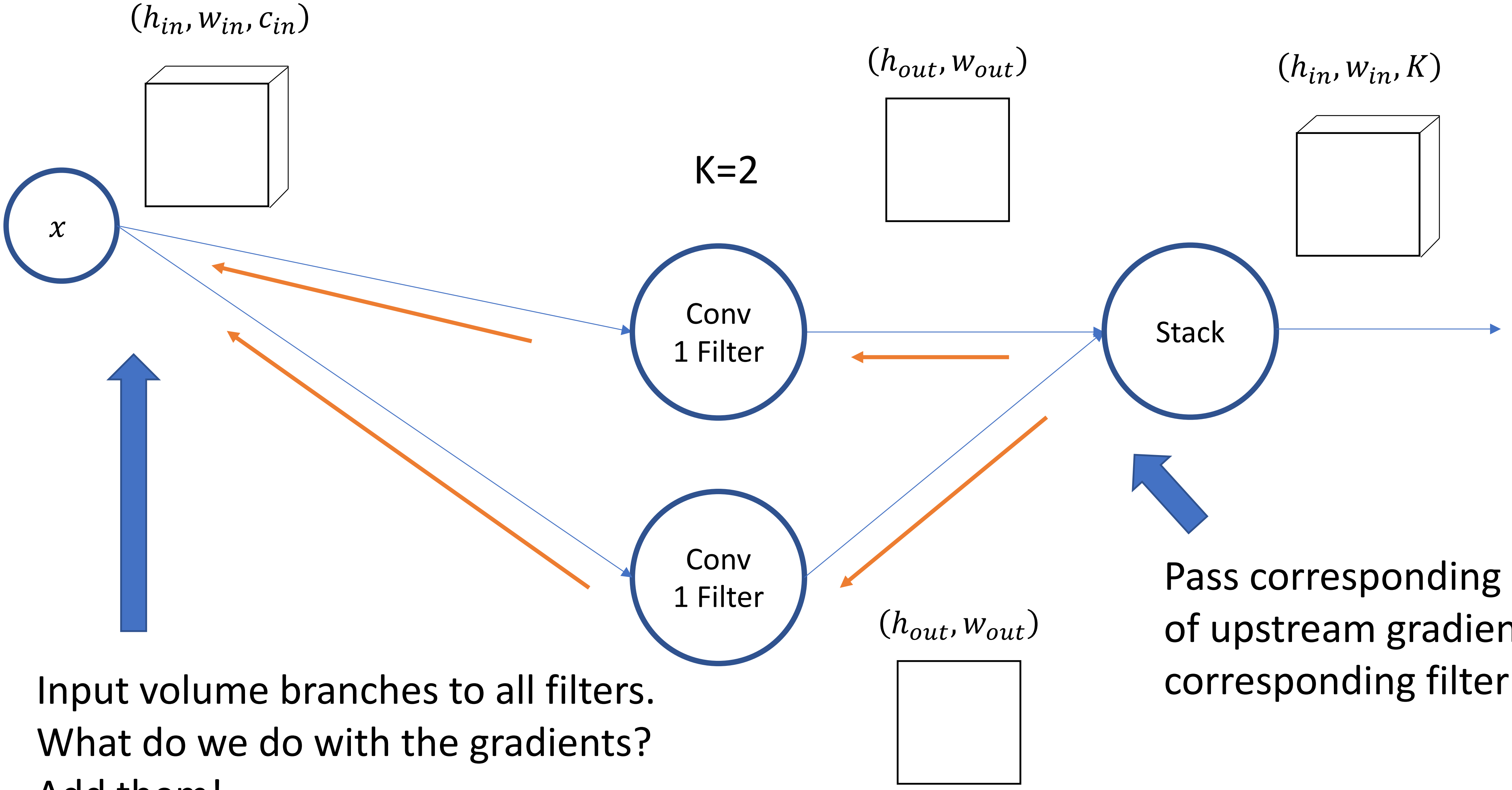
A compute graph for a convolutional layer



Input volume branches to all filters.
What do we do with the gradients?

Pass corresponding channel of upstream gradients to corresponding filter

A compute graph for a convolutional layer



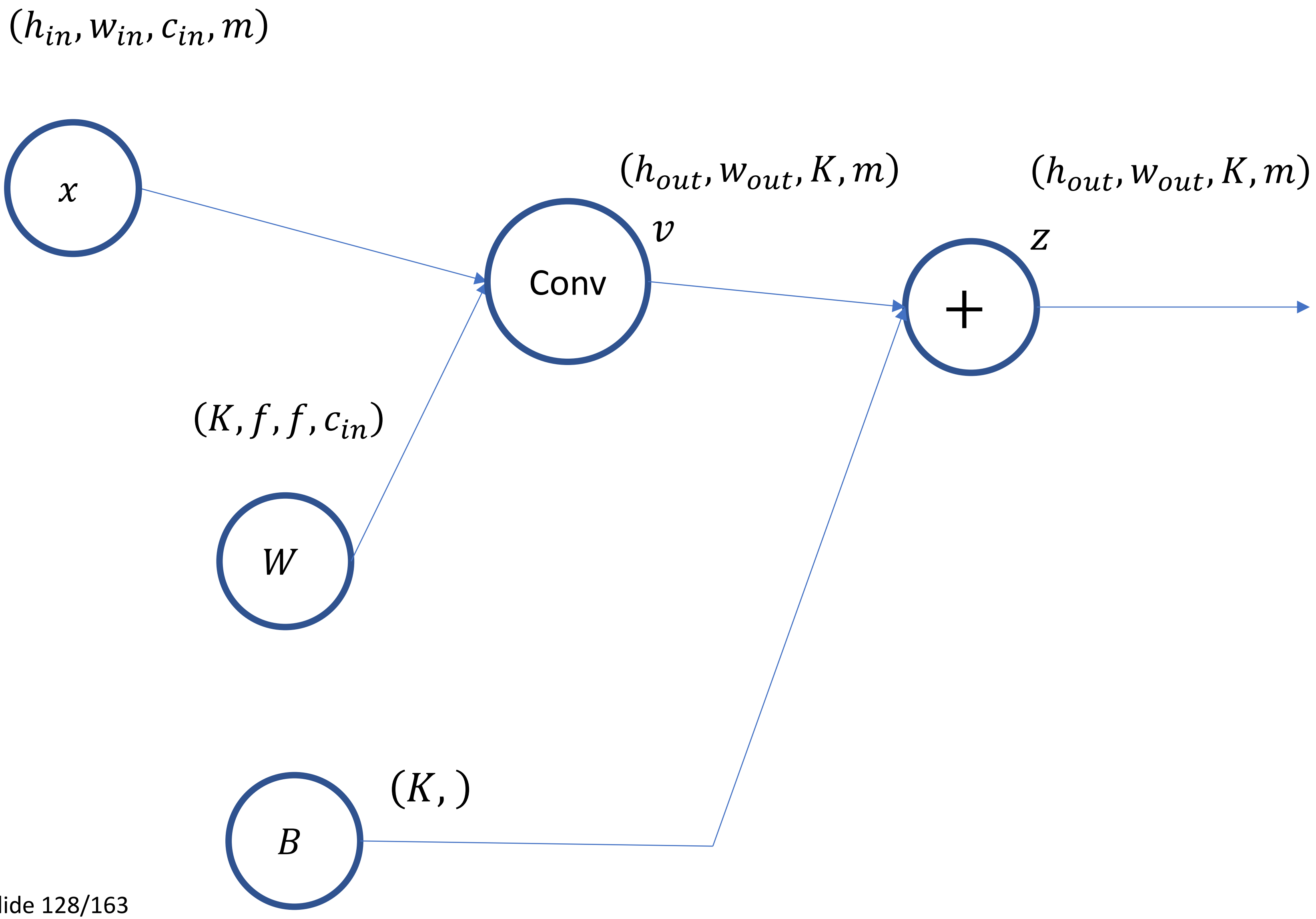
Pass corresponding channel of upstream gradients to corresponding filter

Input volume branches to all filters.
What do we do with the gradients?
Add them!

For Multiple Samples

Brad Quinton, Scott Chin

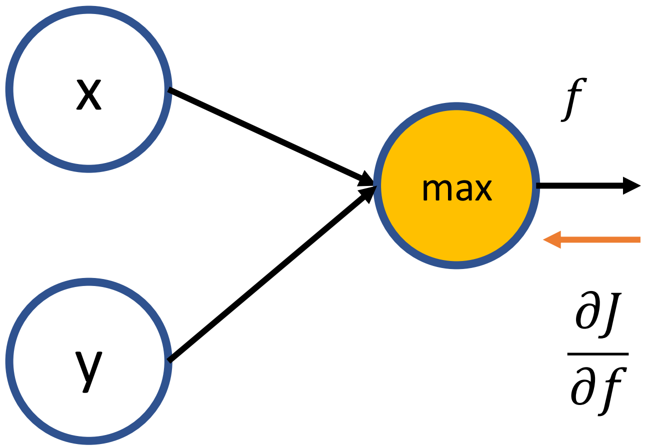
A compute graph for a convolutional layer



Brad Quinton, Scott Chin

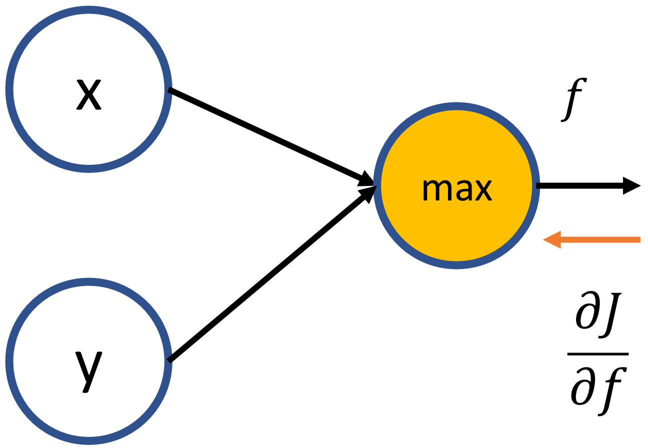
Backpropagation Through MaxPooling

Recall: Backpropagation Through max(x,y)



Brad Quinton, Scott Chin

Recall: Backpropagation Through max(x,y)



- Upstream gradient is routed to larger variable
- Intuition: Only one input can affect the output at any time

Backpropgation in MaxPooling Layer

- Recall that MaxPool, you are taking the max in small regions of the input volume (e.g. 2x2 regions)

1	3	10	0
8	4	2	9
7	2	1	3
1	1	0	4

<u>Max Pooling</u>

$(f, f) = (2,2)$

Stride: 2

→

8	10
7	4

Brad Quinton, Scott Chin

Backpropgation in MaxPooling Layer

- Recall that MaxPool, you are taking the max in small regions of the input volume (e.g. 2x2 regions)

1	3	10	0
8	4	2	9
7	2	1	3
1	1	0	4

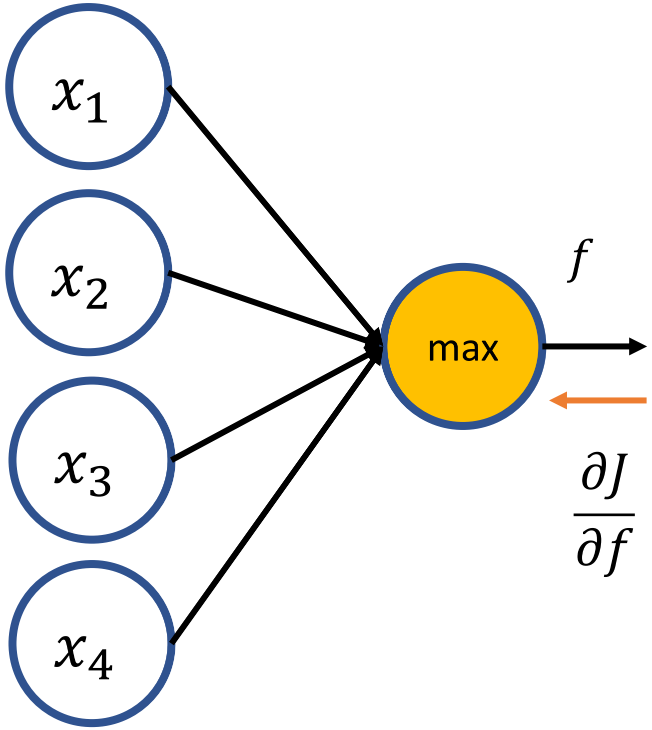
<u>Max Pooling</u>

$(f, f) = (2,2)$

Stride: 2

→

8	10
7	4



Brad Quinton, Scott Chin

Aside: Adversarial Inputs to CNNs

Brad Quinton, Scott Chin

Adversarial Images via Backpropagation

- Pick an input image to modify
- Pick an output class you want to ”trick” the classifier into predicting
- Use a cost function that maximizes that class’ output probability
- Use backpropagation to find changes to the input image (dJ/dx) to maximize cost (i.e.. use gradient ascent)

Brad Quinton, Scott Chin

Adversarial Images via Backpropagation

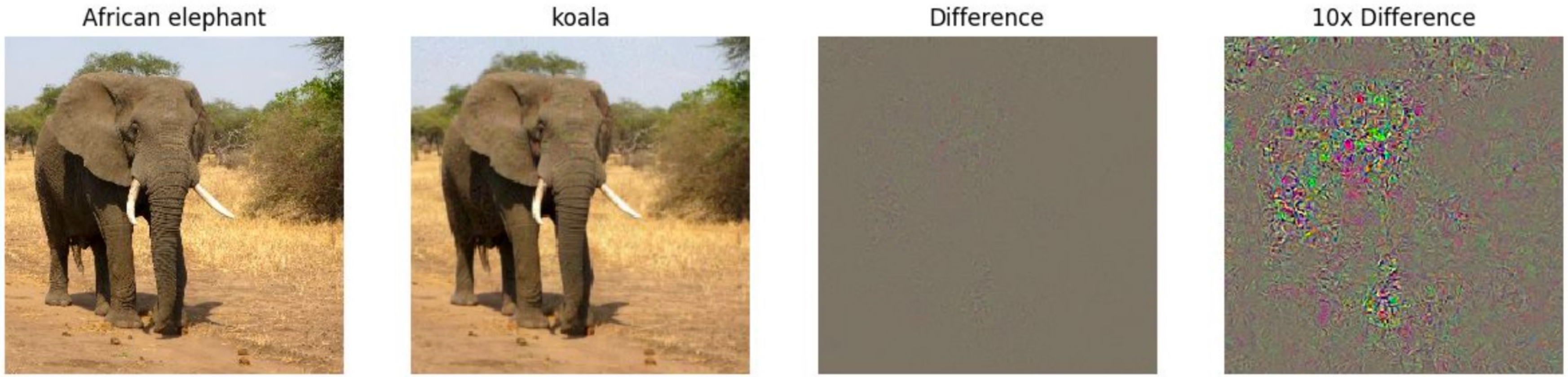


Image Source: CS231n Stanford Course

Adversarial Images via Backpropagation

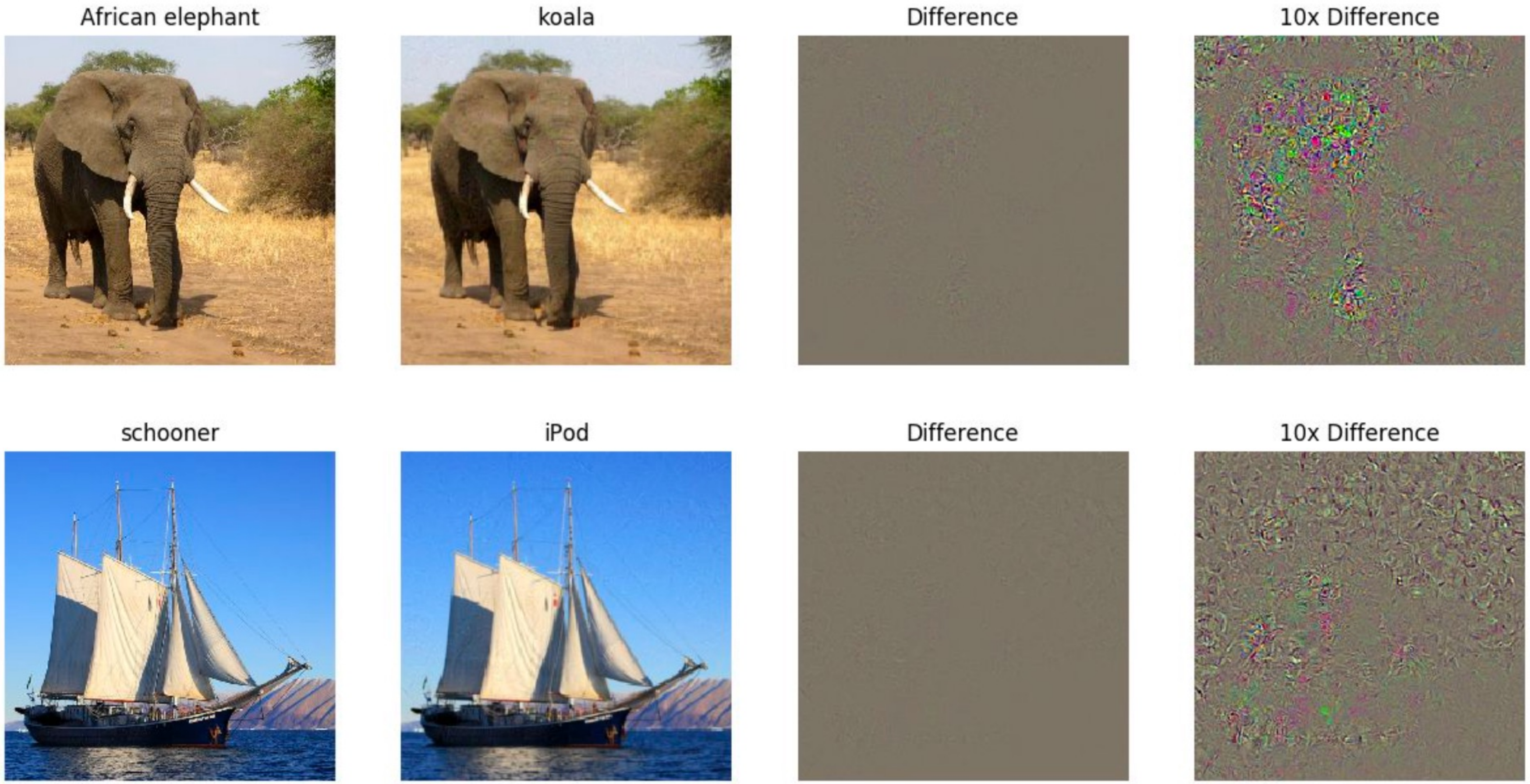
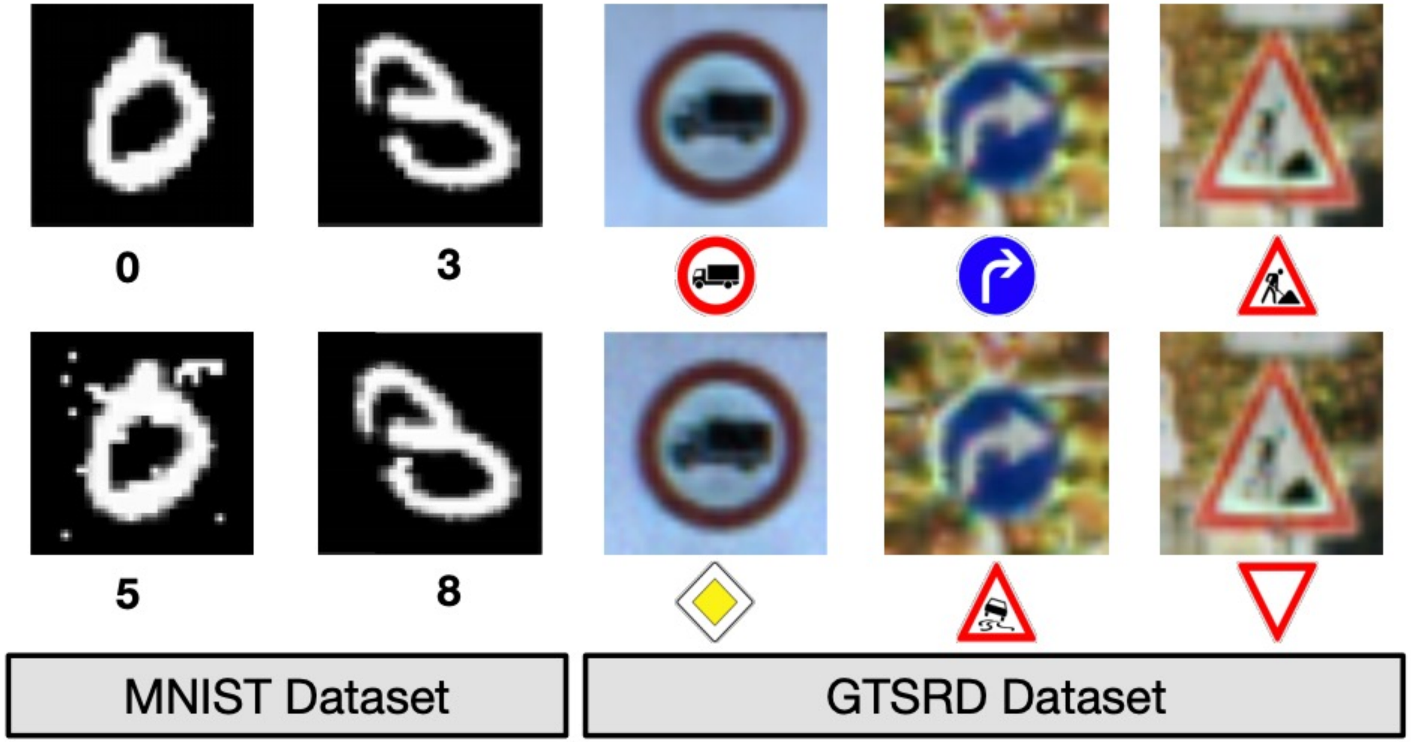


Image Source: CS231n Stanford Course

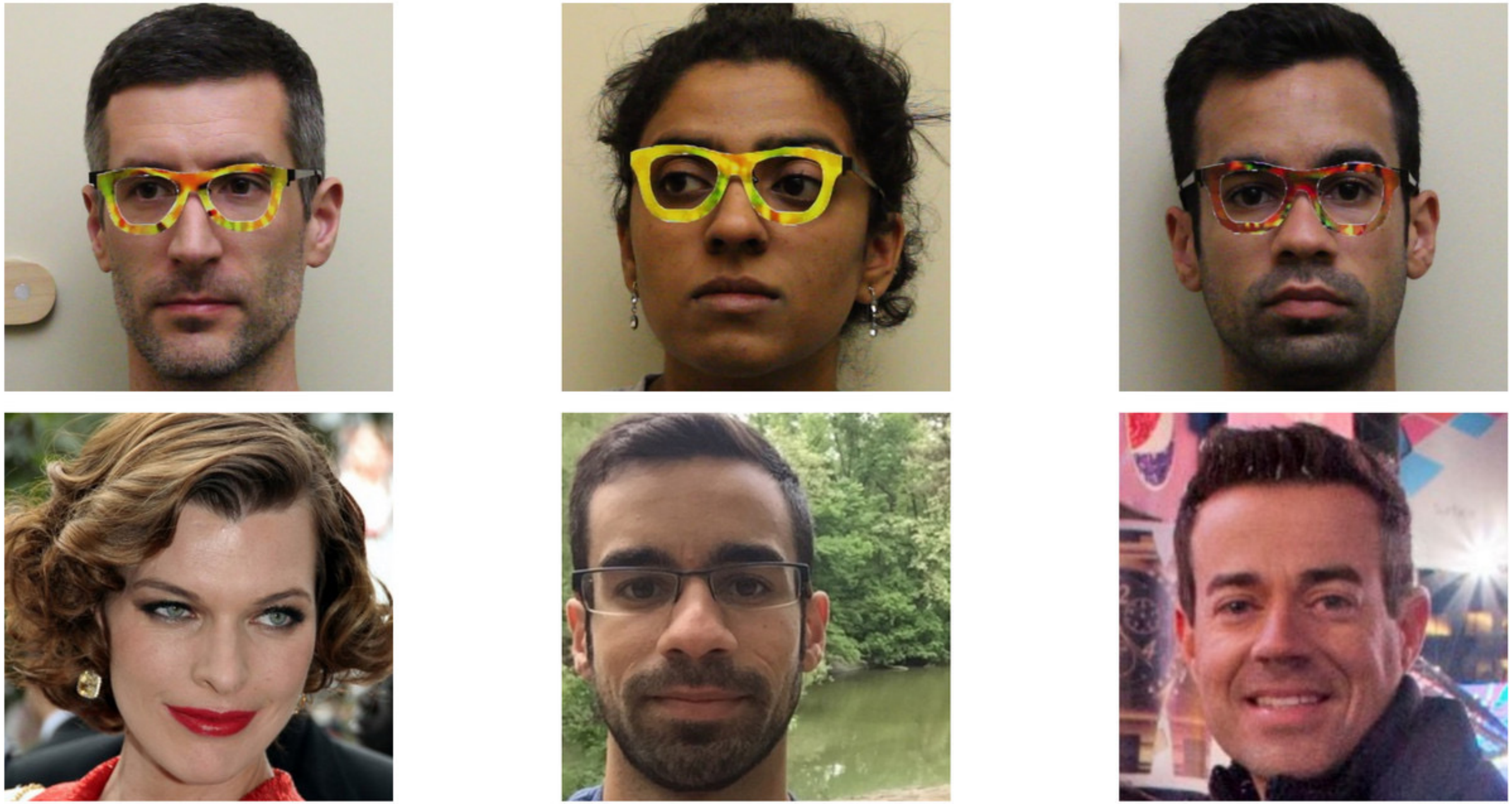
Adversarial Images via Backpropagation



“Practical Black-Box Attacks against Machine Learning”, Papernot et al, 2017, https://arxiv.org/abs/1602.02697

- These were misclassified as a 45mph/hr street sign instead of stop sign

https://arstechnica.com/cars/2017/09/hacking-street-signs-with-stickers-could-confuse-self-driving-cars/



(b) (c) (d)

https://www.theverge.com/2016/11/3/13507542/facial-recognition-glasses-trick-impersonate-fool

Numerical Gradient Checking

- As you saw in assignment 3, easy to make mistakes when implementing backpropagation
- In software development, we want to write tests for our code
- For backprop tests, we need “correct” values
- Computing these is tedious and also error-prone

Test from Assignment 3 for Z=WX+B

```python
# Expected Values
dJ_dx_expected = np.array([[1.32727273, 1.        ],
                           [1.21818182, 1.10909091],
                           [1.10909091, 1.21818182],
                           [1.        , 1.32727273]])
dJ_dw_expected = np.array([[ 1.42857143,  0.51428571, -0.4       , -1.31428571],
                           [ 0.05714286,  0.05714286,  0.05714286,  0.05714286],
                           [-1.31428571, -0.4       ,  0.51428571,  1.42857143]])
dJ_db_expected = np.array([[-1.6],
                           [ 2.22044605e-16],
                           [ 1.6]])

# Test gradient matrix values
assert np.allclose(dJ_dx, dJ_dx_expected), 'Unexpected values for dJ_dx: \n{0}'.format(dJ_dx)
assert np.allclose(dJ_dw, dJ_dw_expected), 'Unexpected values for dJ_dw: \n{0}'.format(dJ_dw)
assert np.allclose(dJ_db, dJ_db_expected), 'Unexpected values for dJ_db: \n{0}'.format(dJ_db)
```


Brad Quinton, Scott Chin

Numerical Gradient Checking

- As you saw in assignment 3, easy to make mistakes when implementing backpropagation
- In software development, we want to write tests for our code
- For backprop tests, we need “correct” values
- Computing these is tedious and also error-prone
- We can instead use Numerical Gradient Checking

Recall from Calculus 101

- Definition of the derivative:

$$\frac{df}{dx} = \lim_{h \to 0} \frac{f(x+h) - f(x-h)}{2h}$$

Brad Quinton, Scott Chin

Recall from Calculus 101

- Definition of the derivative:

$$\frac{df}{dx} = \lim_{h \to 0} \frac{f(x+h) - f(x-h)}{2h}$$

<u>Numerical Gradient Approximation:</u>

- When h is not zero, but small (e.g. 0.00001), we can get a decent approximation to the derivative

Brad Quinton, Scott Chin

For a multivariable function

- y is a scalar
- x is a vector shaped $(n,)$
- $y = f(x)$
- $\frac{\partial y}{\partial x}$ is shaped $(n,)$

Brad Quinton, Scott Chin

For a multivariable function

- y is a scalar
- x is a vector shaped $(n,)$
- $y = f(x)$
- $\frac{\partial y}{\partial x}$ is shaped $(n,)$

$$\frac{\partial y}{\partial x_1} \approx \frac{f(x_1 + h, x_2, \dots, x_n) - f(x_1 - h, x_2, \dots, x_n)}{2h}$$

$$\frac{\partial y}{\partial x_2} \approx \frac{f(x_1, x_2 + h, \dots, x_n) - f(x_1, x_2 - h, \dots, x_n)}{2h}$$

$$\vdots$$

$$\frac{\partial y}{\partial x_n} \approx \frac{f(x_1, x_2, \dots, x_n + h) - f(x_1, x_2, \dots, x_n - h)}{2h}$$

Brad Quinton, Scott Chin

Example

- Let's say: $y = x_1 * x_2 * x_3$ and want gradient at $(x_1, x_2, x_3) = (2,1,3)$

Brad Quinton, Scott Chin

Example – Using "Analytical" Gradient

- Let's say: $y = x_1 * x_2 * x_3$ and want gradient at $(x_1, x_2, x_3) = (2,1,3)$
- Derive the closed-form symbolic derivatives:

$$\frac{\partial y}{\partial x_1} = x_2 * x_3$$

$$\frac{\partial y}{\partial x_2} = x_1 * x_3$$

$$\frac{\partial y}{\partial x_3} = x_1 * x_2$$

Brad Quinton, Scott Chin

Example – Using "Analytical" Gradient

- Let's say: $y = x_1 * x_2 * x_3$ and want gradient at $(x_1, x_2, x_3) = (2,1,3)$
- Derive the closed-form symbolic derivatives:

$$\frac{\partial y}{\partial x_1} = x_2 * x_3 = 1 * 3 = 3$$

$$\frac{\partial y}{\partial x_2} = x_1 * x_3 = 2 * 3 = 6$$

$$\frac{\partial y}{\partial x_3} = x_1 * x_2 = 2 * 1 = 2$$

Example – Using Numerical Gradient Approx.

- Let's say: $y = x_1 * x_2 * x_3$ and want gradient at $(x_1, x_2, x_3) = (2,1,3)$
- Using numerical gradient approximation (h=0.00001)

$$\frac{\partial y}{\partial x_1} \approx \frac{(x_1+h) * x_2 * x_3 - (x_1-h) * x_2 * x_3}{2h}$$

$$\frac{\partial y}{\partial x_2} \approx \frac{x_1 * (x_2 + h) * x_3 - x_1 * (x_2 - h) * x_3}{2h}$$

$$\frac{\partial y}{\partial x_3} \approx \frac{x_1 * x_2 * (x_3 + h) - x_1 * x_2 * (x_3 - h)}{2h}$$

Brad Quinton, Scott Chin

Example – Using Numerical Gradient Approx.

- Let's say: $y = x_1 * x_2 * x_3$ and want gradient at $(x_1, x_2, x_3) = (2,1,3)$
- Using numerical gradient approximation (h=0.00001)

$$\frac{\partial y}{\partial x_1} \approx \frac{(x_1+h) * x_2 * x_3 - (x_1-h) * x_2 * x_3}{2h} = 3$$

$$\frac{\partial y}{\partial x_2} \approx \frac{x_1 * (x_2 + h) * x_3 - x_1 * (x_2 - h) * x_3}{2h} = 6$$

$$\frac{\partial y}{\partial x_3} \approx \frac{x_1 * x_2 * (x_3 + h) - x_1 * x_2 * (x_3 - h)}{2h} = 2$$

Brad Quinton, Scott Chin

Example – Compare

- Let's say: $y = x_1 * x_2 * x_3$ and want gradient at $(x_1, x_2, x_3) = (2,1,3)$

$$\frac{\partial y}{\partial x_1} \approx \frac{(x_1+h) * x_2 * x_3 - (x_1-h) * x_2 * x_3}{2h} = 3 \qquad \frac{\partial y}{\partial x_1} = x_2 * x_3 = 1 * 3 = 3$$

$$\frac{\partial y}{\partial x_2} \approx \frac{x_1 * (x_2 + h) * x_3 - x_1 * (x_2 - h) * x_3}{2h} = 6 \qquad \frac{\partial y}{\partial x_2} = x_1 * x_3 = 2 * 3 = 6$$

$$\frac{\partial y}{\partial x_3} \approx \frac{x_1 * x_2 * (x_3 + h) - x_1 * x_2 * (x_3 - h)}{2h} = 2 \qquad \frac{\partial y}{\partial x_3} = x_1 * x_2 = 2 * 1 = 2$$

Gradient Check For Backpropagation

1. Need to implement these from Assignment 3 (Z=WX+B)
   z = linear_forward_propagate(x, w, b)
   dJ_dw, dJ_b, dJ_dx = linear_backward_propagate(dJ_dz)
2. You typically have the forward prop symbolic equations, so computing expected values for tests is straight-forward
3. So assuming you have correctly implemented the forward prop function, you can use it for numerical gradient checking for your backprop function

Brad Quinton, Scott Chin

Example – Backprop (Scalar)

Forward Propagation:

$$z = f(w_2, x_2, w_1, x_1, b) = \; w_2 x_2 + w_1 x_1 + b$$

```python
def f(x2, x1, w2, w1, b):
    z = w2*x2 + w1*x1 + b
    return z
```

Backward Propagation:

$$\frac{\partial z}{\partial w_2} = \; x_2 \qquad \frac{\partial z}{\partial x_2} = \; w_2$$

$$\frac{\partial z}{\partial w_1} = \; x_1 \qquad \frac{\partial z}{\partial x_1} = w_1$$

$$\frac{\partial z}{\partial b} = 1$$

```python
def df(x2, x1, w2, w1, b):
    dz_dx2 = w2
    dz_dx1 = w1
    dz_dw2 = x2
    dz_dw1 = x1
    dz_db  = 1
    return dz_dx2, dz_dx1, dz_dw2, dz_dw1, dz_db
```


Brad Quinton, Scott Chin

Example – Backprop (Scalar)

Forward Propagation:

$$z = f(w_2, x_2, w_1, x_1, b) = w_2x_2 + w_1x_1 + b$$

```python
def f(x2, x1, w2, w1, b):
    z = w2*x2 + w1*x1 + b
    return z
```

Test: Is this correct?

Backward Propagation:

$$\frac{\partial z}{\partial w_2} = x_2 \qquad \frac{\partial z}{\partial x_2} = w_2$$

$$\frac{\partial z}{\partial w_1} = x_1 \qquad \frac{\partial z}{\partial x_1} = w_1$$

$$\frac{\partial z}{\partial b} = 1$$

```python
def df(x2, x1, w2, w1, b):
    dz_dx2 = w2
    dz_dx1 = w1
    dz_dw2 = x2
    dz_dw1 = x1
    dz_db  = 1
    return dz_dx2, dz_dx1, dz_dw2, dz_dw1, dz_db
```


Brad Quinton, Scott Chin

```python
# For some set of inputs
x2, x1, w2, w1, b = np.random.randn(5)

# Compute analytical gradient
dz_dx2, dz_dx1, dz_dw2, dz_dw1, dz_db = df(x2, x1, w2, w1, b)
```

Are these 5 numbers correct?

Brad Quinton, Scott Chin

```python
# For some set of inputs
x2, x1, w2, w1, b = np.random.randn(5)

# Compute analytical gradient
dz_dx2, dz_dx1, dz_dw2, dz_dw1, dz_db = df(x2, x1, w2, w1, b)
```

Are these 5 numbers correct?

Example: Checking one of the results dz_dx2

```python
# Compute numerical gradient
h = 1e-5
dz_dx2_num = (f(x2+h, x1, w2, w1, b) - f(x2-h, x1, w2, w1, b))/(2*h)
print('dz_dx2 numerical: {0:0.14f} analytic: {1:0.14f} diff: {2}'.format(
    dz_dx2_num, dz_dx2, dz_dx2_num-dz_dx2))
```

```
dz_dx2 numerical: 0.97873798412529 analytic: 0.97873798410574

diff: 1.9549251106809606e-11
```

```python
# For some set of inputs
x2, x1, w2, w1, b = np.random.randn(5)

# Compute analytical gradient
dz_dx2, dz_dx1, dz_dw2, dz_dw1, dz_db = df(x2, x1, w2, w1, b)
```

Are these 5 numbers correct?

Example: Checking one of the results dz_dx2

```python
# Compute numerical gradient
h = 1e-5
dz_dx2_num = (f(x2+h, x1, w2, w1, b) - f(x2-h, x1, w2, w1, b))/(2*h)
print('dz_dx2 numerical: {0:0.14f} analytic: {1:0.14f} diff: {2}'.format(
    dz_dx2_num, dz_dx2, dz_dx2_num-dz_dx2))
```

```
dz_dx2 numerical: 0.97873798412529 analytic: 0.97873798410574
diff: 1.954925106809606e-11
```

Check that difference is within some tolerance

```
dz_dx2_num = (f(x2+h, x1, w2, w1, b) - f(x2-h, x1, w2, w1, b))/(2*h)
dz_dx1_num = (f(x2, x1+h, w2, w1, b) - f(x2, x1-h, w2, w1, b))/(2*h)
dz_dw2_num = (f(x2, x1, w2+h, w1, b) - f(x2, x1, w2-h, w1, b))/(2*h)
dz_dw1_num = (f(x2, x1, w2, w1+h, b) - f(x2, x1, w2, w1-h, b))/(2*h)
dz_db_num  = (f(x2, x1, w2, w1, b+h) - f(x2, x1, w2, w1, b-h))/(2*h)
```

```
dz_dx2 numerical: 0.97873798412529 analytic: 0.97873798410574 diff: 1.9549251106809606e-11
dz_dx1 numerical: 2.24089319917908 analytic: 2.24089319920146 diff: -2.2374102570665855e-11
dz_dw2 numerical: 1.76405234597610 analytic: 1.76405234596766 diff: 8.432810005842839e-12
dz_dw1 numerical: 0.40015720834674 analytic: 0.40015720836722 diff: -2.0482504581309513e-11
dz_db  numerical: 1.00000000000655 analytic: 1.00000000000000 diff: 6.551204023708124e-12
```

```python
def df(x2, x1, w2, w1, b):
    dz_dx2 = w2*2
    dz_dx1 = w1
    dz_dw2 = x2
    dz_dw1 = x1
    dz_db  = 1
    return dz_dx2, dz_dx1, dz_dw2, dz_dw1, dz_db
```

Say I made a mistake in my analytic gradient implementation

Correct:

`dz_dx2 numerical: 0.97873798412529 analytic: 0.97873798410574`

`diff: 1.9549251106809606e-11`

With Mistake:

`dz_dx2 numerical: 0.97873798412529 analytic: 1.95747596821148`

`diff: -0.97873798408619`

Brad Quinton, Scott Chin

Why not just always use this for backprop?

- Too slow
- You need to make two calls to your forward prop function for each gradient value.

Learning Objectives

- More practice with backpropagation
- Understand backprop being applied to convolutional layer
- See some of the “shortcuts” in convolutional layer backprop
- Understand use of numerical gradient