give below. Let the notation '$\star$' represent convolution. Assume that the bias values, $\mathbf{b}$, are appropriately broadcast.

$$\mathbf{X} \rightarrow \boxed{\star\mathbf{W}_0 + \mathbf{b}_0} \rightarrow \boxed{f_{\text{ReLU}}} \rightarrow \boxed{f_{\text{pool}}} \rightarrow \boxed{\star\mathbf{W}_1 + \mathbf{b}_1} \rightarrow \boxed{f_{\text{ReLU}}} \rightarrow \boxed{f_{\text{pool}}} \rightarrow \boxed{f_{\text{flat}}} \rightarrow \boxed{\mathbf{W}_2 + \mathbf{b}_2} \rightarrow \boxed{f_{\text{ReLU}}} \rightarrow \boxed{\mathbf{W}_3 + \mathbf{b}_3} \rightarrow \boxed{f_{\text{smax}}} \rightarrow \hat{\mathbf{P}}$$

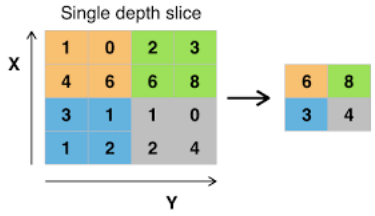
---

98 Example (Max Pooling)

Max pooling is a simple operation used to reduce the output size of a layer.

$$\mathbf{X}_{4\times4} \rightarrow \boxed{f_{\text{pool}}} \rightarrow \mathbf{Y}_{2\times2}$$

Instead of a dot product of the input with the weights of a filter, the maximum value of the input is saved as in the diagram below. A max pooling layer has no trainable parameters.



---

99 Example (LeNet Functional Representation)

The LeNet network in functional form is given below. (Assume that the bias values, $\mathbf{b}$, are appropriately broadcast.)

Layer 1

$$\begin{aligned}
\mathbf{Z}_1 &= \mathbf{X} \star \mathbf{W}_0 + \mathbf{b}_0 \\
\mathbf{A}_{11} &= f_{\text{ReLU}}(\mathbf{Z}_1) \\
\mathbf{A}_{12} &= f_{\text{pool}}(\mathbf{A}_{11})
\end{aligned}$$

Layer 2

$$\begin{aligned}
\mathbf{Z}_2 &= \mathbf{A}_{12} \star \mathbf{W}_1 + \mathbf{b}_1 \\
\mathbf{A}_{21} &= f_{\text{ReLU}}(\mathbf{Z}_1) \\
\mathbf{A}_{22} &= f_{\text{pool}}(\mathbf{A}_{21}) \\
\mathbf{A}_{23} &= f_{\text{flat}}(\mathbf{A}_{22})
\end{aligned}$$

Layer 3

$$\begin{aligned}
\mathbf{Z}_3 &= \mathbf{A}_{23}\mathbf{W}_2 + \mathbf{b}_2 \\
\mathbf{A}_3 &= f_{\text{ReLU}}(\mathbf{Z}_3)
\end{aligned}$$

Layer 4

$$\begin{aligned}
\hat{\mathbf{Y}} &= \mathbf{A}_3\mathbf{W}_3 + \mathbf{b}_3 \\
\hat{\mathbf{P}} &= f_{\text{smax}}(\hat{\mathbf{Y}})
\end{aligned}$$

---

100 Example (LeNet Parameters for CIFAR-10)

Listed below are the LeNet trainable parameters for the CIFAR-10 dataset which consists of inputs that are 32×32 pixel RGB images and 10 class labels as outputs.

Layer 1	$\mathbf{W}_0$	$3 \times 5 \times 5 \times 20$	$1,500$
	$\mathbf{b}_0$	20	20
Layer 2	$\mathbf{W}_1$	$20 \times 5 \times 5 \times 50$	$25,000$
	$\mathbf{b}_1$	50	50
Layer 3	$\mathbf{W}_2$	$(8 \cdot 8 \cdot 50) \times 500$	$1,600,000$
	$\mathbf{b}_2$	500	500
Layer 4	$\mathbf{W}_3$	500×10	$5,000$
	$\mathbf{b}_3$	10	10
		Total	$1,632,080$

Using single precision (32 bits or 4 bytes) we need approximately

$$1,632,080 \times 4 = 6,528,320 \approx 6.5 \text{ MB}$$

6.5 megabytes of memory to store the LeNet parameters.
