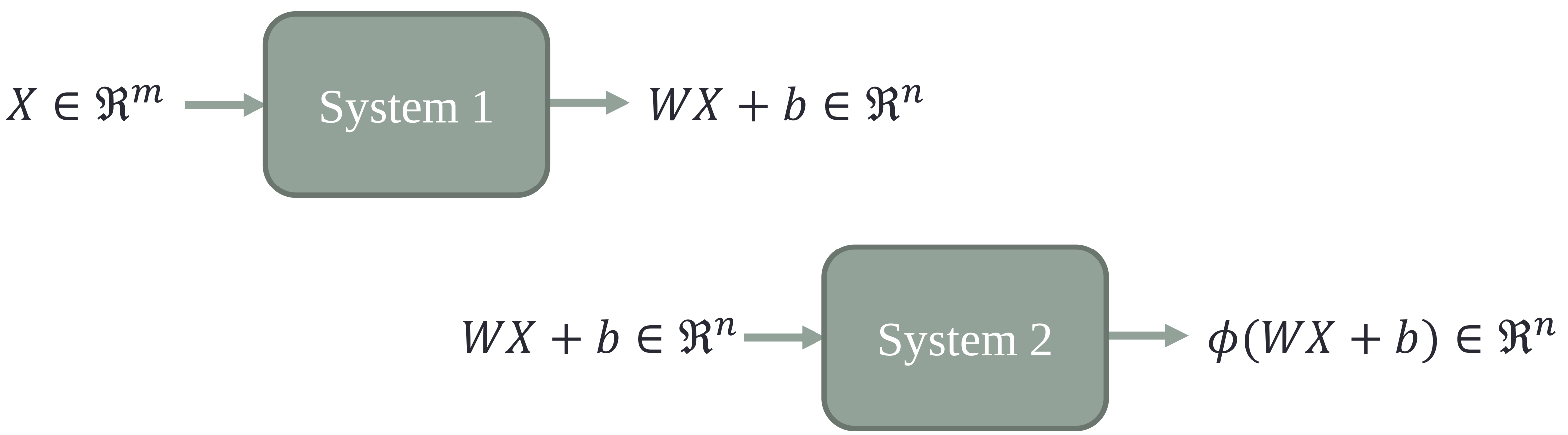
Backpropagation in CNNs

HYUNG IL KOO

General representation

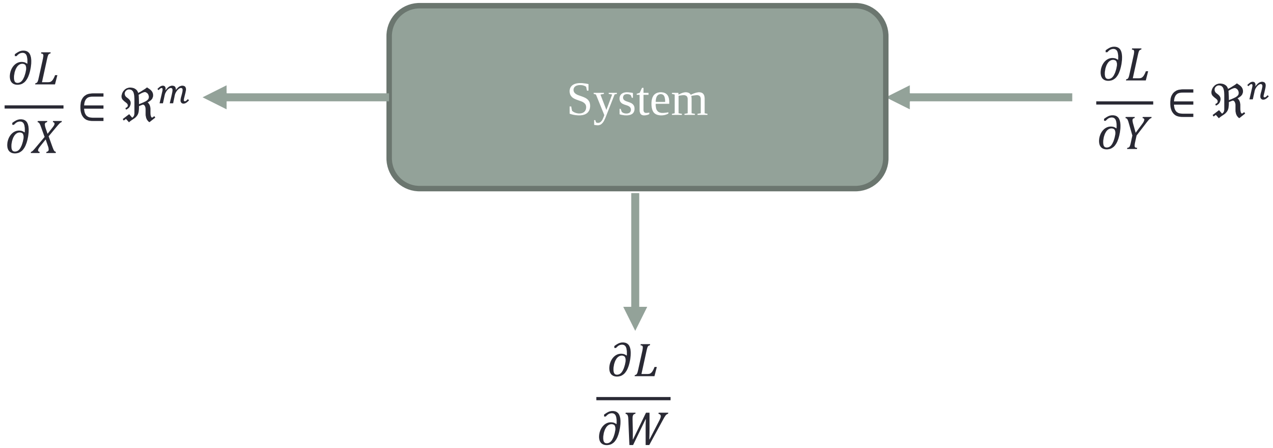
- $Y = \phi(WX + b)$, $\phi(\cdot)$: element-wise nonlinear function



Backpropagation Algorithm - remind

Forward pass:

Backward pass:



Back Propagation as a matrix multiplication

$Y = WX + b$ case

Forward pass:

$$X \in \Re^{m} \rightarrow \boxed{\begin{array}{c} W \\ \text{multiplication} \end{array}} \rightarrow Y = WX + b \in \Re^{n}$$

Backward pass:

$$\left(\frac{\partial L}{\partial X}\right)^{\top} \in \Re^{m} \leftarrow \boxed{\begin{array}{c} W^{T} \\ \text{multiplication} \end{array}} \leftarrow \left(\frac{\partial L}{\partial Y}\right)^{\top} \in \Re^{n}$$

$$\frac{\partial L}{\partial W} \in \Re^{m \times n} \leftarrow \boxed{\begin{array}{c} X \\ \text{multiplication} \end{array}} \leftarrow \frac{\partial L}{\partial Y} \in \Re^{n}$$

$$\frac{\partial L}{\partial b} = \frac{\partial L}{\partial Y} \in \Re^{n}$$

Intuitive derivation

- Want to find: $\Delta L \simeq \boldsymbol{\frac{\partial L}{\partial X}} \Delta X$
  - We know
    - $\Delta L \simeq \frac{\partial L}{\partial Y} \Delta Y$
    - $\Delta Y \simeq \frac{\partial Y}{\partial X} \Delta X = W \Delta X$
  - Hence,
    - $\Delta L \simeq \frac{\partial L}{\partial Y} \Delta Y = \frac{\partial L}{\partial Y} W \Delta X$
  - Finally, putting them all together,
    - $\frac{\partial L}{\partial X} = \frac{\partial L}{\partial Y} W$
    - $\left(\frac{\partial L}{\partial X}\right)^{\top} = W^{\top} \left(\frac{\partial L}{\partial Y}\right)^{\top}$

Intuitive derivation

- Want to find: $\Delta L \simeq \operatorname{tr}\left(\frac{\partial L}{\partial W} \Delta W\right)$
  - We know
    - $\Delta L \simeq \frac{\partial L}{\partial Y} \Delta Y$
    - $\Delta Y \simeq \Delta \mathrm{W} X$
  - Hence,
    - $\Delta L \simeq \frac{\partial L}{\partial Y} \Delta Y = \frac{\partial L}{\partial Y} \Delta W X = \operatorname{tr}\left(\frac{\partial L}{\partial Y} \Delta W X\right) = \operatorname{tr}\left(X \frac{\partial L}{\partial Y} \Delta W\right)$
  - Finally, putting them all together,
    - $\frac{\partial L}{\partial W} = X \frac{\partial L}{\partial Y}$
    - $\left(\frac{\partial L}{\partial W}\right)^{\top} = \left(\frac{\partial L}{\partial Y}\right)^{\top} X^{\top}$

$Y = \phi(X)$ case

Forward pass: $X = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ \vdots \\ x_m \end{bmatrix} \in \Re^m \xrightarrow{} \begin{array}{c} W \\ \text{multiplication} \end{array} \xrightarrow{} Y = \phi(X) = \begin{bmatrix} \phi(x_1) \\ \phi(x_2) \\ \vdots \\ \vdots \\ \phi(x_m) \end{bmatrix} \in \Re^m$

Backward pass: $\left(\frac{\partial L}{\partial X}\right)^{\top} \in \Re^m \xleftarrow{} \begin{array}{c} W^T \\ \text{multiplication} \end{array} \xleftarrow{} \left(\frac{\partial L}{\partial Y}\right)^{\top} \in \Re^m$

$$\left(\frac{\partial L}{\partial X}\right)^{\top} = \begin{bmatrix} \phi'(x_1)\frac{\partial L}{\partial y_1} \\ \phi'(x_2)\frac{\partial L}{\partial y_2} \\ \vdots \\ \vdots \\ \phi'(x_m)\frac{\partial L}{\partial y_m} \end{bmatrix} \in \Re^m$$

Back Propagation as a Convolution operation

Backpropagation in Conv. layer

- Idea
  - To represent a convolution layer with a matrix multiplication
  - Taking a matrix transpose yields the backpropagation algorithm for the convolution layer

Forward pass:

$$\begin{array}{|c|c|c|}\hline a & b & c \\ \hline d & e & f \\ \hline g & h & i \\ \hline \end{array} * \begin{array}{|c|c|c|}\hline w_{11} & w_{12} & w_{13} \\ \hline w_{21} & w_{22} & w_{23} \\ \hline w_{31} & w_{32} & w_{33} \\ \hline \end{array} = \begin{array}{|c|c|c|}\hline \alpha & \beta & \gamma \\ \hline \delta & \epsilon & \zeta \\ \hline \eta & \theta & \kappa \\ \hline \end{array}$$

$$\begin{pmatrix} \alpha \\ \beta \\ \gamma \\ \delta \\ \epsilon \\ \zeta \\ \eta \\ \theta \\ \kappa \end{pmatrix} = \begin{pmatrix} w_{22} & w_{23} & 0 & w_{32} & w_{33} & 0 & 0 & 0 & 0 \\ w_{21} & w_{22} & w_{23} & w_{31} & w_{32} & w_{33} & 0 & 0 & 0 \\ 0 & w_{21} & w_{22} & 0 & w_{31} & w_{32} & 0 & 0 & 0 \\ w_{12} & w_{13} & 0 & w_{22} & w_{23} & 0 & w_{32} & w_{33} & 0 \\ w_{11} & w_{12} & w_{13} & w_{21} & w_{22} & w_{23} & w_{31} & w_{32} & w_{33} \\ 0 & w_{11} & w_{12} & 0 & w_{21} & w_{22} & 0 & w_{31} & w_{32} \\ 0 & 0 & 0 & w_{12} & w_{13} & 0 & w_{22} & w_{23} & 0 \\ 0 & 0 & 0 & w_{11} & w_{12} & w_{13} & w_{21} & w_{22} & w_{23} \\ 0 & 0 & 0 & 0 & w_{11} & w_{12} & 0 & w_{21} & w_{22} \end{pmatrix} \begin{pmatrix} a \\ b \\ c \\ d \\ e \\ f \\ g \\ h \\ i \end{pmatrix}$$

Backward pass as a matrix mult.

$$\begin{pmatrix} \frac{\partial L}{\partial a} \\ \frac{\partial L}{\partial b} \\ \frac{\partial L}{\partial c} \\ \frac{\partial L}{\partial d} \\ \frac{\partial L}{\partial e} \\ \frac{\partial L}{\partial f} \\ \frac{\partial L}{\partial g} \\ \frac{\partial L}{\partial h} \\ \frac{\partial L}{\partial i} \end{pmatrix} = \begin{pmatrix} w_{22} & w_{23} & 0 & w_{32} & w_{33} & 0 & 0 & 0 & 0 \\ w_{21} & w_{22} & w_{23} & w_{31} & w_{32} & w_{33} & 0 & 0 & 0 \\ 0 & w_{21} & w_{22} & 0 & w_{31} & w_{32} & 0 & 0 & 0 \\ w_{12} & w_{13} & 0 & w_{22} & w_{23} & 0 & w_{32} & w_{33} & 0 \\ w_{11} & w_{12} & w_{13} & w_{21} & w_{22} & w_{23} & w_{31} & w_{32} & w_{33} \\ 0 & w_{11} & w_{12} & 0 & w_{21} & w_{22} & 0 & w_{31} & w_{32} \\ 0 & 0 & 0 & w_{12} & w_{13} & 0 & w_{22} & w_{23} & 0 \\ 0 & 0 & 0 & w_{11} & w_{12} & w_{13} & w_{21} & w_{22} & w_{23} \\ 0 & 0 & 0 & 0 & w_{11} & w_{12} & 0 & w_{21} & w_{22} \end{pmatrix}^\top \begin{pmatrix} \frac{\partial L}{\partial \alpha} \\ \frac{\partial L}{\partial \beta} \\ \frac{\partial L}{\partial \gamma} \\ \frac{\partial L}{\partial \delta} \\ \frac{\partial L}{\partial \epsilon} \\ \frac{\partial L}{\partial \zeta} \\ \frac{\partial L}{\partial \eta} \\ \frac{\partial L}{\partial \theta} \\ \frac{\partial L}{\partial \kappa} \end{pmatrix}$$

Backward pass as a convolution

$$\begin{array}{|c|c|c|}\hline \frac{\partial L}{\partial \alpha} & \frac{\partial L}{\partial \beta} & \frac{\partial L}{\partial \gamma} \\ \hline \frac{\partial L}{\partial \delta} & \frac{\partial L}{\partial \epsilon} & \frac{\partial L}{\partial \zeta} \\ \hline \frac{\partial L}{\partial \eta} & \frac{\partial L}{\partial \theta} & \frac{\partial L}{\partial \kappa} \\ \hline \end{array} * \begin{array}{|c|c|c|}\hline g_{11} & g_{12} & g_{13} \\ \hline g_{21} & g_{22} & g_{23} \\ \hline g_{31} & g_{32} & g_{33} \\ \hline \end{array} = \begin{array}{|c|c|c|}\hline \frac{\partial L}{\partial a} & \frac{\partial L}{\partial b} & \frac{\partial L}{\partial c} \\ \hline \frac{\partial L}{\partial d} & \frac{\partial L}{\partial e} & \frac{\partial L}{\partial f} \\ \hline \frac{\partial L}{\partial g} & \frac{\partial L}{\partial h} & \frac{\partial L}{\partial i} \\ \hline \end{array}$$

$$
\begin{pmatrix}
w_{22} & w_{23} & 0 & w_{32} & w_{33} & 0 & 0 & 0 & 0 \\
w_{21} & w_{22} & w_{23} & w_{31} & w_{32} & w_{33} & 0 & 0 & 0 \\
0 & w_{21} & w_{22} & 0 & w_{31} & w_{32} & 0 & 0 & 0 \\
w_{12} & w_{13} & 0 & w_{22} & w_{23} & 0 & w_{32} & w_{33} & 0 \\
w_{11} & w_{12} & w_{13} & w_{21} & w_{22} & w_{23} & w_{31} & w_{32} & w_{33} \\
0 & w_{11} & w_{12} & 0 & w_{21} & w_{22} & 0 & w_{31} & w_{32} \\
0 & 0 & 0 & w_{12} & w_{13} & 0 & w_{22} & w_{23} & 0 \\
0 & 0 & 0 & w_{11} & w_{12} & w_{13} & w_{21} & w_{22} & w_{23} \\
0 & 0 & 0 & 0 & w_{11} & w_{12} & 0 & w_{21} & w_{22}
\end{pmatrix}^{\top}
$$

$$
= \begin{pmatrix}
w_{22} & w_{21} & 0 & w_{12} & w_{11} & 0 & 0 & 0 & 0 \\
w_{23} & w_{22} & w_{21} & w_{13} & w_{12} & w_{11} & 0 & 0 & 0 \\
0 & w_{23} & w_{22} & 0 & w_{13} & w_{12} & 0 & 0 & 0 \\
w_{32} & w_{31} & 0 & w_{22} & w_{21} & 0 & w_{12} & w_{11} & 0 \\
w_{33} & w_{32} & w_{31} & w_{23} & w_{22} & w_{21} & w_{13} & w_{12} & w_{11} \\
0 & w_{33} & w_{32} & 0 & w_{23} & w_{22} & 0 & w_{13} & w_{12} \\
0 & 0 & 0 & w_{32} & w_{31} & 0 & w_{22} & w_{21} & 0 \\
0 & 0 & 0 & w_{33} & w_{32} & w_{31} & w_{23} & w_{22} & w_{21} \\
0 & 0 & 0 & 0 & w_{33} & w_{32} & 0 & w_{23} & w_{22}
\end{pmatrix}
$$

Just another convolution!

$$\begin{bmatrix} \frac{\partial L}{\partial \alpha} & \frac{\partial L}{\partial \beta} & \frac{\partial L}{\partial \gamma} \\ \frac{\partial L}{\partial \delta} & \frac{\partial L}{\partial \epsilon} & \frac{\partial L}{\partial \zeta} \\ \frac{\partial L}{\partial \eta} & \frac{\partial L}{\partial \theta} & \frac{\partial L}{\partial \kappa} \end{bmatrix} * \begin{bmatrix} w_{33} & w_{32} & w_{31} \\ w_{23} & w_{22} & w_{21} \\ w_{13} & w_{12} & w_{11} \end{bmatrix} = \begin{bmatrix} \frac{\partial L}{\partial a} & \frac{\partial L}{\partial b} & \frac{\partial L}{\partial c} \\ \frac{\partial L}{\partial d} & \frac{\partial L}{\partial e} & \frac{\partial L}{\partial f} \\ \frac{\partial L}{\partial g} & \frac{\partial L}{\partial h} & \frac{\partial L}{\partial i} \end{bmatrix}$$

Weight Update

Weight update – basic math

- $M^{new} = M^{old} - \mu \left(\frac{\partial L}{\partial M}\right)^{\top}$
  - μ: learning rate

$$M = \begin{bmatrix} w_{11} \\ w_{12} \\ w_{13} \\ w_{21} \\ w_{22} \\ w_{23} \\ w_{31} \\ w_{32} \\ w_{33} \end{bmatrix}$$

Matrix representation

$$Y = \begin{bmatrix} \alpha \\ \beta \\ \gamma \\ \delta \\ \epsilon \\ \zeta \\ \eta \\ \theta \\ \kappa \end{bmatrix} = \begin{bmatrix} 0 & 0 & 0 & 0 & a & b & 0 & d & e \\ 0 & 0 & 0 & a & b & c & d & e & f \\ 0 & 0 & 0 & b & c & 0 & e & f & 0 \\ 0 & a & b & 0 & d & e & 0 & g & h \\ a & b & c & d & e & f & g & h & i \\ b & c & 0 & e & f & 0 & h & i & 0 \\ 0 & d & e & 0 & g & h & 0 & 0 & 0 \\ d & e & f & g & h & i & 0 & 0 & 0 \\ e & f & 0 & h & i & 0 & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} w_{11} \\ w_{12} \\ w_{13} \\ w_{21} \\ w_{22} \\ w_{23} \\ w_{31} \\ w_{32} \\ w_{33} \end{bmatrix} = QM$$

Note that

$$\begin{array}{|c|c|c|} \hline a & b & c \\ \hline d & e & f \\ \hline g & h & i \\ \hline \end{array} * \begin{array}{|c|c|c|} \hline w_{11} & w_{12} & w_{13} \\ \hline w_{21} & w_{22} & w_{23} \\ \hline w_{31} & w_{32} & w_{33} \\ \hline \end{array} = \begin{array}{|c|c|c|} \hline \alpha & \beta & \gamma \\ \hline \delta & \epsilon & \zeta \\ \hline \eta & \theta & \kappa \\ \hline \end{array}$$

Weight update

- $\left(\frac{\partial L}{\partial M}\right)^{\top} = Q^{\top} \times \left(\frac{\partial L}{\partial Y}\right)^{\top}$

$$\begin{bmatrix} \frac{\partial L}{\partial w_{11}} \\ \frac{\partial L}{\partial w_{12}} \\ \frac{\partial L}{\partial w_{13}} \\ \frac{\partial L}{\partial w_{21}} \\ \frac{\partial L}{\partial w_{22}} \\ \frac{\partial L}{\partial w_{23}} \\ \frac{\partial L}{\partial w_{31}} \\ \frac{\partial L}{\partial w_{32}} \\ \frac{\partial L}{\partial w_{33}} \end{bmatrix} = \begin{bmatrix} 0 & 0 & 0 & 0 & a & b & 0 & d & e \\ 0 & 0 & 0 & a & b & c & d & e & f \\ 0 & 0 & 0 & b & c & 0 & e & f & 0 \\ 0 & a & b & 0 & d & e & 0 & g & h \\ a & b & c & d & e & f & g & h & i \\ b & c & 0 & e & f & 0 & h & i & 0 \\ 0 & d & e & 0 & g & h & 0 & 0 & 0 \\ d & e & f & g & h & i & 0 & 0 & 0 \\ e & f & 0 & h & i & 0 & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} \frac{\partial L}{\partial \alpha} \\ \frac{\partial L}{\partial \beta} \\ \frac{\partial L}{\partial \gamma} \\ \frac{\partial L}{\partial \delta} \\ \frac{\partial L}{\partial \epsilon} \\ \frac{\partial L}{\partial \zeta} \\ \frac{\partial L}{\partial \eta} \\ \frac{\partial L}{\partial \theta} \\ \frac{\partial L}{\partial \kappa} \end{bmatrix}$$

$$\begin{bmatrix} \frac{\partial L}{\partial \alpha} & \frac{\partial L}{\partial \beta} & \frac{\partial L}{\partial \gamma} \\ \frac{\partial L}{\partial \delta} & \frac{\partial L}{\partial \epsilon} & \frac{\partial L}{\partial \zeta} \\ \frac{\partial L}{\partial \eta} & \frac{\partial L}{\partial \theta} & \frac{\partial L}{\partial \kappa} \end{bmatrix} \times \begin{bmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & a & b & c & 0 \\ 0 & d & e & f & 0 \\ 0 & g & h & i & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix} = \begin{bmatrix} \boxed{\frac{\partial L}{\partial w_{11}}} & \frac{\partial L}{\partial w_{12}} & \frac{\partial L}{\partial w_{13}} \\ \frac{\partial L}{\partial w_{21}} & \frac{\partial L}{\partial w_{22}} & \frac{\partial L}{\partial w_{23}} \\ \frac{\partial L}{\partial w_{31}} & \frac{\partial L}{\partial w_{32}} & \frac{\partial L}{\partial w_{33}} \end{bmatrix}$$

$$\begin{bmatrix} \frac{\partial L}{\partial \alpha} & \frac{\partial L}{\partial \beta} & \frac{\partial L}{\partial \gamma} \\ \frac{\partial L}{\partial \delta} & \frac{\partial L}{\partial \epsilon} & \frac{\partial L}{\partial \zeta} \\ \frac{\partial L}{\partial \eta} & \frac{\partial L}{\partial \theta} & \frac{\partial L}{\partial \kappa} \end{bmatrix} \times \begin{bmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & a & b & c & 0 \\ 0 & d & e & f & 0 \\ 0 & g & h & i & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix} = \begin{bmatrix} \frac{\partial L}{\partial w_{11}} & \boxed{\frac{\partial L}{\partial w_{12}}} & \frac{\partial L}{\partial w_{13}} \\ \frac{\partial L}{\partial w_{21}} & \frac{\partial L}{\partial w_{22}} & \frac{\partial L}{\partial w_{23}} \\ \frac{\partial L}{\partial w_{31}} & \frac{\partial L}{\partial w_{32}} & \frac{\partial L}{\partial w_{33}} \end{bmatrix}$$

$$\begin{bmatrix} \frac{\partial L}{\partial \alpha} & \frac{\partial L}{\partial \beta} & \frac{\partial L}{\partial \gamma} \\ \frac{\partial L}{\partial \delta} & \frac{\partial L}{\partial \epsilon} & \frac{\partial L}{\partial \zeta} \\ \frac{\partial L}{\partial \eta} & \frac{\partial L}{\partial \theta} & \frac{\partial L}{\partial \kappa} \end{bmatrix} \times \begin{bmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & a & b & c & 0 \\ 0 & d & e & f & 0 \\ 0 & g & h & i & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix} = \begin{bmatrix} \frac{\partial L}{\partial w_{11}} & \frac{\partial L}{\partial w_{12}} & \boxed{\frac{\partial L}{\partial w_{13}}} \\ \frac{\partial L}{\partial w_{21}} & \frac{\partial L}{\partial w_{22}} & \frac{\partial L}{\partial w_{23}} \\ \frac{\partial L}{\partial w_{31}} & \frac{\partial L}{\partial w_{32}} & \frac{\partial L}{\partial w_{33}} \end{bmatrix}$$

$$\begin{bmatrix} \frac{\partial L}{\partial \alpha} & \frac{\partial L}{\partial \beta} & \frac{\partial L}{\partial \gamma} \\ \frac{\partial L}{\partial \delta} & \frac{\partial L}{\partial \epsilon} & \frac{\partial L}{\partial \zeta} \\ \frac{\partial L}{\partial \eta} & \frac{\partial L}{\partial \theta} & \frac{\partial L}{\partial \kappa} \end{bmatrix} \times \begin{bmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & a & b & c & 0 \\ 0 & d & e & f & 0 \\ 0 & g & h & I & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix} = \begin{bmatrix} \frac{\partial L}{\partial w_{11}} & \frac{\partial L}{\partial w_{12}} & \frac{\partial L}{\partial w_{13}} \\ \boxed{\frac{\partial L}{\partial w_{21}}} & \frac{\partial L}{\partial w_{22}} & \frac{\partial L}{\partial w_{23}} \\ \frac{\partial L}{\partial w_{31}} & \frac{\partial L}{\partial w_{32}} & \frac{\partial L}{\partial w_{33}} \end{bmatrix}$$

$$\begin{bmatrix} \frac{\partial L}{\partial \alpha} & \frac{\partial L}{\partial \beta} & \frac{\partial L}{\partial \gamma} \\ \frac{\partial L}{\partial \delta} & \frac{\partial L}{\partial \epsilon} & \frac{\partial L}{\partial \zeta} \\ \frac{\partial L}{\partial \eta} & \frac{\partial L}{\partial \theta} & \frac{\partial L}{\partial \kappa} \end{bmatrix} \times \begin{bmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & a & b & c & 0 \\ 0 & d & e & f & 0 \\ 0 & g & h & I & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix} = \begin{bmatrix} \frac{\partial L}{\partial w_{11}} & \frac{\partial L}{\partial w_{12}} & \frac{\partial L}{\partial w_{13}} \\ \frac{\partial L}{\partial w_{21}} & \boxed{\frac{\partial L}{\partial w_{22}}} & \frac{\partial L}{\partial w_{23}} \\ \frac{\partial L}{\partial w_{31}} & \frac{\partial L}{\partial w_{32}} & \frac{\partial L}{\partial w_{33}} \end{bmatrix}$$

$$\begin{bmatrix} \frac{\partial L}{\partial \alpha} & \frac{\partial L}{\partial \beta} & \frac{\partial L}{\partial \gamma} \\ \frac{\partial L}{\partial \delta} & \frac{\partial L}{\partial \epsilon} & \frac{\partial L}{\partial \zeta} \\ \frac{\partial L}{\partial \eta} & \frac{\partial L}{\partial \theta} & \frac{\partial L}{\partial \kappa} \end{bmatrix} \times \begin{bmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & a & b & c & 0 \\ 0 & d & e & f & 0 \\ 0 & g & h & I & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix} = \begin{bmatrix} \frac{\partial L}{\partial w_{11}} & \frac{\partial L}{\partial w_{12}} & \frac{\partial L}{\partial w_{13}} \\ \frac{\partial L}{\partial w_{21}} & \frac{\partial L}{\partial w_{22}} & \boxed{\frac{\partial L}{\partial w_{23}}} \\ \frac{\partial L}{\partial w_{31}} & \frac{\partial L}{\partial w_{32}} & \frac{\partial L}{\partial w_{33}} \end{bmatrix}$$

$$\begin{bmatrix} \frac{\partial L}{\partial \alpha} & \frac{\partial L}{\partial \beta} & \frac{\partial L}{\partial \gamma} \\ \frac{\partial L}{\partial \delta} & \frac{\partial L}{\partial \epsilon} & \frac{\partial L}{\partial \zeta} \\ \frac{\partial L}{\partial \eta} & \frac{\partial L}{\partial \theta} & \frac{\partial L}{\partial \kappa} \end{bmatrix} \times \begin{bmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & a & b & c & 0 \\ 0 & d & e & f & 0 \\ 0 & g & h & i & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix} = \begin{bmatrix} \frac{\partial L}{\partial w_{11}} & \frac{\partial L}{\partial w_{12}} & \frac{\partial L}{\partial w_{13}} \\ \frac{\partial L}{\partial w_{21}} & \frac{\partial L}{\partial w_{22}} & \frac{\partial L}{\partial w_{23}} \\ \boxed{\frac{\partial L}{\partial w_{31}}} & \frac{\partial L}{\partial w_{32}} & \frac{\partial L}{\partial w_{33}} \end{bmatrix}$$

$$\begin{bmatrix} \frac{\partial L}{\partial \alpha} & \frac{\partial L}{\partial \beta} & \frac{\partial L}{\partial \gamma} \\ \frac{\partial L}{\partial \delta} & \frac{\partial L}{\partial \epsilon} & \frac{\partial L}{\partial \zeta} \\ \frac{\partial L}{\partial \eta} & \frac{\partial L}{\partial \theta} & \frac{\partial L}{\partial \kappa} \end{bmatrix} \times \begin{bmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & a & b & c & 0 \\ 0 & d & e & f & 0 \\ 0 & g & h & i & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix} = \begin{bmatrix} \frac{\partial L}{\partial w_{11}} & \frac{\partial L}{\partial w_{12}} & \frac{\partial L}{\partial w_{13}} \\ \frac{\partial L}{\partial w_{21}} & \frac{\partial L}{\partial w_{22}} & \frac{\partial L}{\partial w_{23}} \\ \frac{\partial L}{\partial w_{31}} & \boxed{\frac{\partial L}{\partial w_{32}}} & \frac{\partial L}{\partial w_{33}} \end{bmatrix}$$

$$\begin{bmatrix} \frac{\partial L}{\partial \alpha} & \frac{\partial L}{\partial \beta} & \frac{\partial L}{\partial \gamma} \\ \frac{\partial L}{\partial \delta} & \frac{\partial L}{\partial \epsilon} & \frac{\partial L}{\partial \zeta} \\ \frac{\partial L}{\partial \eta} & \frac{\partial L}{\partial \theta} & \frac{\partial L}{\partial \kappa} \end{bmatrix} \times \begin{bmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & a & b & c & 0 \\ 0 & d & e & f & 0 \\ 0 & g & h & i & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix} = \begin{bmatrix} \frac{\partial L}{\partial w_{11}} & \frac{\partial L}{\partial w_{12}} & \frac{\partial L}{\partial w_{13}} \\ \frac{\partial L}{\partial w_{21}} & \frac{\partial L}{\partial w_{22}} & \frac{\partial L}{\partial w_{23}} \\ \frac{\partial L}{\partial w_{31}} & \frac{\partial L}{\partial w_{32}} & \boxed{\frac{\partial L}{\partial w_{33}}} \end{bmatrix}$$

Summary

Summary

- Forward pass: convolution
- Backward pass: (transposed) convolution
- Weight update:
  - (Large-size) convolution

Multi-channel cases

Convolution

- f: input channels
- f': output channels
- S: minibatch

Forward pass:

$$y_{(s,j)} = \sum_{i \in f} x_{(s,i)} \star w_{(j,i)}$$

Backward pass:

$$\frac{\partial L}{\partial x_{(s,i)}} = \sum_{j \in f'} \frac{\partial L}{\partial y_{(s,j)}} * w_{(j,i)}$$

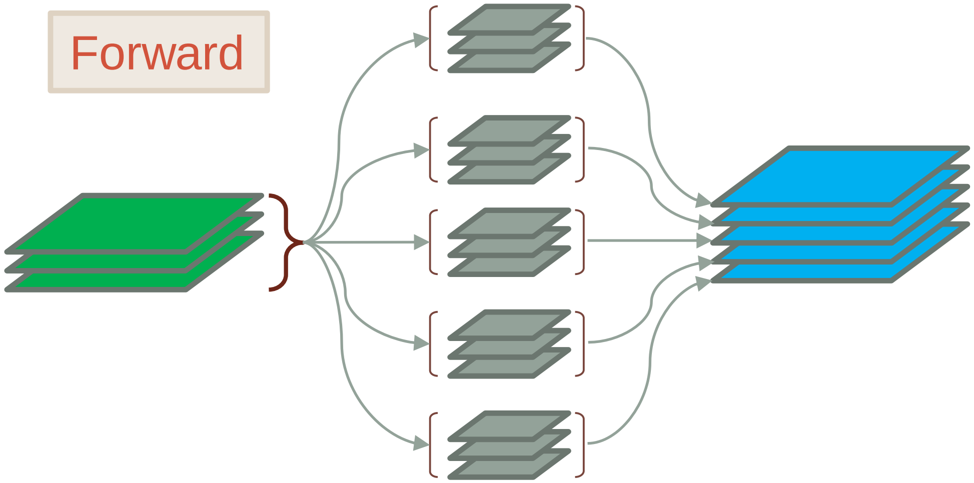
Weight update:

$$\frac{\partial L}{\partial w_{(j,i)}} = \sum_{s \in S} \frac{\partial L}{\partial y_{(s,j)}} \star x_{(s,i)}$$

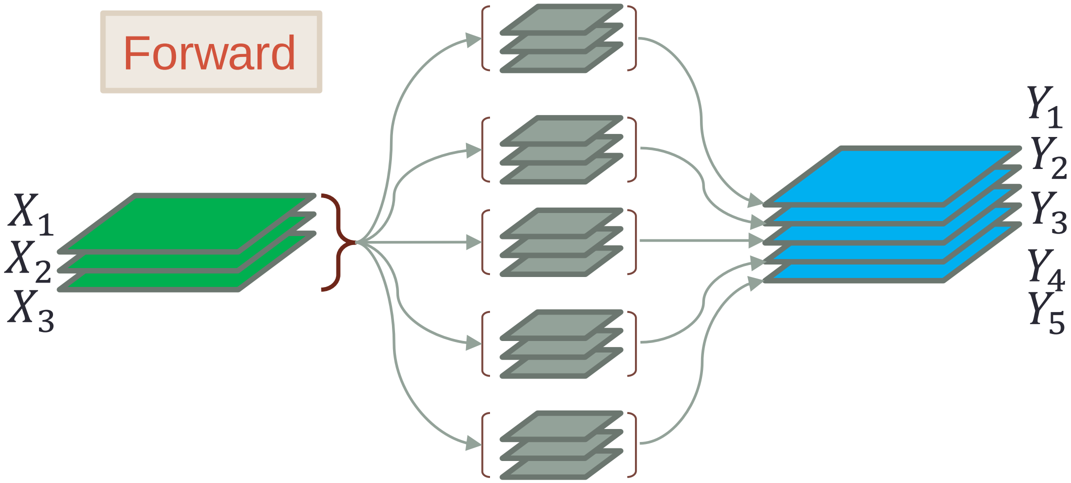
$\star$: $cross - correation$

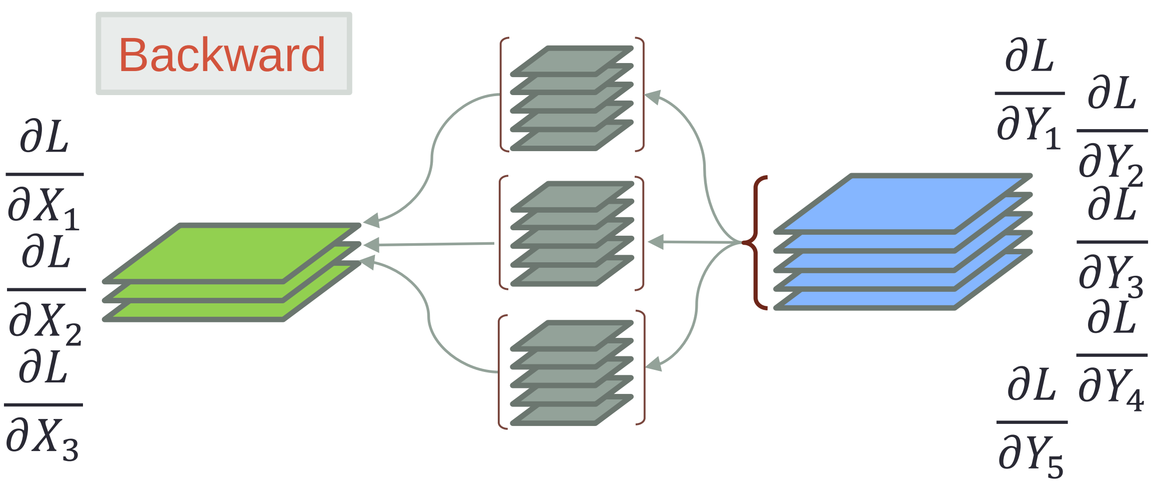
FW, BW and Weight updates at a glance

Forward



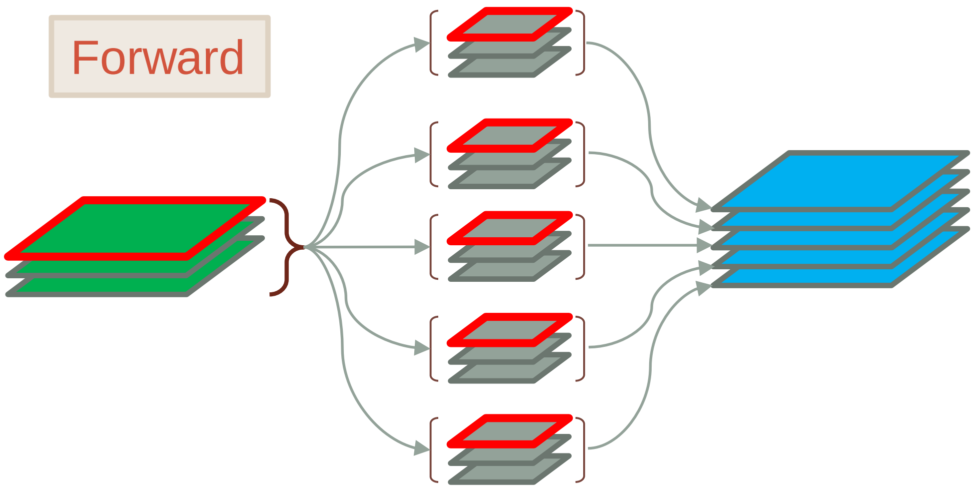
FW, BW and Weight updates at a glance



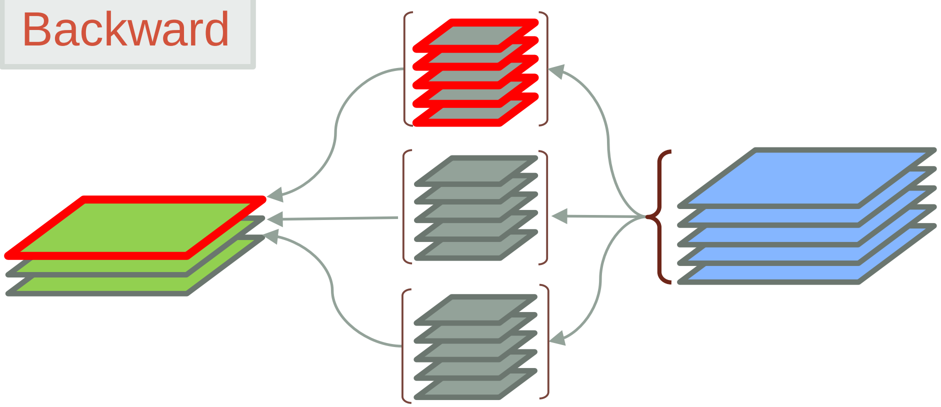


FW, BW and Weight updates at a glance

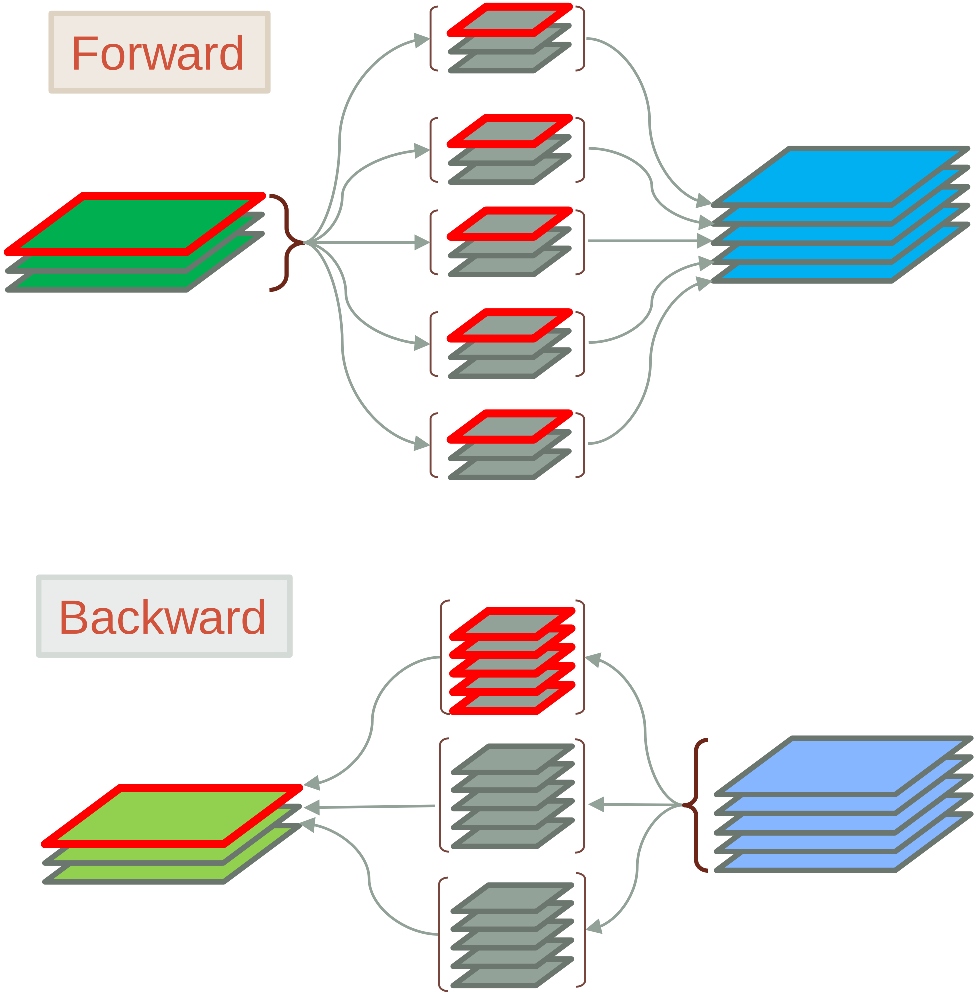
Forward

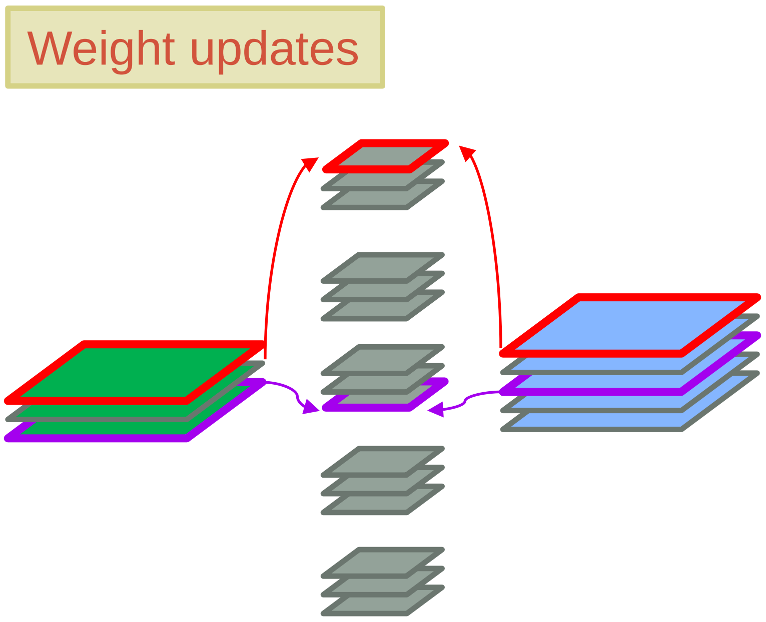


Backward



FW, BW and Weight updates





Backups

Weight update – intuitive derivation

- Want to find: $\Delta L \simeq \frac{\partial L}{\partial M} \Delta M$
  - We know
    - $\Delta L \simeq \frac{\partial L}{\partial Y} \Delta Y$
    - $\Delta Y \simeq \frac{\partial Y}{\partial M} \Delta M = Q \Delta \mathrm{M}$
  - Hence,
    - $\Delta L \simeq \frac{\partial L}{\partial Y} \Delta Y = \frac{\partial L}{\partial Y} Q \Delta M$
  - Finally, we have
    - $\frac{\partial L}{\partial M} = \frac{\partial L}{\partial Y} Q$
    - $\left(\frac{\partial L}{\partial M}\right)^{\top} = Q^{\top} \times \left(\frac{\partial L}{\partial Y}\right)^{\top}$

Differentiation

- First order approximation of a system

$$\Delta Y \simeq \boxed{}\,\Delta X$$

Jacobian, J_F, $\frac{\partial Y}{\partial X}$, ….