

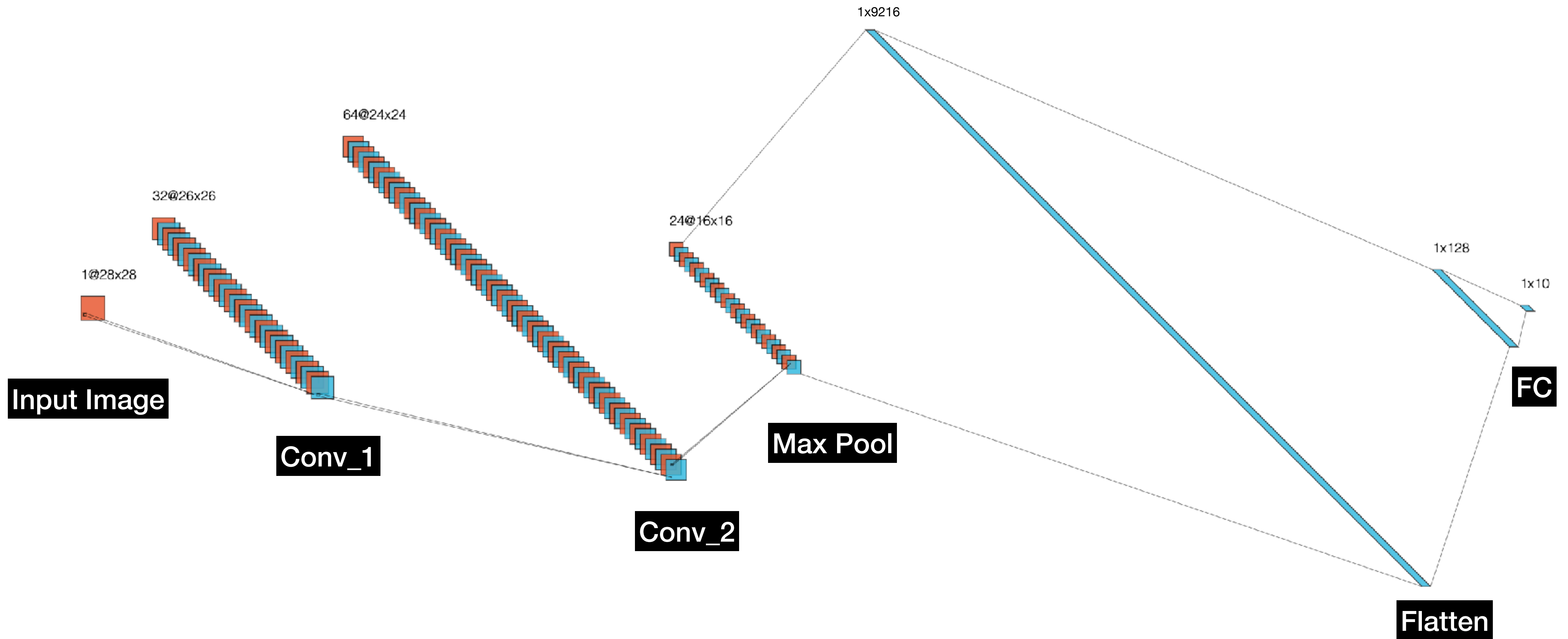
MODERN COMPUTER VISION

BY RAJEEV RATAN

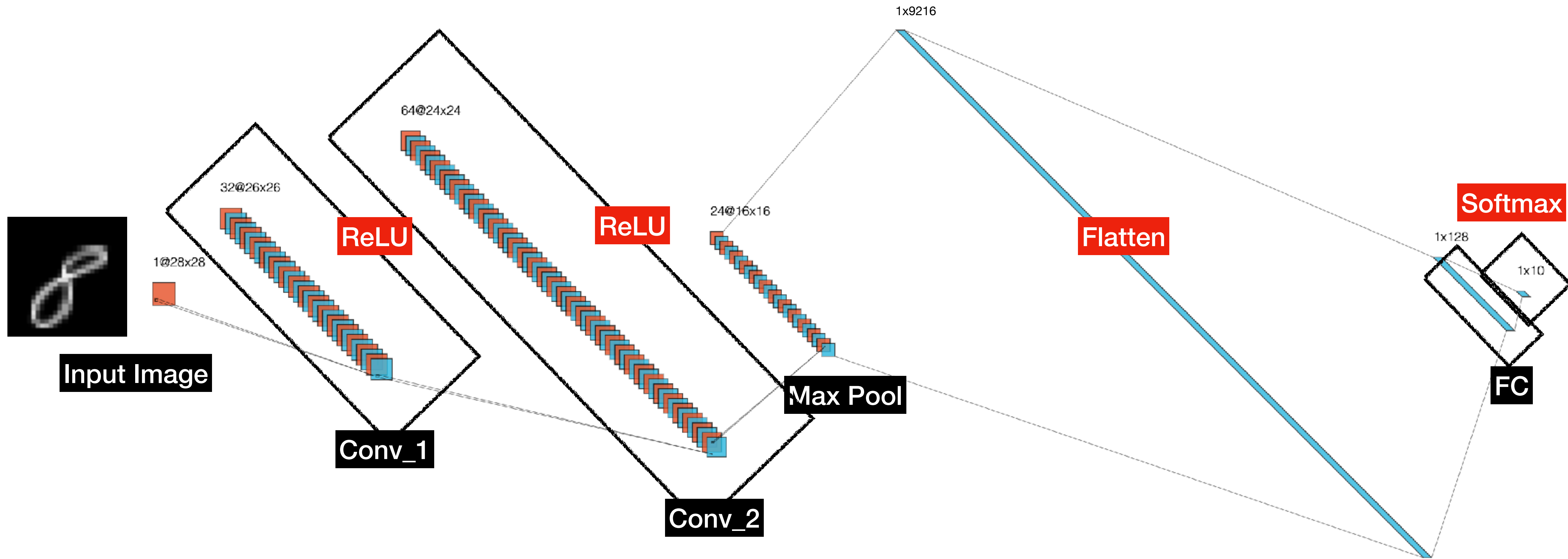
Building a CNN

Let's put all the pieces together and build a CNN

A Simple CNN

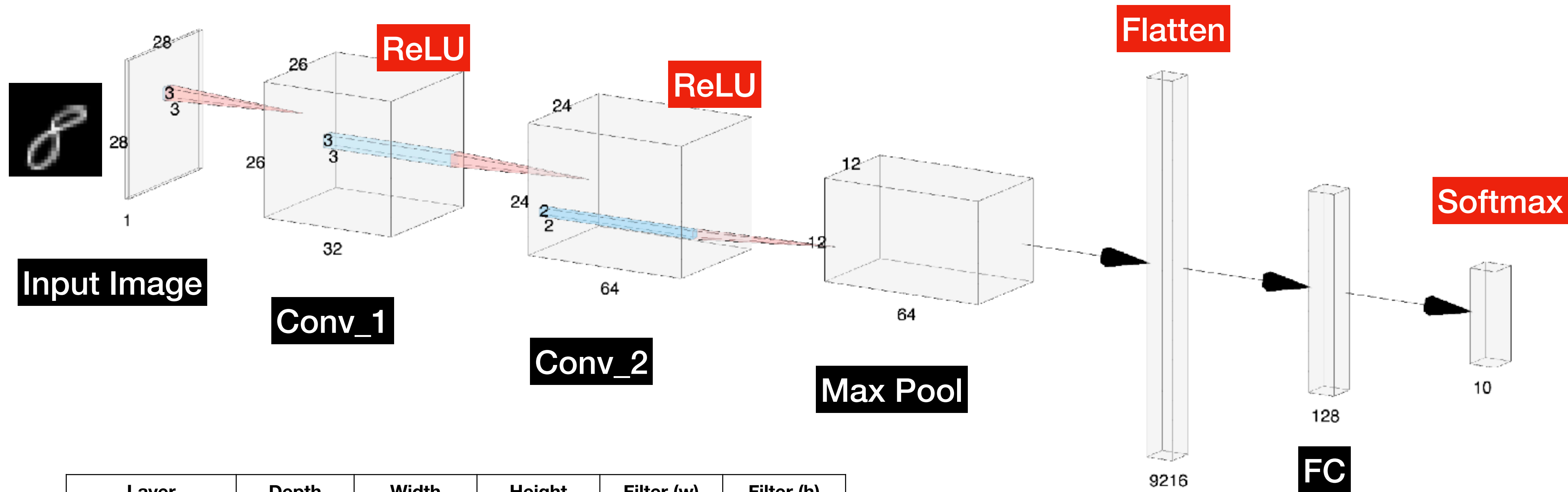


A 4 Layer Deep CNN for MNIST



Example of a 4 Layer CNN we can use for the MNIST Dataset

Another Representation

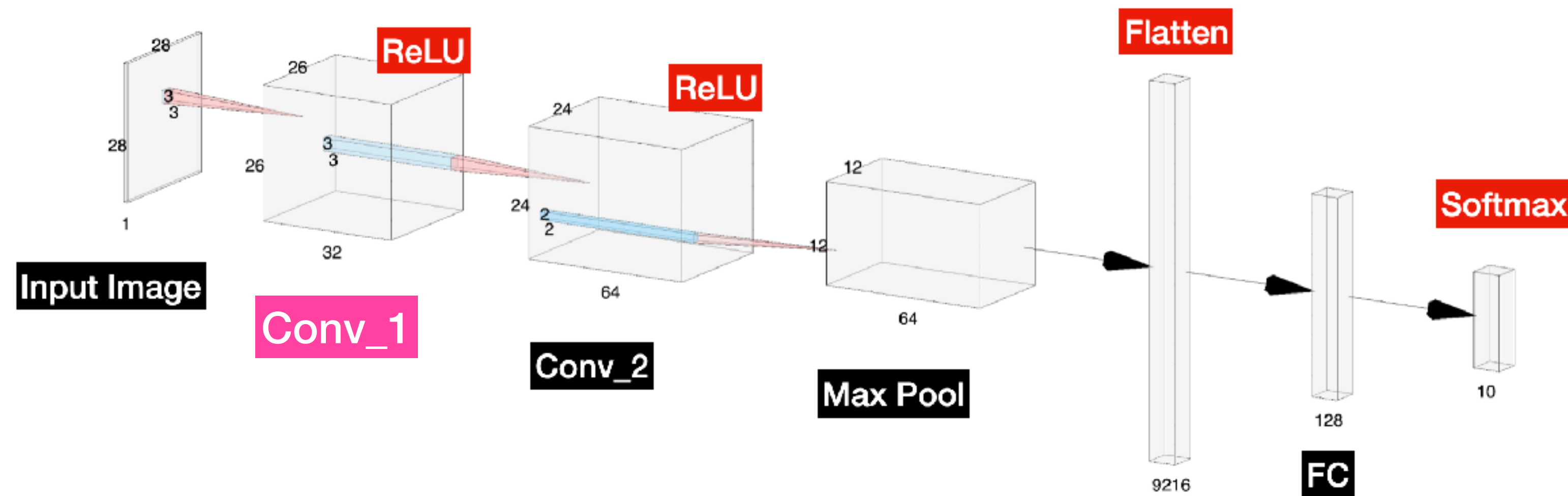


Layer	Depth	Width	Height	Filter (w)	Filter (h)
Input	1	28	28		
Conv_1	32	26	26	3	3
Conv_2	64	24	24	3	3
Max Pool	64	12	12	2	2
Flatten	9216	1	1		
Fully Connected	128	1	1		
Output	10	1	1		

Notes:

- We choose 32 & 64 Filters or Kernels for Conv_1 & Conv_2
- We choose to
- The Feature Maps are shown as Conv_1 and Conv_2
- Stride is 1
- Padding is 0 (not used)
- Max Pool Stride is 2

Calculating the Output Size of Conv_1



Notes:

- We choose 32 Filters or Kernels for Conv_1
- The Feature Maps are shown as Conv_1 & Conv_2
- Stride is 1
- Padding is 0 (not used)
- Max Pool Stride is 2

$$(n \times n) * (f \times f) = \left(\frac{n + 2p - f}{s} + 1 \right) \times \left(\frac{n + 2p - f}{s} + 1 \right) = \left(\frac{28 + (2 \times 0) - 3}{1} + 1 \right) \times \left(\frac{28 + (2 \times 0) - 3}{1} + 1 \right) = 26 \times 26$$

Where

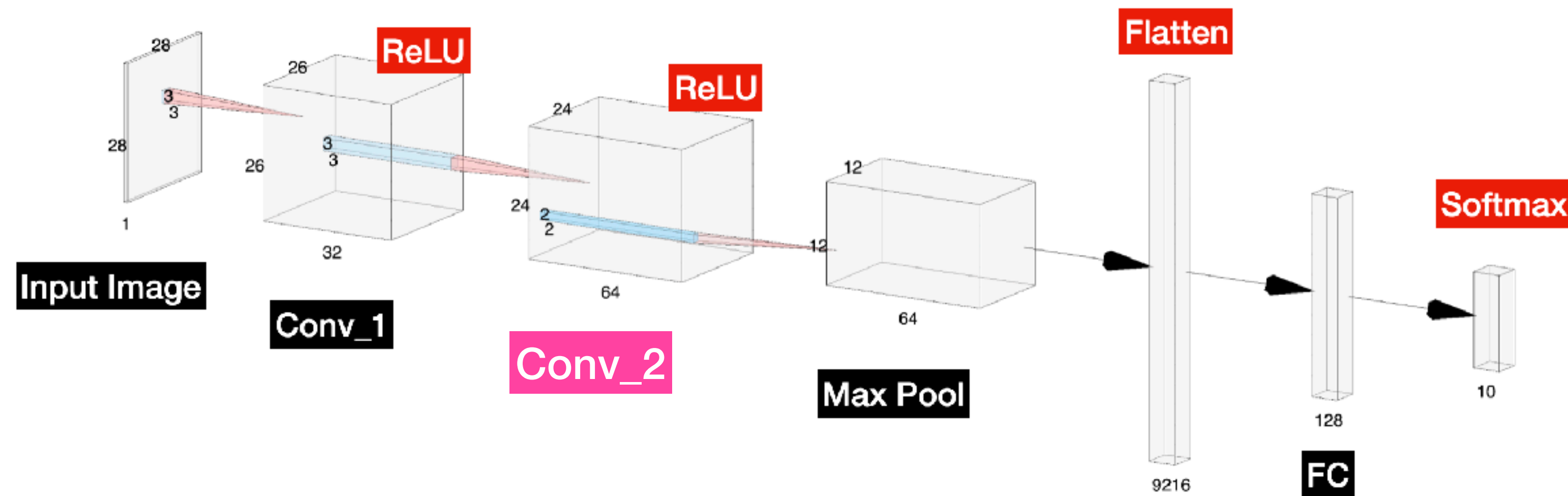
$$n = 28$$

$$f = 3$$

$$s = 1$$

$$p = 0$$

Calculating the Output Size of Conv_2



Notes:

- We choose 64 Filters or Kernels for Conv_2
- The Feature Maps are shown as Conv_1 & Conv_2
- Stride is 1
- Padding is 0 (not used)
- Max Pool Stride is 2

$$(n \times n) * (f \times f) = \left(\frac{n + 2p - f}{s} + 1 \right) \times \left(\frac{n + 2p - f}{s} + 1 \right) = \left(\frac{26 + (2 \times 0) - 3}{1} + 1 \right) \times \left(\frac{26 + (2 \times 0) - 3}{1} + 1 \right) = 24 \times 24$$

Where

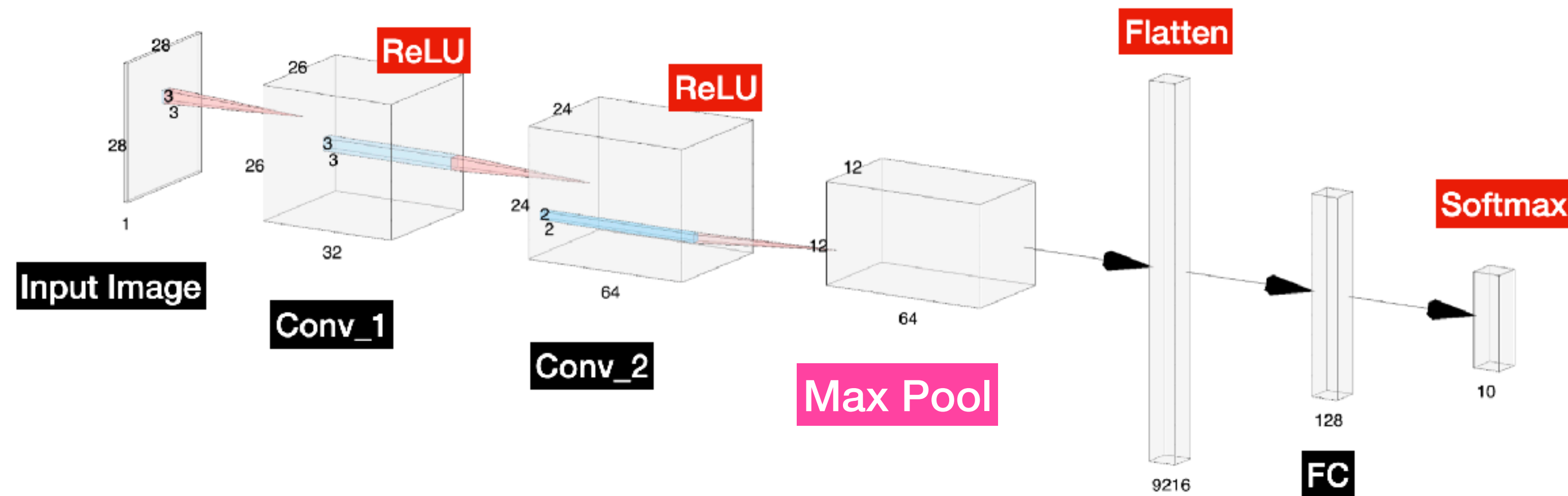
$$n = 26$$

$$f = 3$$

$$s = 1$$

$$p = 0$$

Calculating the Output Size of the Max Pool Layer



Notes:

- We choose 64 Filters or Kernels for Conv_2
- The Feature Maps are shown as Conv_1 & Conv_2
- Stride is 1
- Padding is 0 (not used)
- Max Pool Stride is 2

$$(n \times n) * (f \times f) = \left(\frac{n + 2p - f}{s} + 1\right) \times \left(\frac{n + 2p - f}{s} + 1\right) = \left(\frac{24 + (2 \times 0) - 2}{2} + 1\right) \times \left(\frac{24 + (2 \times 0) - 2}{2} + 1\right) = 12 \times 12$$

12 x 12

Where

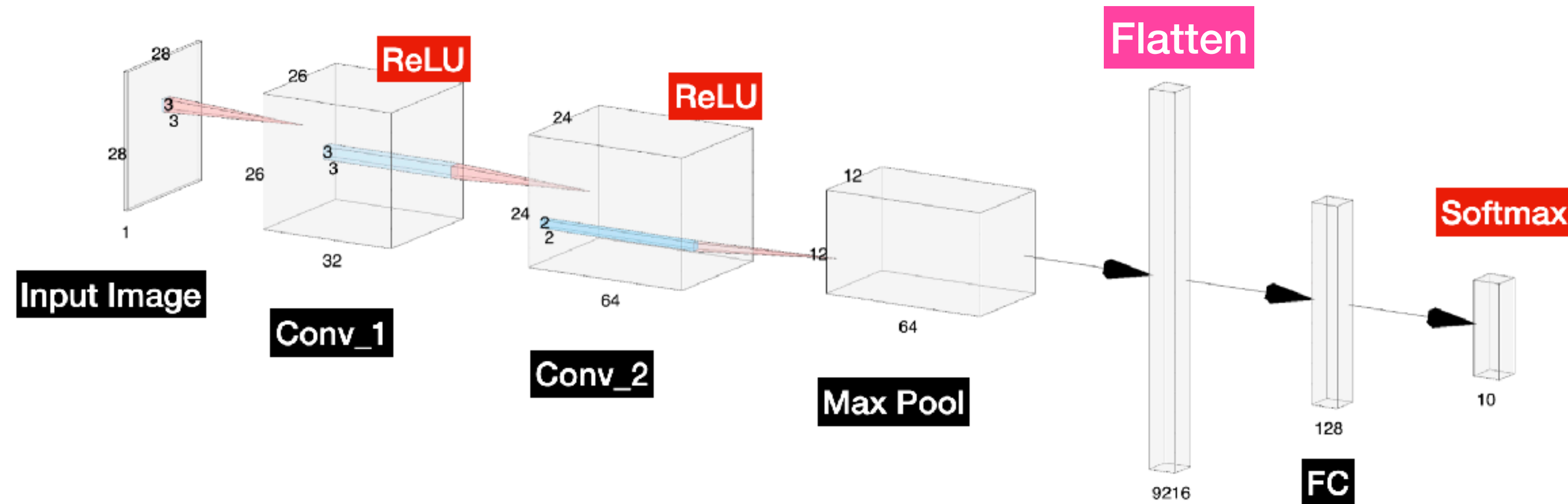
$$n = 24$$

$$f = 2$$

$$s = 2$$

$$p = 0$$

Calculating the Output Size of Flattened Layer



Notes:

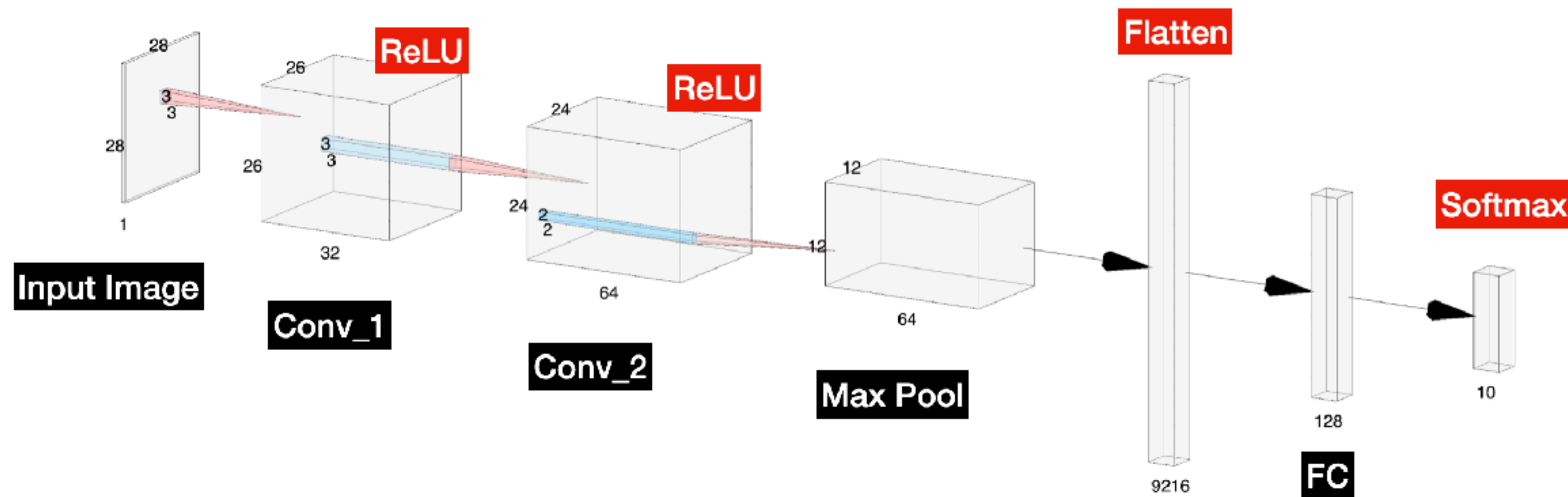
- We choose 64 Filters or Kernels for Conv_2
- The Feature Maps are shown as Conv_1 & Conv_2
- Stride is 1
- Padding is 0 (not used)
- Max Pool Stride is 2

$$12 \times 12 \times 64 = 9216$$

Where

$$n = 12$$

The Rest of the Our CNN



Notes:

- We choose 64 Filters or Kernels for Conv_2
- The Feature Maps are shown as Conv_1 & Conv_2
- Stride is 1
- Padding is 0 (not used)
- Max Pool Stride is 2
- **We choose 128 nodes in our FC Layer**
- **Our Dataset has 10 classes, hence why the final layer has 10 nodes**



MODERN COMPUTER VISION

BY RAJEEV RATAN

Next...

Parameter Counts in CNNs