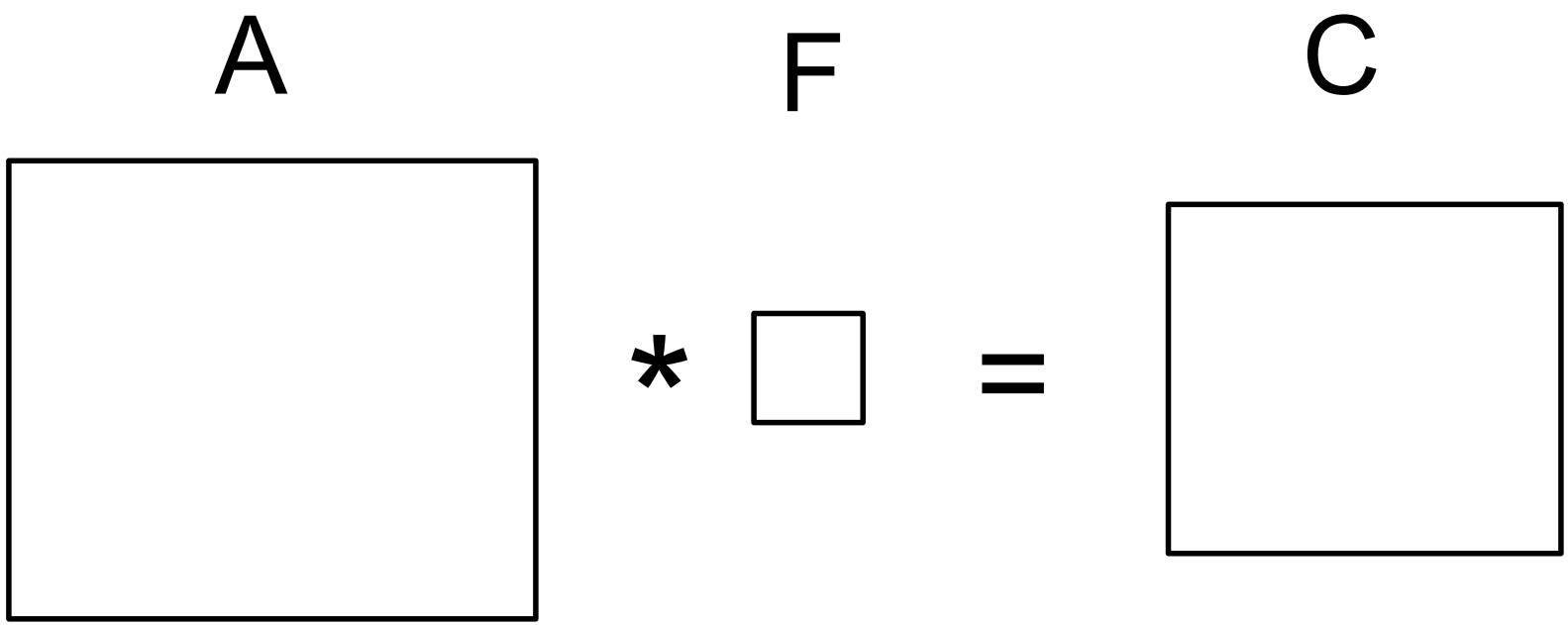
Convolutions for neural networks

Arnaud Bergeron

August 10, 2015

The basic convolution operation



Where A is the input channel, F is the filter and C is the output channel.

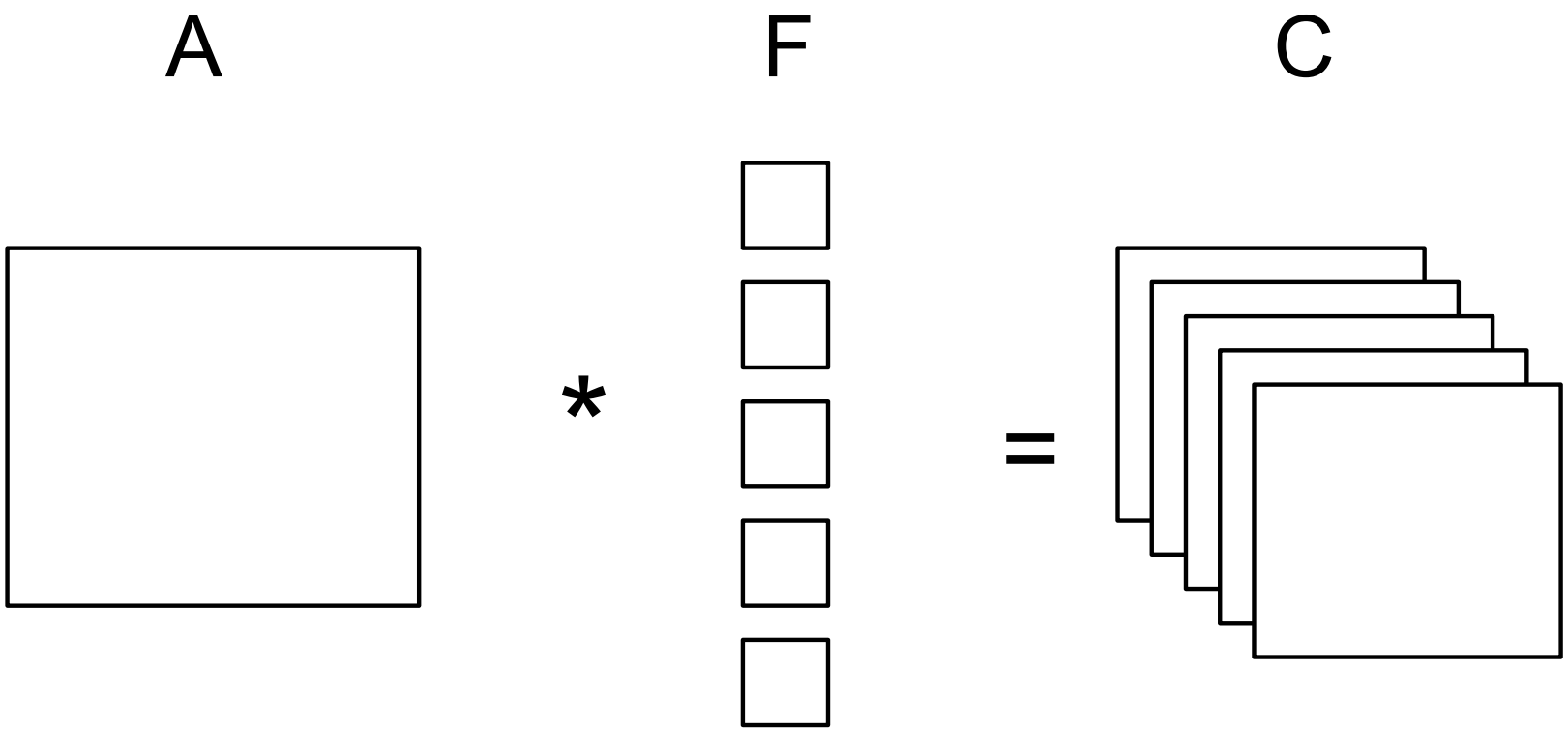
The basic convolution operation

$$C(i,j) = \sum_{m-1}^{x=0} \sum_{n-1}^{y=0} A(x+i, y+j) F(x,y)$$

- (m, n) is the size of the filter.

Where A is the input channel, F is the filter and C is the output channel.

The basic convolution operation with multiple outputs



where A is the input channel, F are the filters and C are output channels.

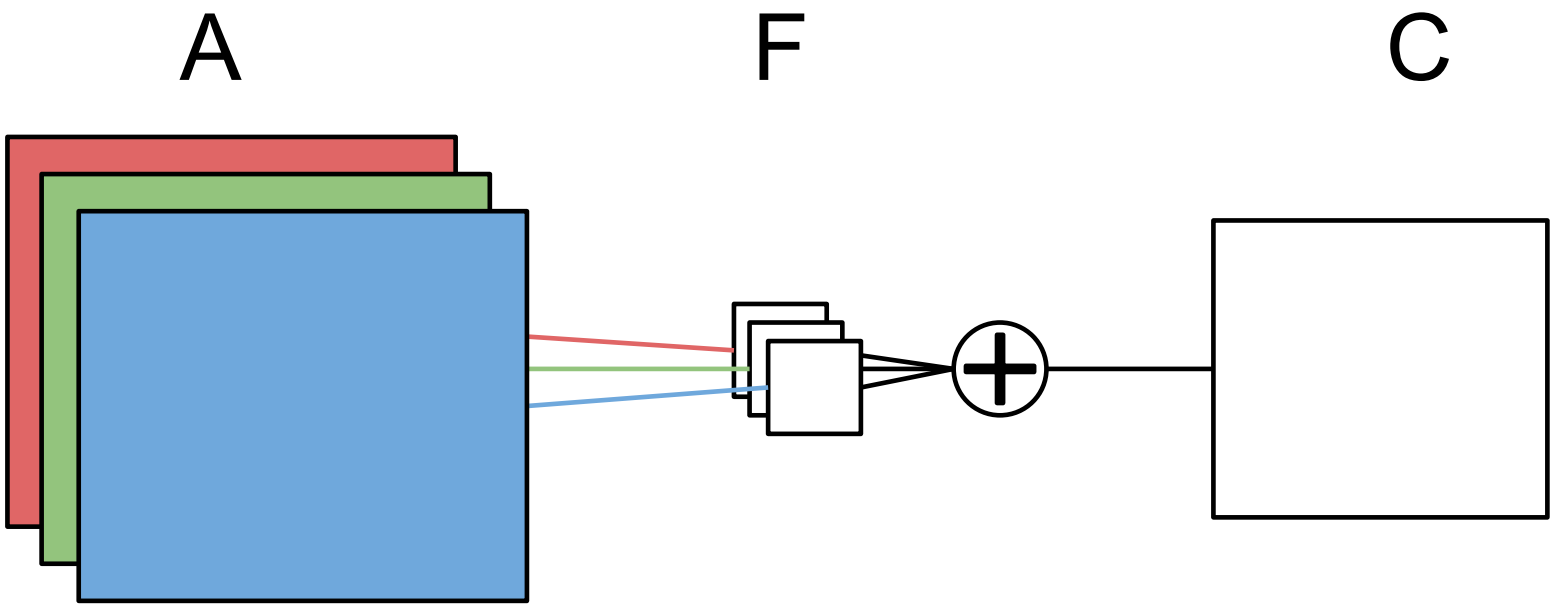
The basic convolution operation with multiple outputs

$$C_o(i,j) = \sum_{m-1}^{x=0} \sum_{n-1}^{y=0} A(x+i, y+j) F_o(x,y)$$

- (m, n) is the size of the filters.
- o is the output channel.

where A is the input channel, F are the filters and C are output channels.

The basic convolution operation with multiple inputs



where A are the input channels, F is the filter bank and C is the output channel.

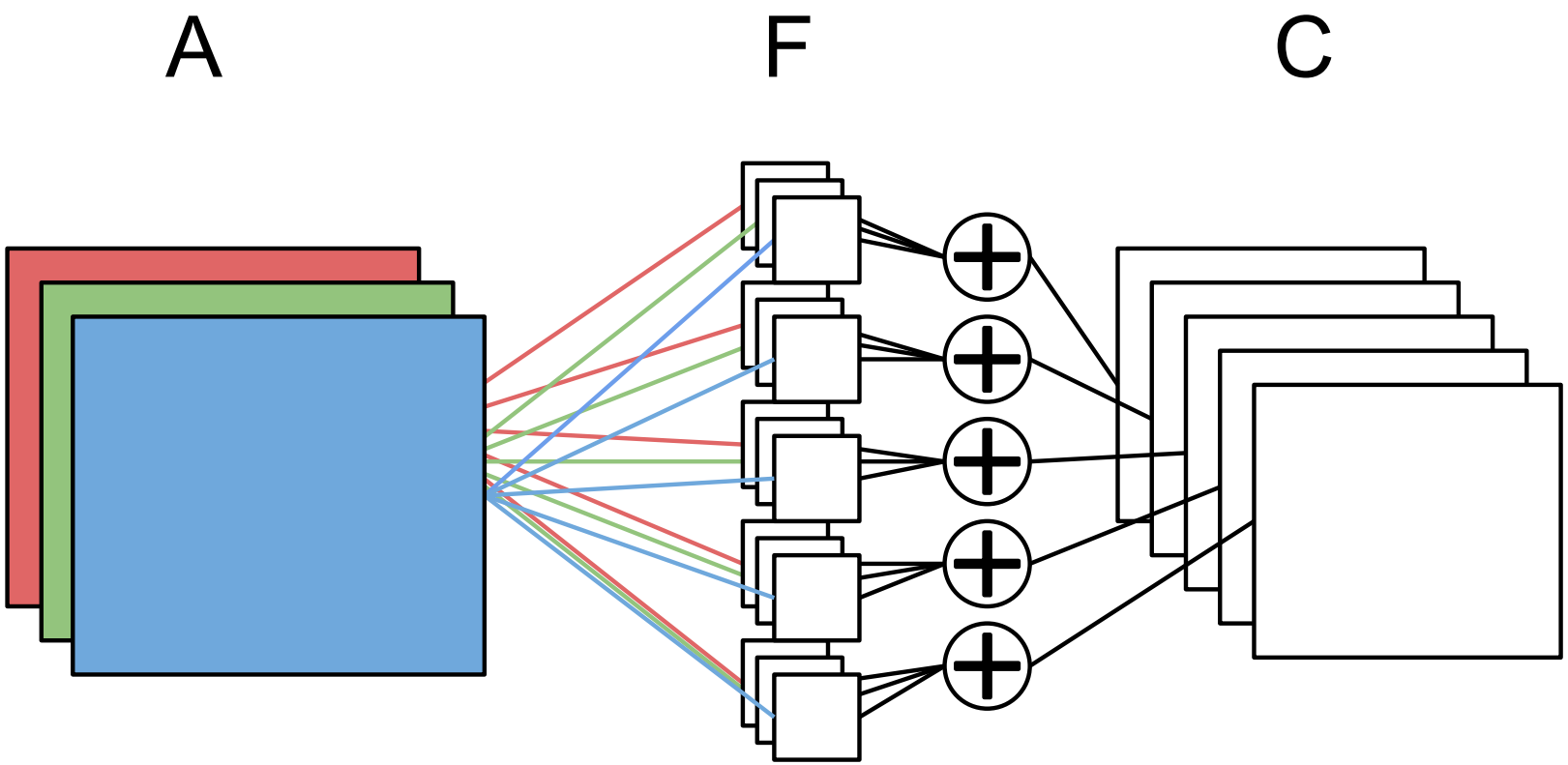
The basic convolution operation with multiple inputs

$$C(i,j) = \sum_{l}^{k=0} \sum_{m-1}^{x=0} \sum_{n-1}^{y=0} A_k(x+i, y+j) F_k(x,y)$$

- (m, n) is the size of the filters.
- k is the input channel.
- l is the number of input channels.

where A are the input channels, F is the filter bank and C is the output channel.

The basic convolution operation with multiple inputs and outputs



where A are the input channels, F are the filter banks and C are the output channels.

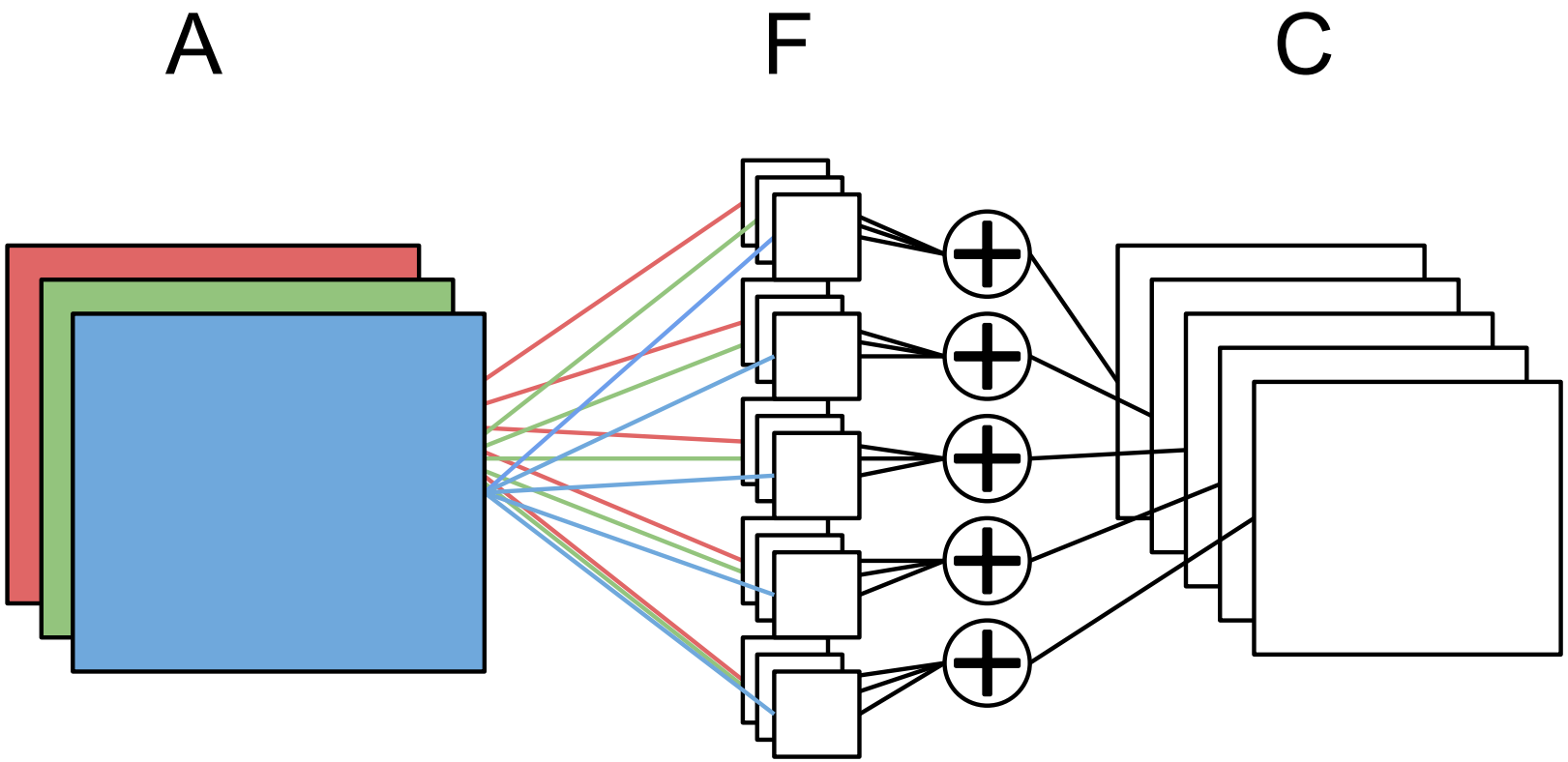
The basic convolution operation with multiple inputs and outputs

$$C_o(i,j) = \sum_{l-1}^{k=0} \sum_{m-1}^{x=0} \sum_{n-1}^{y=0} A_k(x+i, y+j) F_{ko}(x,y)$$

- (m, n) is the size of the filters.
- o is the output channel.
- k is the input channel.
- l is the number of input channels.

where A are the input channels, F are the filter banks and C are the output channels.

The basic convolution operation with batches



where A are the input channels, F are the filter banks and C are the output channels.

The basic convolution operation with batches

$$C_{bo}(i,j) = \sum_{l-1}^{k=0} \sum_{m-1}^{x=0} \sum_{n-1}^{y=0} A_{bk}(x+i, y+j) F_{ko}(x,y)$$

- (m, n) is the size of the filters.
- o is the output channel.
- k is the input channel.
- l is the number of input channels.
- b is the batch.

where A are the input channels, F are the filter banks and C are the output channels.

Some vocabulary:

filter what we call the smaller or "learned" 2D matrix in a traditional convolution.

channel what we call input or output 2D matrices in a traditional convolution. Sometimes called feature map.

filter bank group of filters whose convolution output will be summed to form one output channel. Sometimes called filter stack.

Memory layout for images: 'bc01'

- first dimension is the batch ('b')
- second dimension is the channel ('c')
- last two dimensions are the data ('0', '1')

Memory layout for filters: 'nc01'

- first dimension is the output channel ('n')
- second dimension is the input channel ('c')
- last two dimensions are the data ('0', '1')

Memory layout for images: 'bc01'

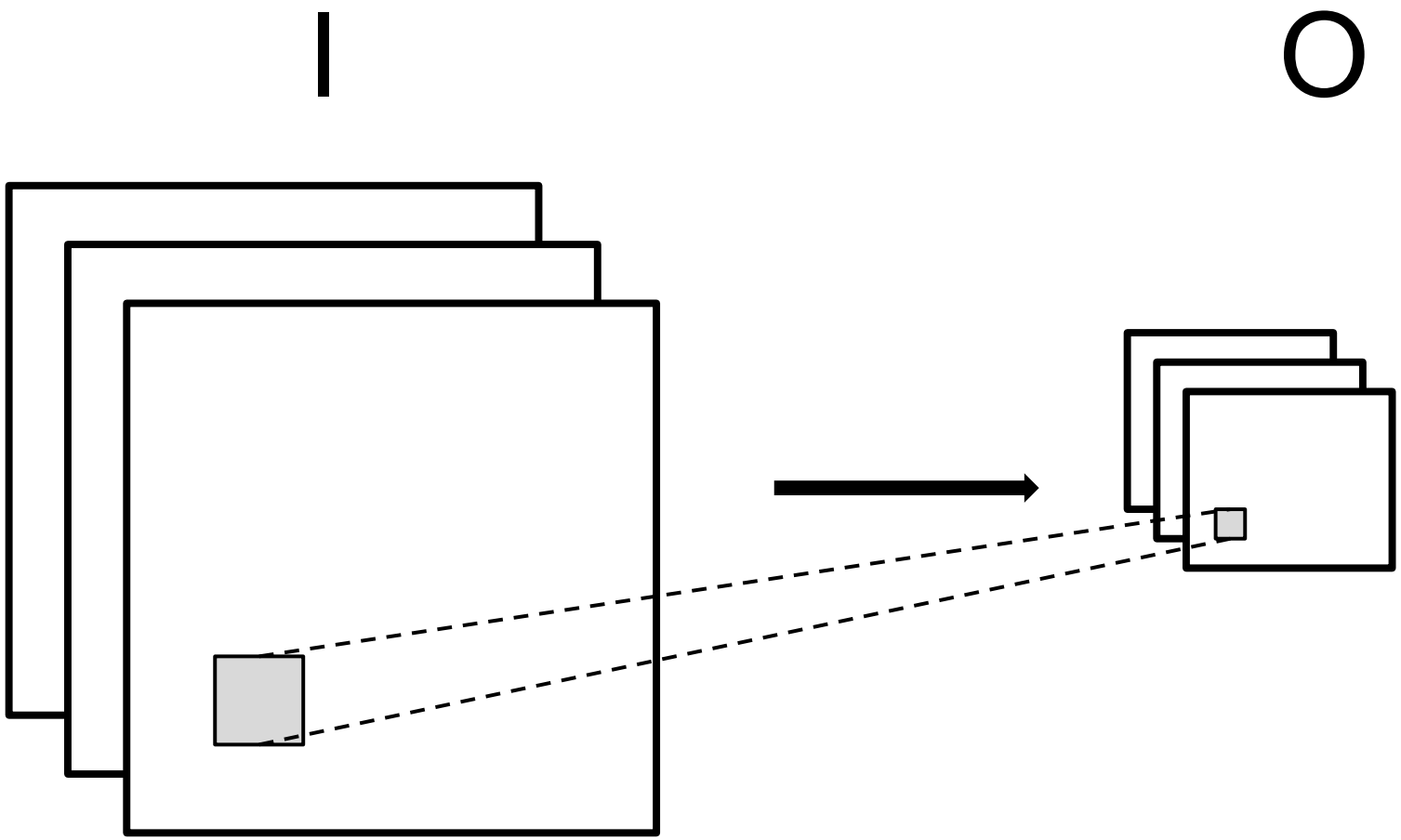
- first dimension is the batch ('b')
- second dimension is the channel ('c')
- last two dimensions are the data ('0', '1')

Memory layout for filters: 'nc01'

- first dimension is the output channel ('n')
- second dimension is the input channel ('c')
- last two dimensions are the data ('0', '1')

Some other packages may use different conventions.

Basic pooling operation



Max Pooling

$$O_k(i,j) = \max_{\substack{0 \le x < m \\ 0 \le y < n}} I_k(x+i, y+j)$$

Average Pooling

$$O_k(i,j) = \frac{1}{mn} \sum_{m-1}^{x=0} \sum_{n-1}^{y=0} I_k(x+i, y+j)$$

- k is the channel.
- (m, n) is the size of the filters.