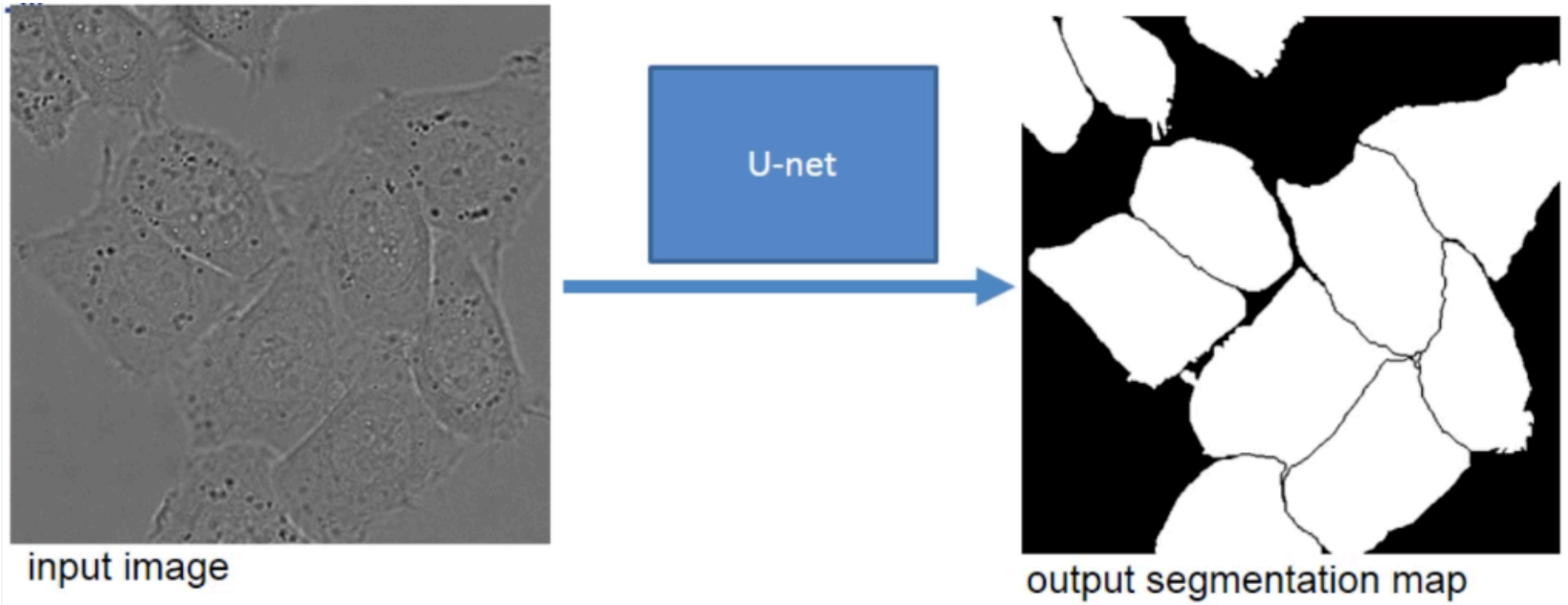
U-Net: Convolutional Networks for Biomedical Image Segmentation

ST7 Image&Sound - Theory Video

Xinjian OUYANG

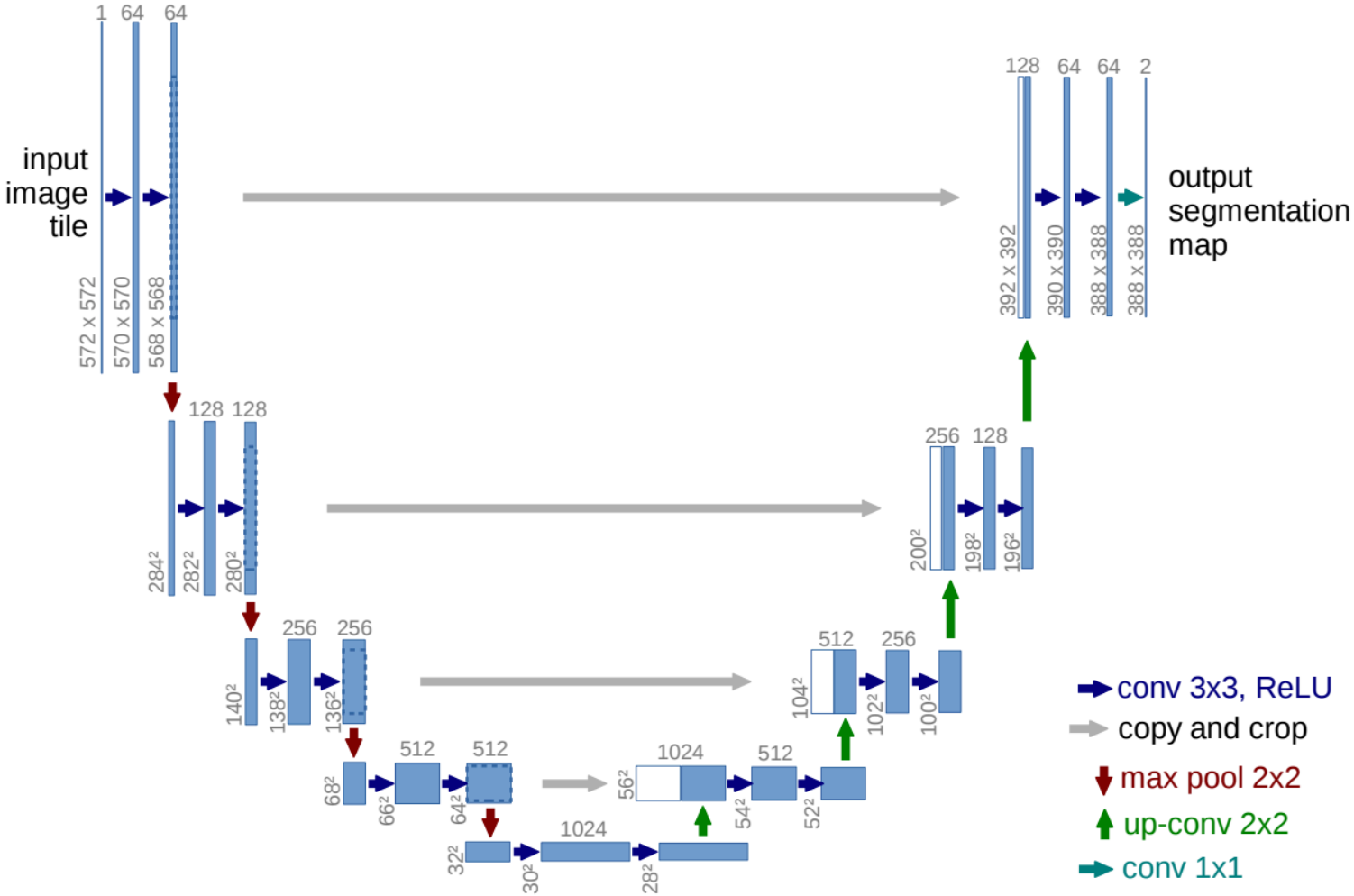
The general idea of U-net



End-to-end learning:

U-net learns segmentation in an end-to-end setting

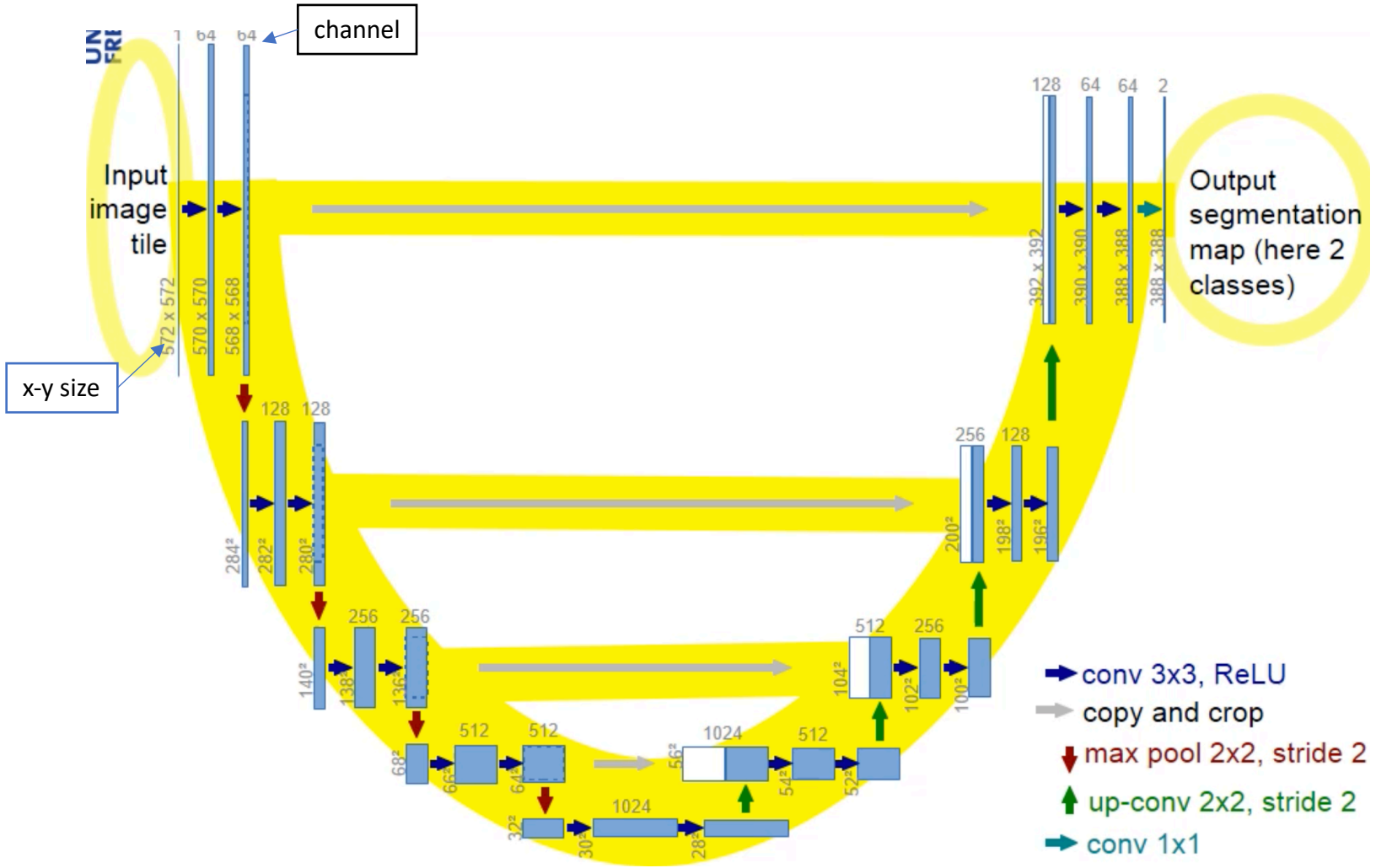
U-net architecture



Encoder-Decoder + skip connection

1. Encoder: the contracting path (left side, downsampling)
2. Decoder: the expansive path (right side, upsampling)
3. Skip connection: copy and crop operation at each stage(the gray arrows)

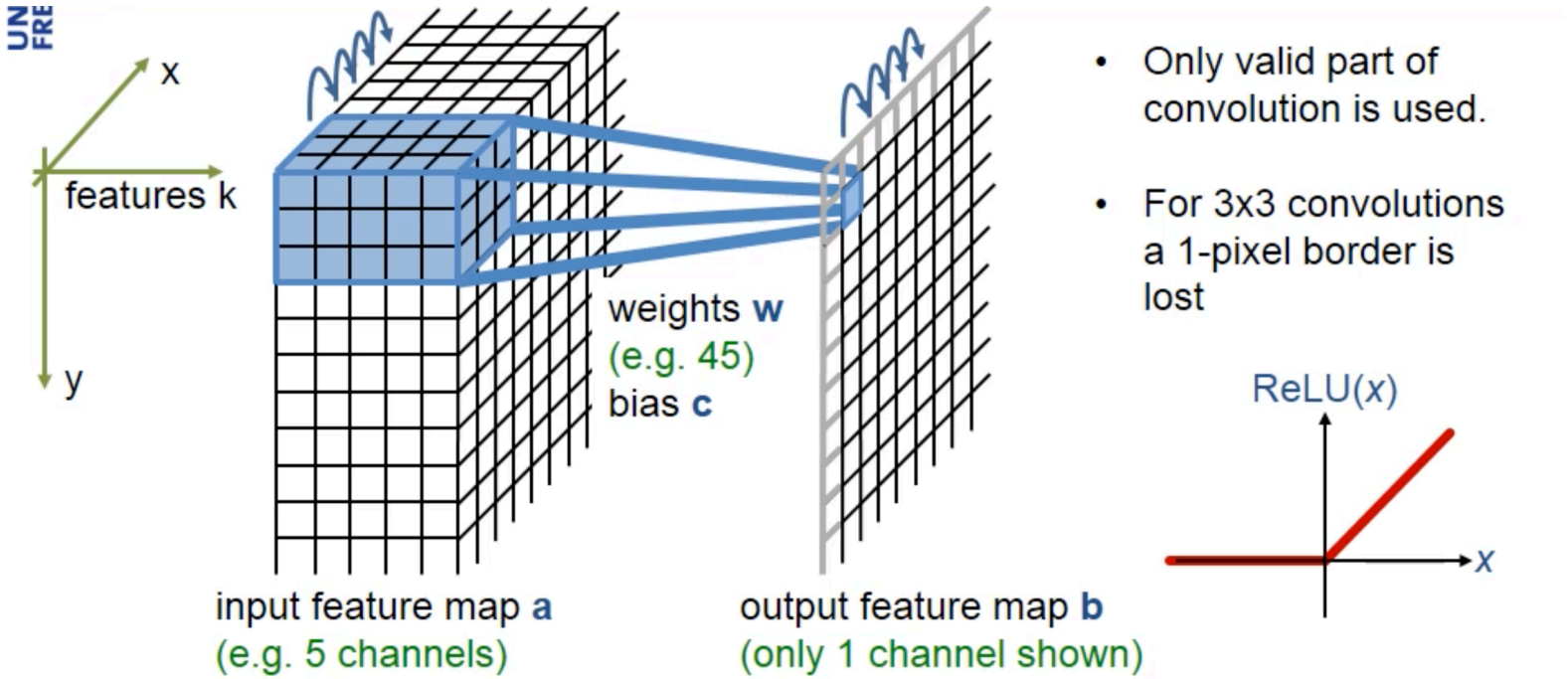
U-net architecture



Encoder-Decoder + skip connection

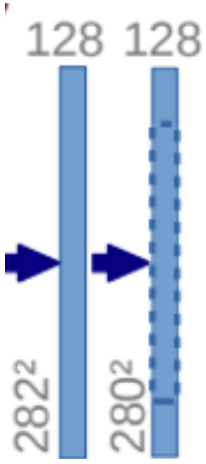
1. Encoder: downsampling, decrease the size of data but increase the number of channels.
2. Decoder: upsampling, increase the x-y-size and decrease the number of channels.
3. Skip connection: copy and crop operation at each stage(the gray arrows)

Operations: 3x3 convolution + ReLU



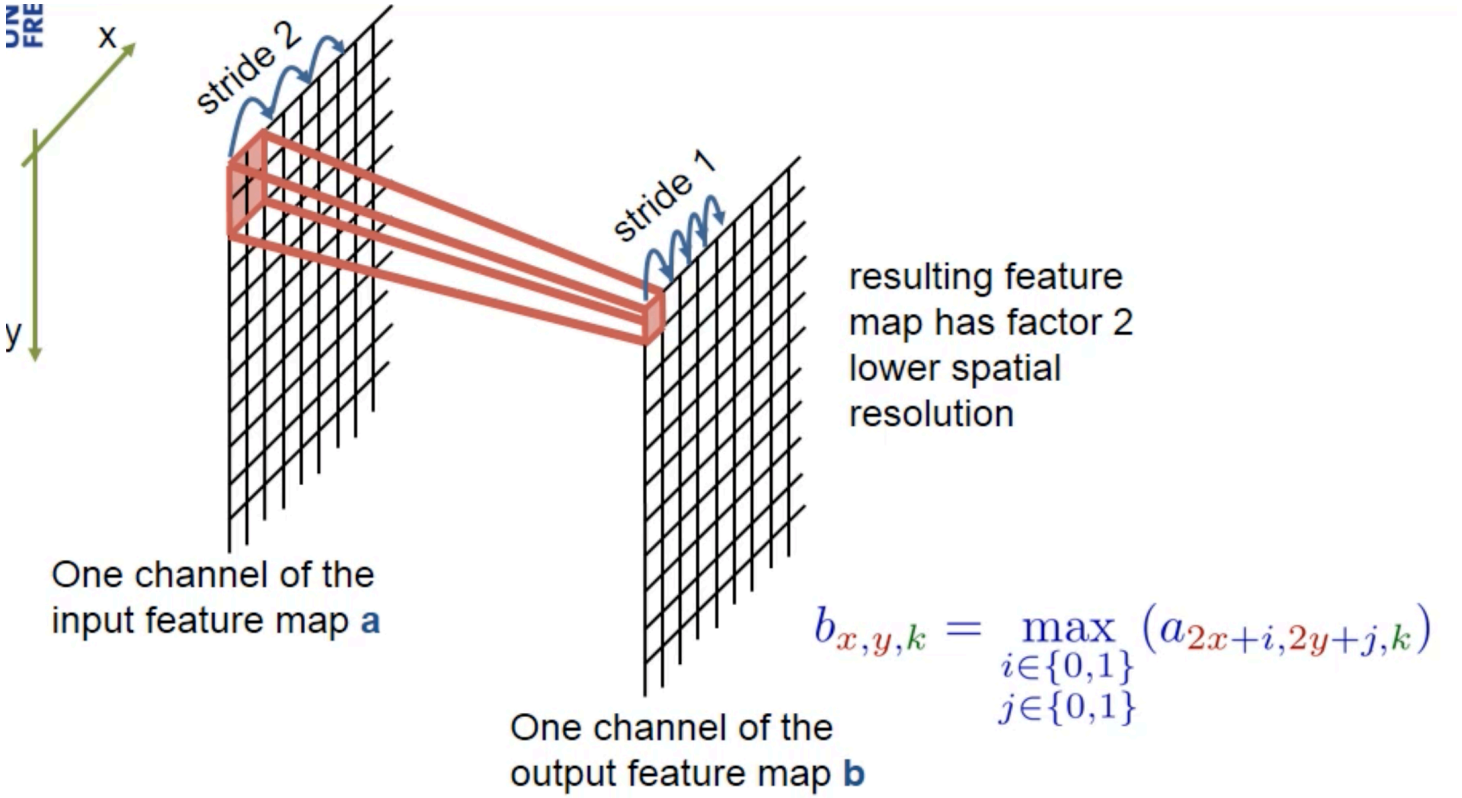
- Only valid part of convolution is used.
- For 3x3 convolutions a 1-pixel border is lost

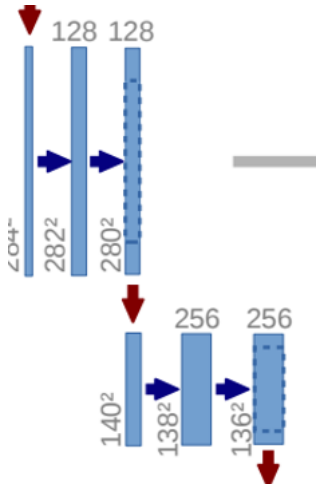
$$b_{x,y,l} = \mathrm{ReLU}\Big(\sum_{\substack{i\in\{-1,0,1\} \\ j\in\{-1,0,1\} \\ k\in\{1,\ldots,K\}}} w_{i,j,k,l} \cdot a_{x+i,y+j,k} + c_l \Big)$$



- Most of the operations are convolutions followed by non-linear activation functions.
- Only valid part of convolution is used(1-pixel border is lost): allows later to process large images in individual tiles

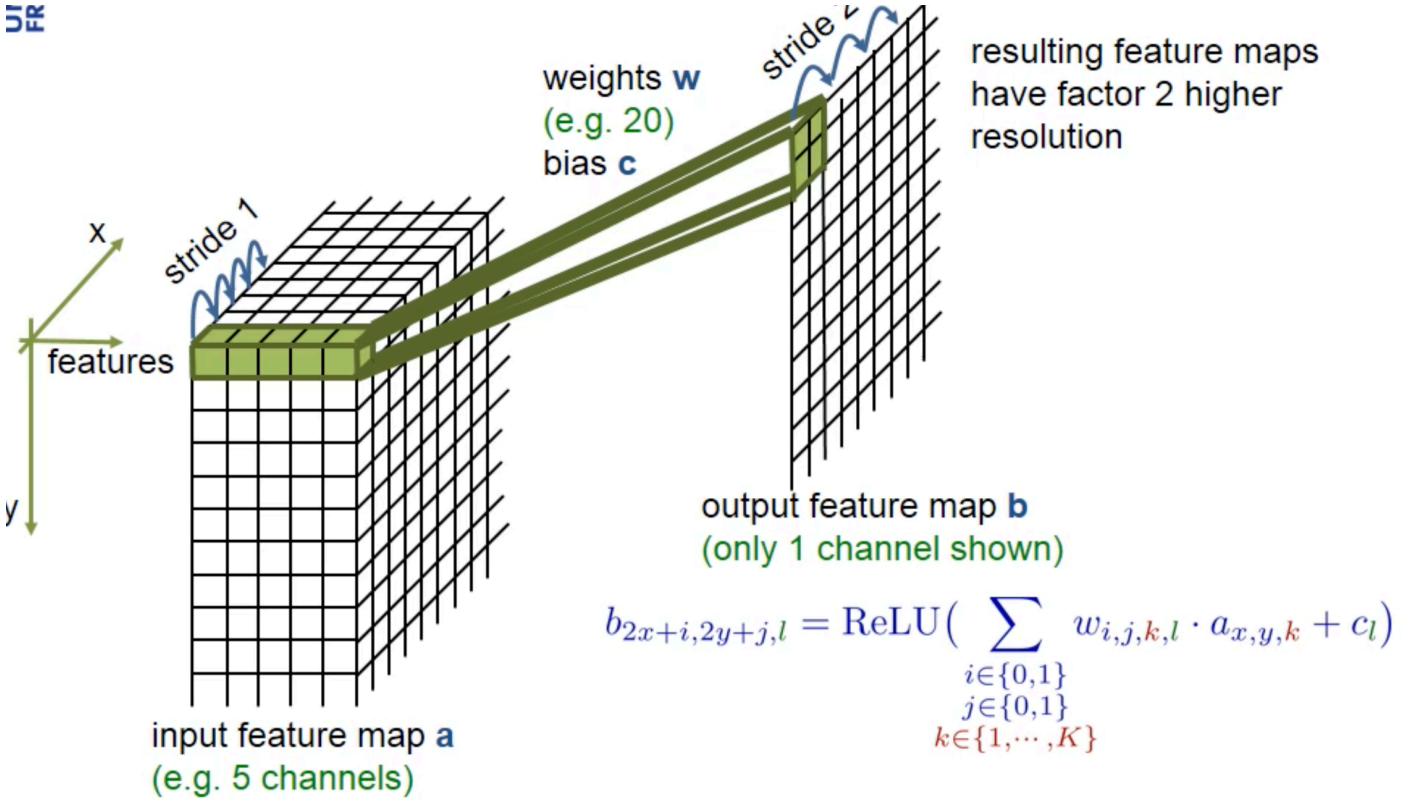
Operations: 2x2 max-pooling

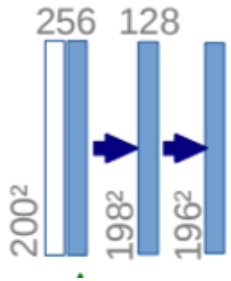




- Reduce the x-y-size of the feature map.
- Propagate the maximum activation from each 2 by 2 window to the next feature map.

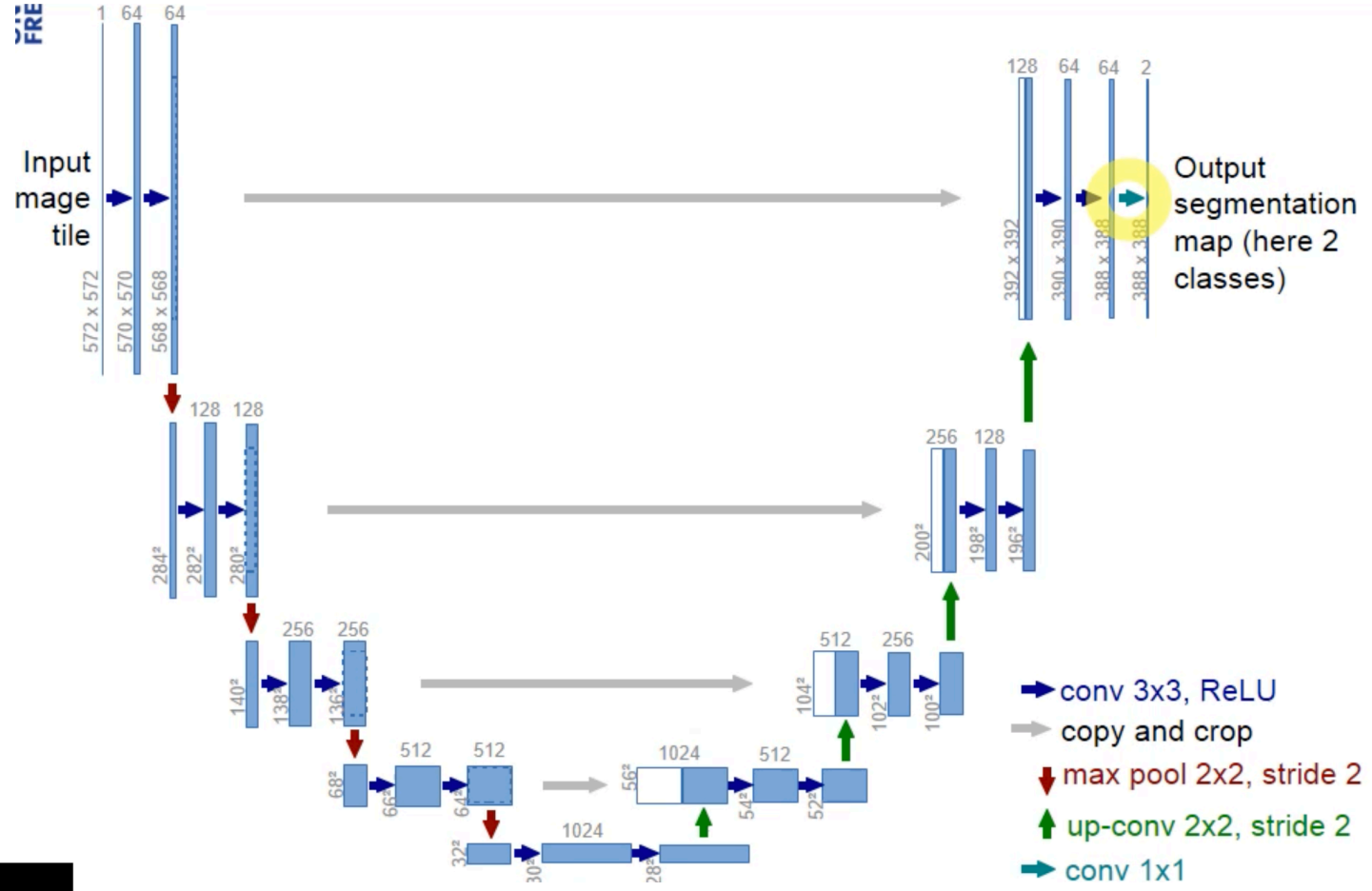
Operations: 2x2 up-convolution





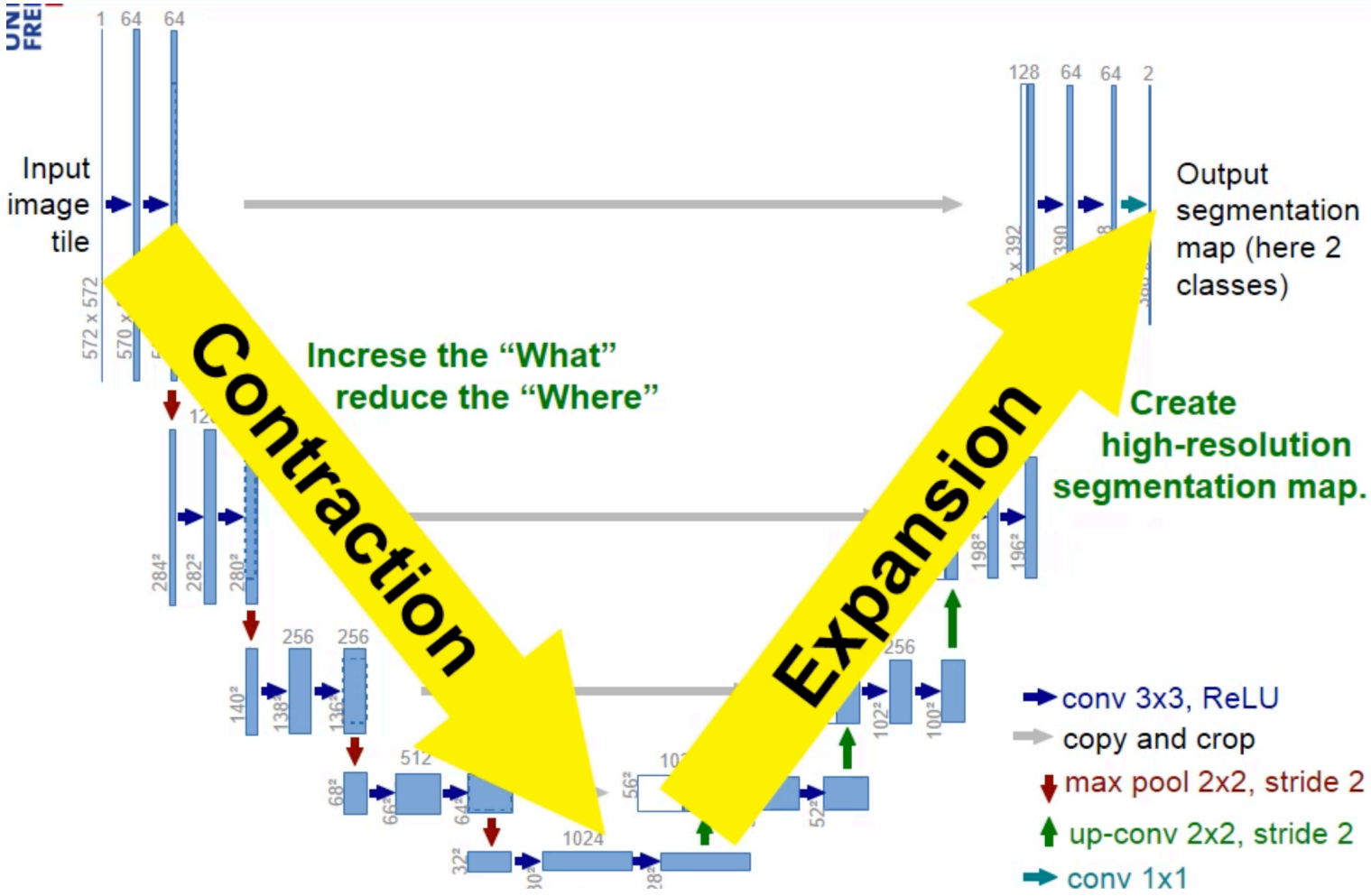
- This up-convolution uses a learned corner to map each feature vector to the 2x2 pixel window

Operations: 1x1 convolution (Expansion path)



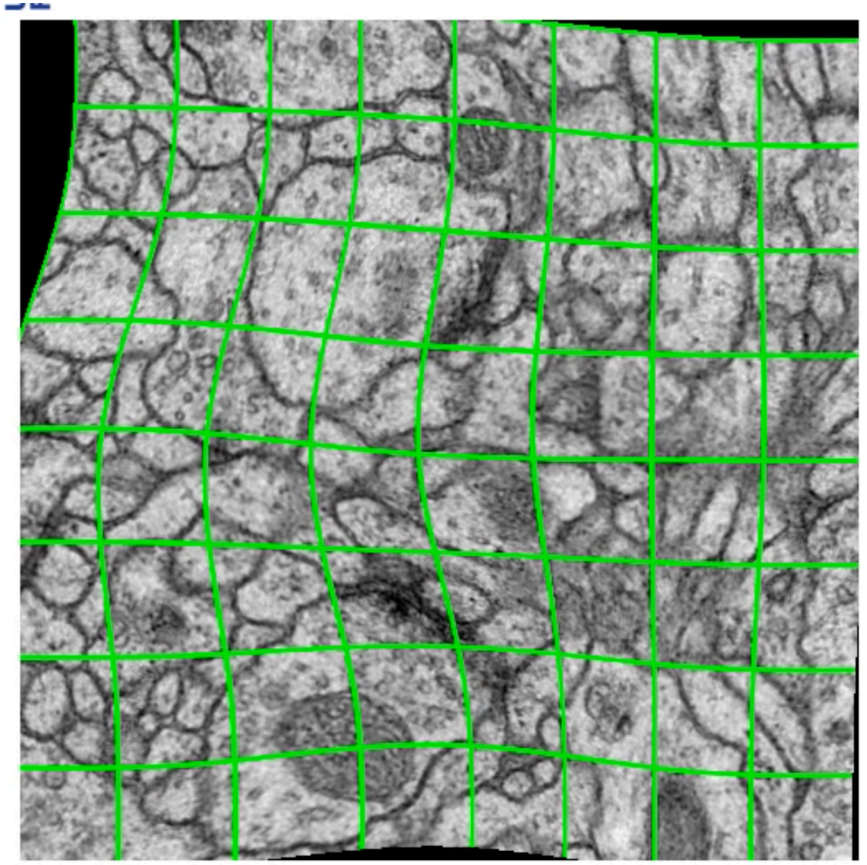
Output segmentation map has 2 channels: **one for the foreground and the other for the background.**

U-net architecture

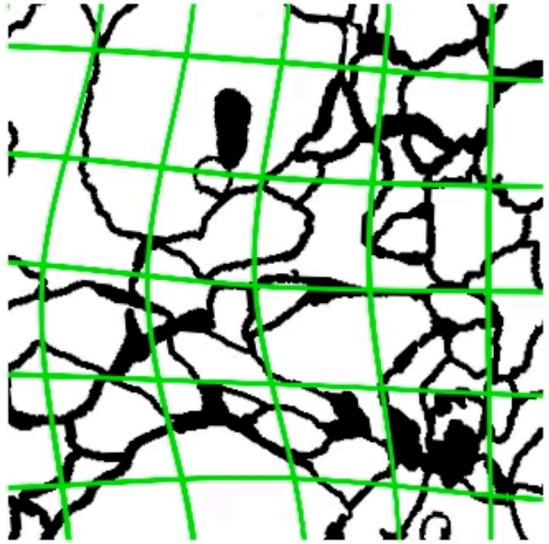


1. **Encoder(Contraction**): A sequence of convolutions and max-pooling operations, in order to get high-resolution features
2. **Decoder(Expansion):** A sequence of up-convolutions and concatenation with high-resolution features from contracting path, in order to propagate context information to higher resolution layers through a large number of feature channels..
3. **Skip connection(copy and crop):** compensate the loss of border pixels in every convolution.

Two main challenges: very few annotated images



resulting deformed image
(for visualization: no rotation, no shift, no extrapolation)



correspondingly deformed manual labels

- Very few annotated images
- Solution: **data augmentation using random elastic deformations =>** teach the network the desired invariance and robustness properties

Two main challenges: **touching objects of the same class**

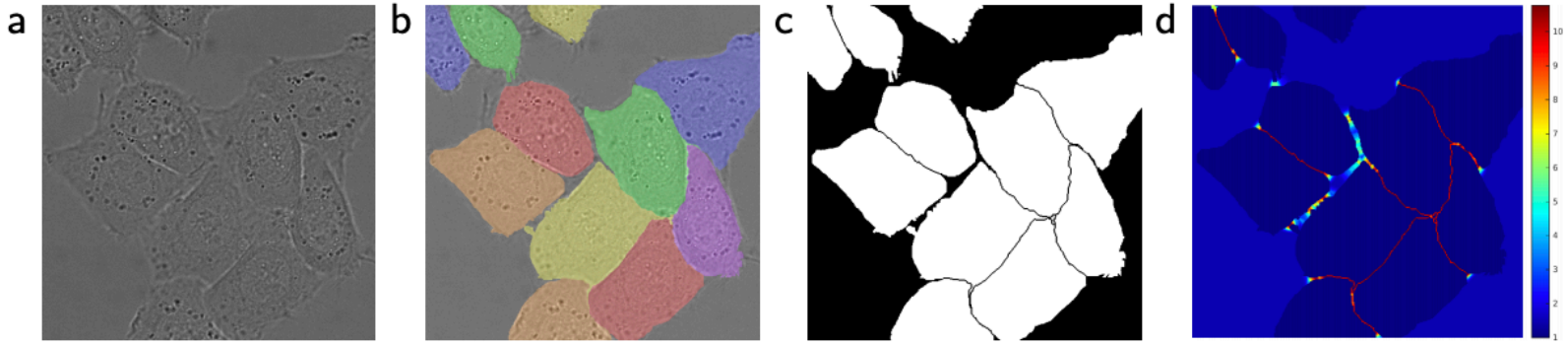
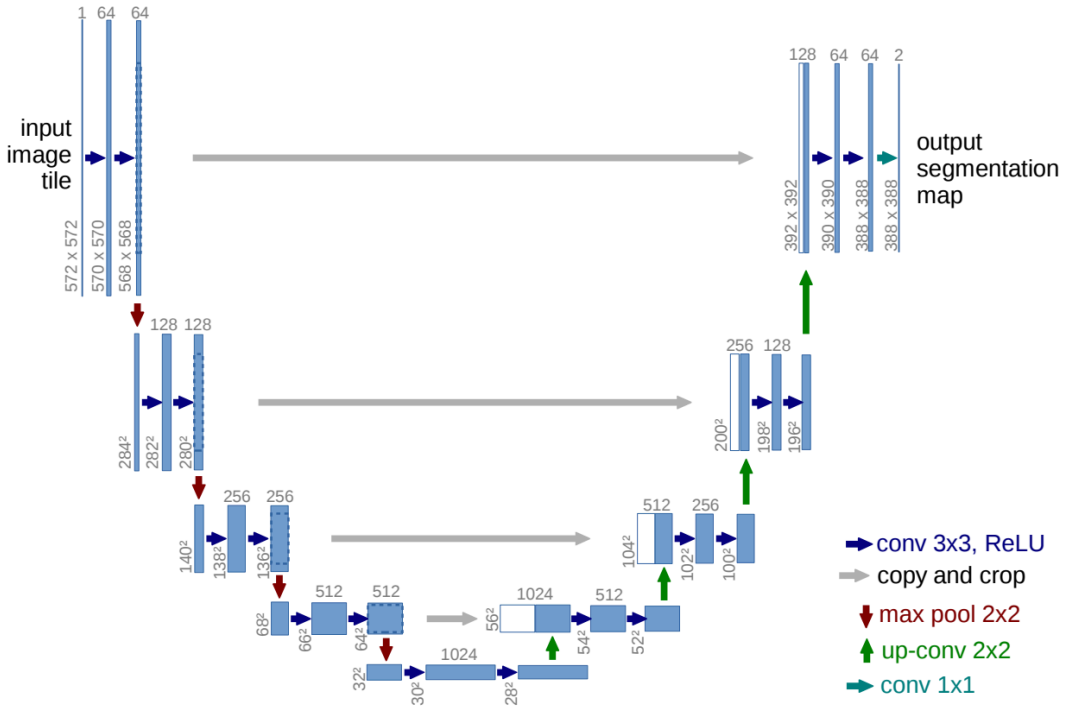


Fig. 3. HeLa cells on glass recorded with DIC (differential interference contrast) microscopy. (**a**) raw image. (**b**) overlay with ground truth segmentation. Different colors indicate different instances of the HeLa cells. (**c**) generated segmentation mask (white: foreground, black: background). (**d**) map with a pixel-wise loss weight to force the network to learn the border pixels.

- the separation of touching objects of the same class;
- Solution: the use of a **weighted loss**, where the separating background labels between touching cells obtain **a large weight** in the loss function.

U-net v.s. Standard convolutional networks



- Standard convolutional networks: the output to an image is a single class label
- U-net: the output includes localization = a class label is assigned to **each pixel**.

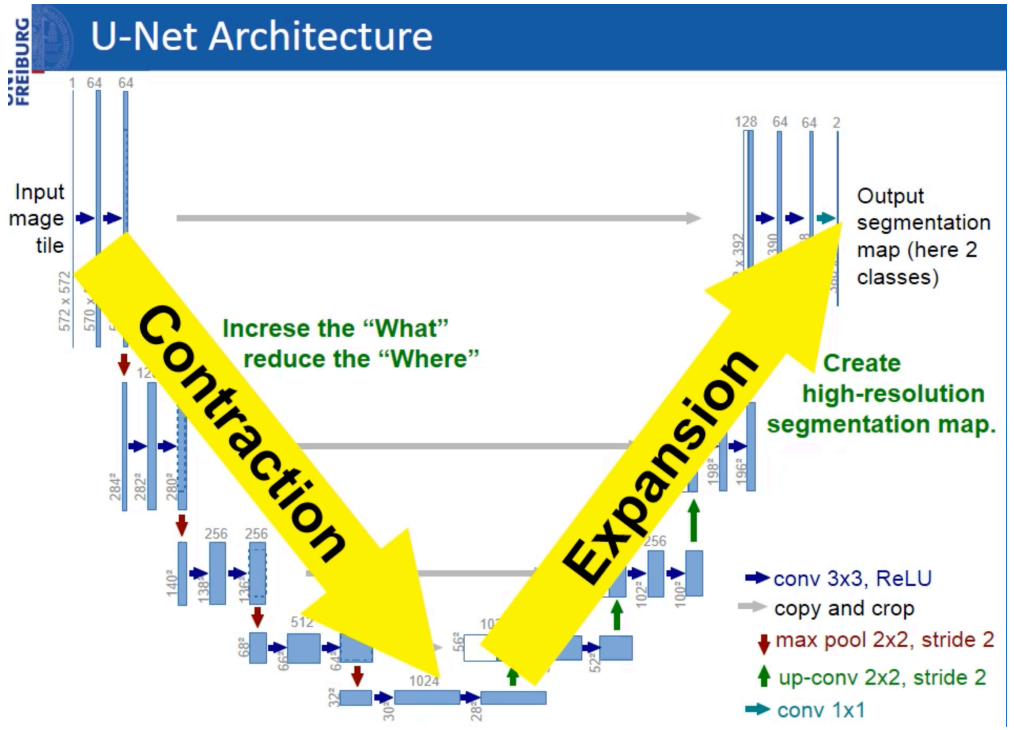
- A U-shaped architecture:

contracting path + expansive path

Pooling operators are replaced by upsampling operators => increase the resolution of the output.

Summary: U-net for Biomedical Image Segmentation

- It works with very few training images and yields more precise segmentations;
- U-net achieves **good localization and the use of context** at the same time.
- **Localization**: high-resolution features from the contracting path are combined with the upsampled output. A successive convolution layer can then learn to assemble a more precise output based on this information.
- **Context**: a large number of feature channels in the upsampling part allow the network to propagate context information to higher resolution layers.



References

- 1. U-Net: *Convolutional Networks for Biomedical Image Segmentation*. Olaf Ronneberger, Philipp Fischer, Thomas Brox
- 2. WWW: web page of *U-net implementation, trained networks and supplementary material*, http://lmb.informatik.uni-freiburg.de/people/ronneber/u-net