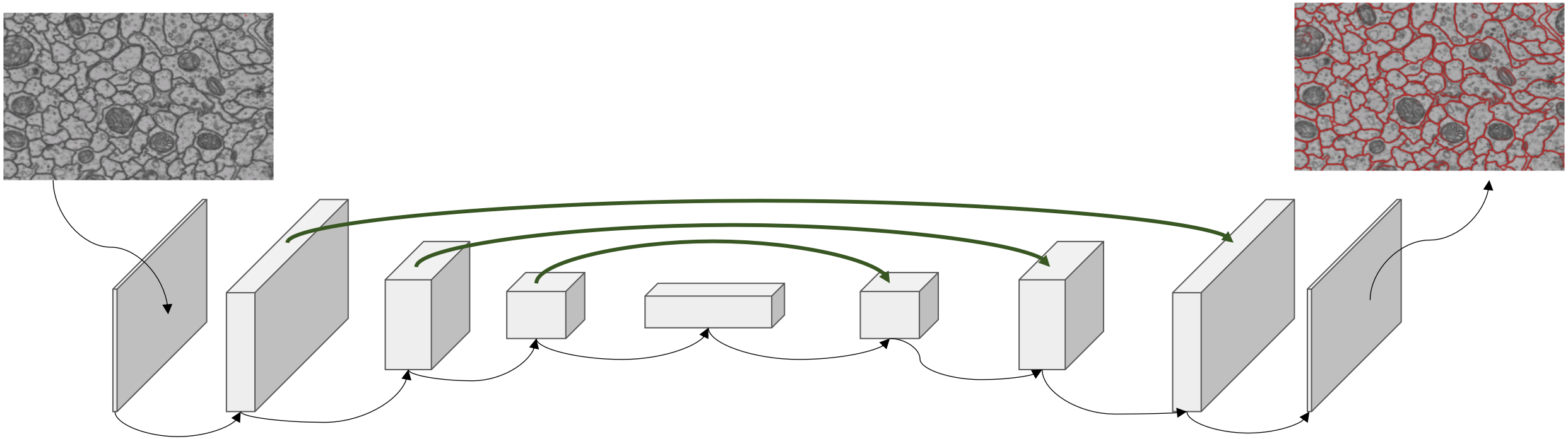
U-Net

Shalin Mehta, CZ Biohub
Anna Kreshuk, EMBL

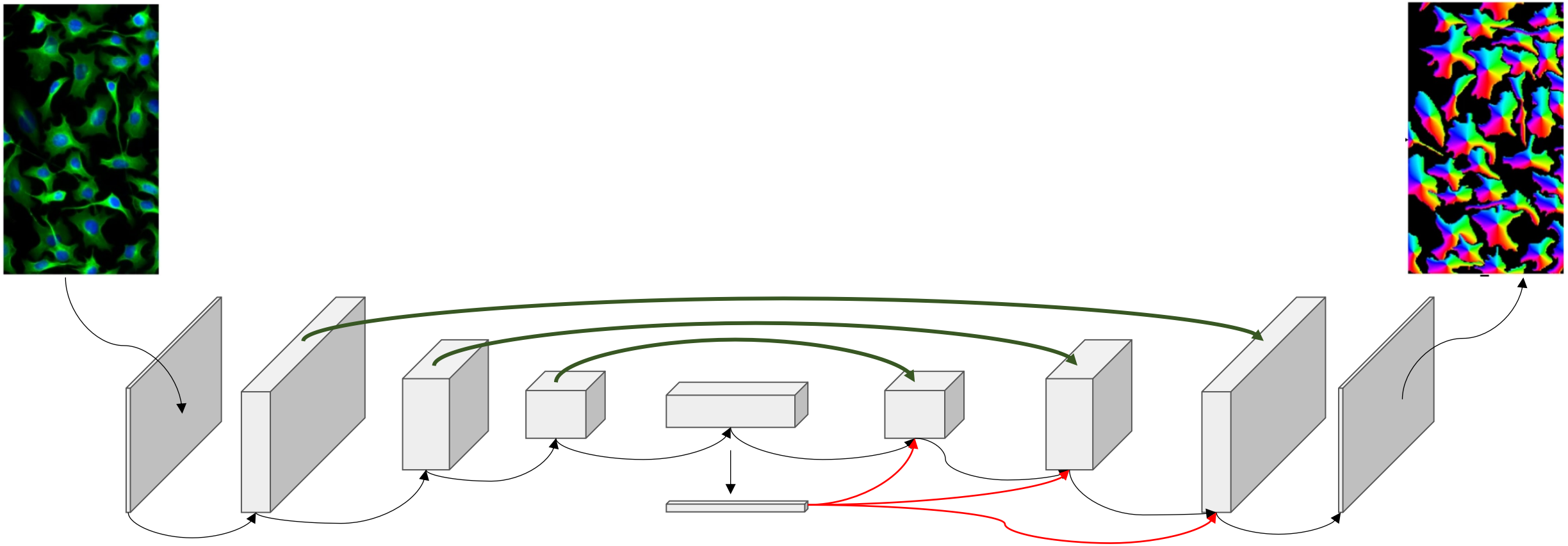
DL@MBL 2024

U-net – the baseline for image transforms



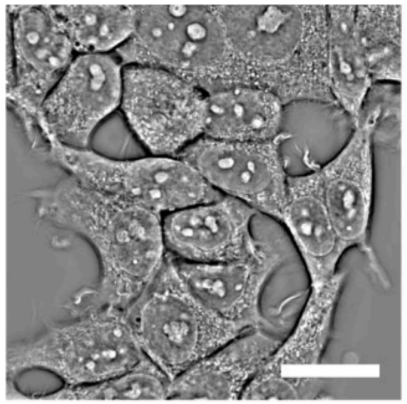
[Ronneberger et al., 2015]

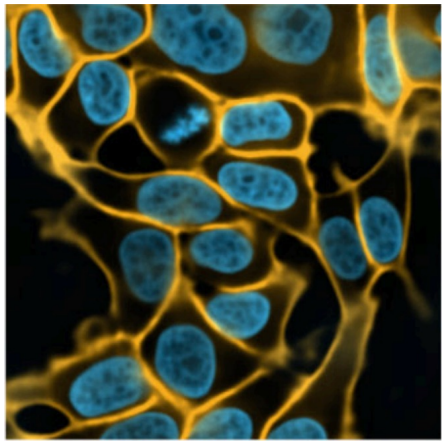
U-net – the baseline for image transforms

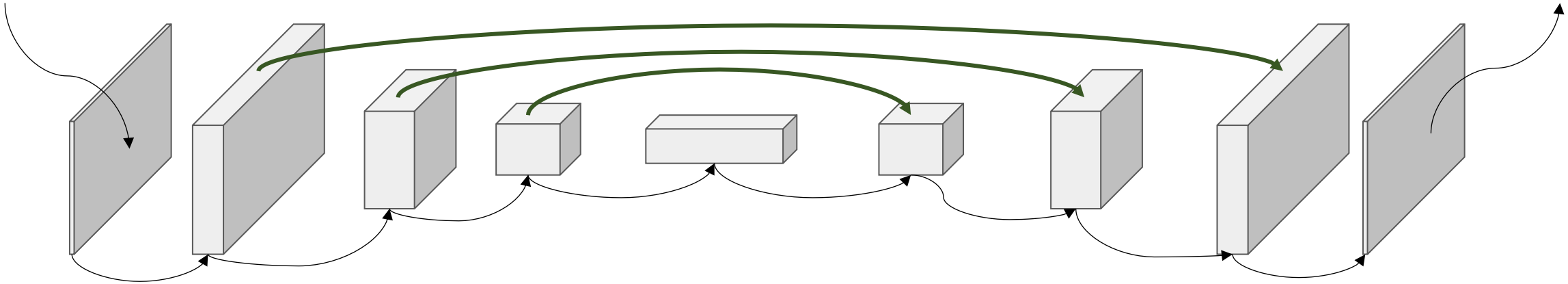


[Stringer et al., 2021]

U-net – the baseline for image transforms







[Liu et al., 2024]

image transforms

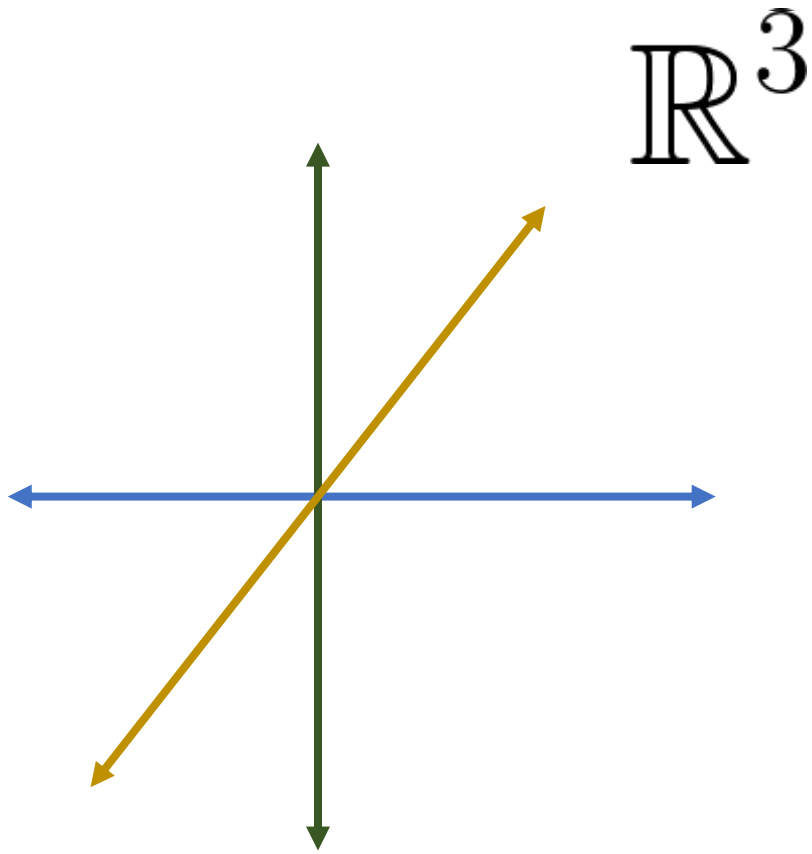
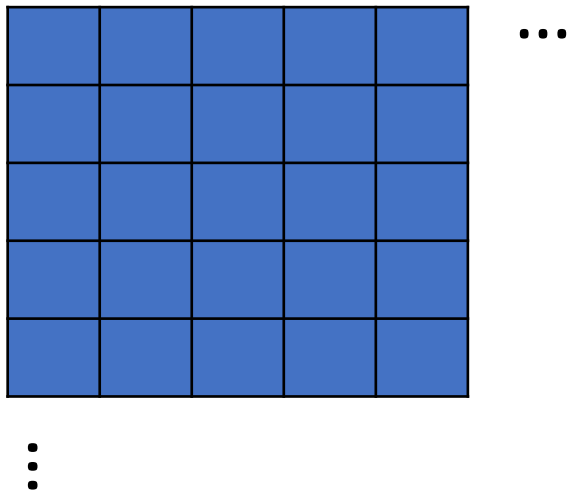


image transforms

$$\mathbb{R}^{H \times W}$$



...

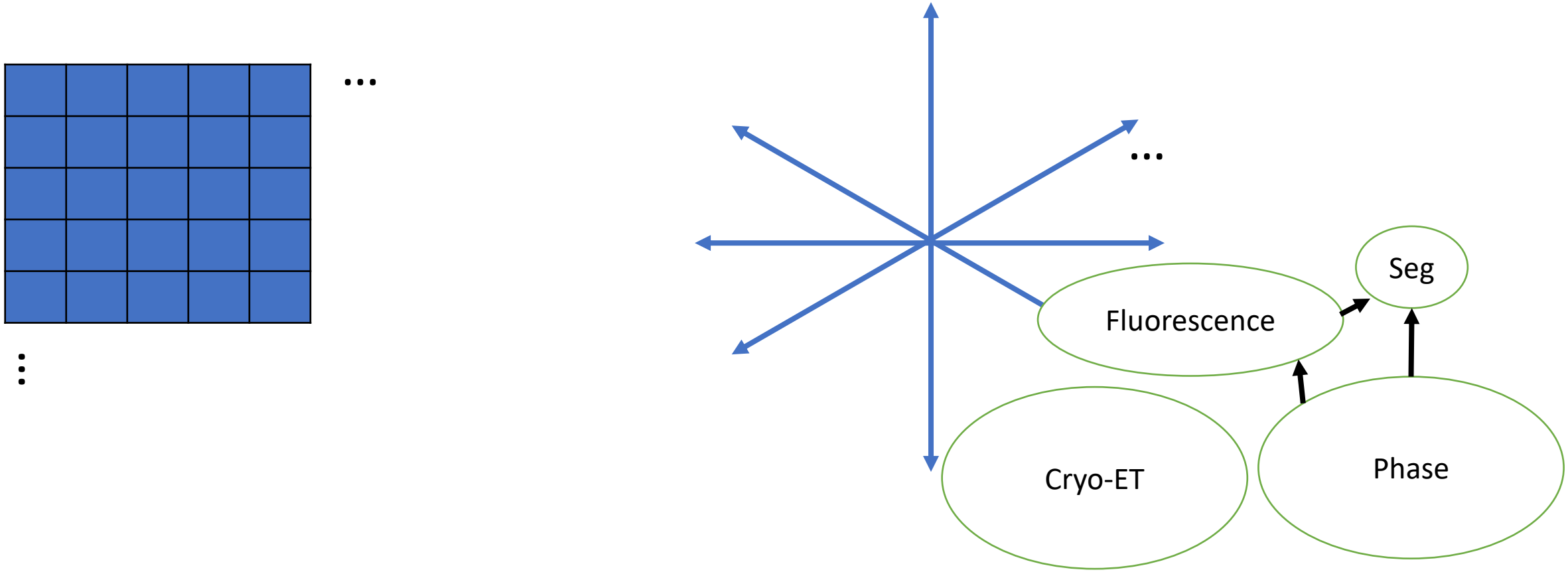
...

Natural images

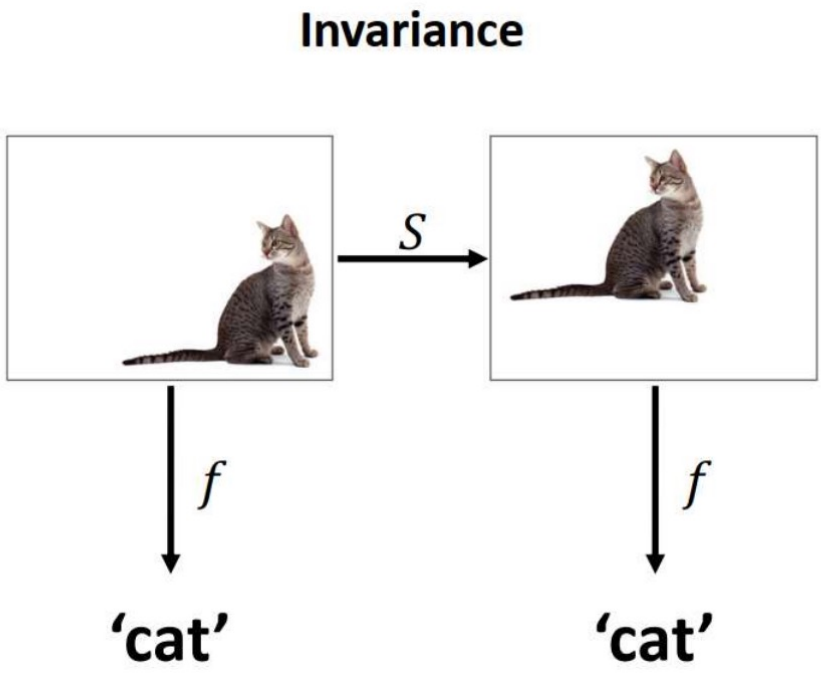
Images of molecules, cells, and tissues

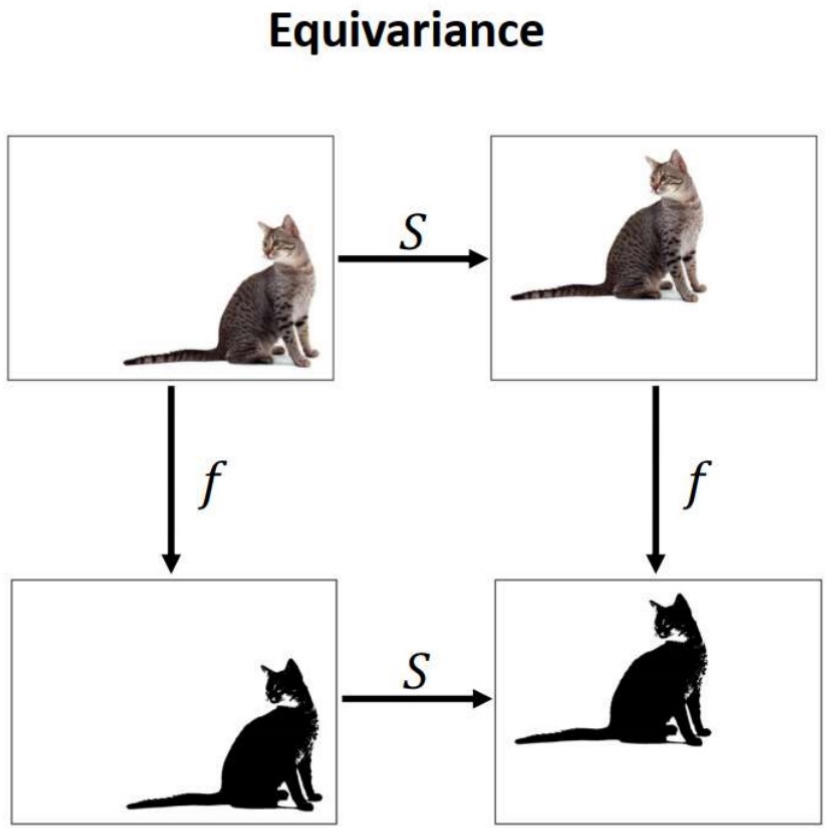
image transforms

$$\mathbb{R}^{H \times W}$$

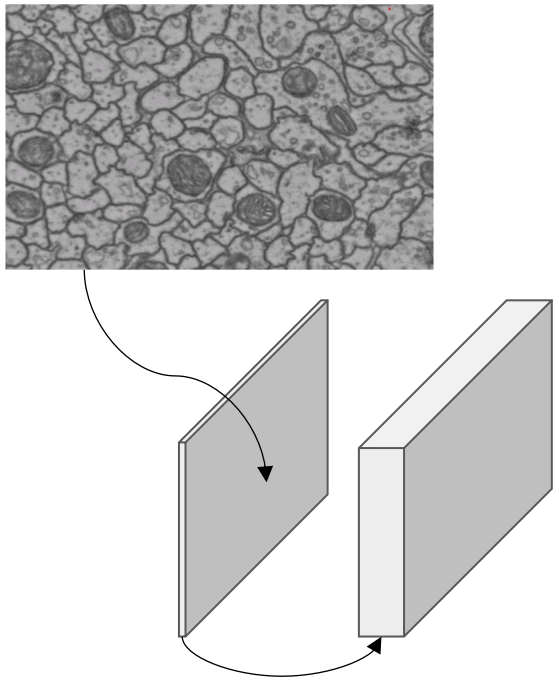


Shift equivariance





Unpacking the boxes:



- Convolutions
- Activation
- Downsampling /Upsampling

Convolutional layer

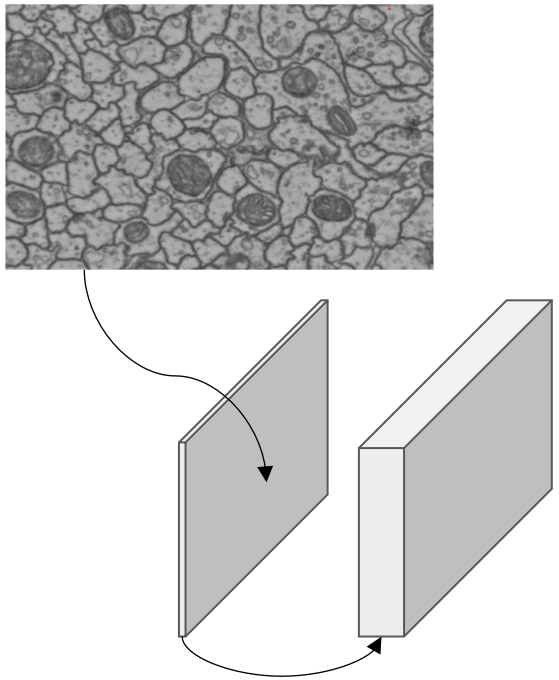
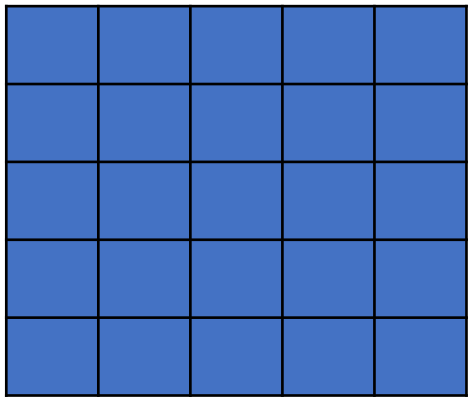
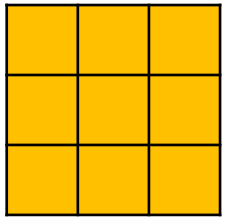


Image transforms, parametrized by a matrix of weights (usually 3x3)

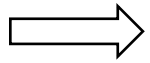


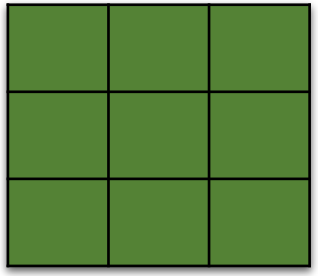
Input image

*



Weights





One channel

Convolutional layer

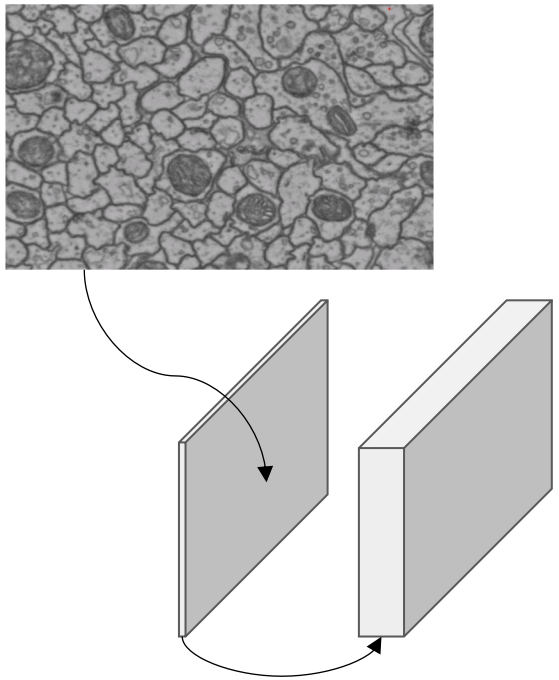
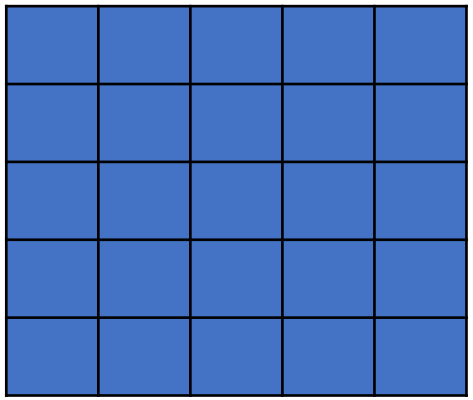
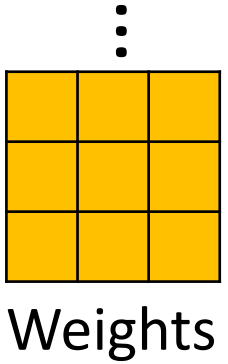


Image transforms, parametrized by a matrix of weights (usually 3x3)

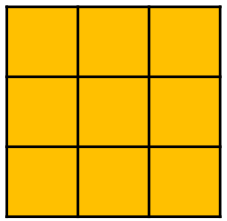


Input image

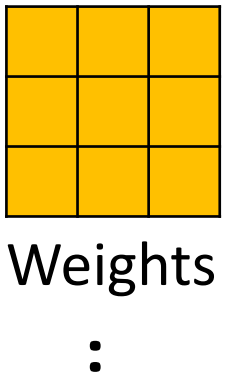
*



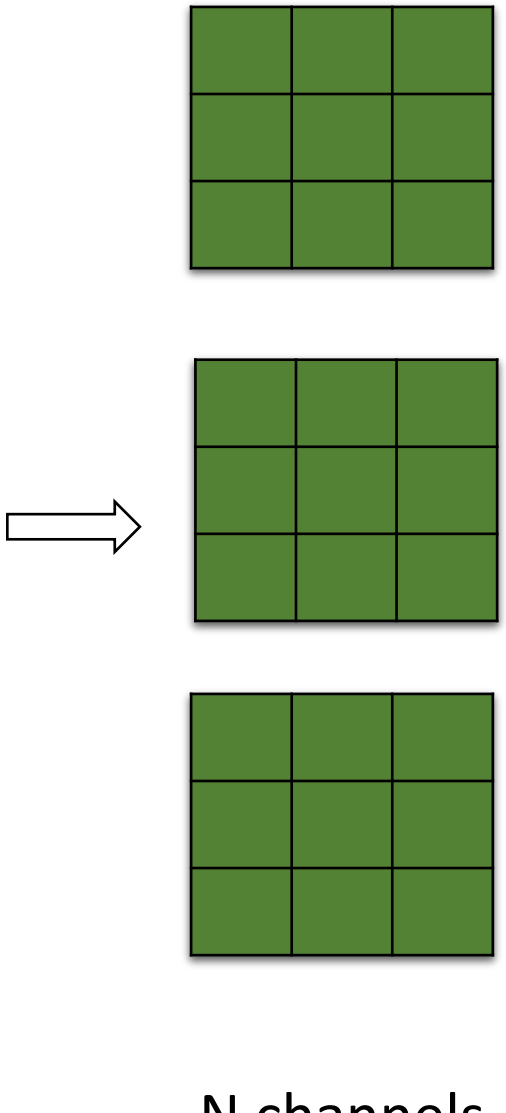
Weights



Weights

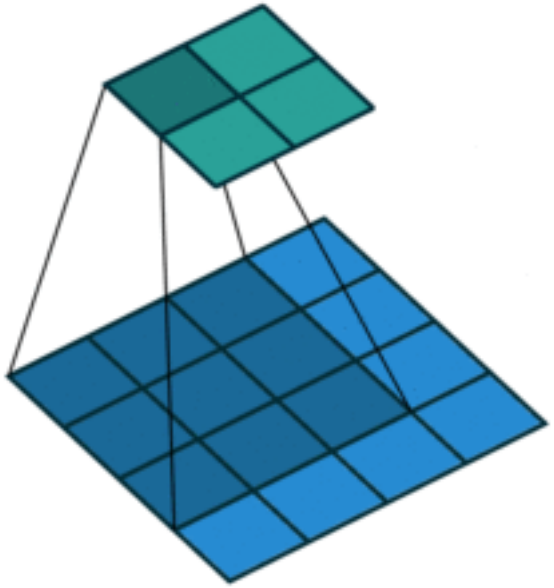


Weights

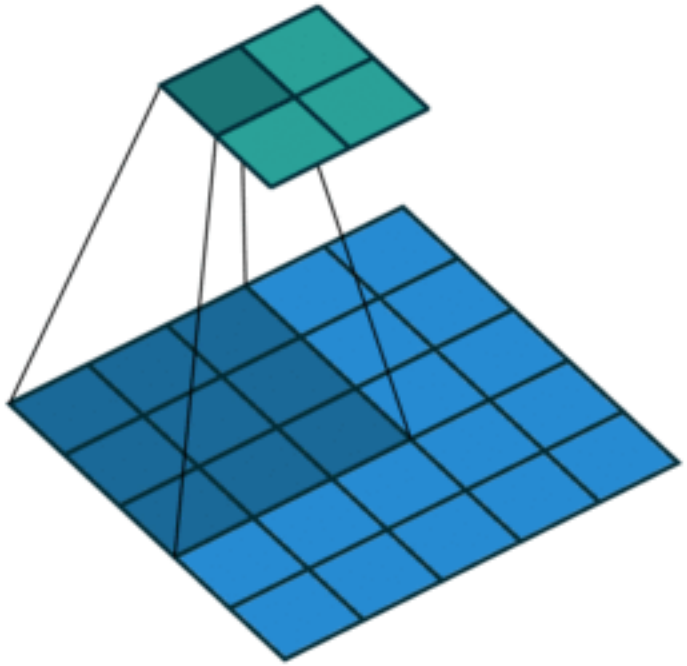


N channels

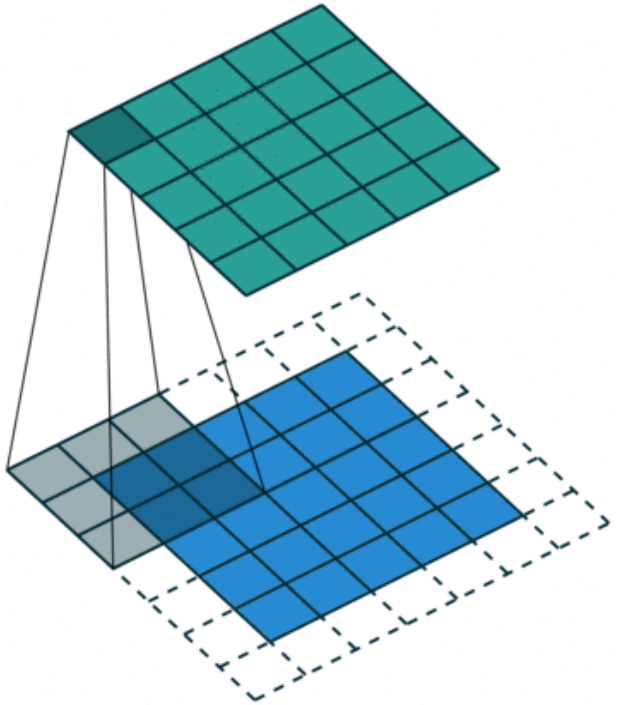
Parameters of convolution



padding = 0, stride = 1



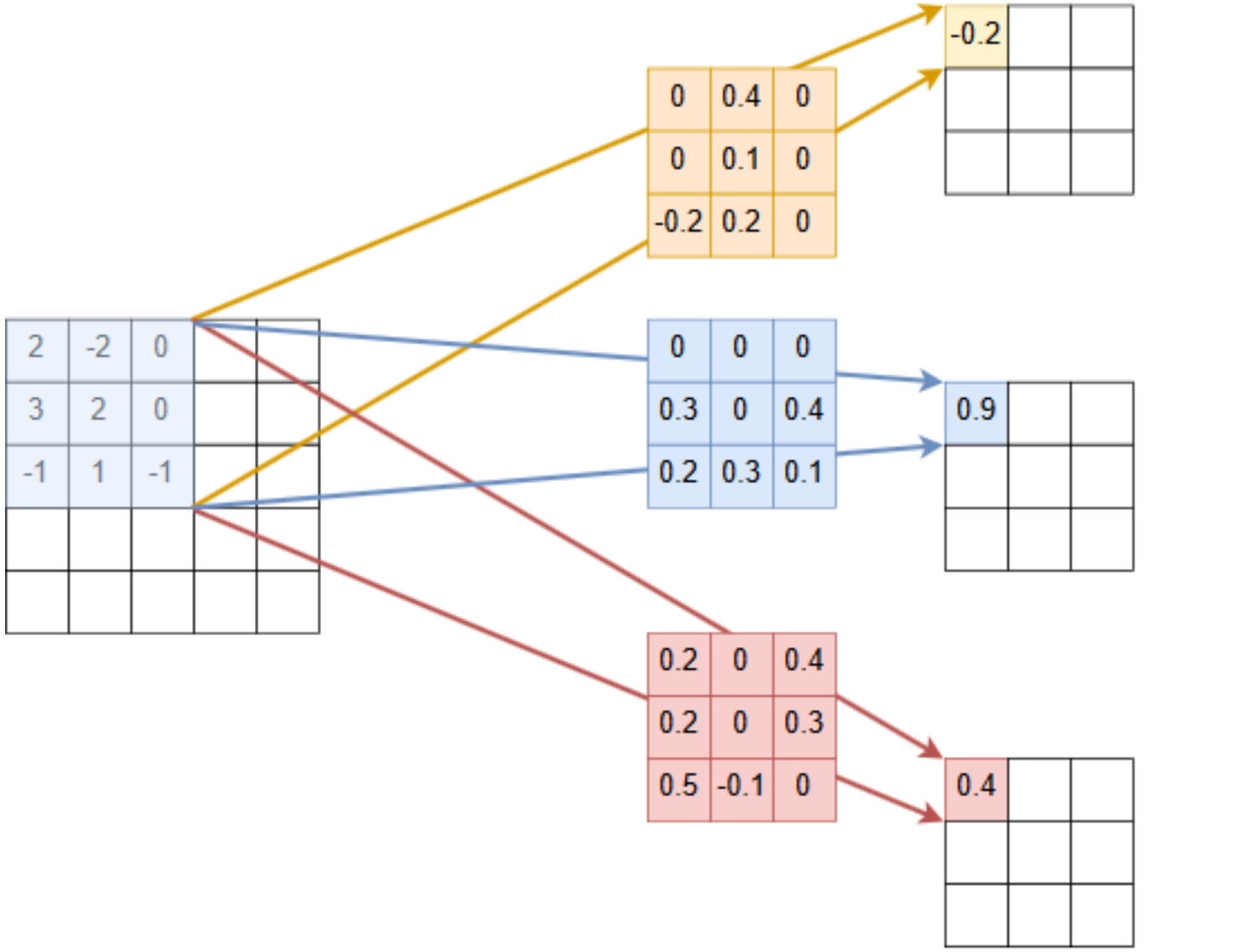
padding = 0, stride = 2



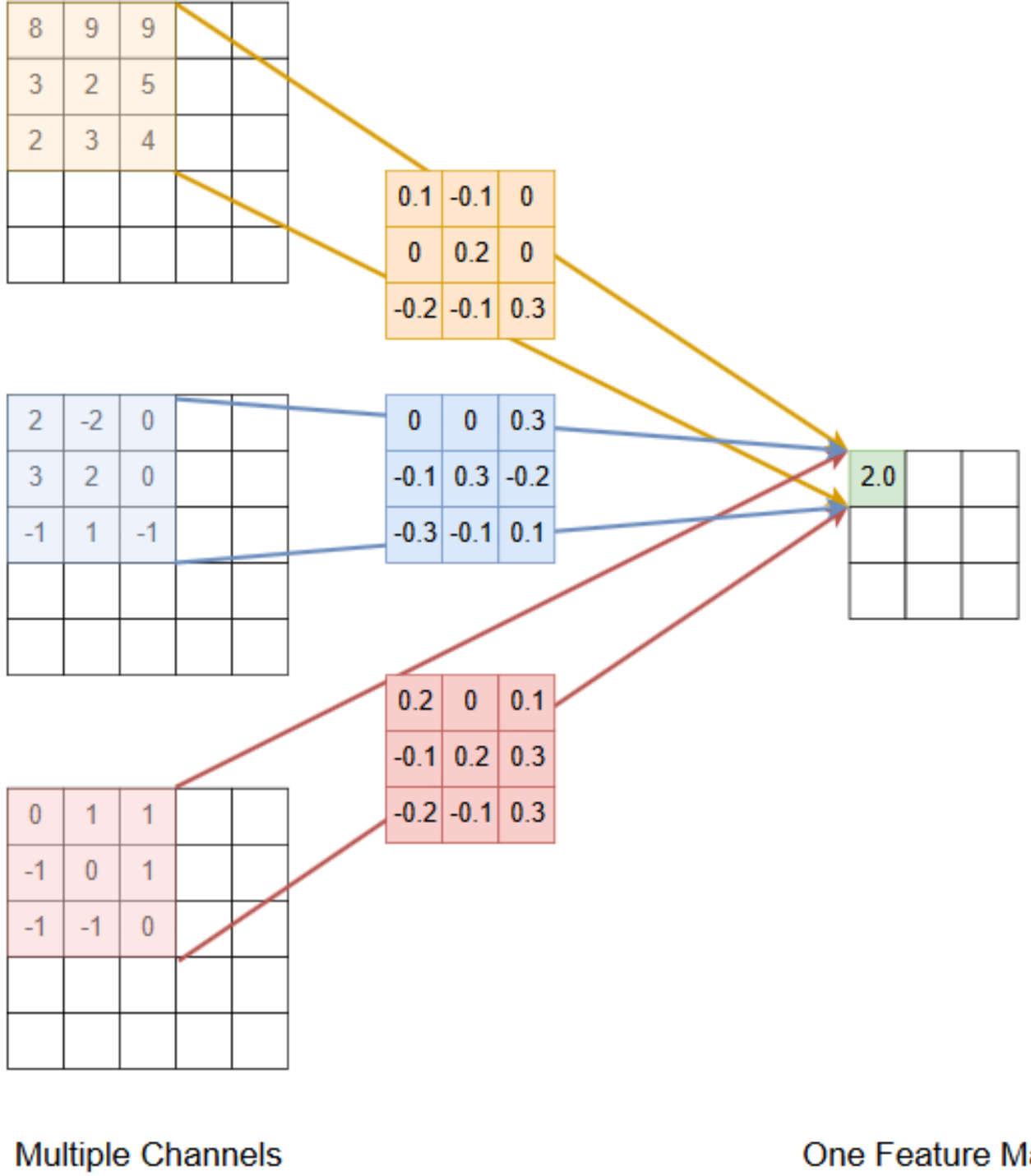
padding = 1, stride = 1

https://hannibunny.github.io/mlbook/neuralnetworks/convolutionDemos.html

Multi-channel output

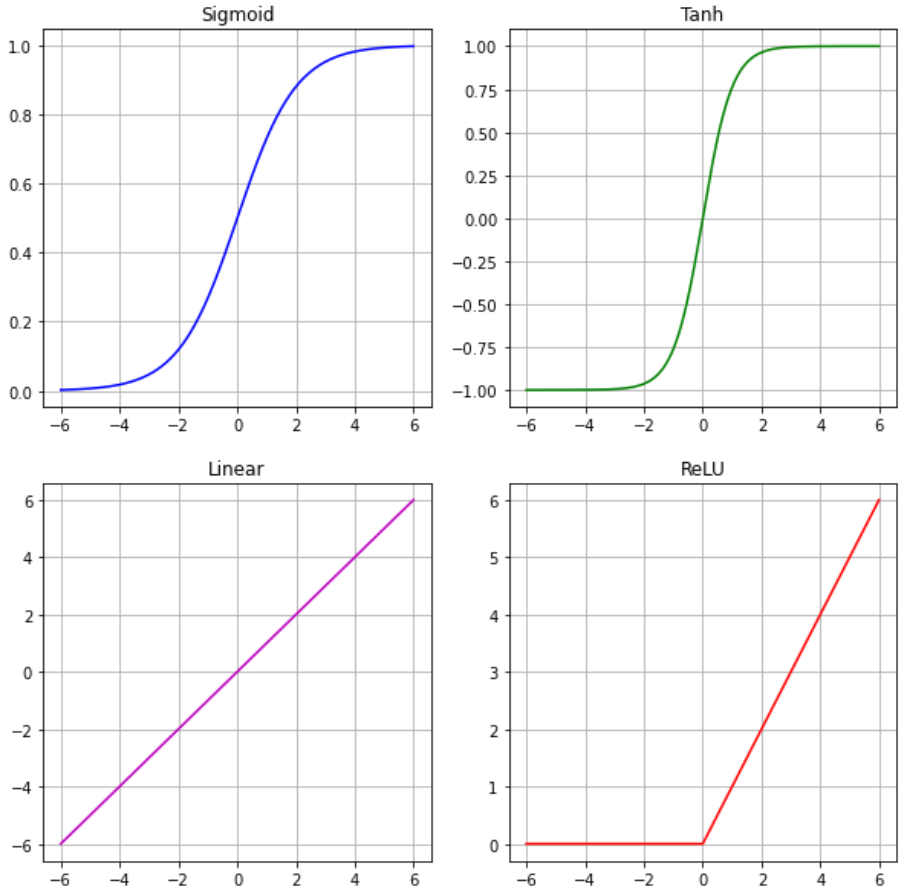


Multi-channel input



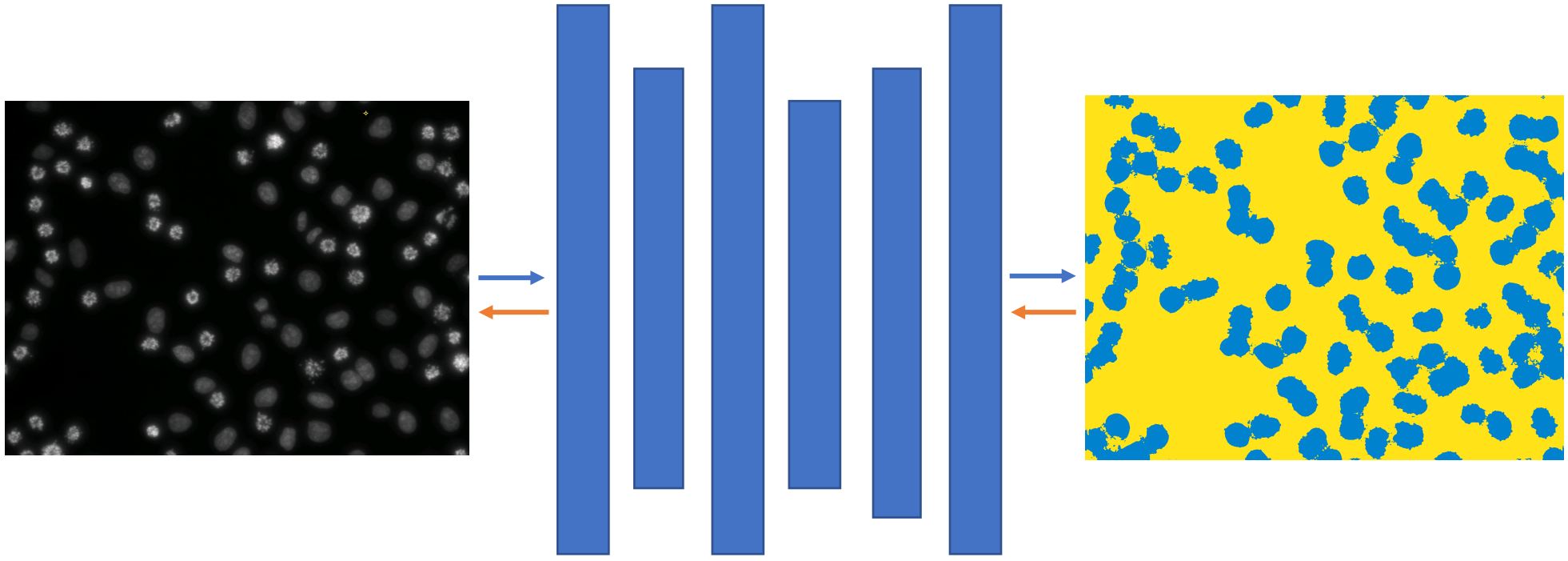
Activation functions

Update feature-maps element-wise



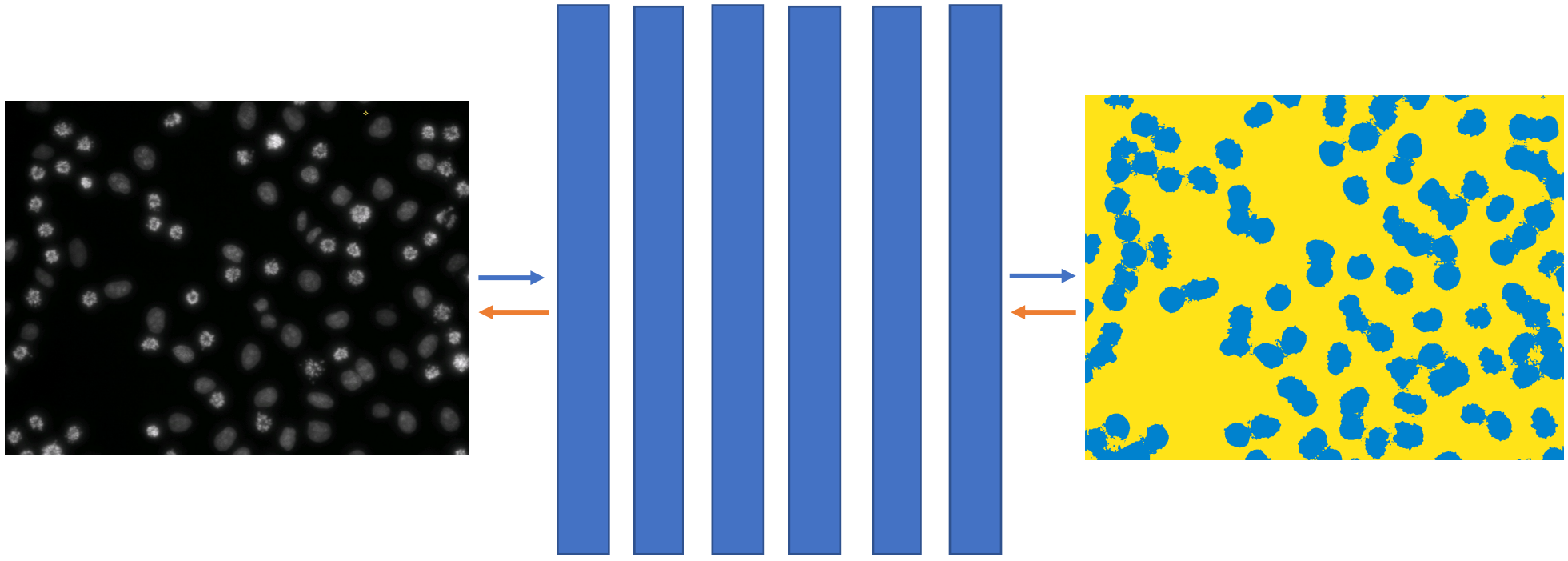
Network architecture

Why are these different size?



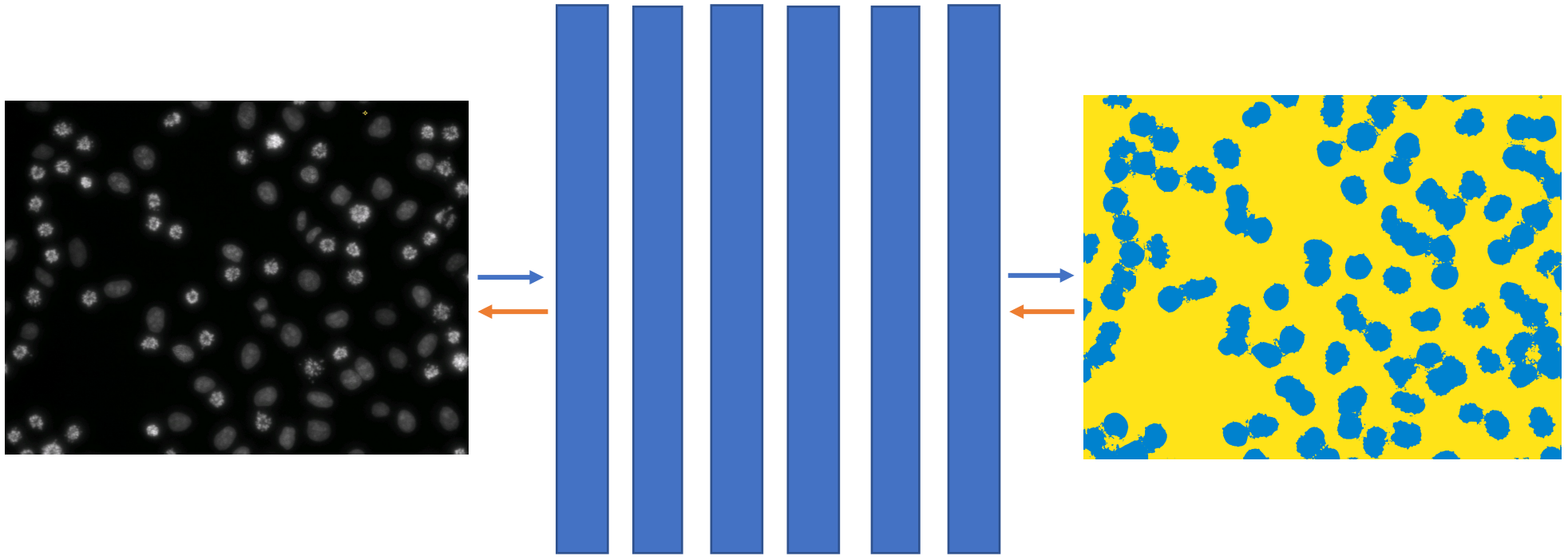
Network architecture

What if they were all the same?



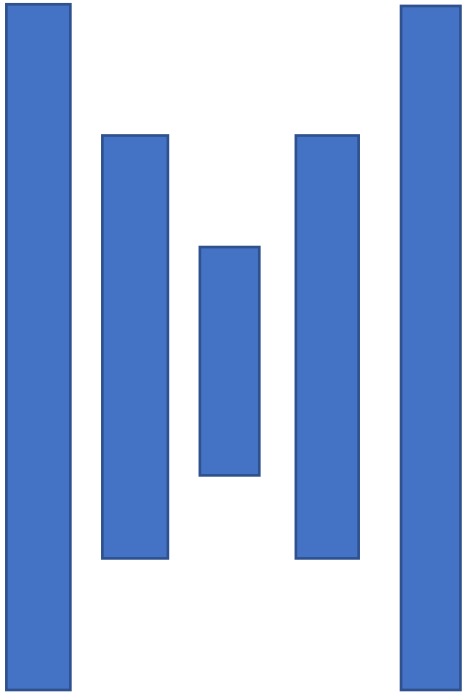
Network architecture

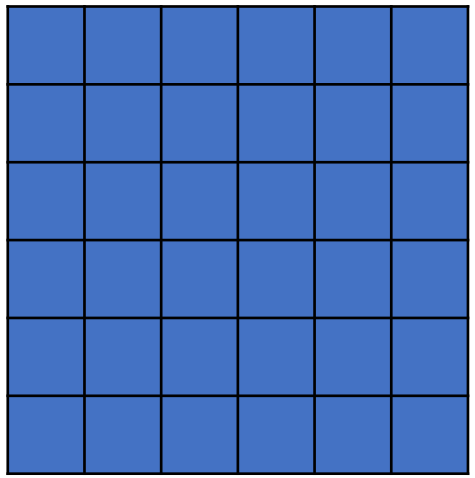
What if they were all the same?



If they are all convolutional, **no context**

Downsampling

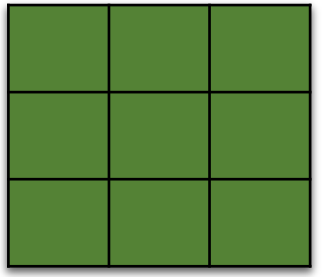




Some layer

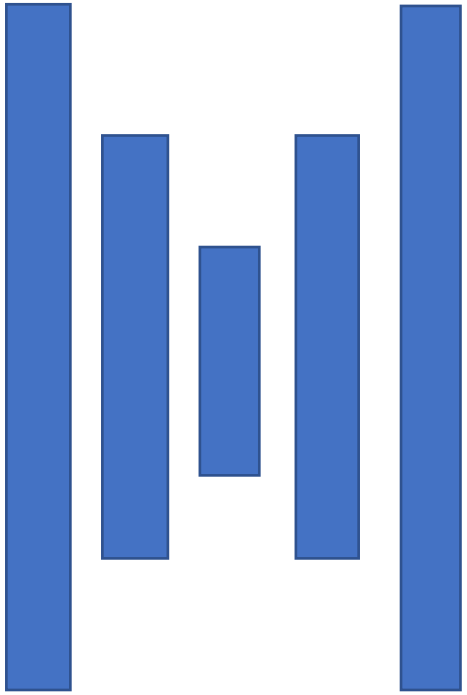
*

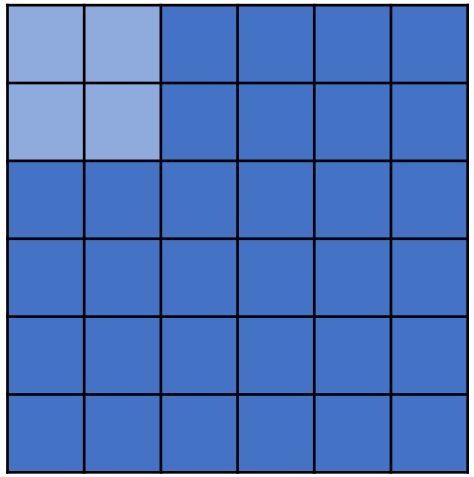
Max/Avg
Pooling



Next layer

Downsampling

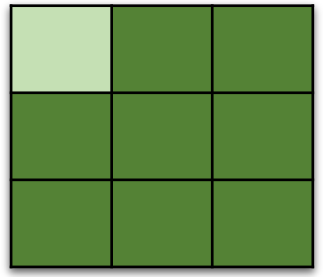




Some layer

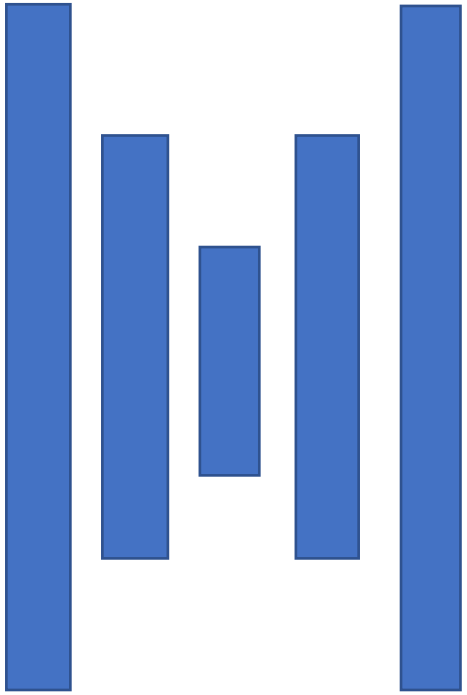
*

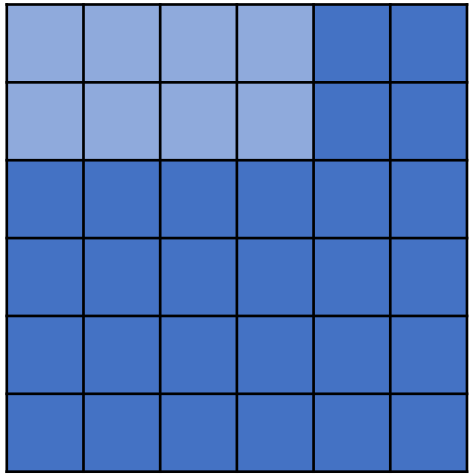
Max/Avg
Pooling



Next layer

Downsampling

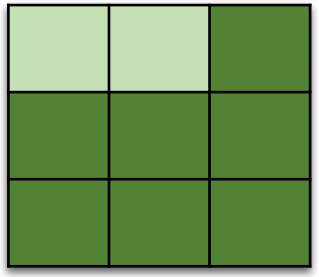




Some layer

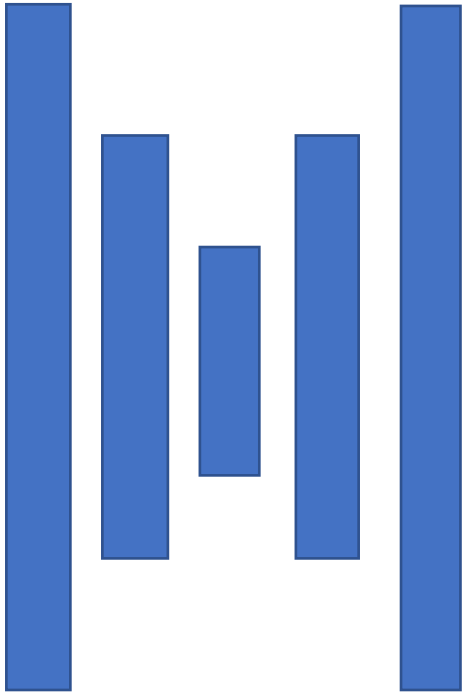
*

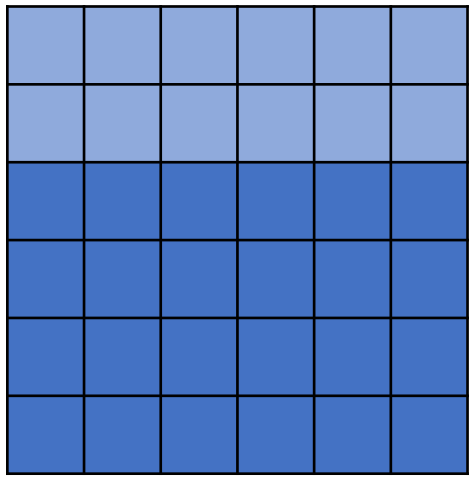
Max/Avg
Pooling



Next layer

Downsampling

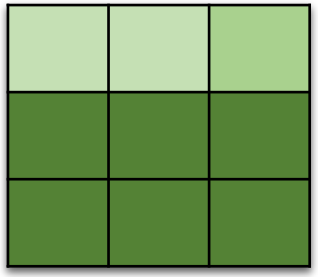




Some layer

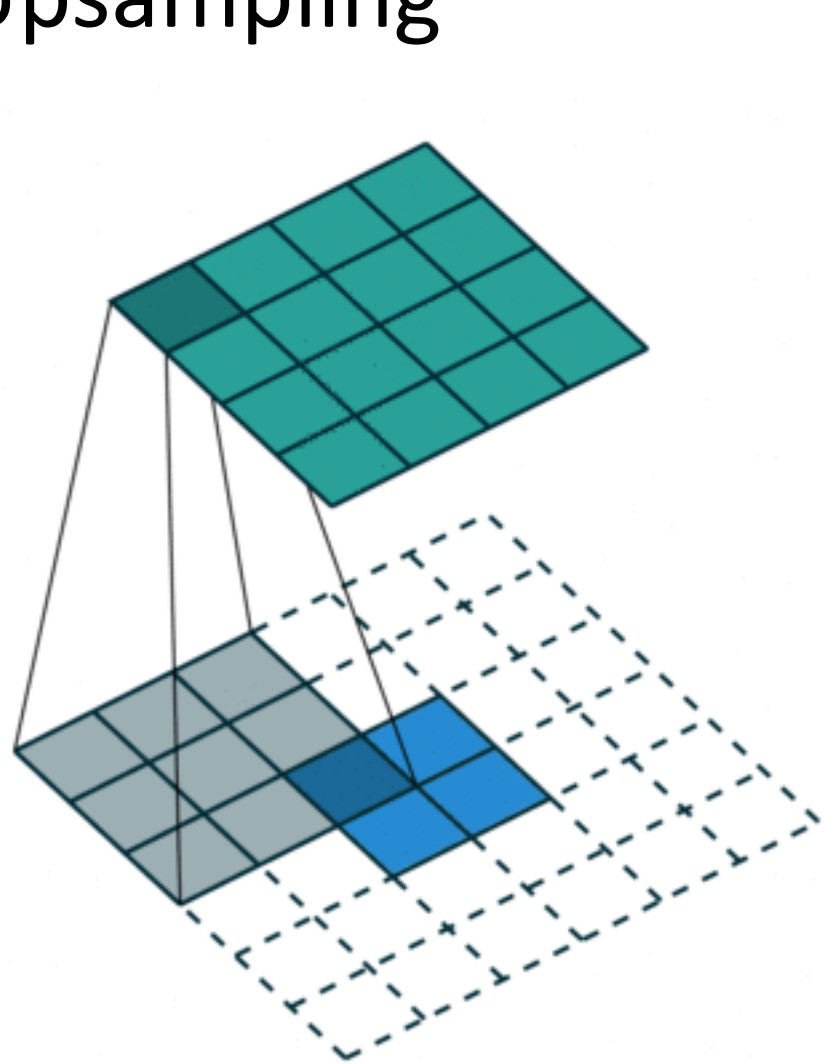
*

Max/Avg
Pooling

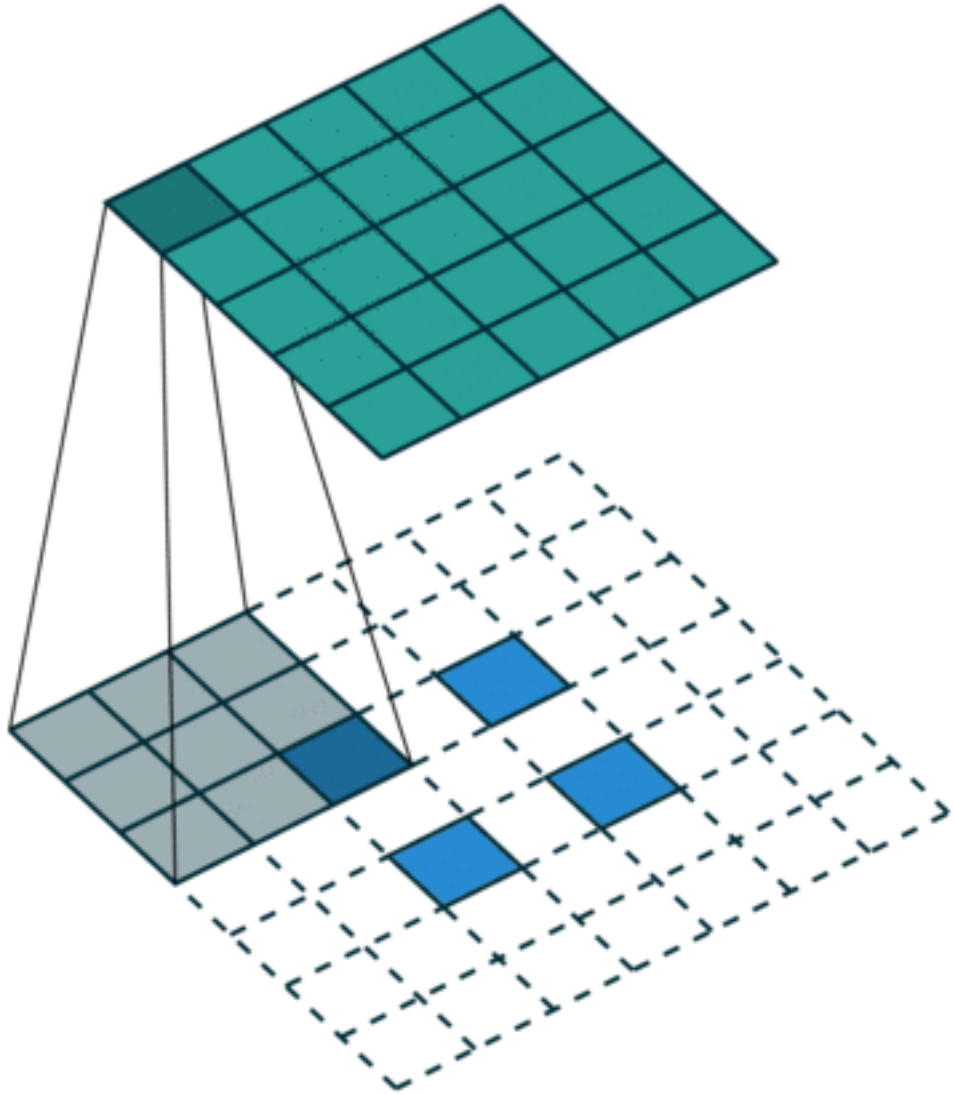


Next layer

Upsampling

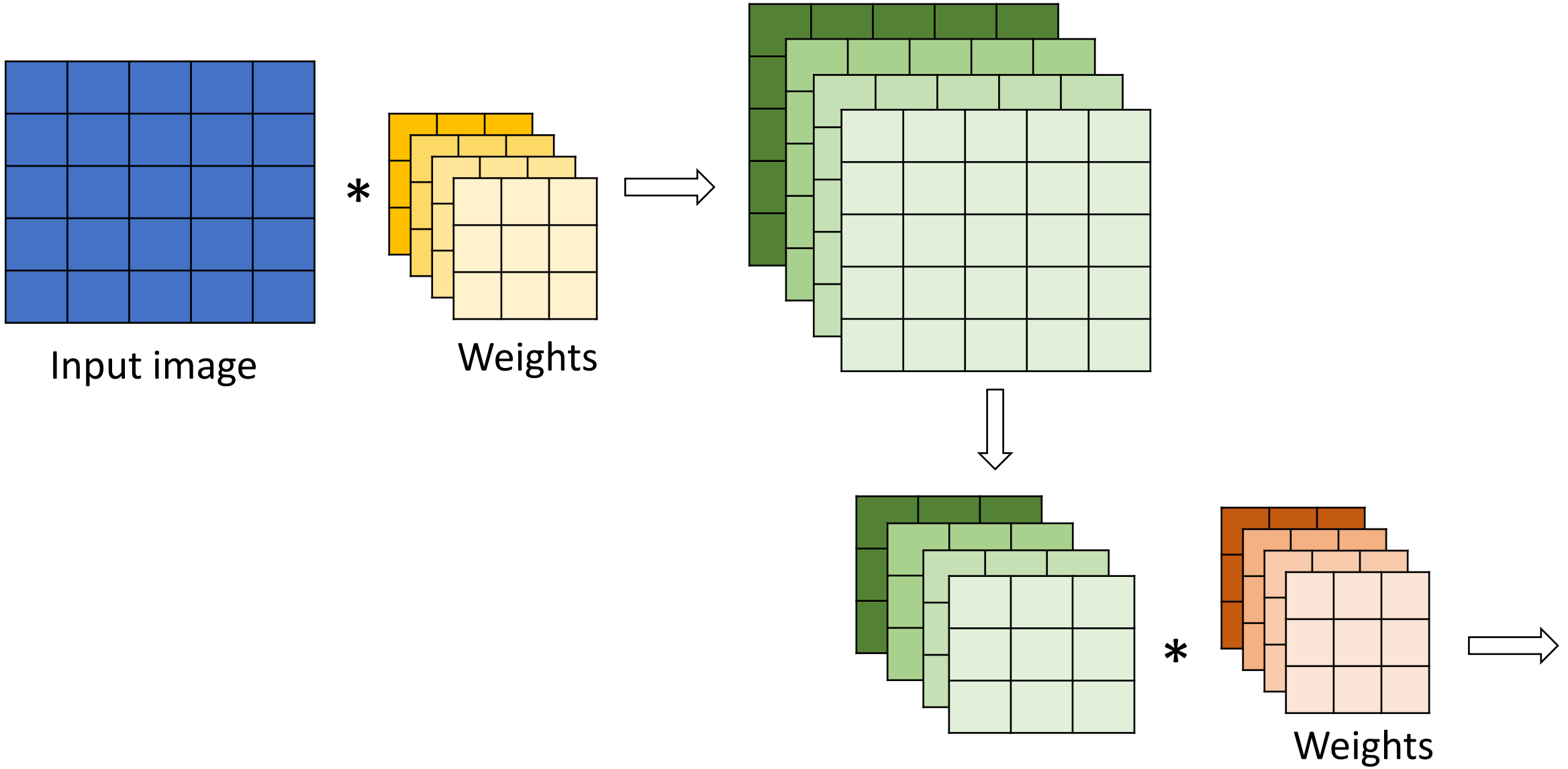


stride = 1

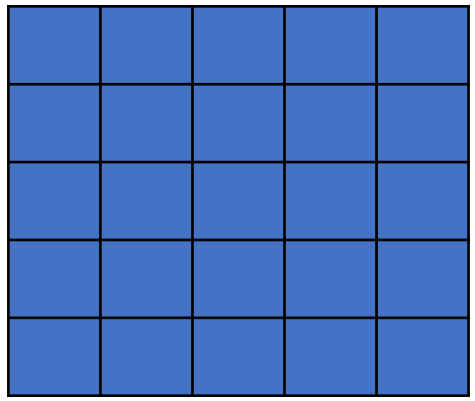


stride = 2

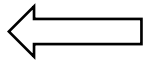
Building an encoder

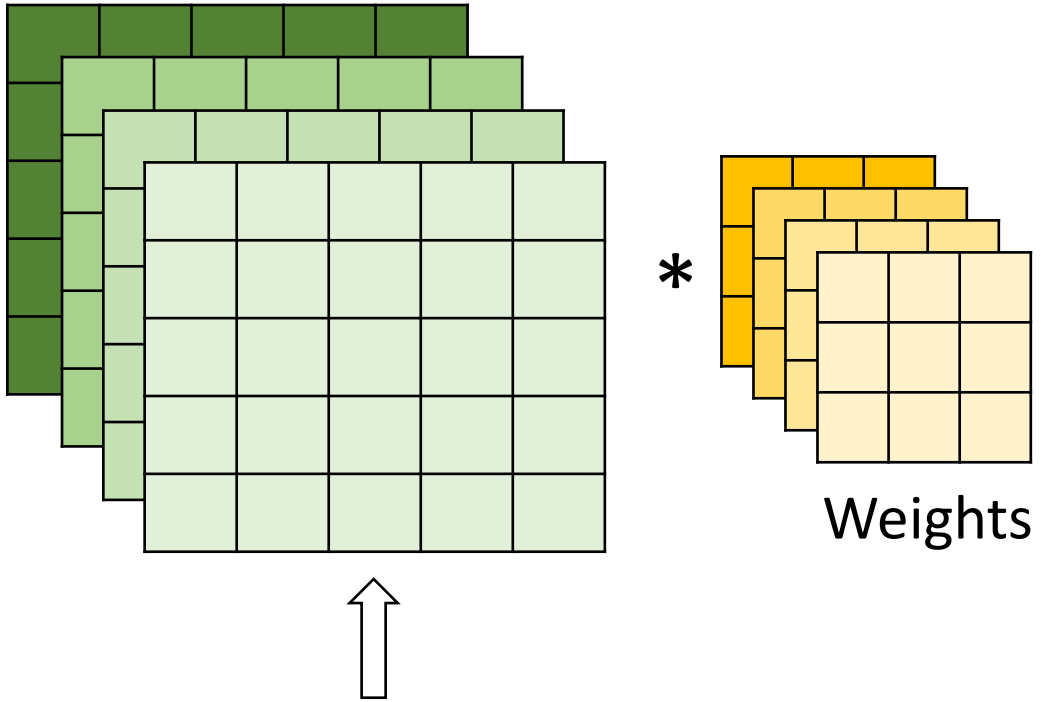


Building a decoder

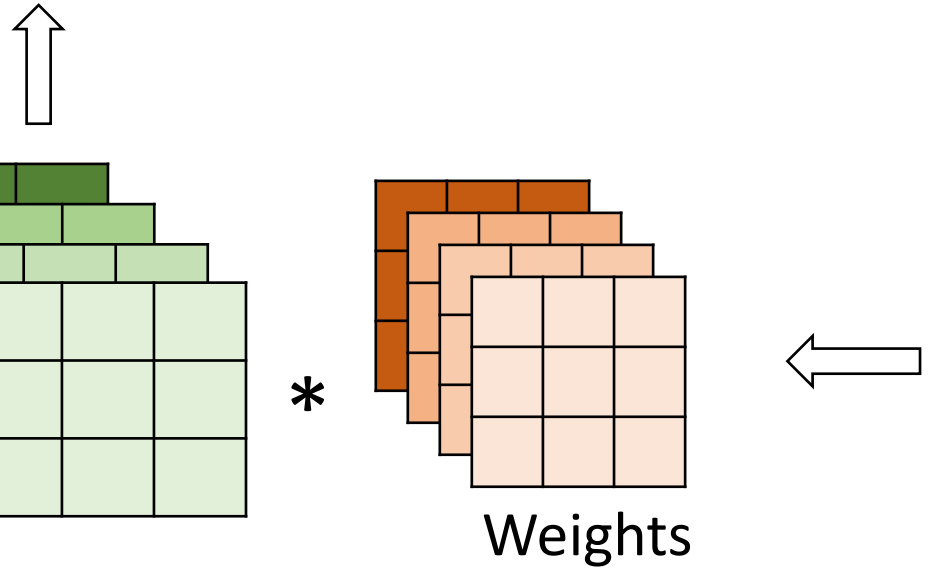


Output image





Weights

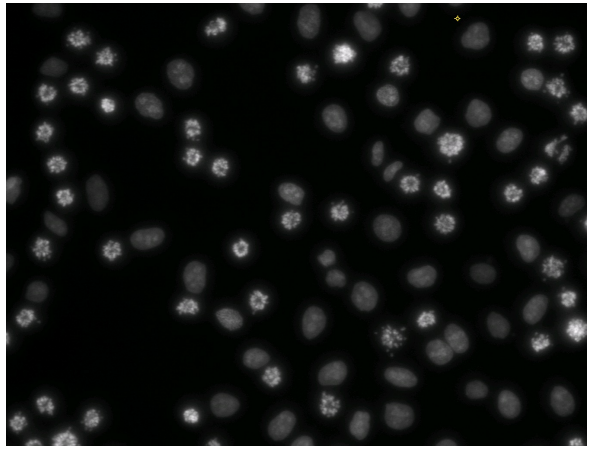


Weights

Image-to-image

Forward pass

Backpropagation



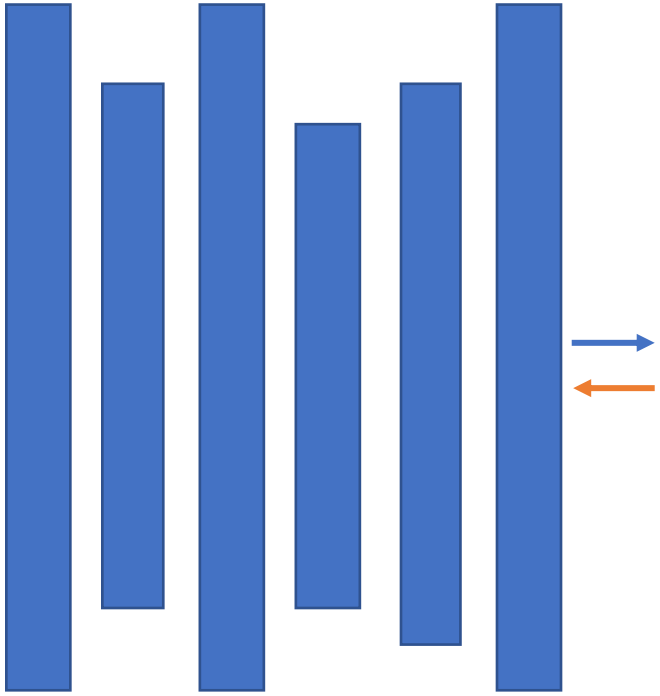


Image transforms: convolutions and pooling layers

Groundtruth

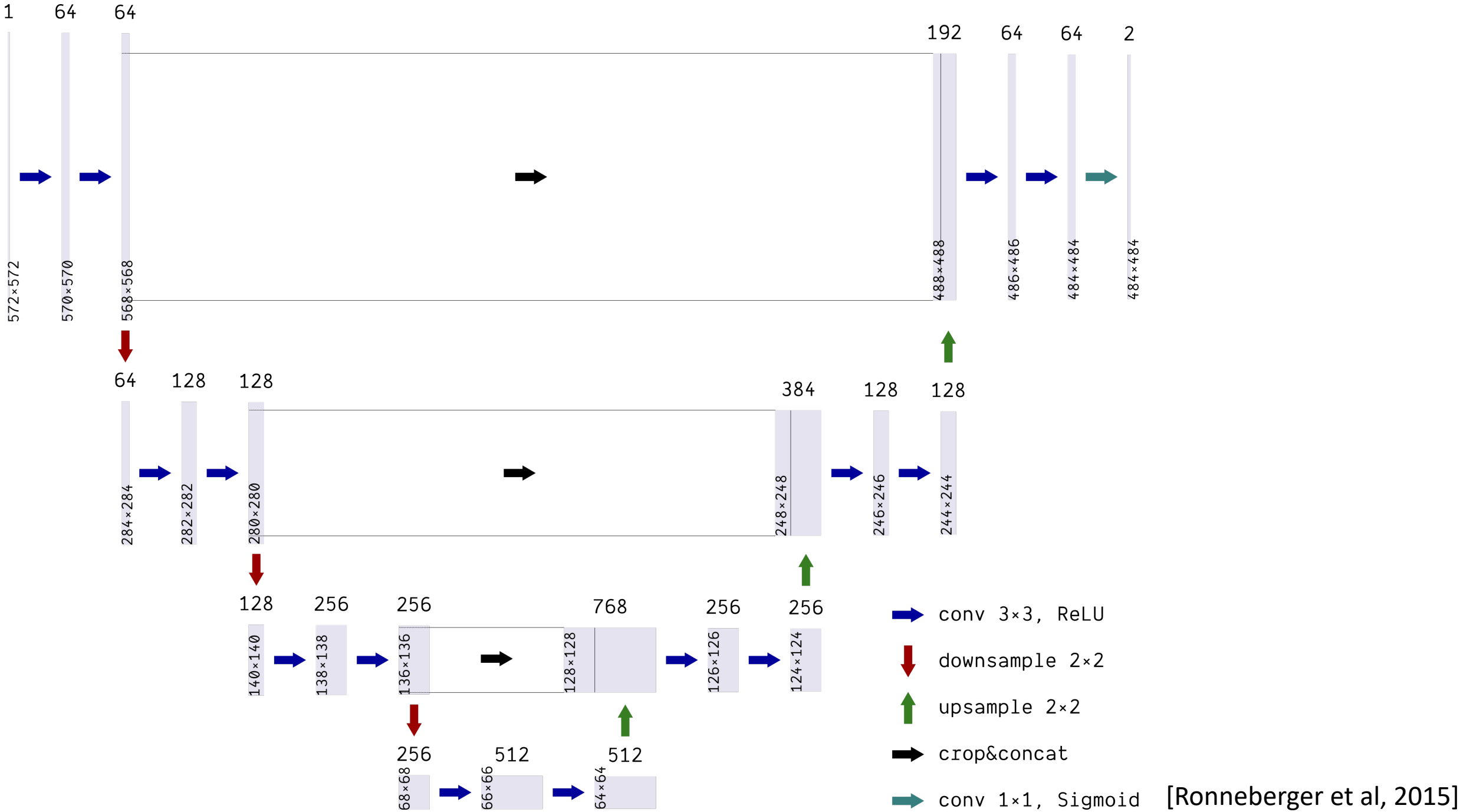
Some other image

Loss for training

Trying to become some other image

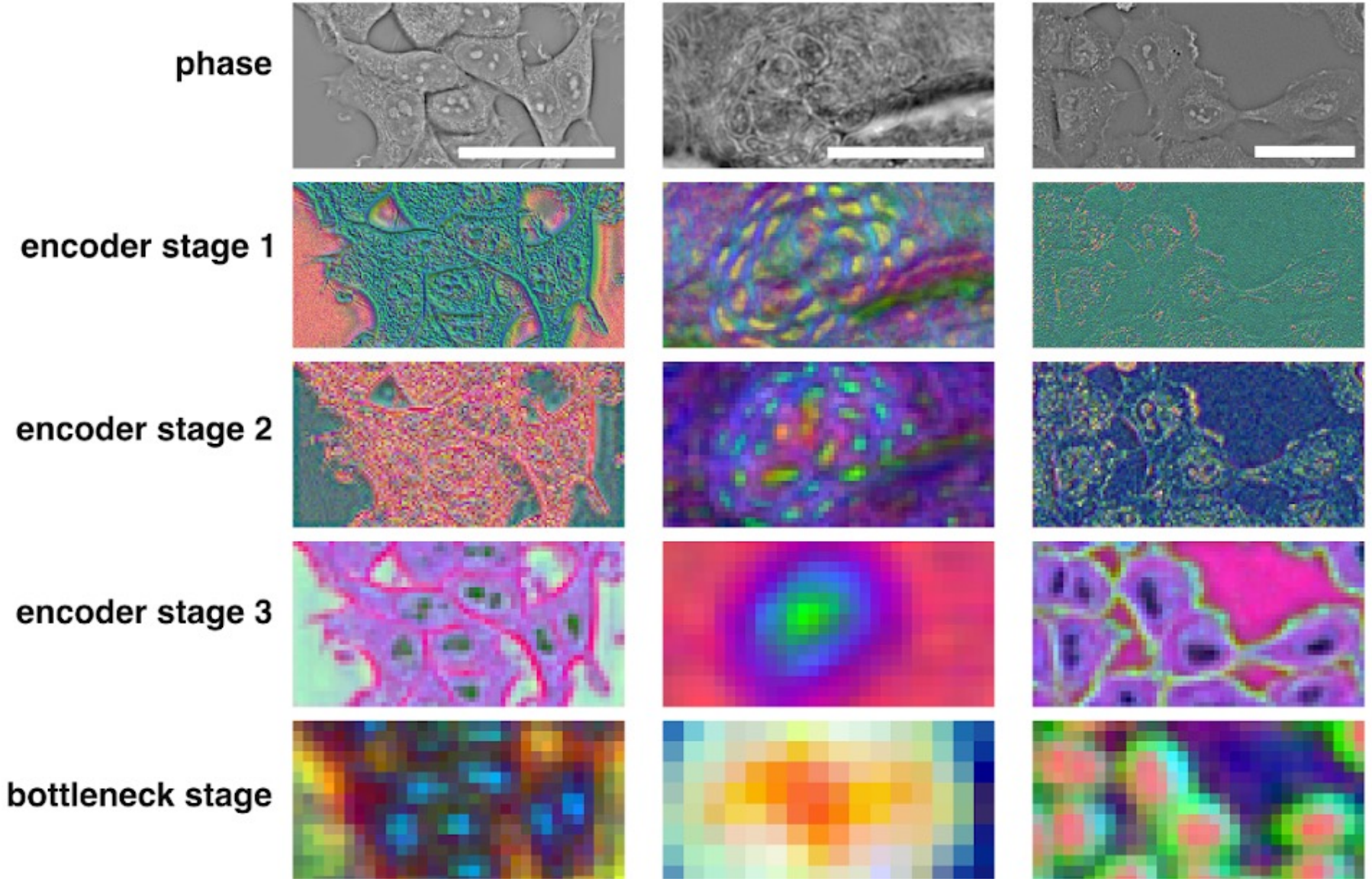
Current output: transformed image

Metric for validation



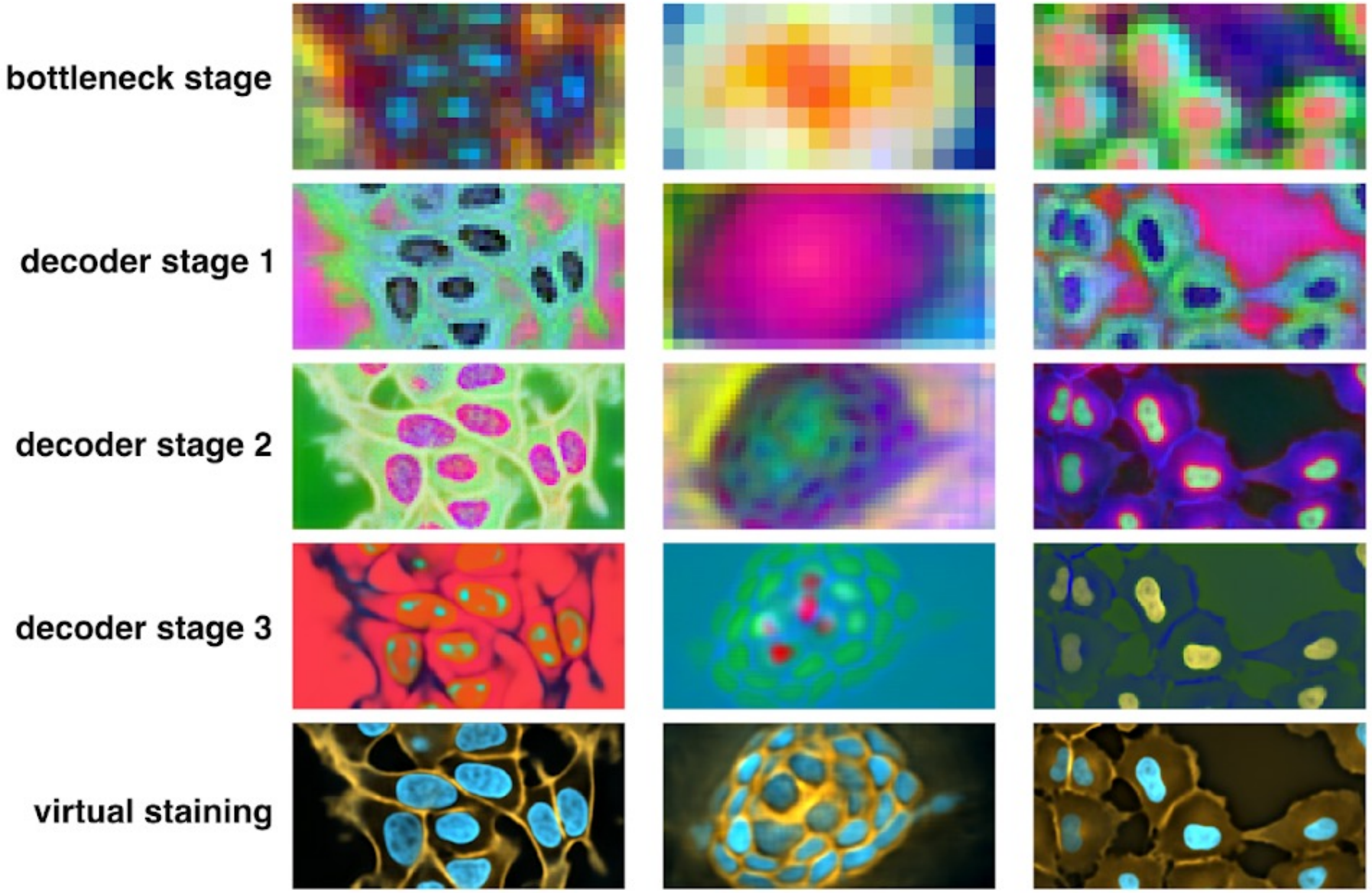
[Ronneberger et al, 2015]

Multi-scale representation



[Liu et al., 2024]

Multi-scale representation



[Liu et al., 2024]

Other things to consider

- Data
  - Normalization
  - Completeness of annotation
  - Sampling of patches
  - Batch-size
- Loss
  - Regression (denoising, translation): Mean square error, mean absolute error
  - Segmentation: Cross-entropy between classes
- Training protocol
  - Learning rate
  - Augmentations
  - End-to-end vs pre-training + fine-tuning

Go build your own U-Net!