U-net: Convolutional Networks for Biomedical Image Segmentation

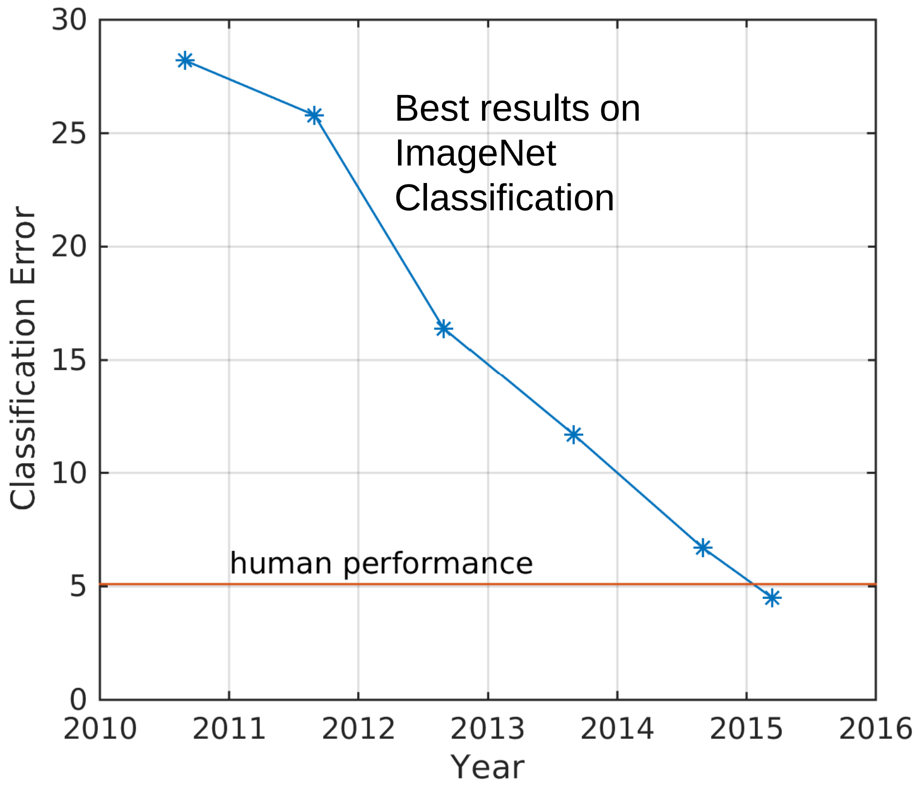
Olaf Ronneberger

Computer Science Department and
BIOSS Centre for Biological Signalling Studies
University of Freiburg, Germany

O. Ronneberger, P. Fischer, T. Brox: U-net: Deep Convolutional Networks for Biomedical Image Segmentation.
MICCAI 2015, Springer, LNCS, Vol.9351: 234--241, 2015

Image Classification State of the Art

- Deep convolutional neural networks surpassed human-level performance in Feb 2015



GT: mountain tent
1: sleeping bag
2: mountain tent
3: parachute
4: ski
5: flagpole

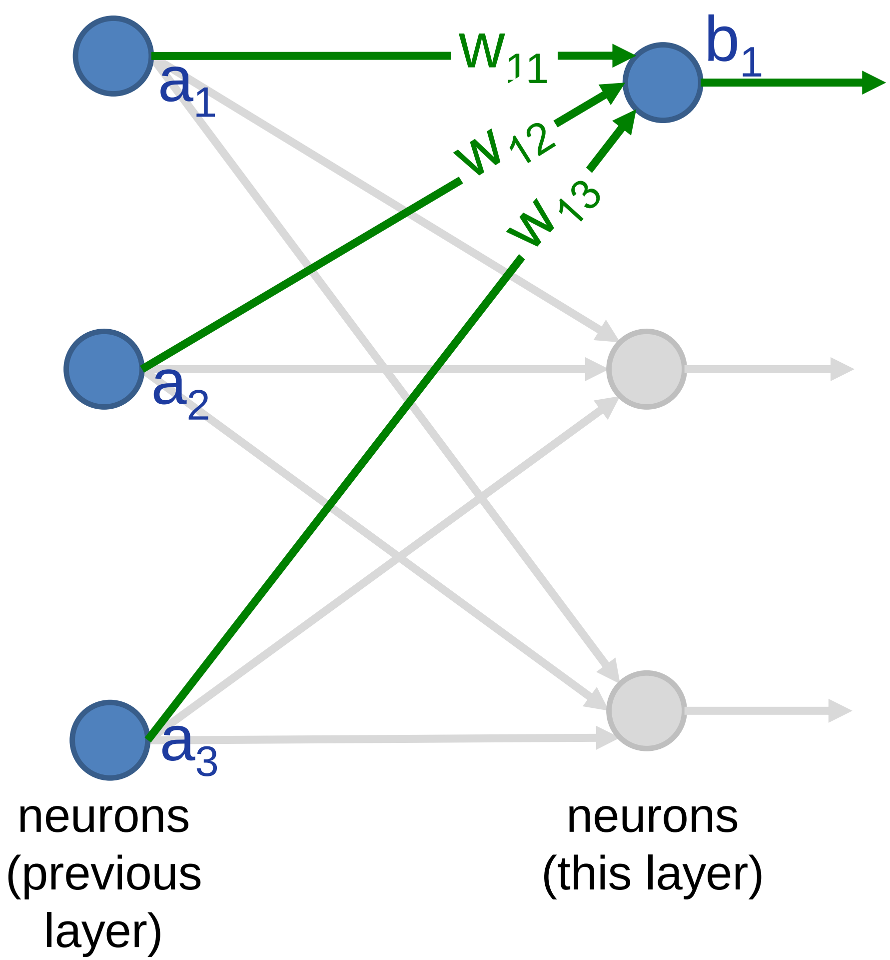
GT: geyser
1: geyser
2: volcano
3: sandbar
4: breakwater
5: leatherback turtle

K. He, X. Zhang, S. Ren, J. Sun: "Delving Deep into Rectifiers: Surpassing Human-Level Performance on ImageNet Classification." (2015) arXiv:1502.01852 [cs.CV]

Neural Networks

- Neural networks are a very old concept. Idea: simulate a brain in the computer

$$b_1 = f(w_{11}a_1 + w_{12}a_2 + w_{13}a_3)$$

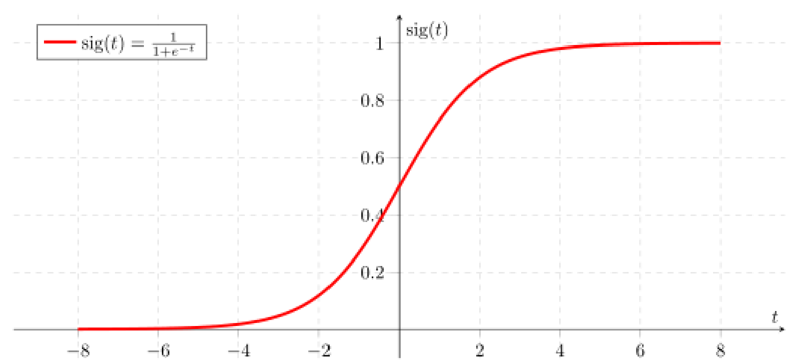


$a_1 - a_3$: output of neurons in previous layer

$w_{11} - w_{13}$: weights

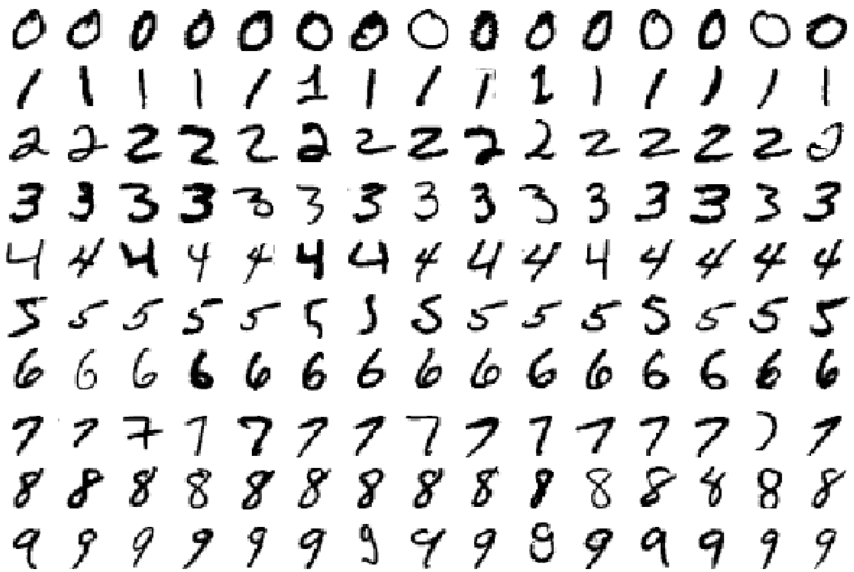
b_1: output of neuron in this layer

f(x): activation function



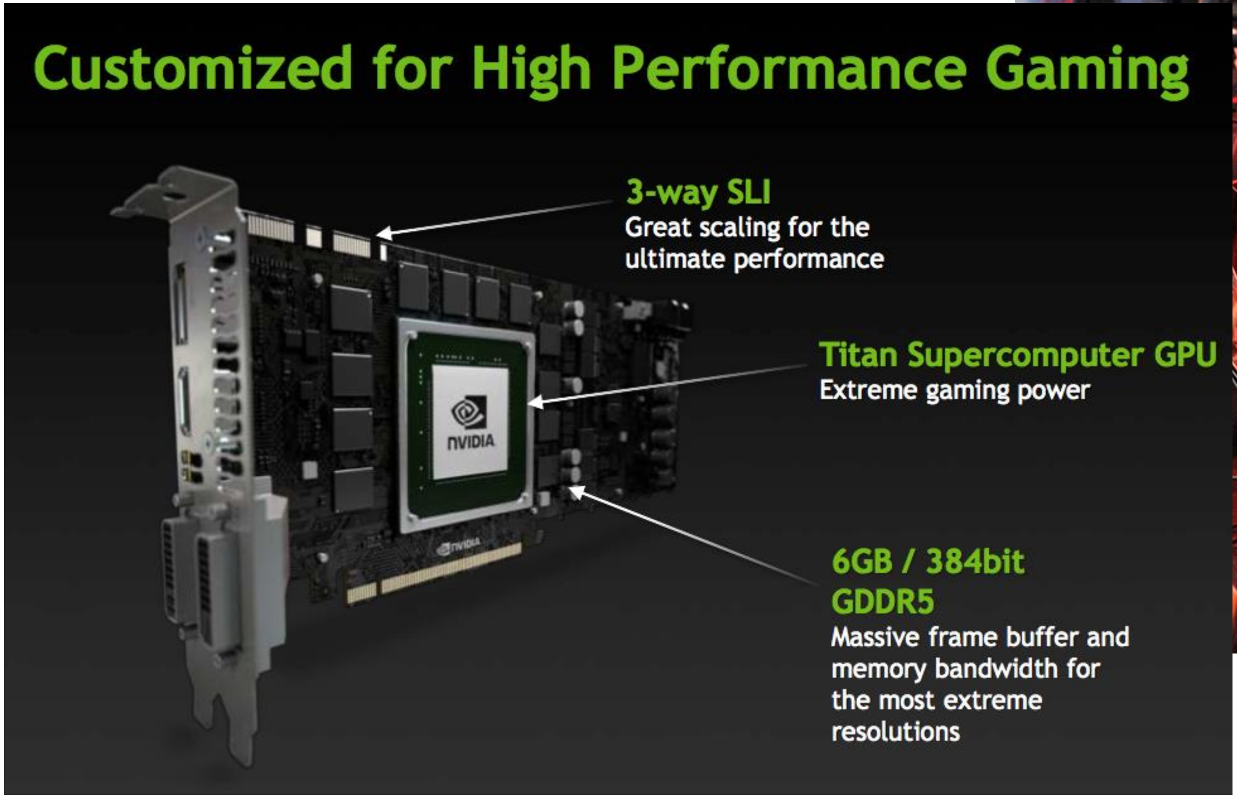
Neural Networks

- Several hypes
- Only successful on very specific problems: e.g. “handwritten digits recognition”



- **Why do they work now?**

Availability of Enormous Computing Power



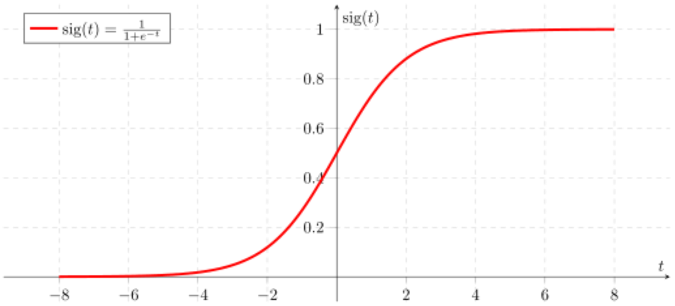
- Thanks to all people who support the development of theses supercomputers!

Availability of Millions of Annotated Images

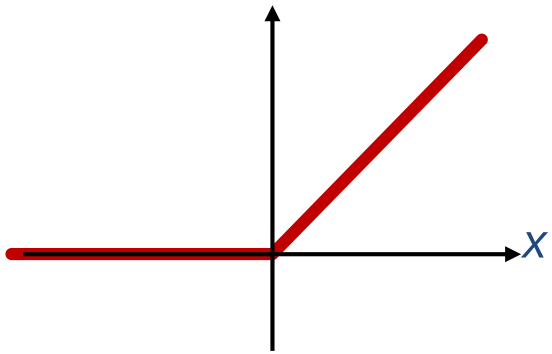
[O. Russakovsky et al. “ImageNet Large Scale Visual Recognition Challenge.” (2014) arXiv:1409.0575 [cs.CV]]

Slight Improvements in the Architecture

Sigmoid(x)



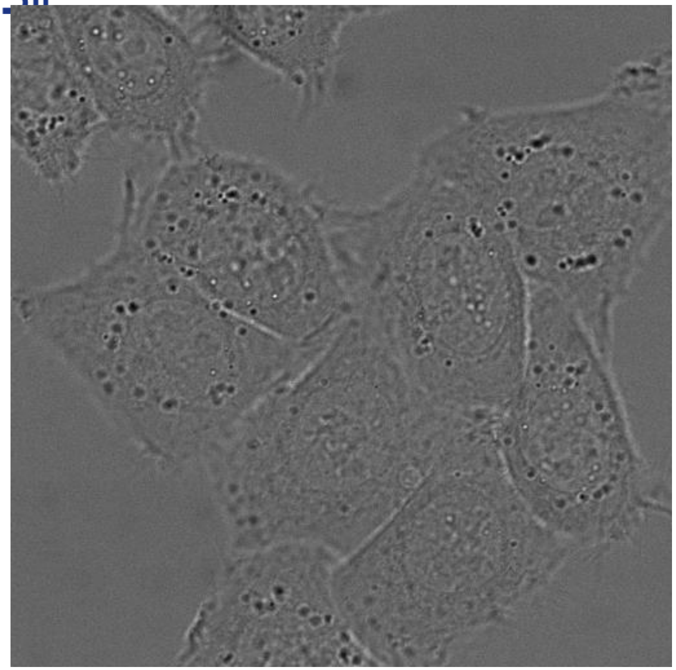
ReLU(*x*)



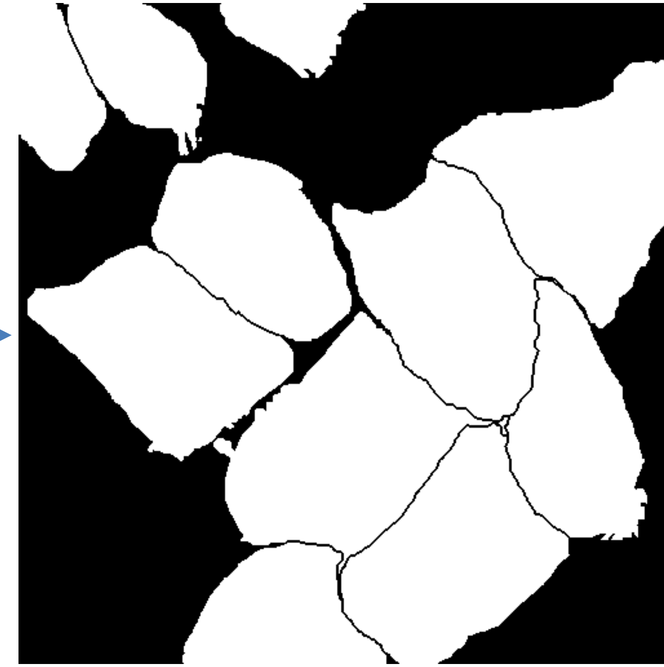
- Non-saturating neurons: no “vanishing gradients”
- Drop-Out Layers: less overfitting
- Efficient GPU implementations
- Very deep networks (millions of neurons)

- It is definitely time to apply these networks to something really useful

Biomedical Image Segmentation with U-net



input image

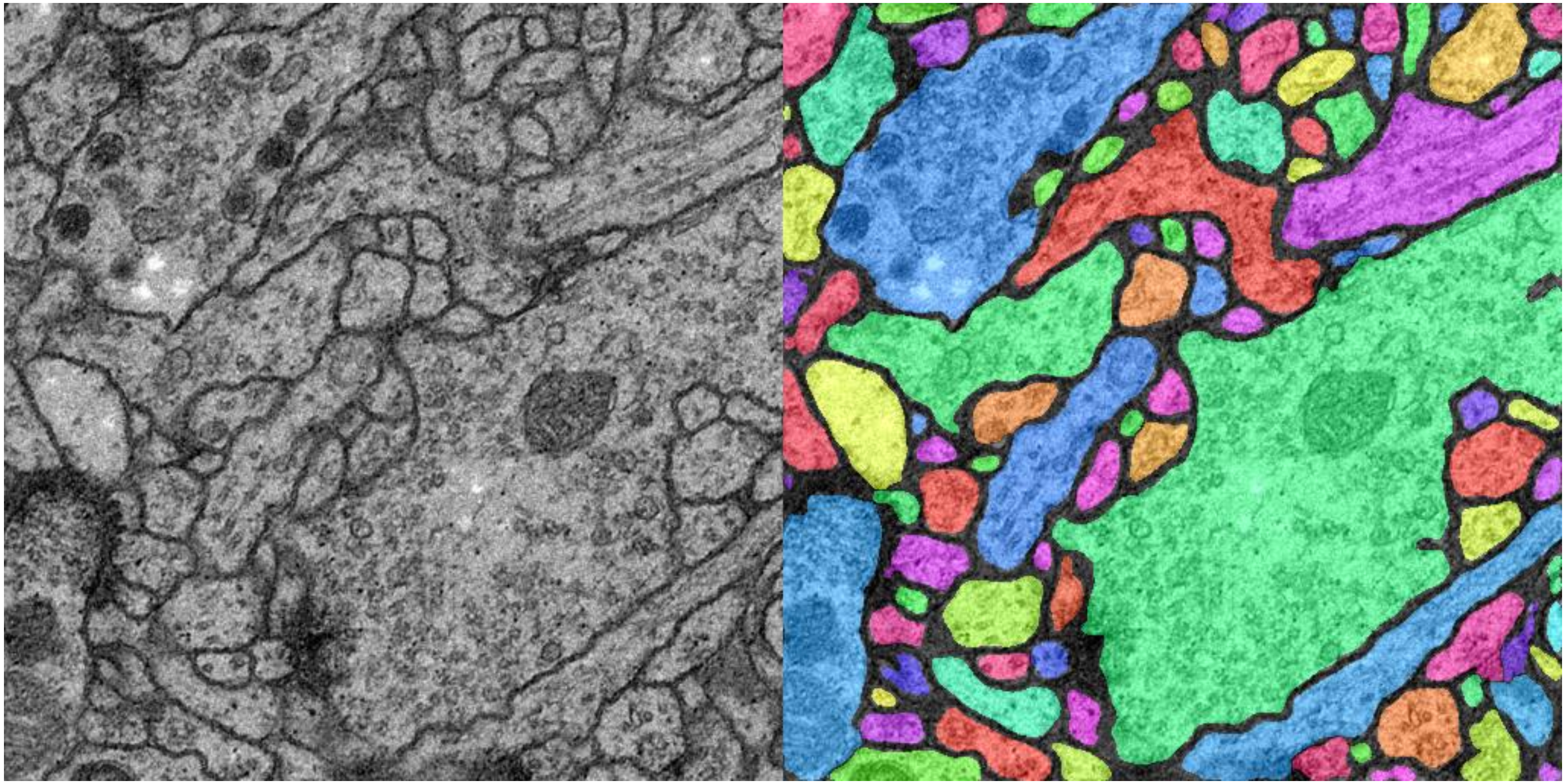


output segmentation map

- U-net **learns segmentation** in an end-to-end setting
- **Very few annotated images** (approx. 30 per application)
- **Touching objects** of the same class

Data provided by Dr. Gert van Cappellen, Erasmus Medical Center. Rotterdam. The Netherlands

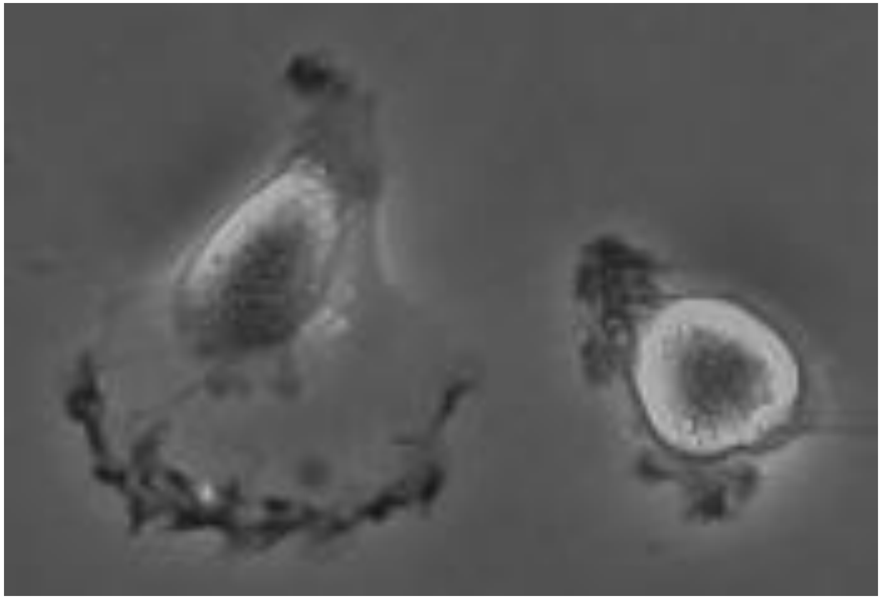
Neuronal Structures in EM

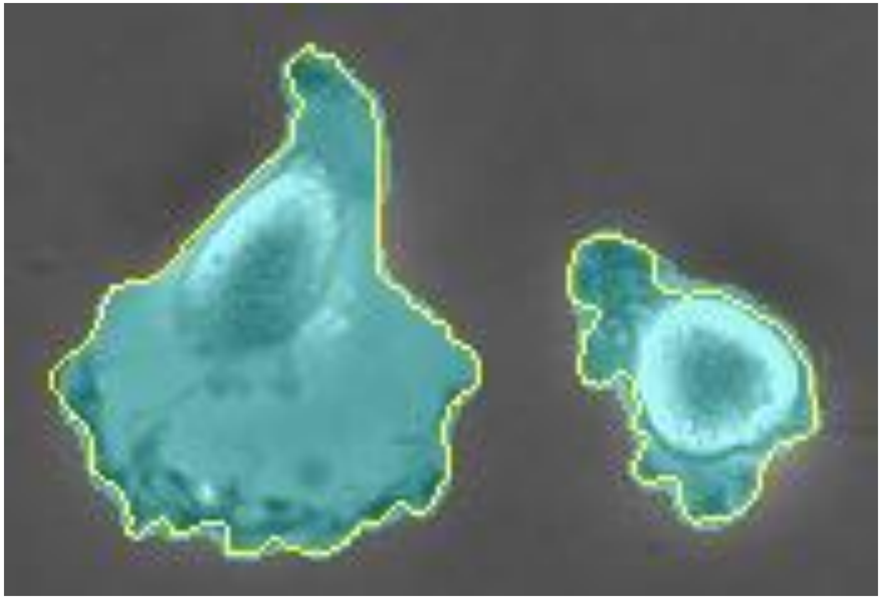


- EM segmentation challenge (ongoing since ISBI 2012):
  - Our result: 0.000353 warping error
    (**New best score** at submission march 6th, 2015)
  - Sliding-window CNN: 0.000420

Data provided by the ISBI 2012 EM segmentation challenge.
Cardona, A. et al.: . PLoS Biol 8(10), e1000502 (2010)

Cells in Phase-Contrast Microscopy

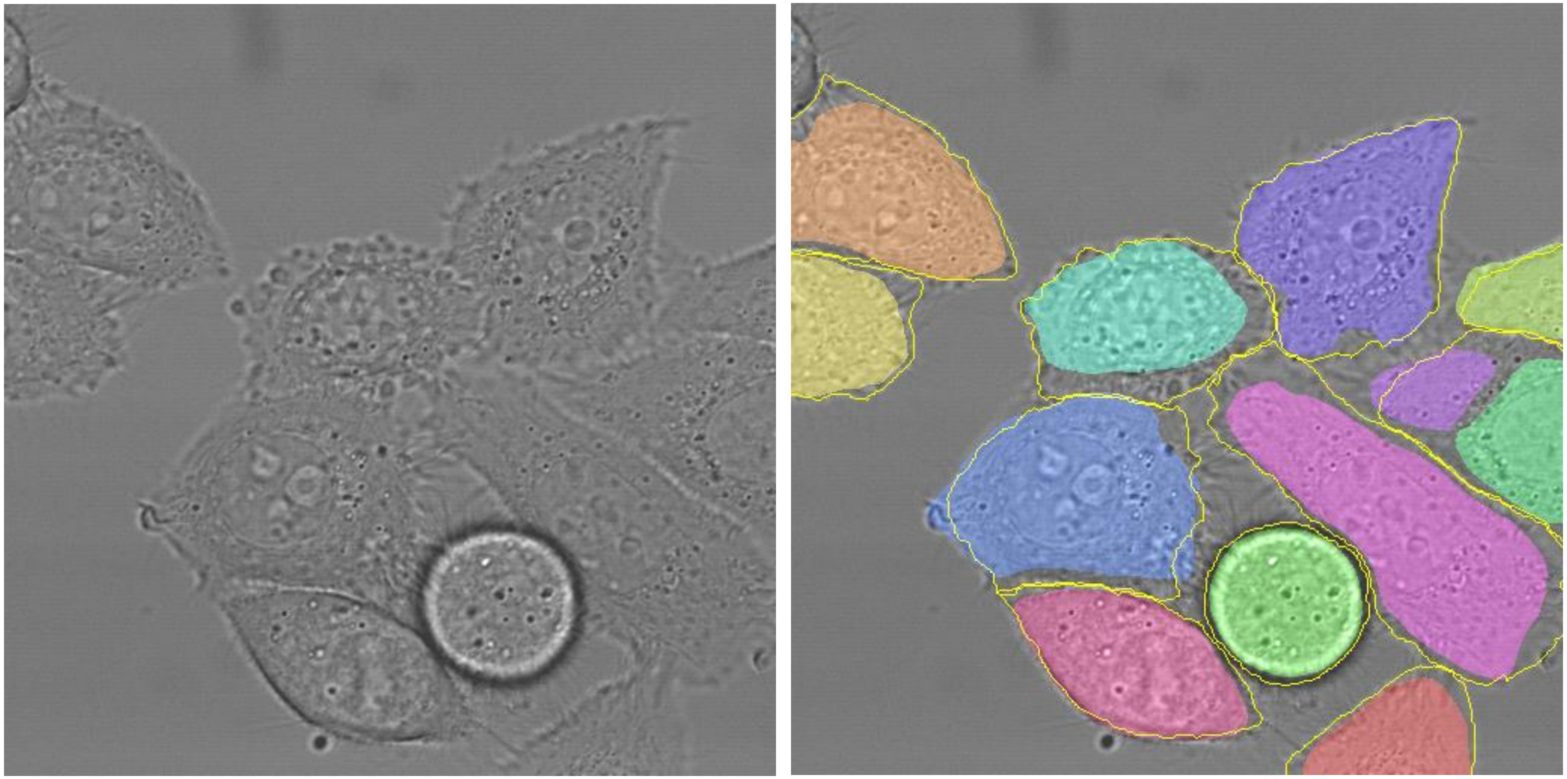




- ISBI Cell Tracking Challenge 2015
  Dataset „PhC-U373“
  - Our Result: 92% Intersection over Union, **Winner**
  - Second best result: 83%

Data provided by Dr. Sanjay Kumar. Department of Bioengineering
University of California at Berkeley. Berkeley CA (USA)

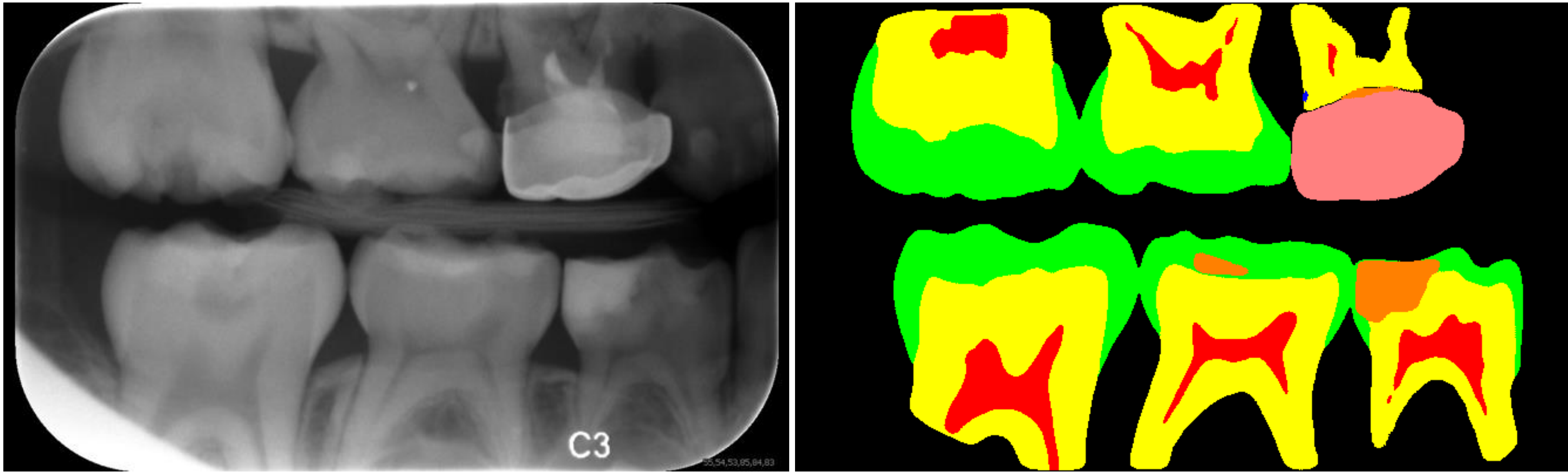
Cells in DIC Microscopy



- ISBI Cell Tracking Challenge 2015
  Dataset „DIC-Hela“
  - Our Result: 77.6% Intersection over Union, **Winner**
  - Second best result: 46,0%

Data provided by Dr. Gert van Cappellen,
Erasmus Medical Center. Rotterdam. The Netherlands

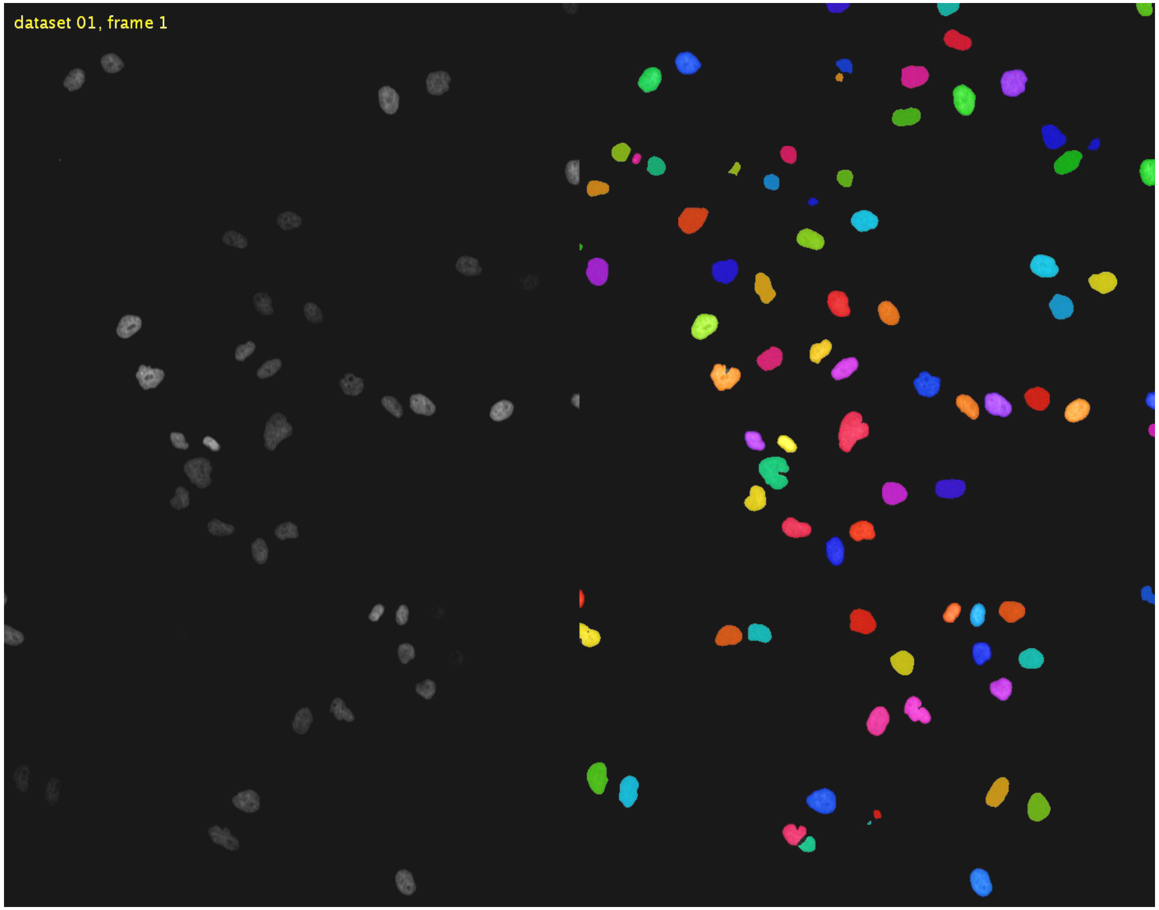
Dental X-Ray Images



- Grand Challenge for Computer-Automated Detection of Caries in Bitewing Radiography at ISBI 2015
  - Our Result: 0.525 (average F-score), **Winner**
  - Second best result: 0.287

Data provided by Prof Ching-Wei Wang, PhD, Graduate Institute of Biomedical Engineering, National Taiwan University of Science and Technology, Taiwan,

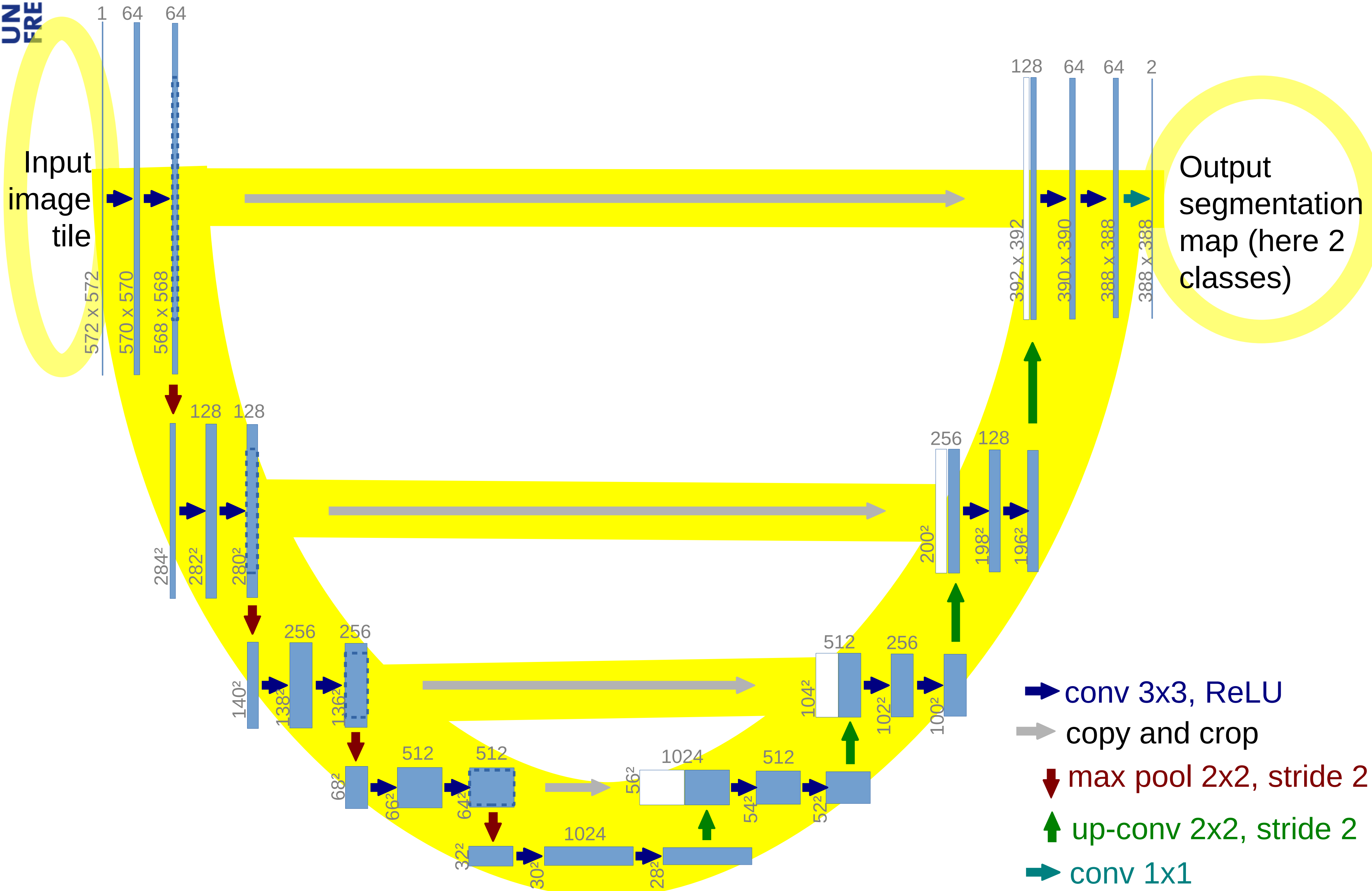
Fluorescence Microscopy



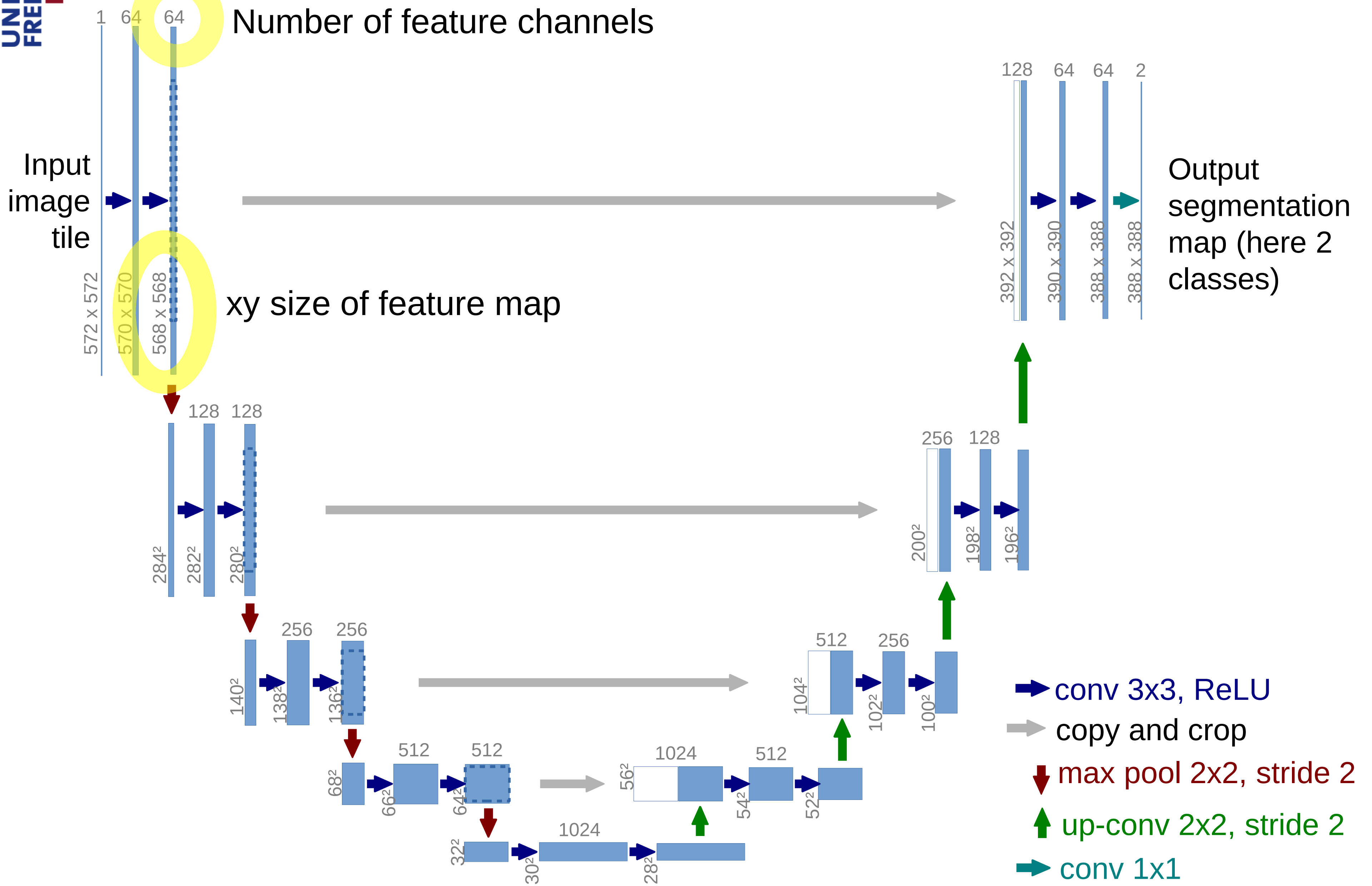
- ISBI Cell Tracking Challenge 2015 Dataset „Fluo-HeLa“
  - Our Result: 90% IoU, (**Winner)**
  - Second best result: 89%

Data provided by Mitocheck Consortium

U-net Architecture

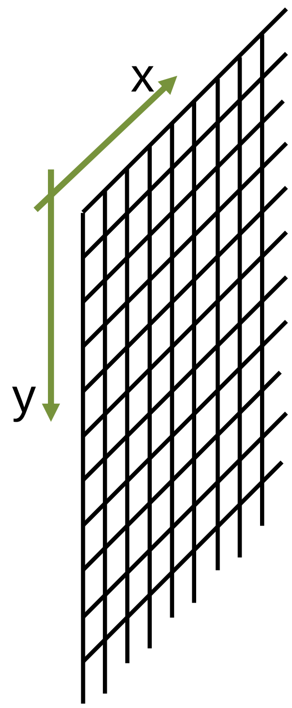


U-net Architecture

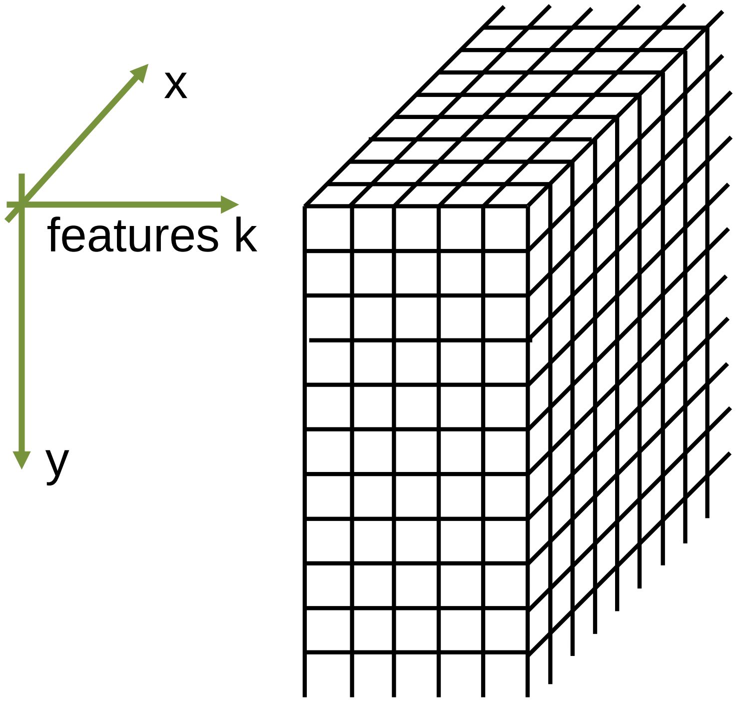


Feature maps

Feature maps are many neurons arranged in a grid

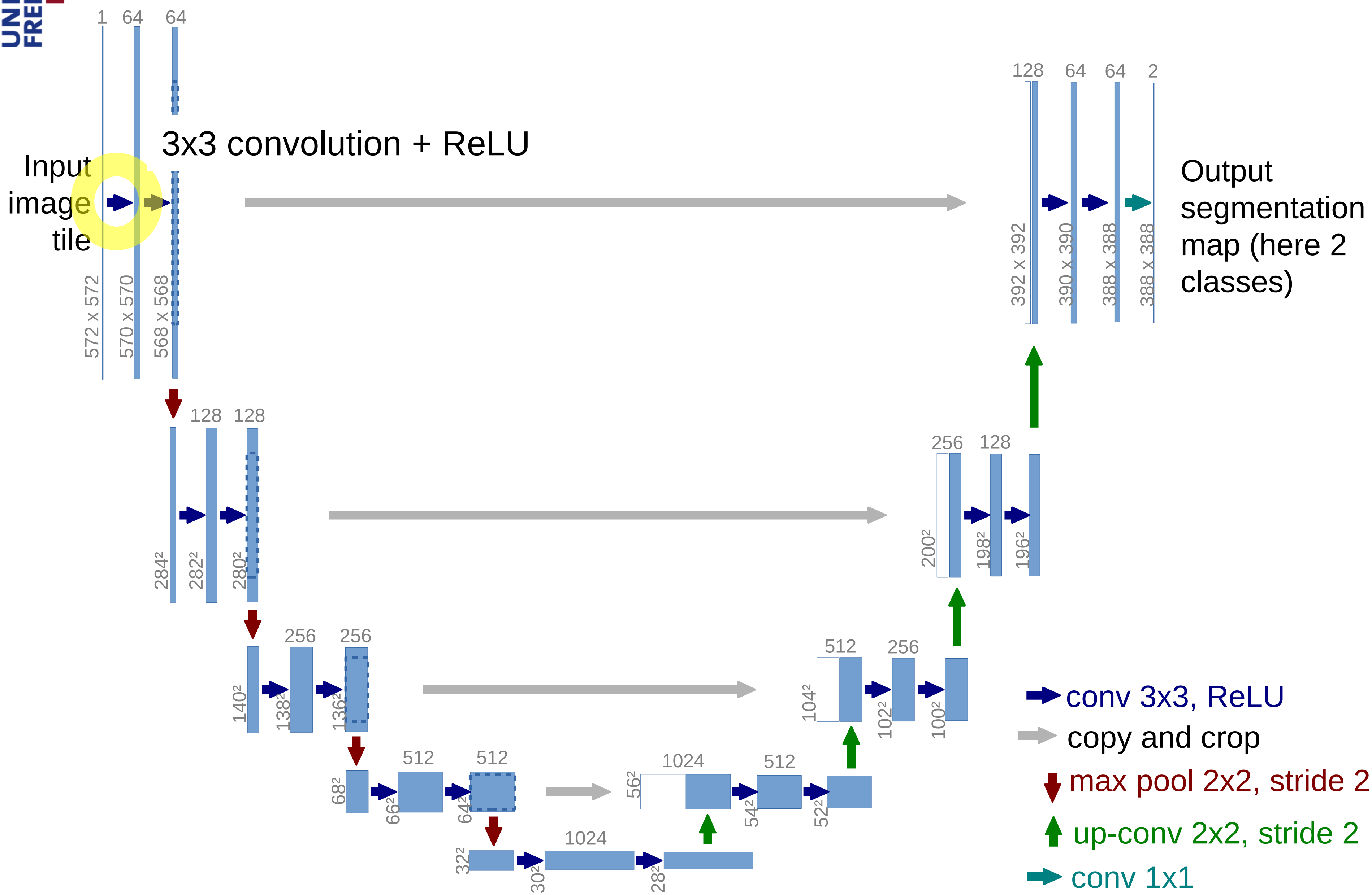


Example feature map
(1 channel)
e.g. gray image

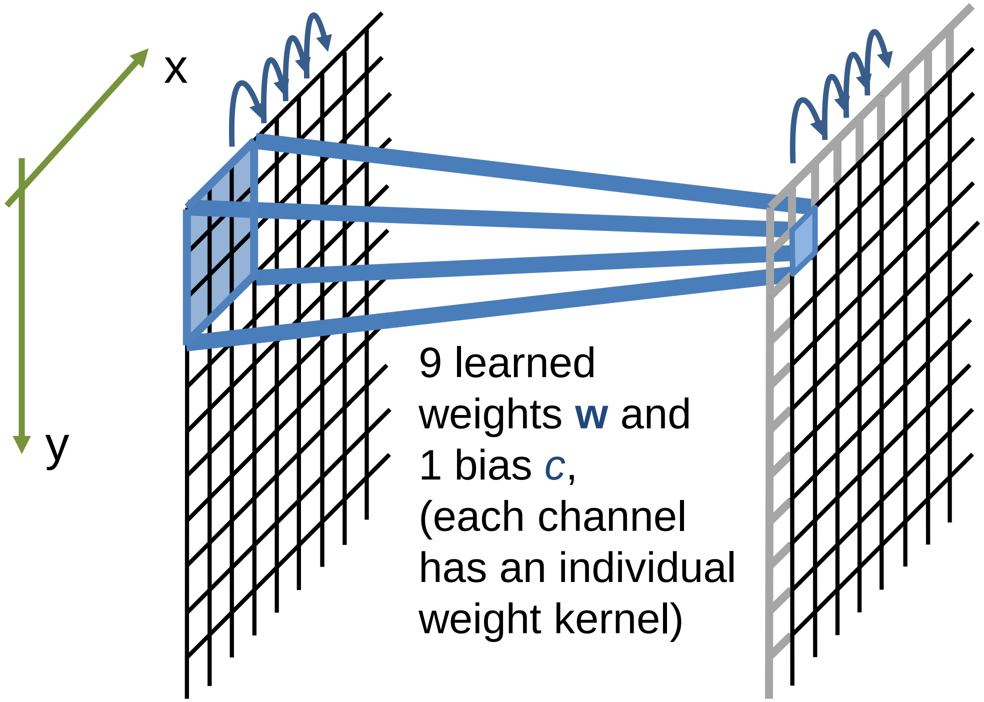


Example feature map
(5 channels)

U-net Architecture



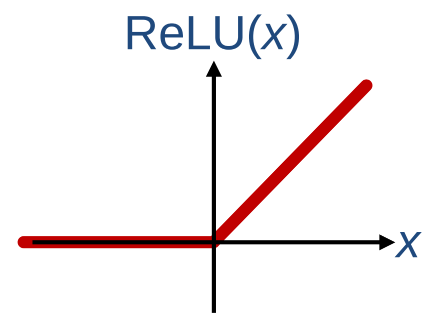
3x3 convolution + ReLU (first Layer)



input image
(1 channel)

output feature map
(only 1 channel of 64 shown)

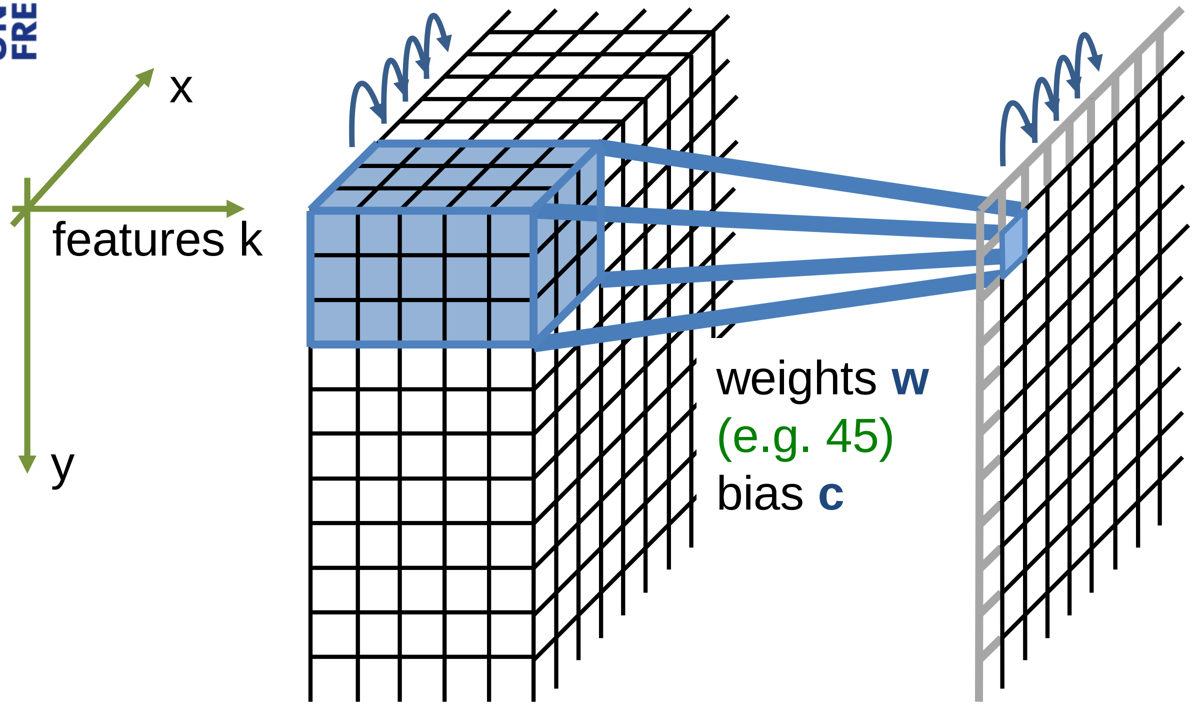
- Only valid part of convolution is used.
- For 3x3 convolutions a 1-pixel border is lost



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

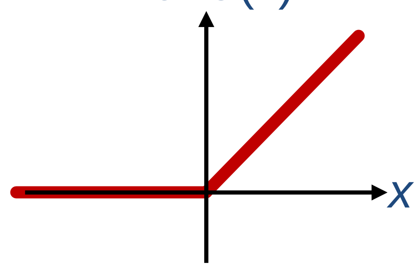
Intuition: Output neuron fires, when a certain input structure is seen
(in the first layer tiny parts, like edges, corners, spots, texture elements, …)

3x3 convolution + ReLU (all Subsequent Layers)



input feature map **a**
(e.g. 5 channels)

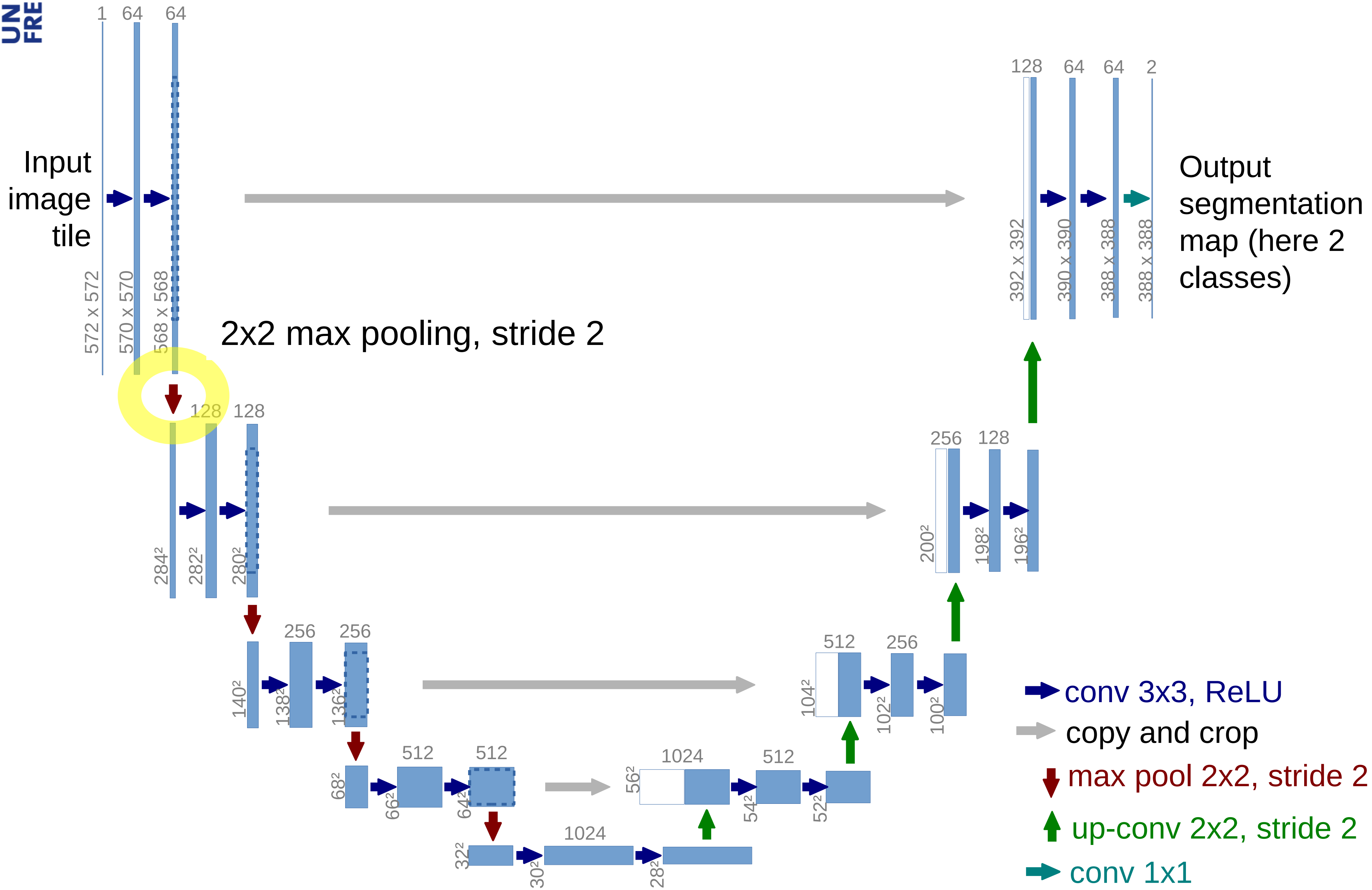
output feature map **b**
(only 1 channel shown)



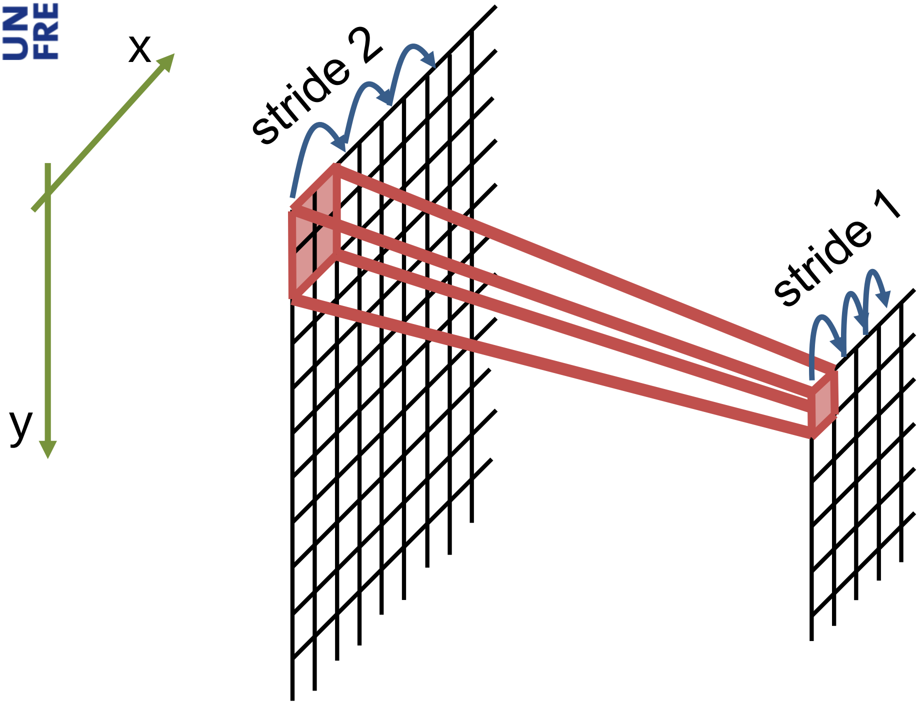
$$b_{x,y,l} = \mathrm{ReLU}\left(\sum_{\substack{i\in\{-1,0,1\} \\ j\in\{-1,0,1\} \\ k\in\{1,\ldots,K\}}} w_{i,j,k,l} \cdot a_{x+i,y+j,k} + c_l \right)$$

Intuition: Neurons fire, when constellation of small parts build a certain larger part

U-net Architecture



2x2 max-pooling



resulting feature map has factor 2 lower spatial resolution

One channel of the input feature map **a**

One channel of the output feature map **b**

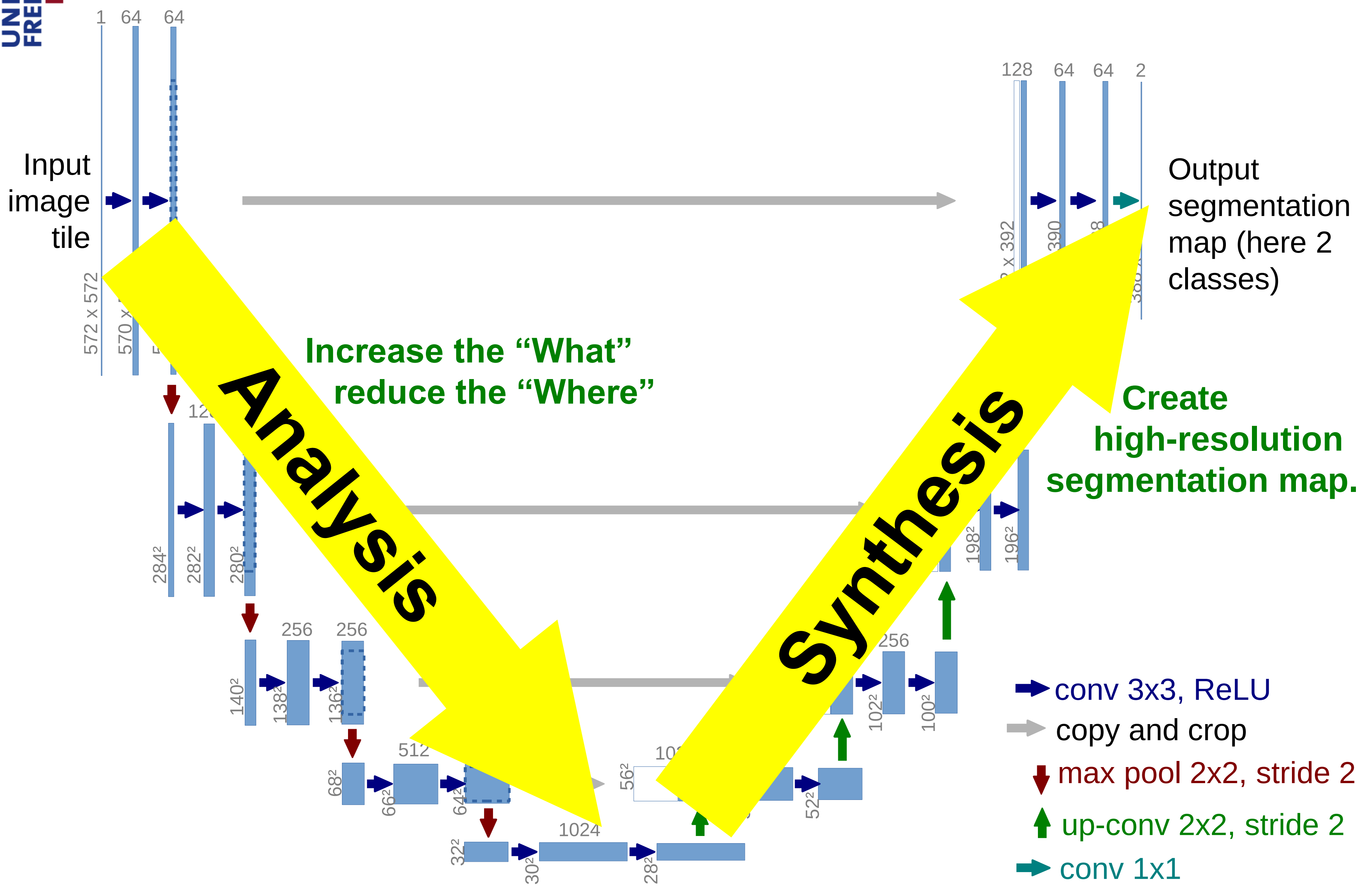
$$b_{x,y,k} = \max_{\substack{i\in\{0,1\}\\ j\in\{0,1\}}} \left(a_{2x+i,2y+j,k}\right)$$

Intuition: strongest activation (in local surrounding) is propagated

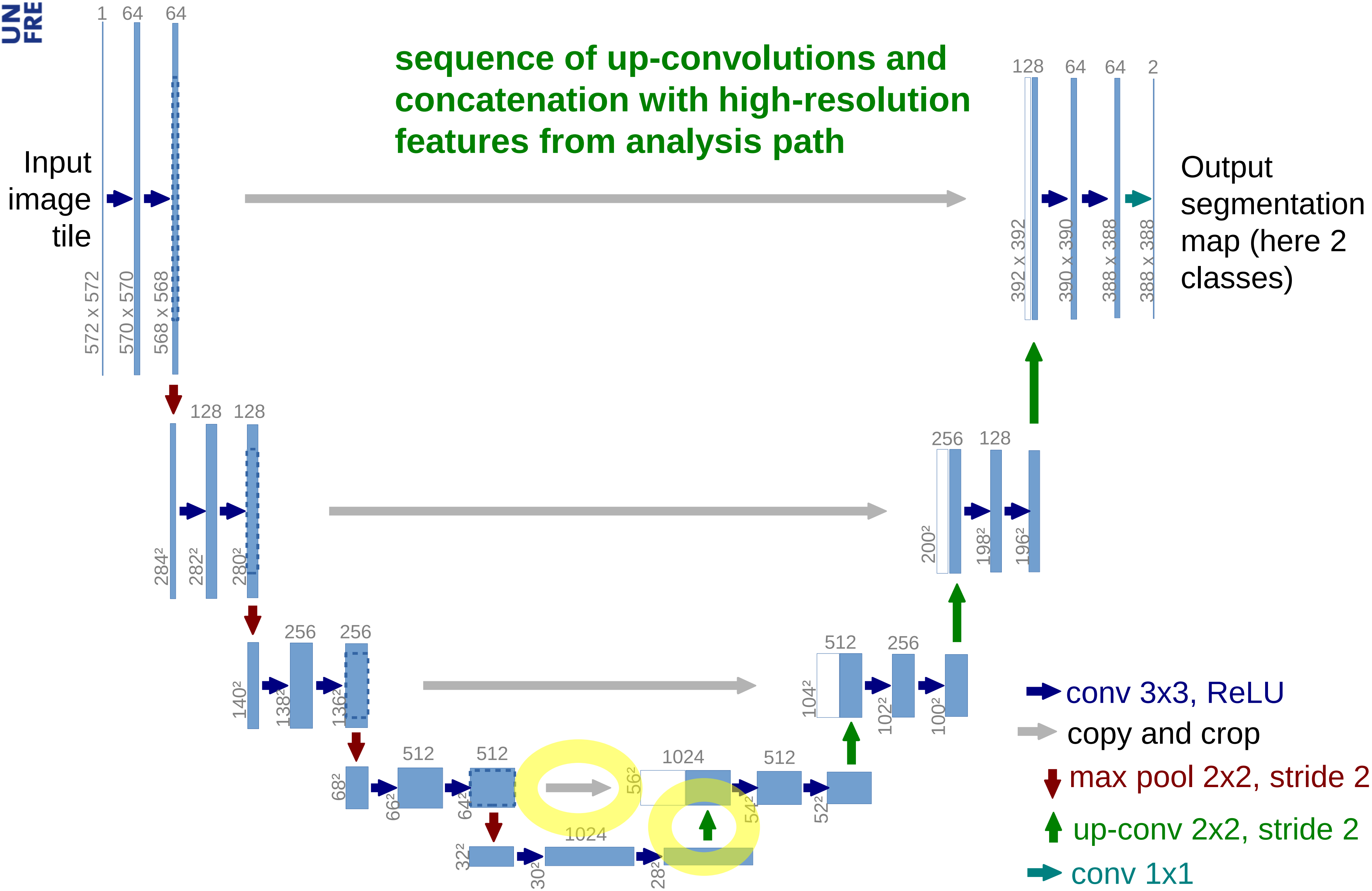
→ Robustness to spatial variations

→ reduce spatial resolution to increase context

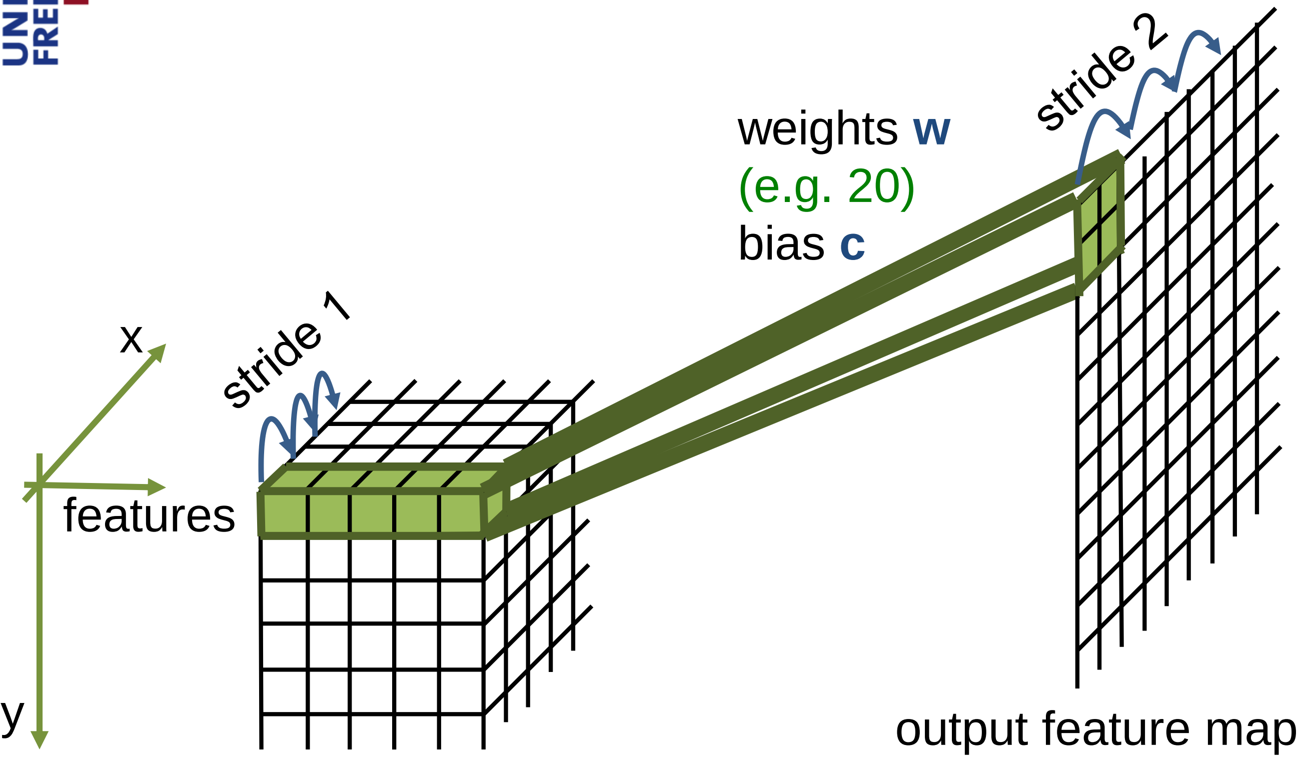
U-Net Architecture



Synthesis Path



2x2 up-convolution



resulting feature maps have factor 2 higher resolution

output feature map **b**
(only 1 channel shown)

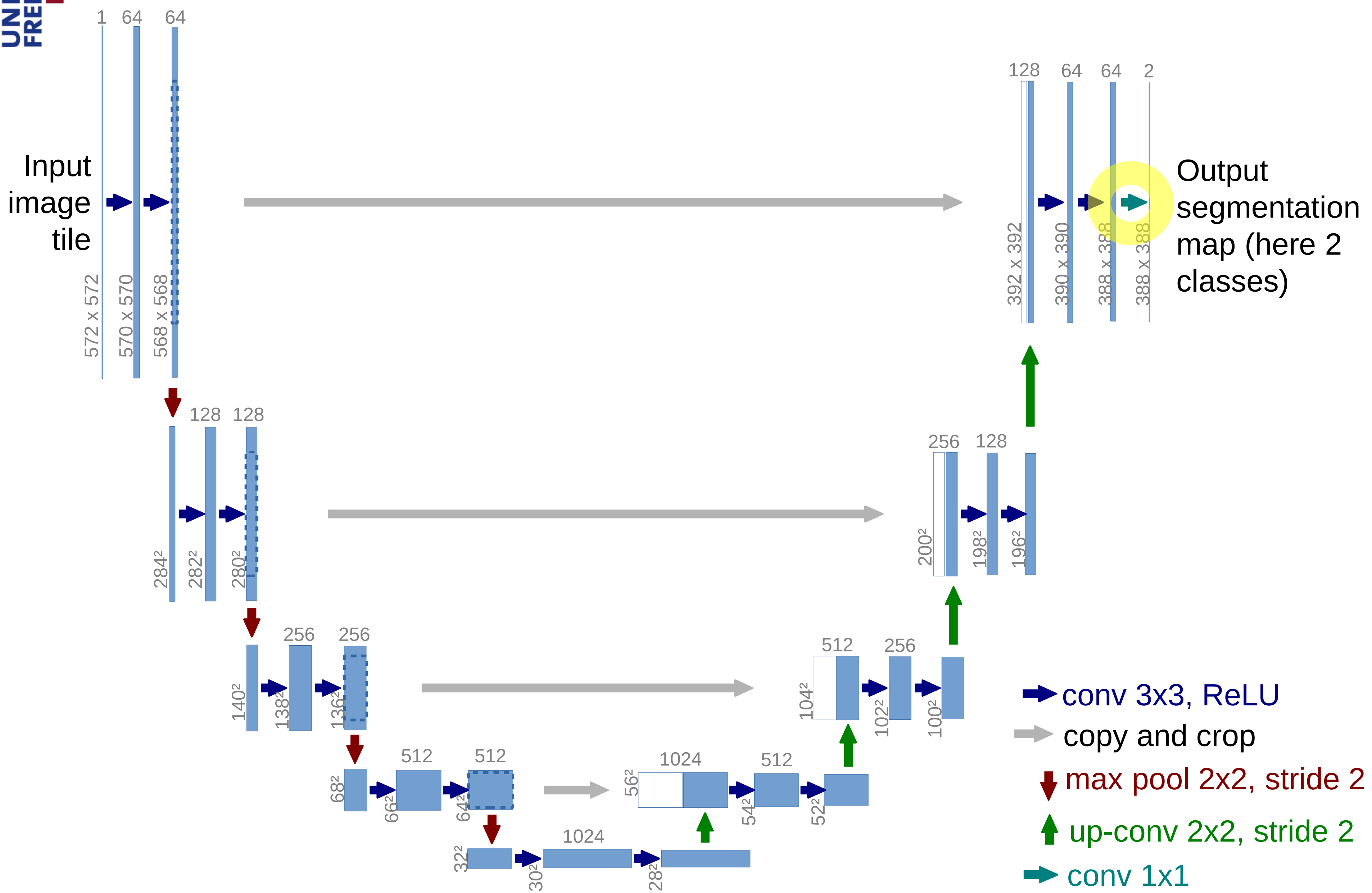
input feature map **a**
(e.g. 5 channels)

$$b_{2x+i,2y+j,l} = \mathrm{ReLU}\Big(\sum_{\substack{i\in\{0,1\}\\ j\in\{0,1\}\\ k\in\{1,\cdots,K\}}} w_{i,j,k,l} \cdot a_{x,y,k} + c_l \Big)$$

Intuition: Learned "Upsampling"

→ Synthesize high-res images from low-res feature vectors

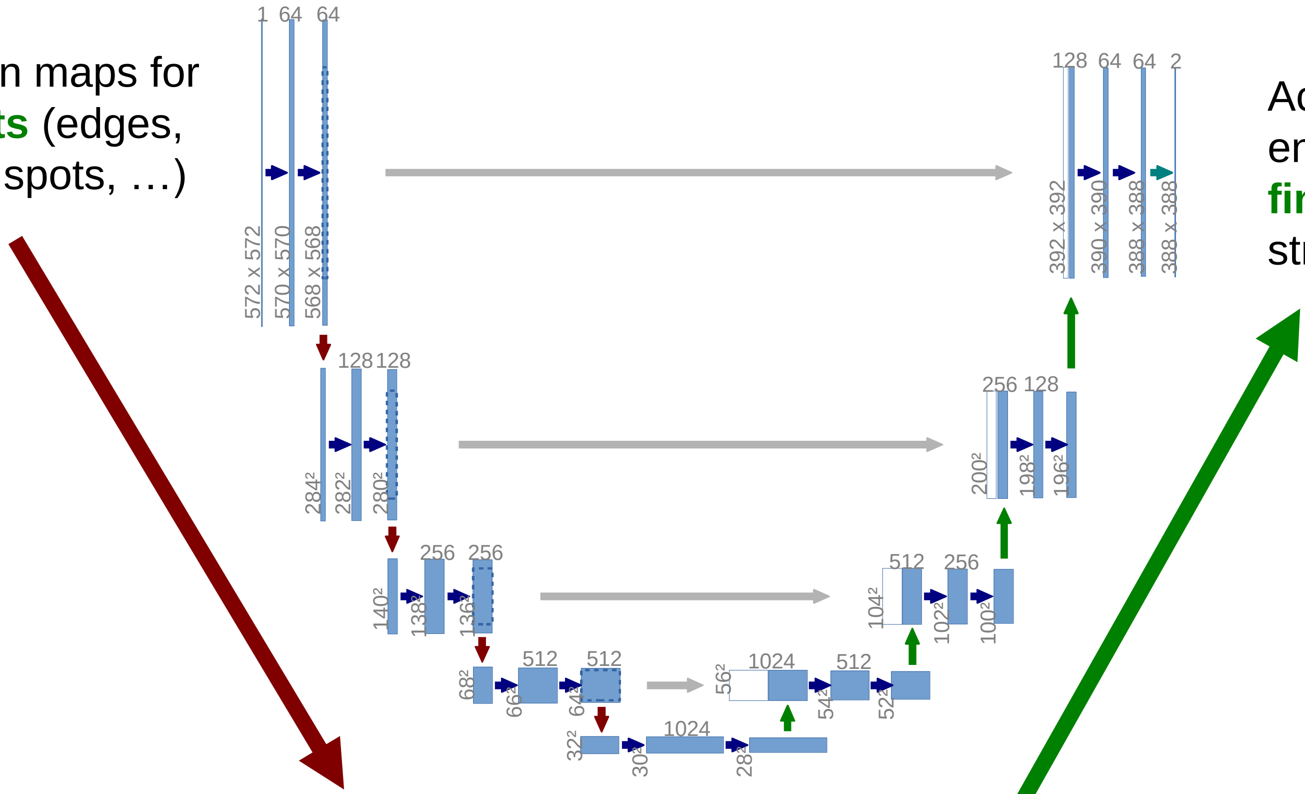
Synthesis Path



What is in the Feature Maps?

Activation maps for **tiny parts** (edges, corners, spots, …)

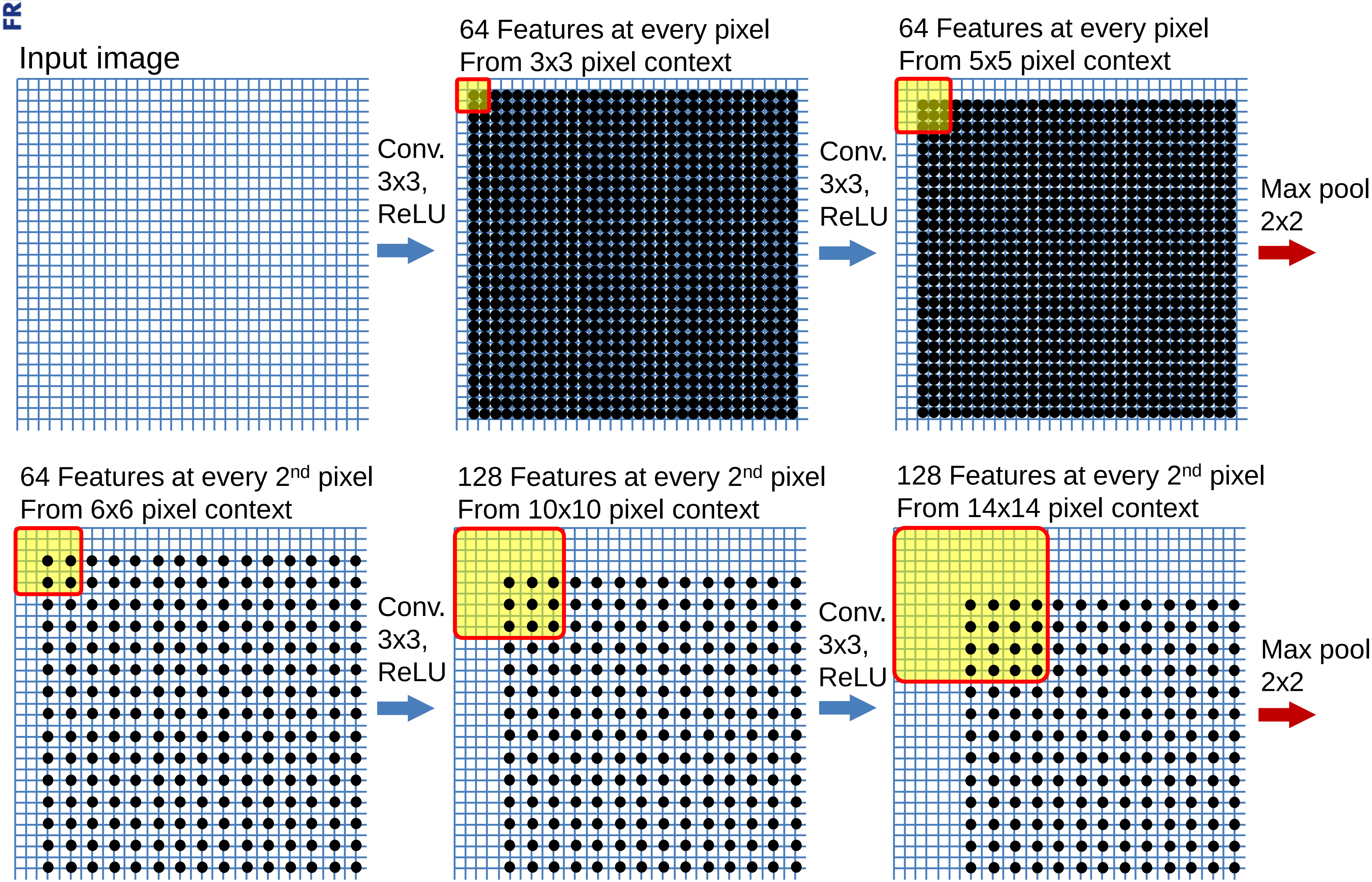
Activation maps that encode the final **fine segmentation** structures



Activation maps for **large parts** (full cells, mitochondria, cell constellations, …)

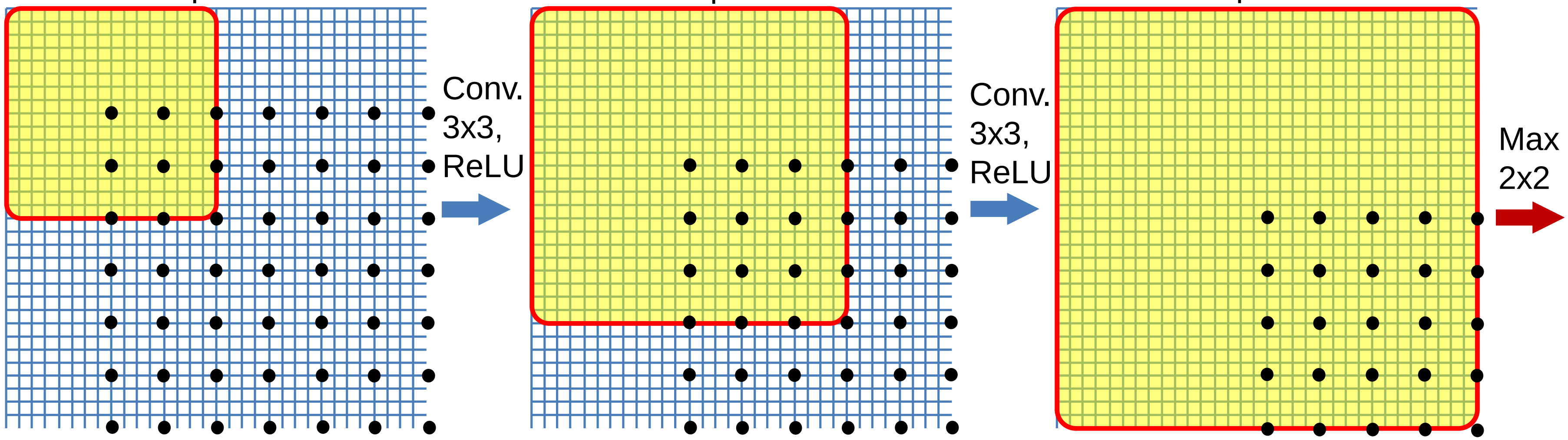
Activation maps that encode **coarse segmentation** structures

Analysis Path: Resulting Feature Vectors



Anaylsis Path: Resulting Feature Vectors

128 Features at every 4^{th} pixel From 16x16 pixel context | 256 Features at every 4^{th} pixel From 24x24 pixel context | 256 Features at every 4^{th} pixel From 32x32 pixel context

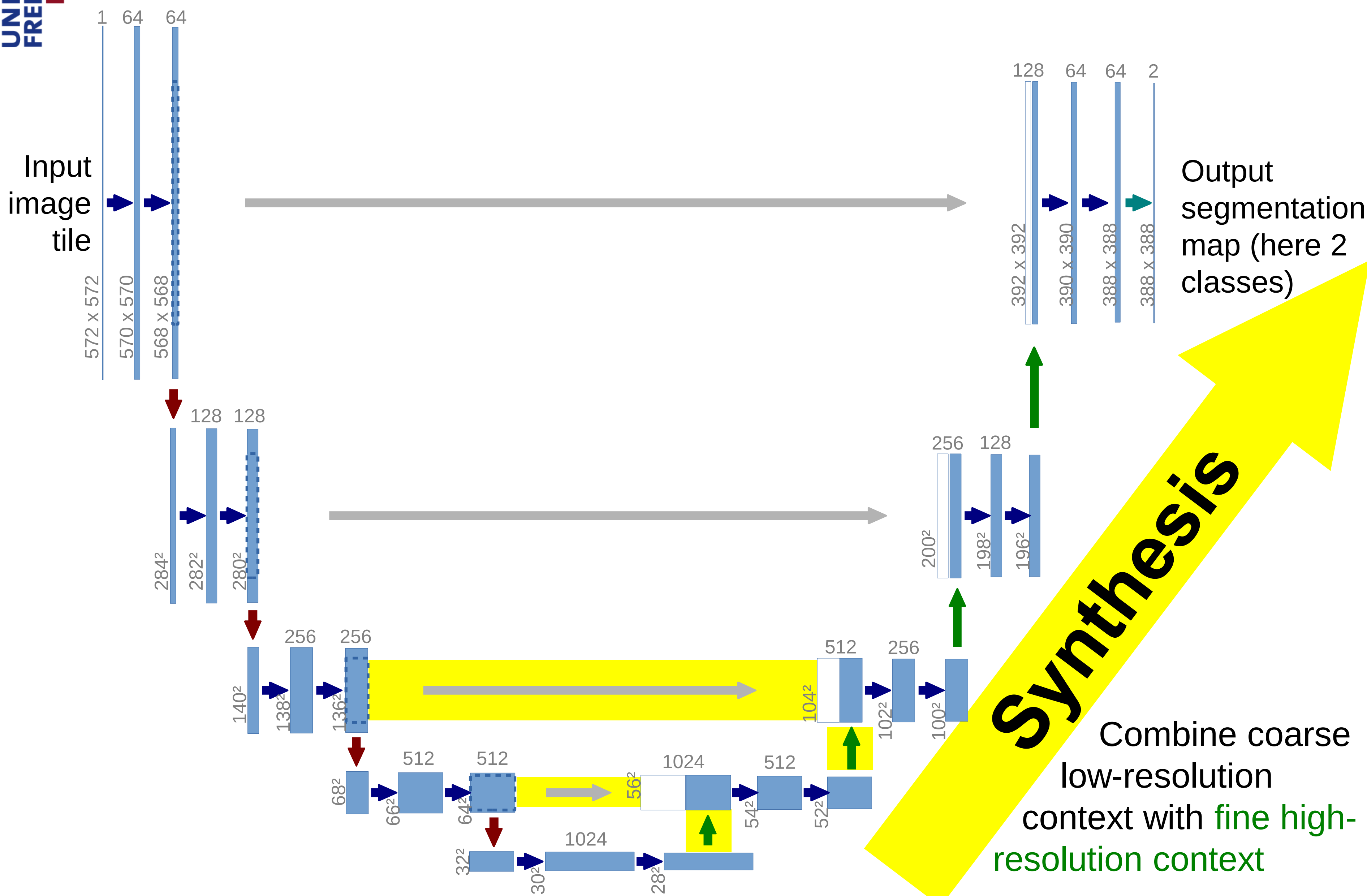


and so on: Increase the **"What"**, reduce the **"Where"**

Features	256	512	512	512	1024	1024
Sampling	8	8	8	16	16	16
Context	36	52	68	76	108	140

… until 1024 features at every 16^{th} pixel from 140x140 pixel context

U-net Architecture



Synthesis Path

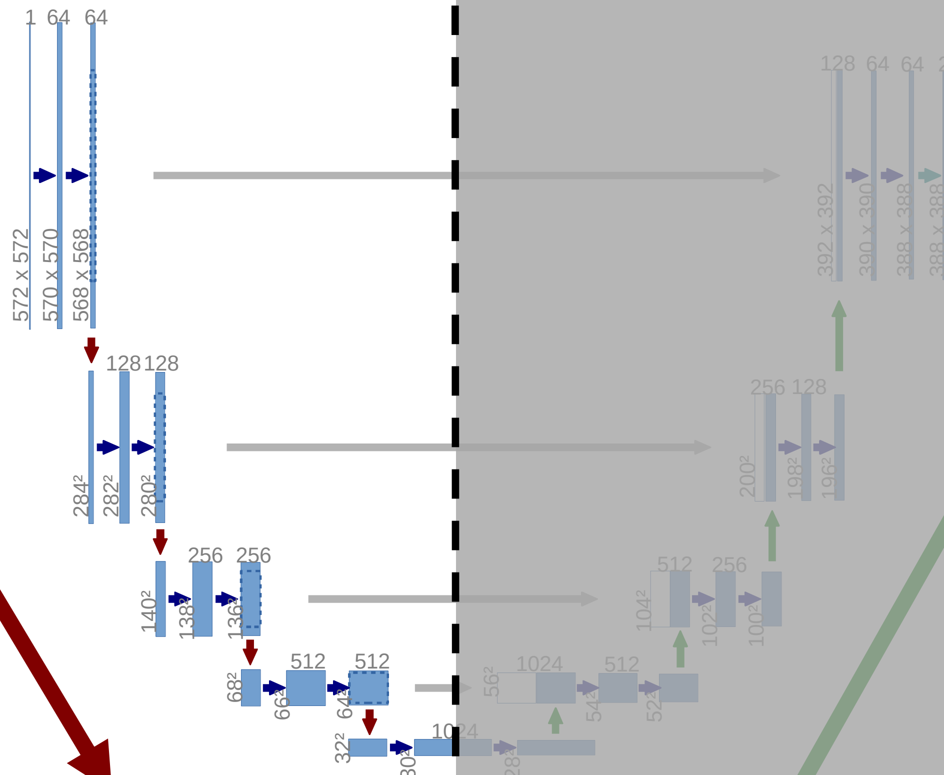
- Combine coarse low-resolution context with fine high-resolution context

Features	512 +512	512	512	256 +256	256	256	128 +128	128	128	64 +64	64	64	7
Sampling	8	8	8	4	4	4	2	2	2	1	1	1	1
Context	132 (68)	148	164	160 (32)	168	176	174 (14)	178	182	181 (5)	183	185	185

- … until 7 features (scores for the 7 classes) at every pixel from 185x185 pixel context.

Visualization of Standard Classification Networks

Activation maps for **tiny parts** (edges, corners, spots, …)

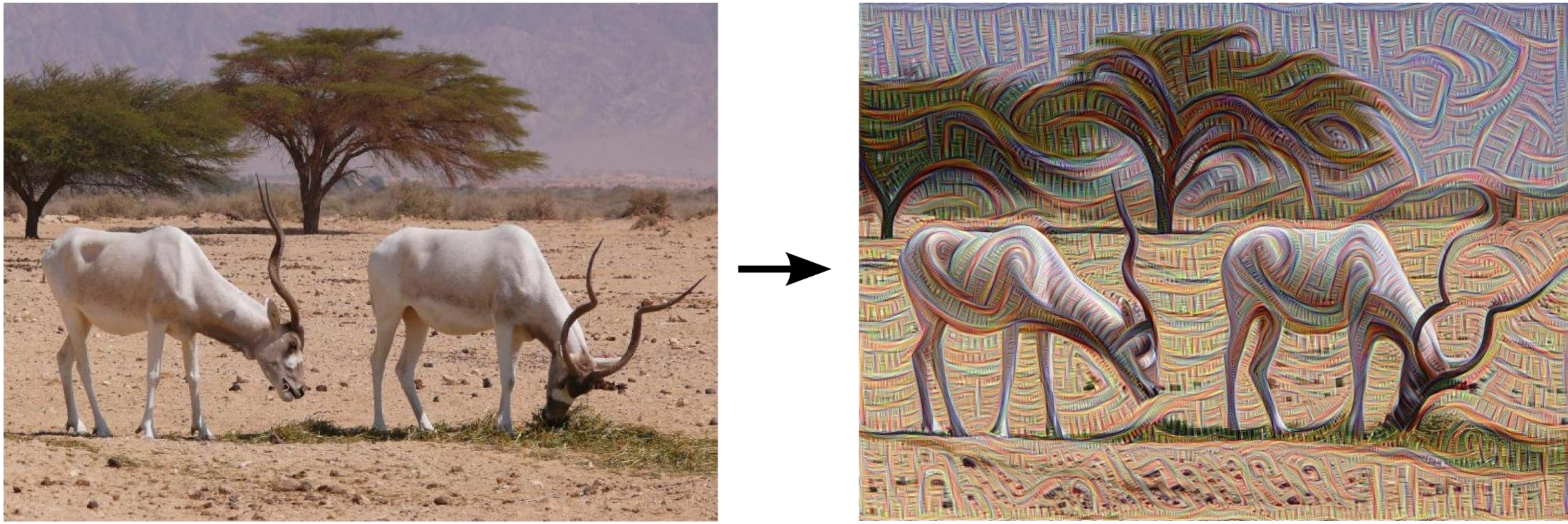


Activation maps for **large parts** (full cells, mitochondria, cell constellations, …)

Activation maps that encode the final **fine segmentation** structures

Activation maps that encode **coarse segmentation** structures

Visualization of Neuron Activations



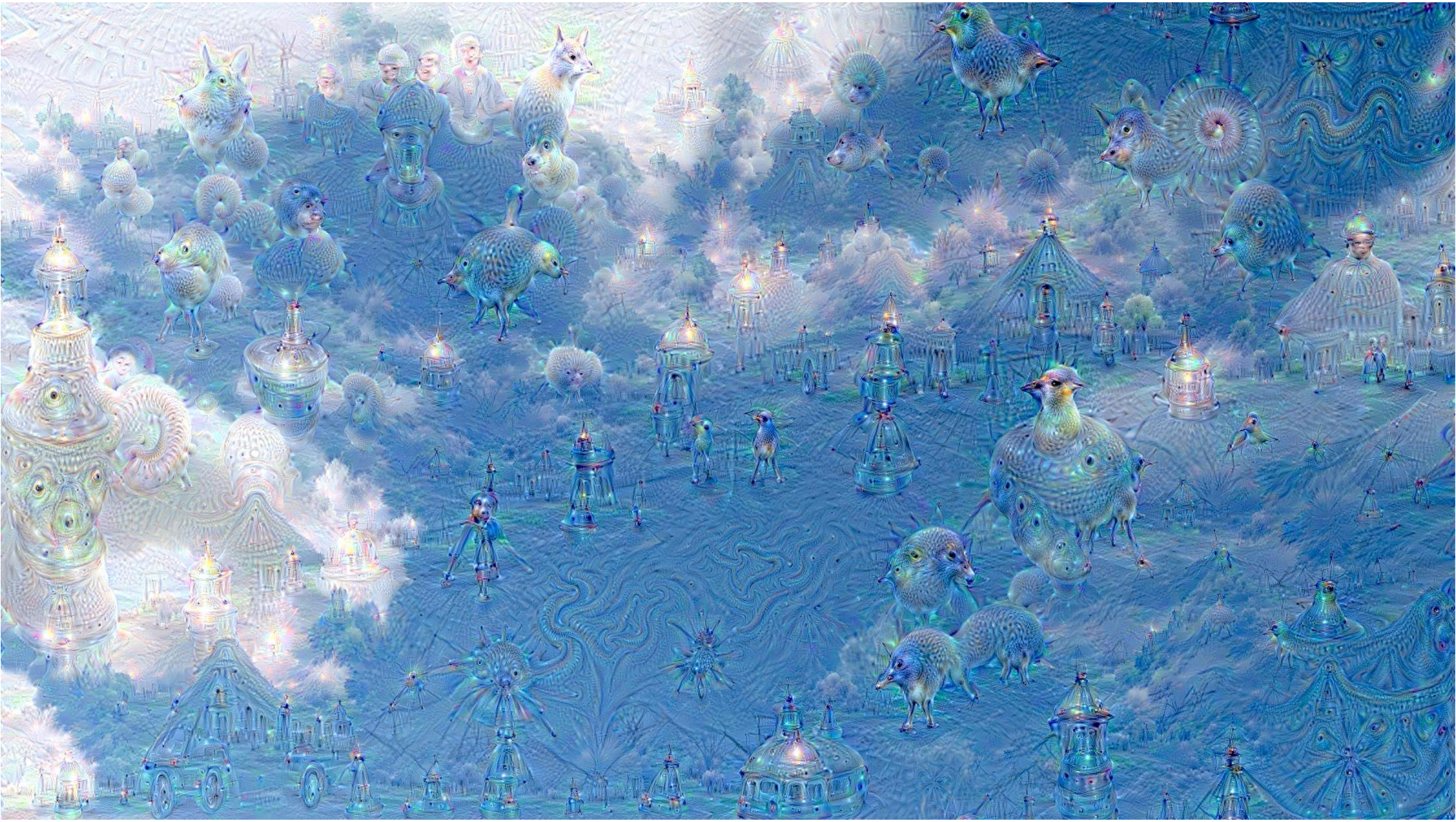
- propagate the image through the network
- check with neurons get the **strongest activation**
- change the image, such that these activations are further **amplified**
- **Here**: Neurons in **early layers** (encoding small-scale structures)

http://googleresearch.blogspot.de/2015/06/inceptionism-going-deeper-into-neural.html

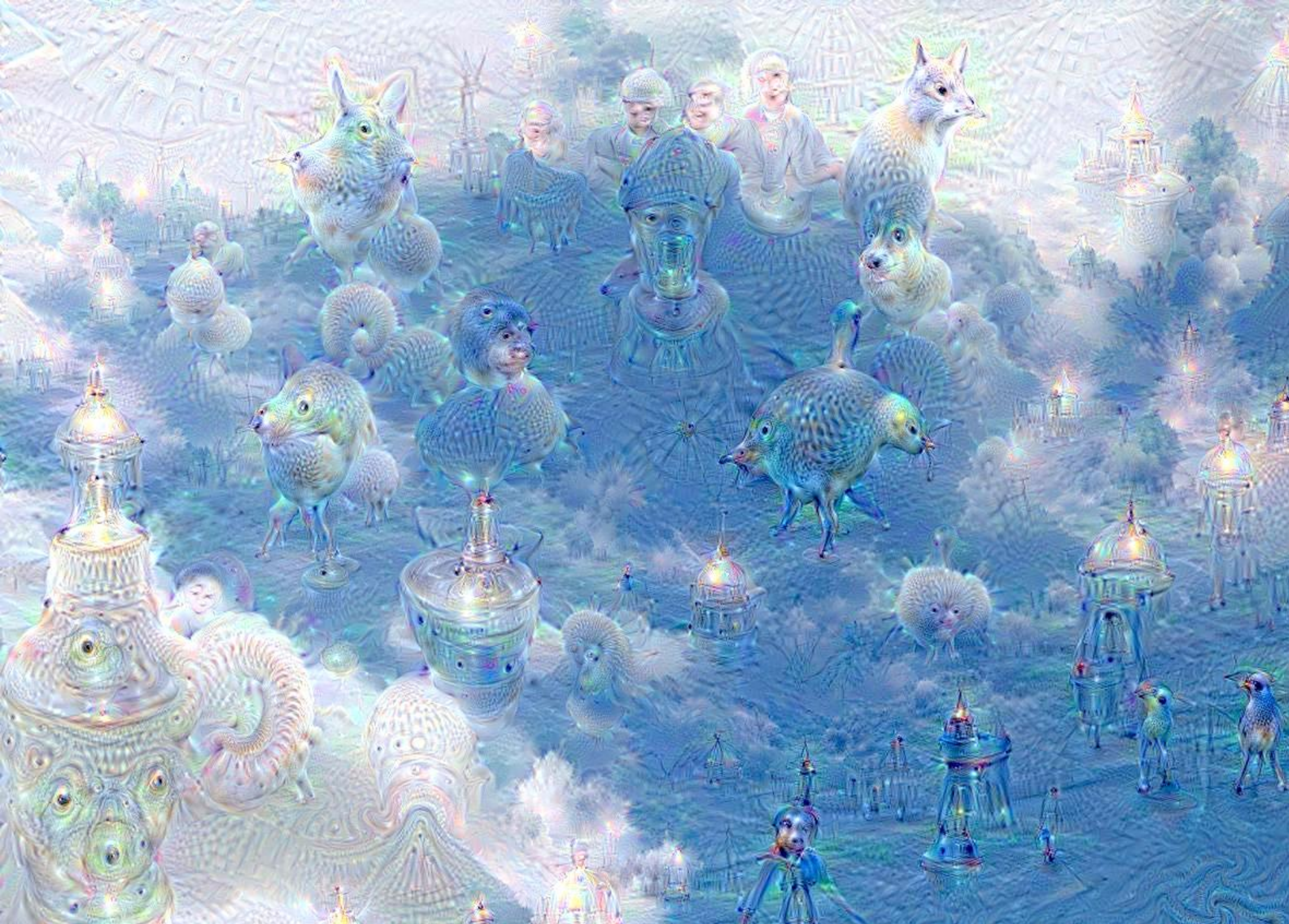
Neuron Activations in Late Layers

http://googleresearch.blogspot.de/2015/06/inceptionism-going-deeper-into-neural.html

Amplification of large-scale structures



http://googleresearch.blogspot.de/2015/06/inceptionism-going-deeper-into-neural.html

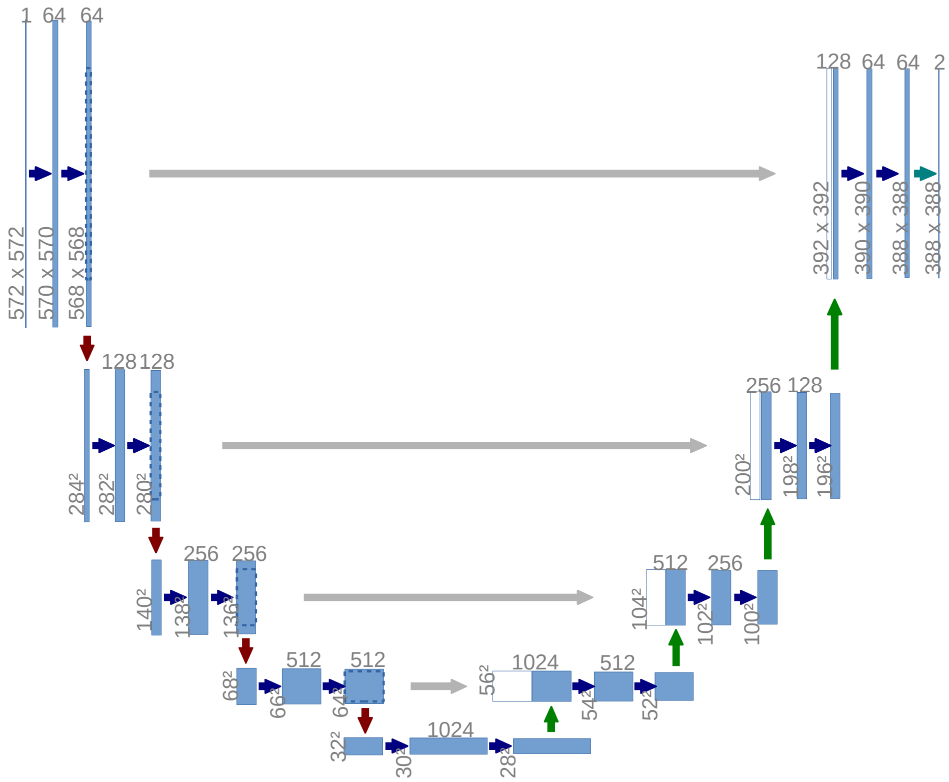


http://googleresearch.blogspot.de/2015/06/inceptionism-going-deeper-into-neural.html

http://googleresearch.blogspot.de/2015/06/inceptionism-going-deeper-into-neural.html

Overlap-tile strategy for arbitrary large images

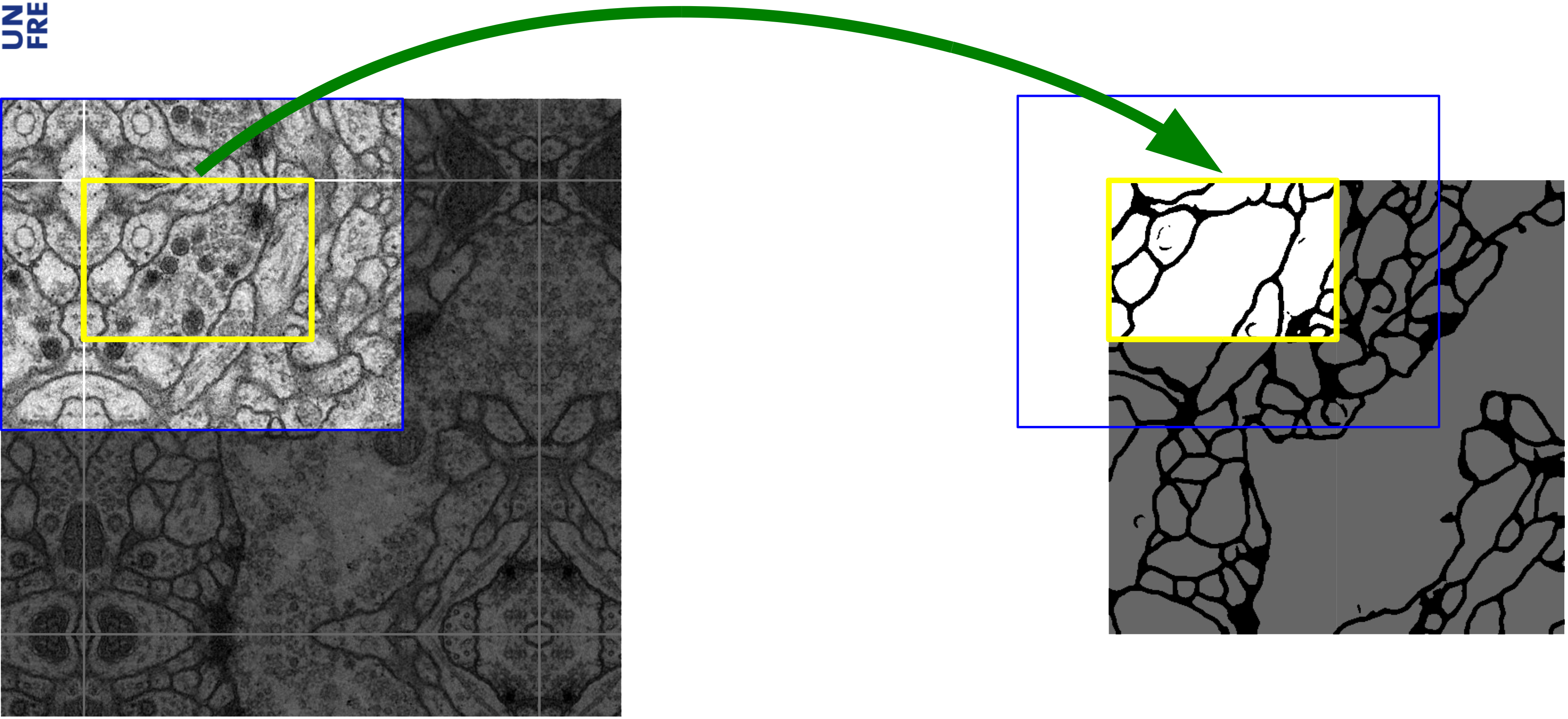
Input image tile,
572 x 572 pixel



Output segmentation map:
388 x 388 pixel

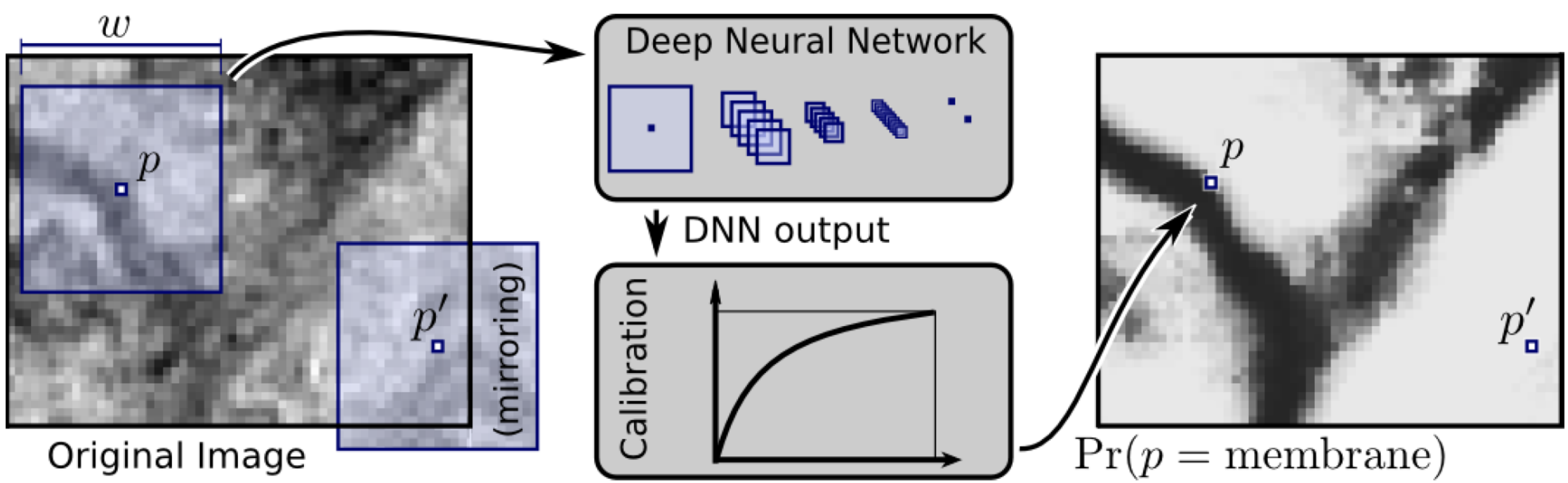
Use only the valid part of each convolution

Overlap-tile strategy for arbitrary large images



- Segmentation of the yellow area needs input data of the blue area
- Raw data extrapolation my mirroring

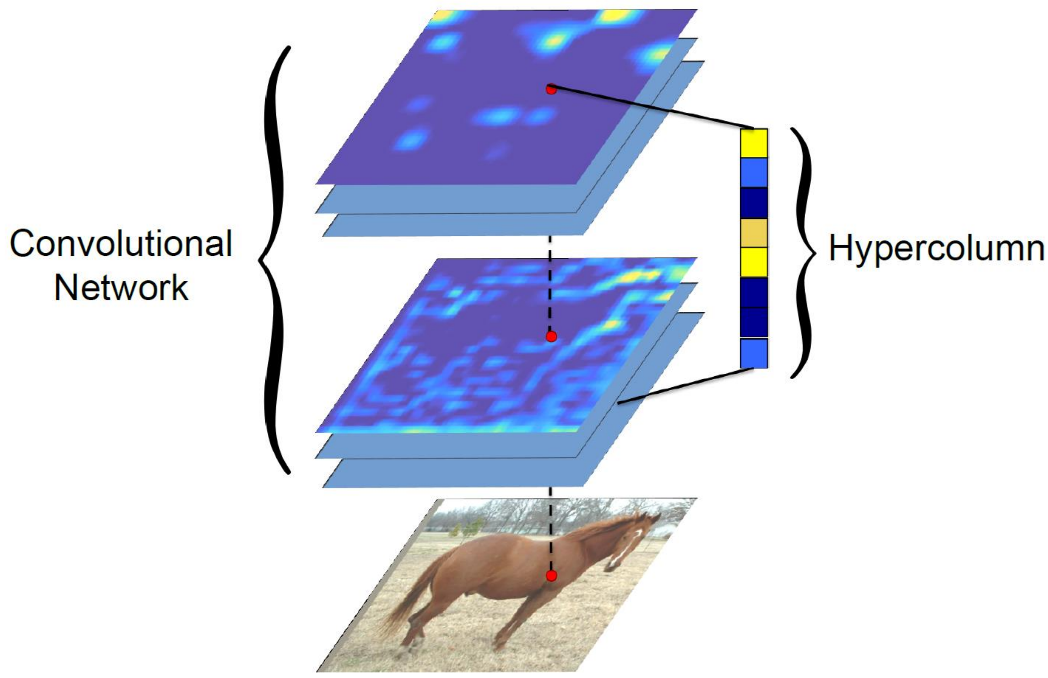
Related Work: Sliding-Window Approach



D. Ciresan, A. Giusti, L. M. Gambardella, J. Schmidhuber - Deep Neural Networks Segment Neuronal Membranes in Electron Microscopy Images (NIPS 2012)

- **Sliding-window approach** with a standard classification network
- Winner of **EM segmentation challenge** at ISBI 2012
- Needs **full network evaluation** for each pixel
- **Trade-off** between precise localization and large context (larger windows need **more max pooling layers**)
- (u-net outperformed them now)

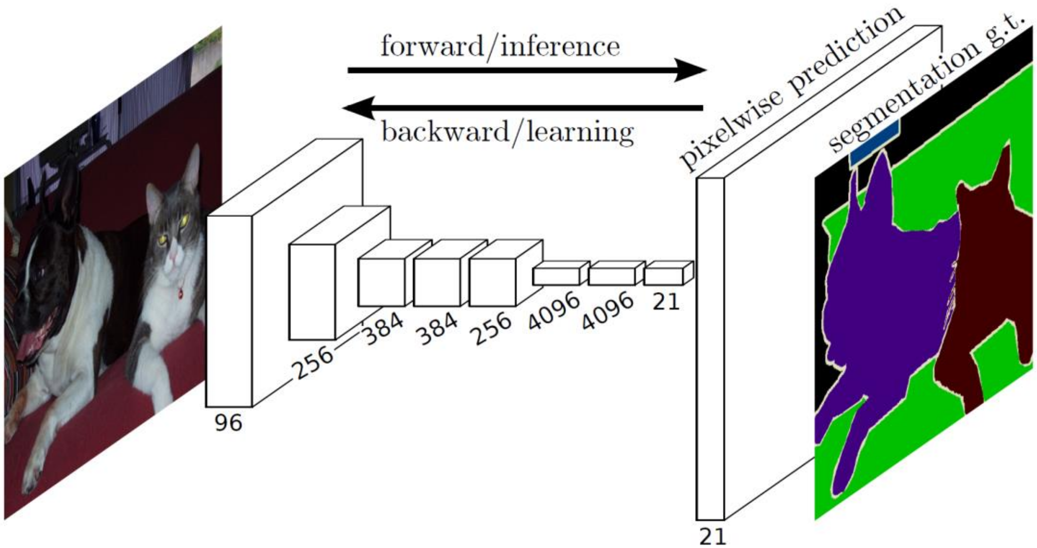
Related Work: Hypercolumns

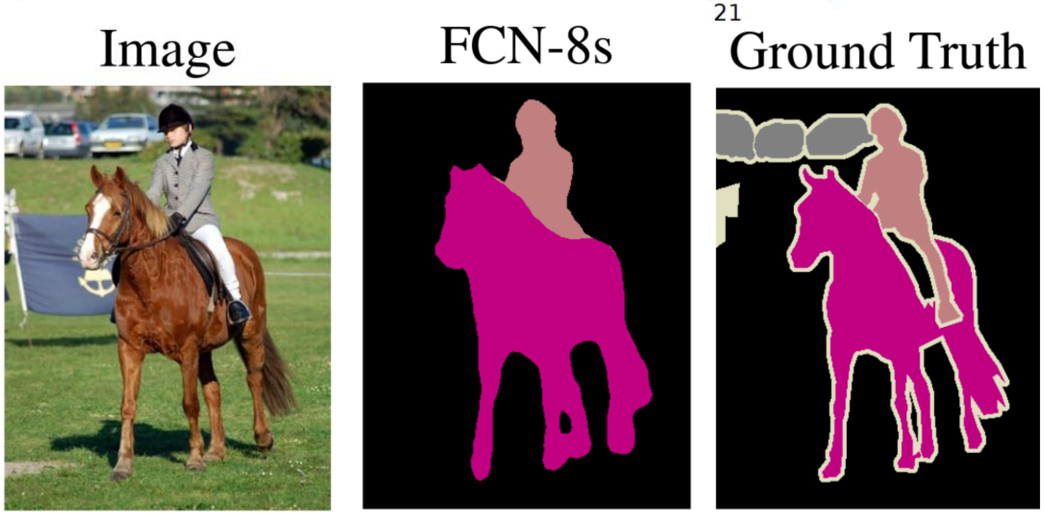


- **Rescale all feature maps** to original image size.
- Use **support vector machine** to predict label at each position.
- Needs a **pre-trained classification network.**
- **No end-to-end** training possible.

B. Hariharan, P. Arbeláez, R. Girshick, J. Malik: “Hypercolumns for Object Segmentation and Fine-grained Localization” *CVPR*, 2015. arXiv:1411.5752 [cs.CV]

Related Work: Fully Convolutional Network

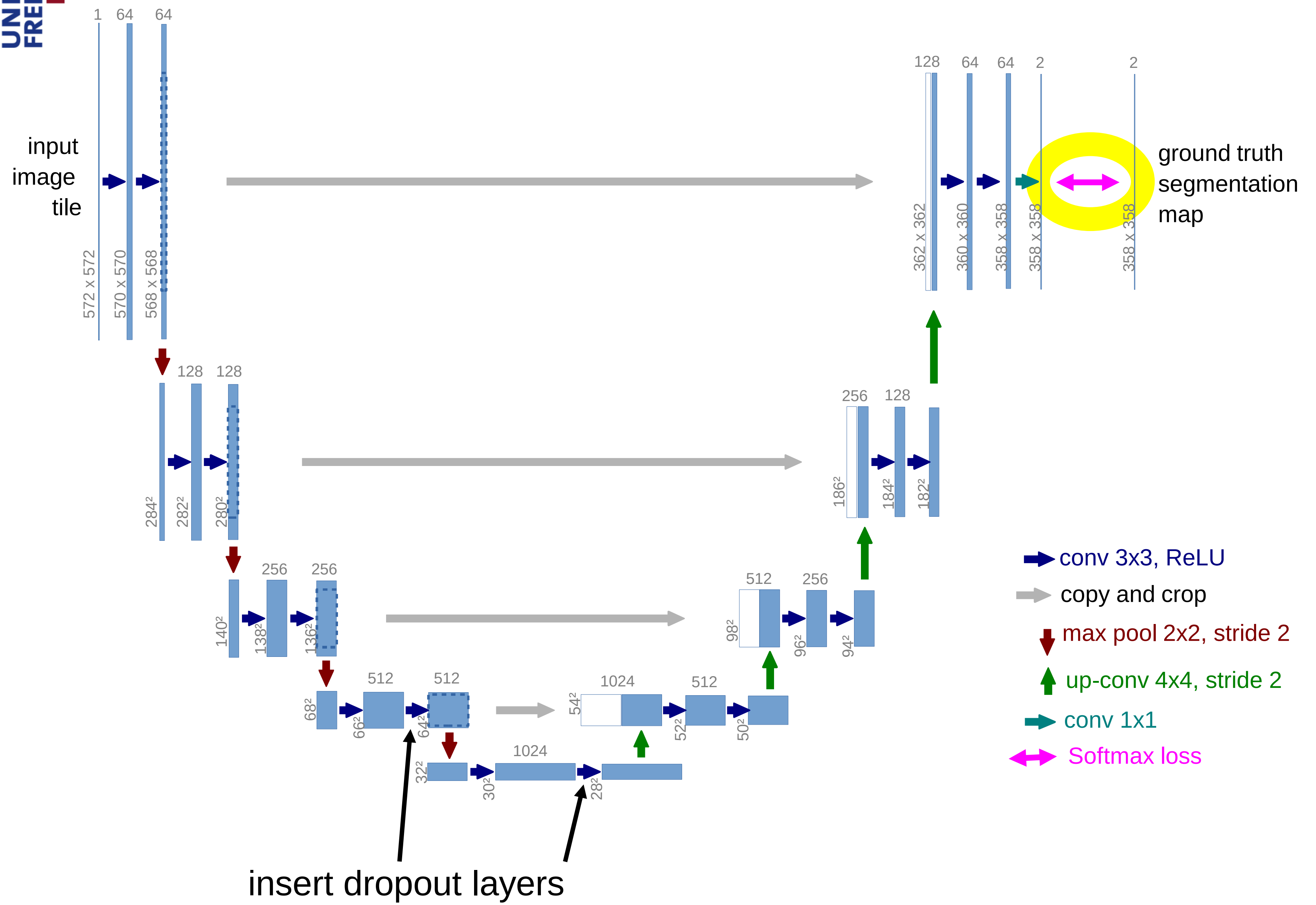




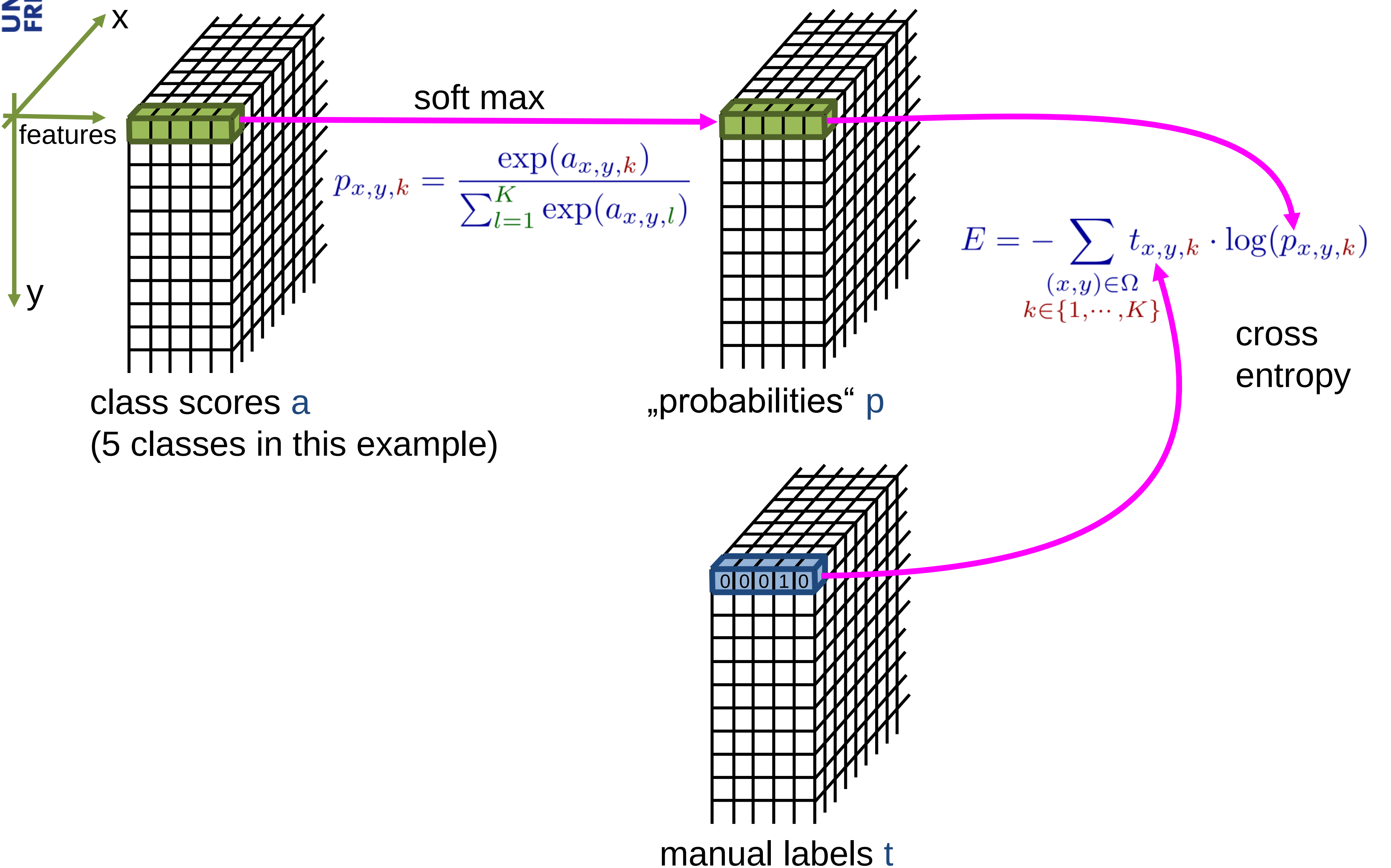
- **Nearly identical architecture** like u-net
- Uses a **pre-trained classification network**
- Uses only **1 feature channel per class** in lowest resolution
- Does not go back to full resolution in **synthesis path**
- Uses **padded convolutions**, no tiling possible

J. Long, E. Shelhamer, T. Darrell: “Fully Convolutional Networks for Semantic Segmentation”. CVPR 2015, arXiv:1411.4038 [cs.CV]

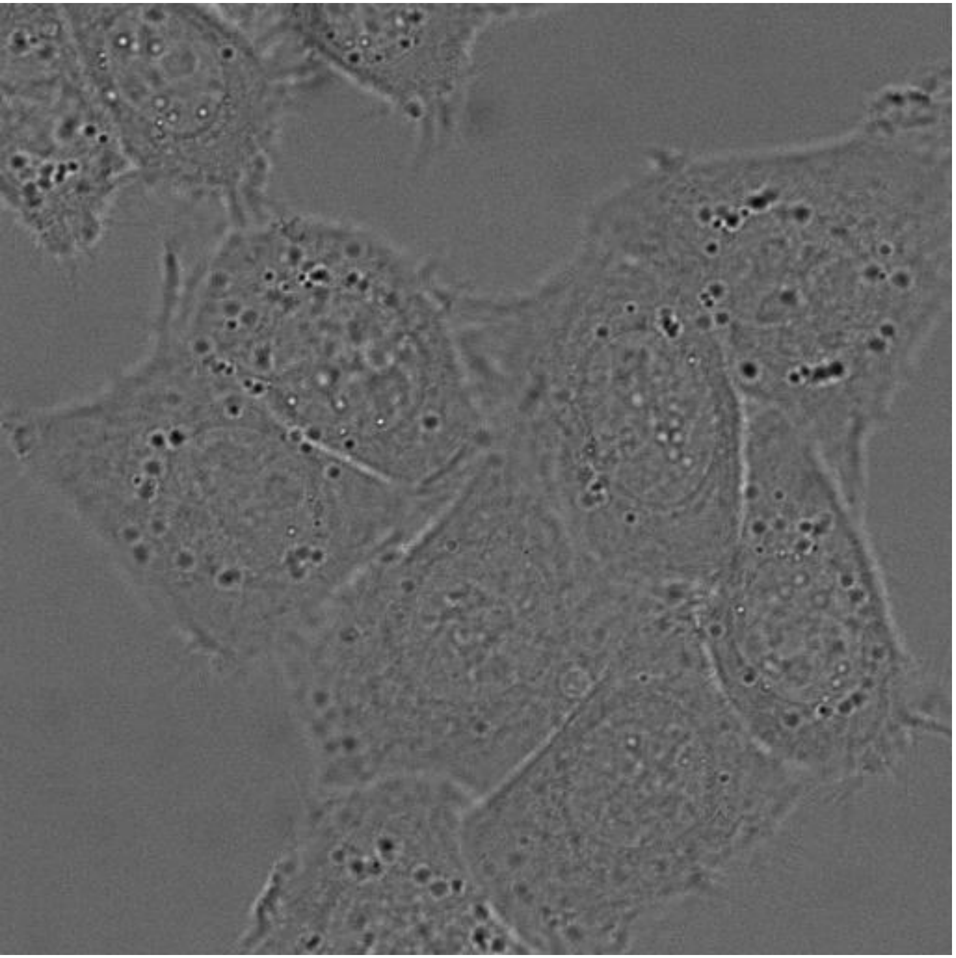
Training the U-net



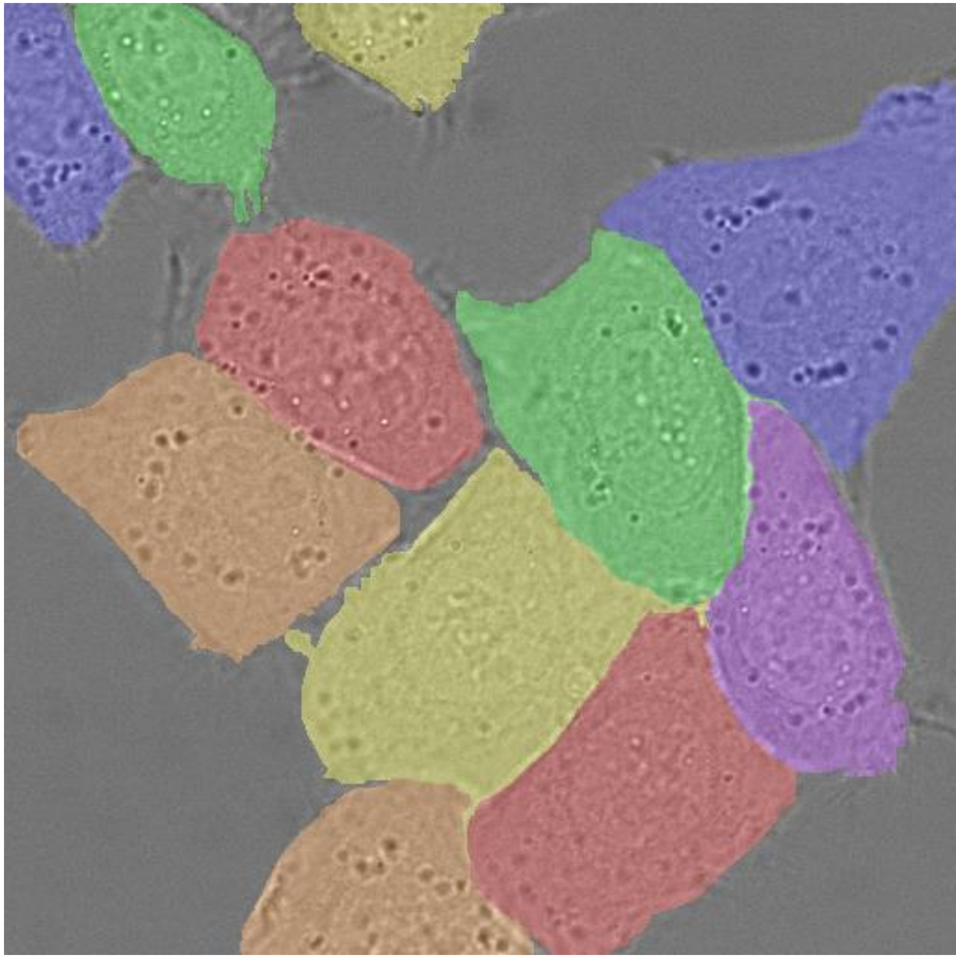
Softmax-loss layer



Segmentation of Touching Objects of the Same Class

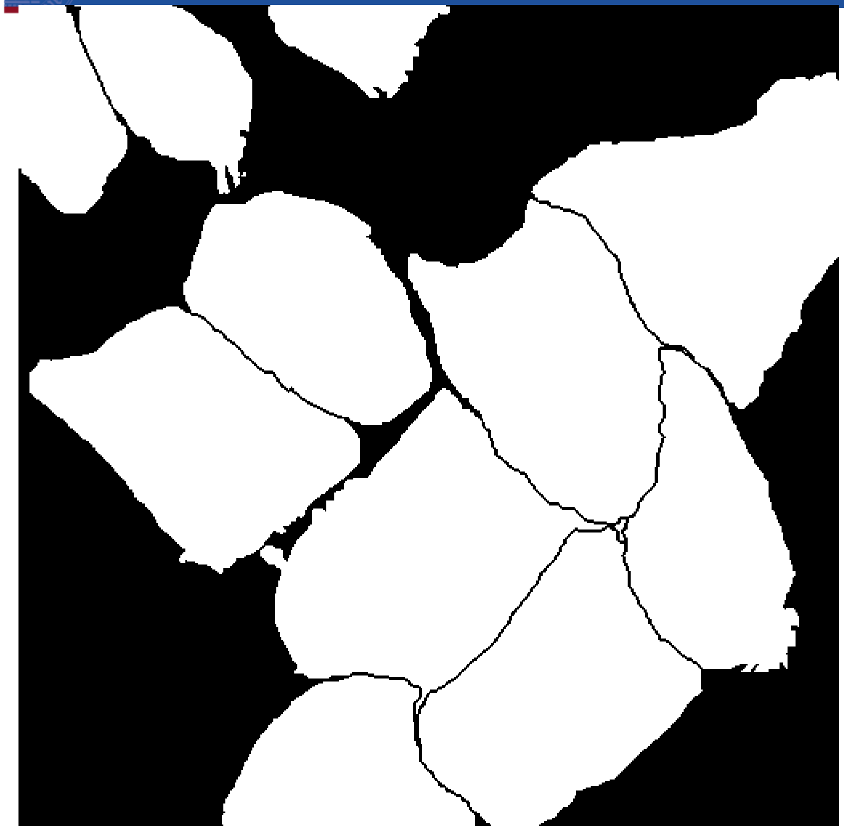


HeLa cells recorded with DIC microscopy

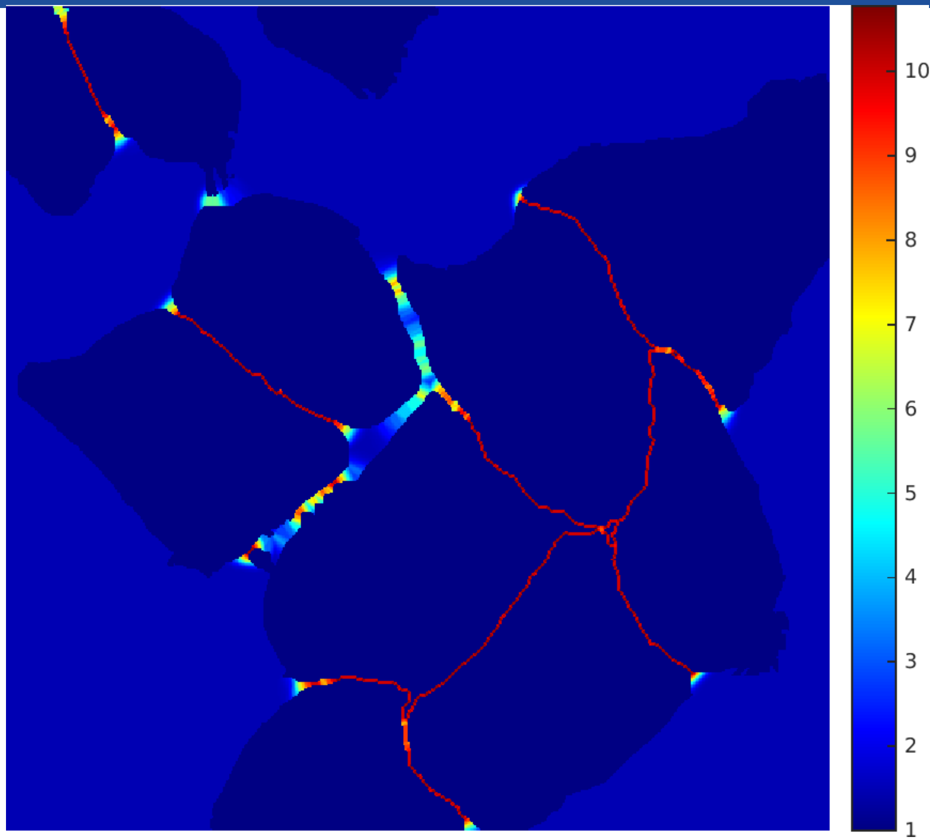


manual segmentation
(colors: different instances)

Ensure Separation of Touching Objects



Segmentation mask for training (inserted background between touching objects)

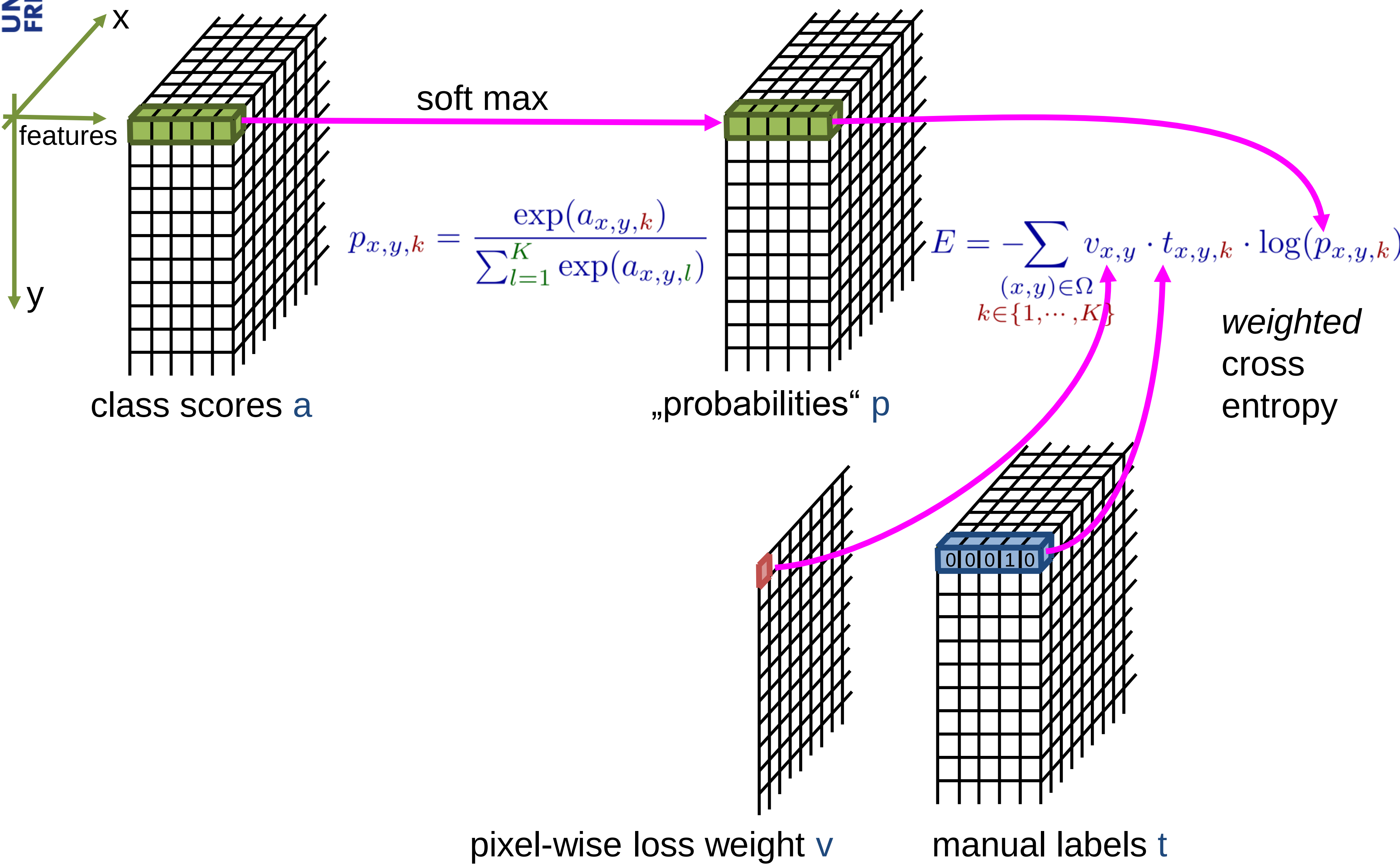


Loss-weight for each pixel

$$v(\mathbf{x}) = v_{\text{bal}}(\mathbf{x}) + v_0 \cdot \exp\left(-\frac{(d_1(\mathbf{x}) + d_2(\mathbf{x}))^2}{2\sigma^2}\right)$$

v_{bal}: weight map to balance class frequency
d_1: distance to border of closest object
d_2: distance to border of second closest object

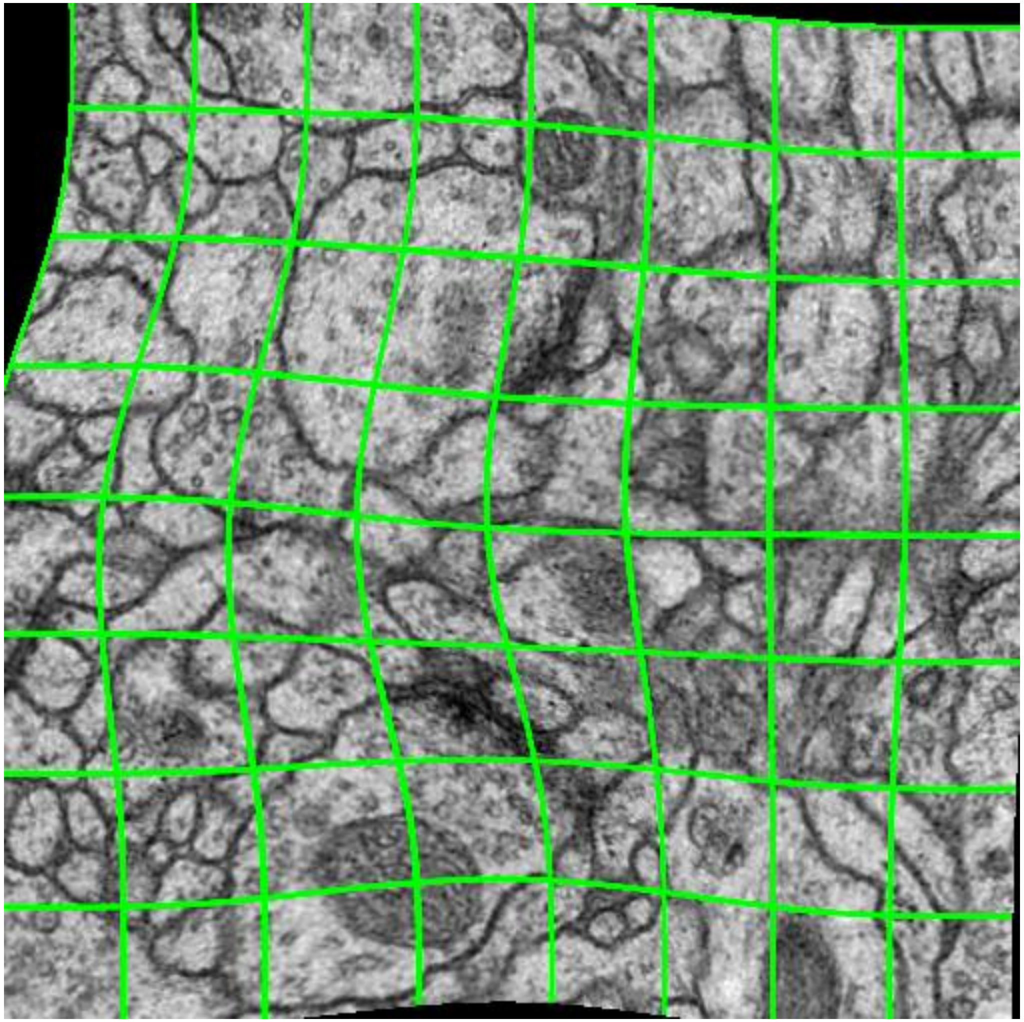
Weighted Softmax-loss layer



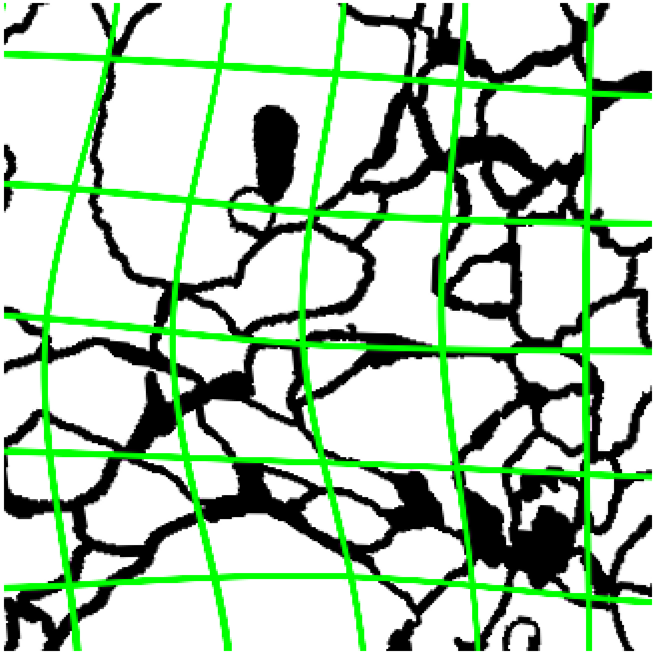
$$p_{x,y,k} = \frac{\exp(a_{x,y,k})}{\sum_{l=1}^{K} \exp(a_{x,y,l})}$$

$$E = -\sum_{\substack{(x,y)\in\Omega \\ k\in\{1,\cdots,K\}}} v_{x,y} \cdot t_{x,y,k} \cdot \log(p_{x,y,k})$$

Augment Training Data using Deformations

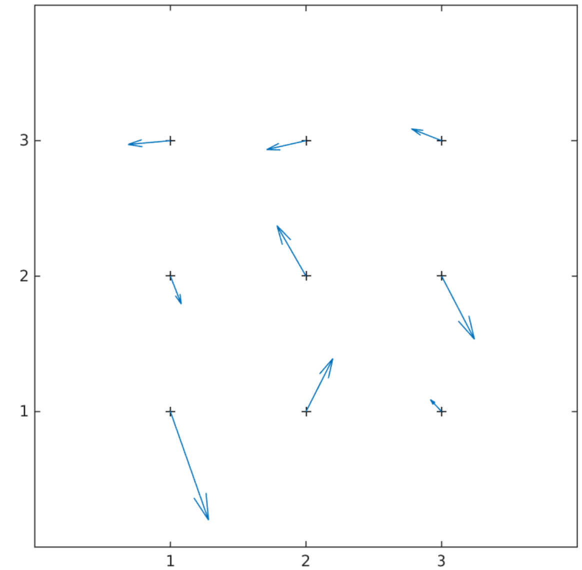


Randomly deformed image
(for visualization: no rotation, no shift, no extrapolation)

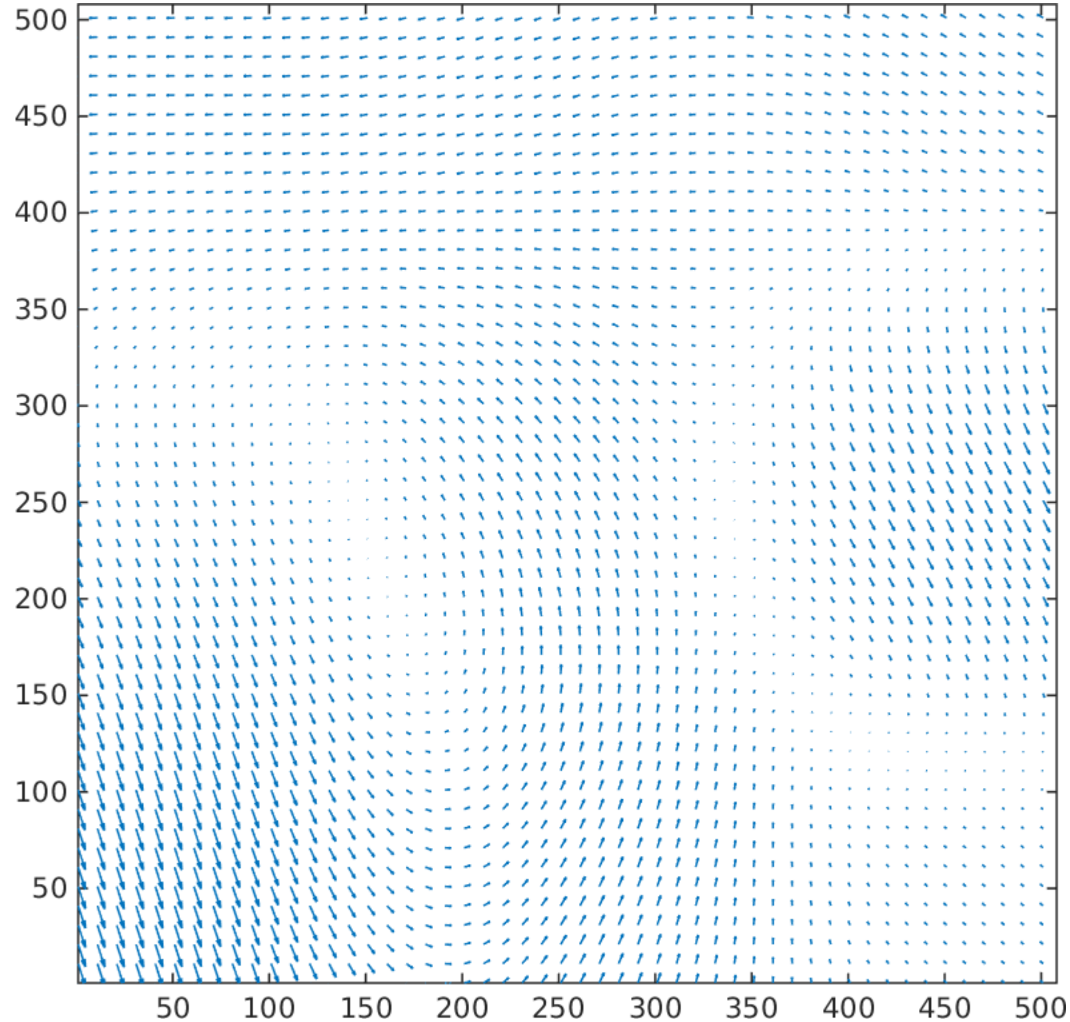


correspondingly deformed
manual labels

Creating Random Deformations



3x3 random deformation vectors



Full deformation field by **bicubic interpolation**

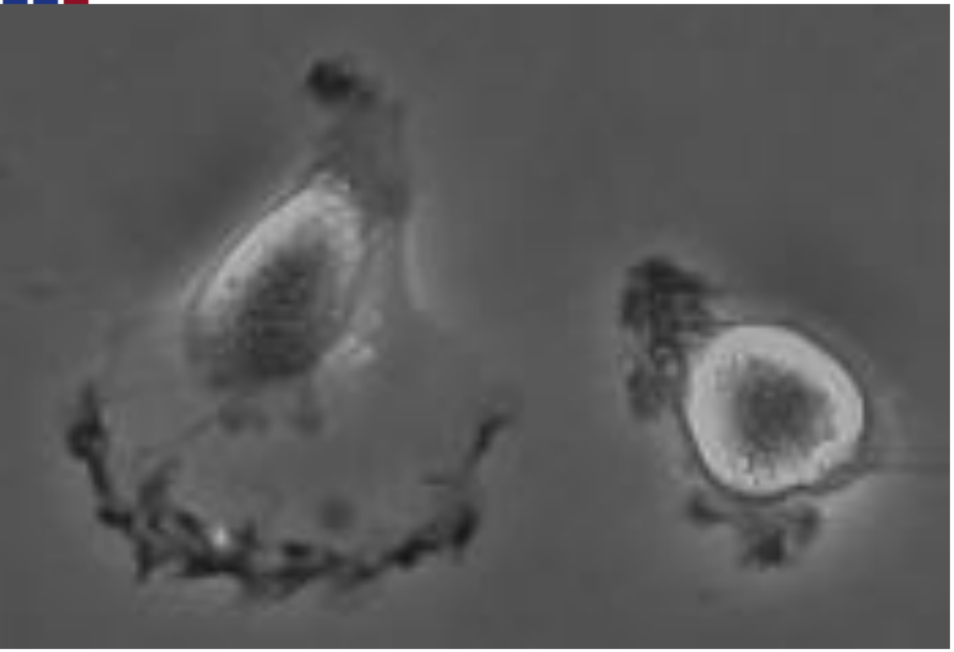
Implementation Details

- **Caffe neural network library** with the „Deconvolution“ and Cropping Layer from the FCN
- Added **sqrt(2/N)** initialization, and **weighted softmaxloss** layer
- **Data augmentation** and weight computation in matlab
  - Create 20,000 augmented image tiles (with corresponding segmentation and weight maps)
  - Rotation augmentation: 360° (25° for dental x-ray)
  - Random cropping position (shift augmentation)
  - Random mirroring
  - Random deformation: 3x3 normally distributed displacement vectors with 10 pixels standard deviation
  - Intensity augmentation 10%: resulting factor: $(1 + N(0,0.1))$

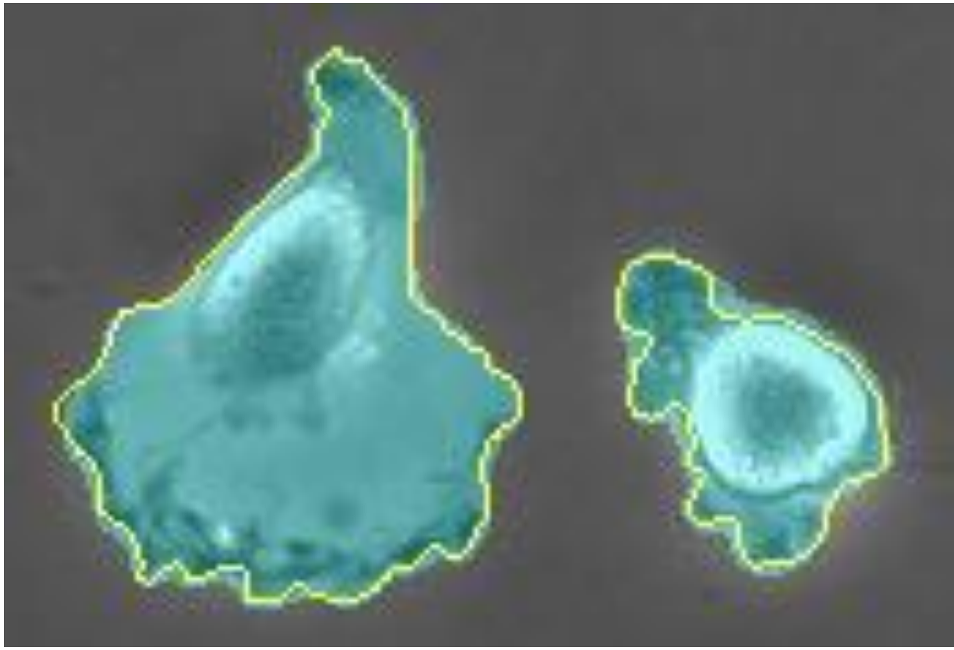
Implementation Details

- **Overlap-tile classification** in matlab (using caffe‘s matlab interface). **Averaging** over 7 rotated versions (EM data) or 4 mirrored versions (all other applications)
- **Training**: Stochastic gradient descent with momentum
  - „batch“-size: 1 image
  - Initial learning rate: 0.001
  - Momentum: 0.99 (very high to virtually increase batch size),
  - All 20,000 iterations, reduce learning rate by 1/10
  - 60,000 iterations total
  - Training time: approx 10 hours on a Nvidia Titan GPU. First meaningful results already after a few 1000 iterations (half an hour or less)

ISBI cell tracking challenge: dataset **PhC-U373**



Glioblastoma-astrocytoma U373 cells on a polyacrylimide substrate
Phase contrast microscopy

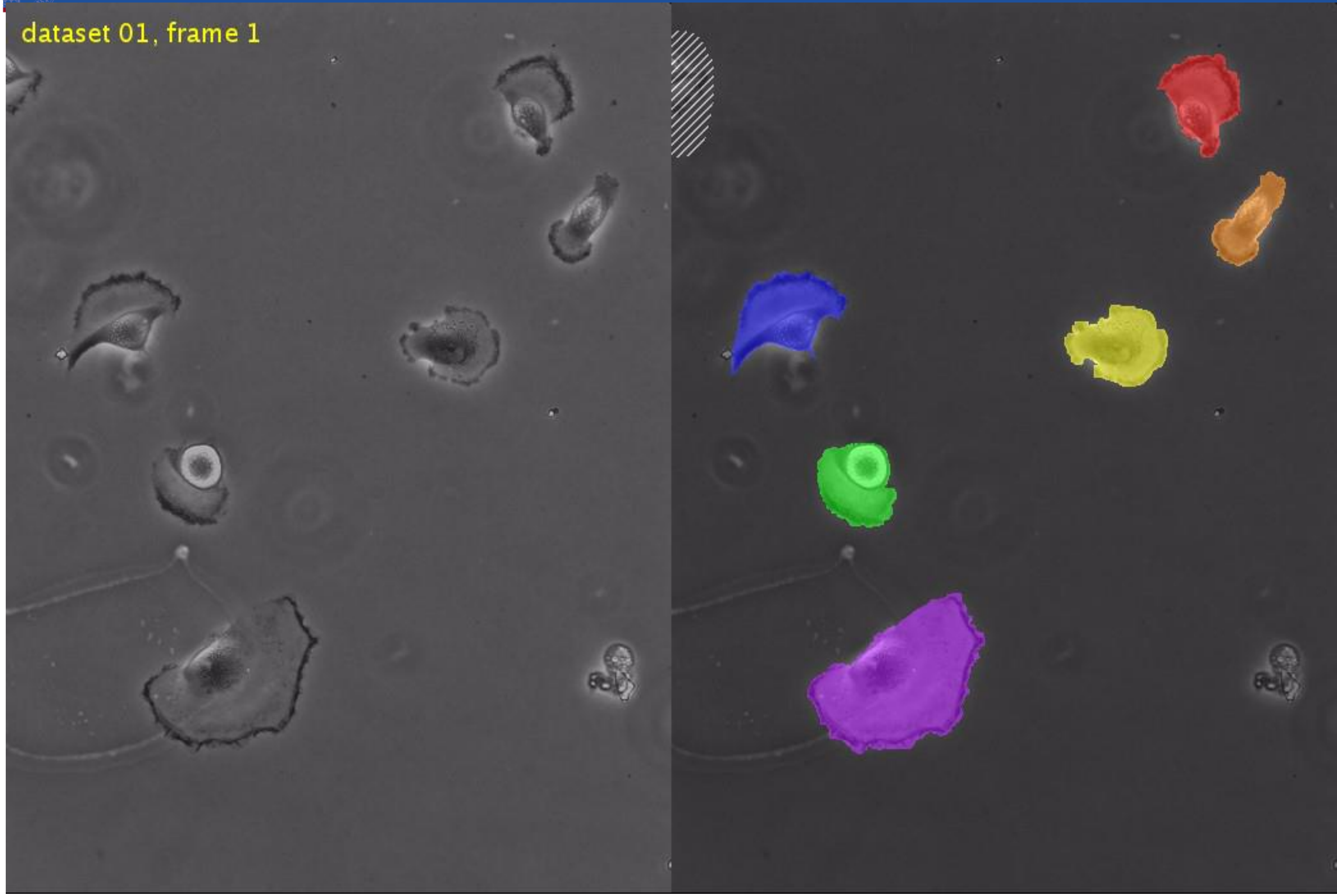


Cyan: segmentation by u-net
Yellow borders: our manual ground truth

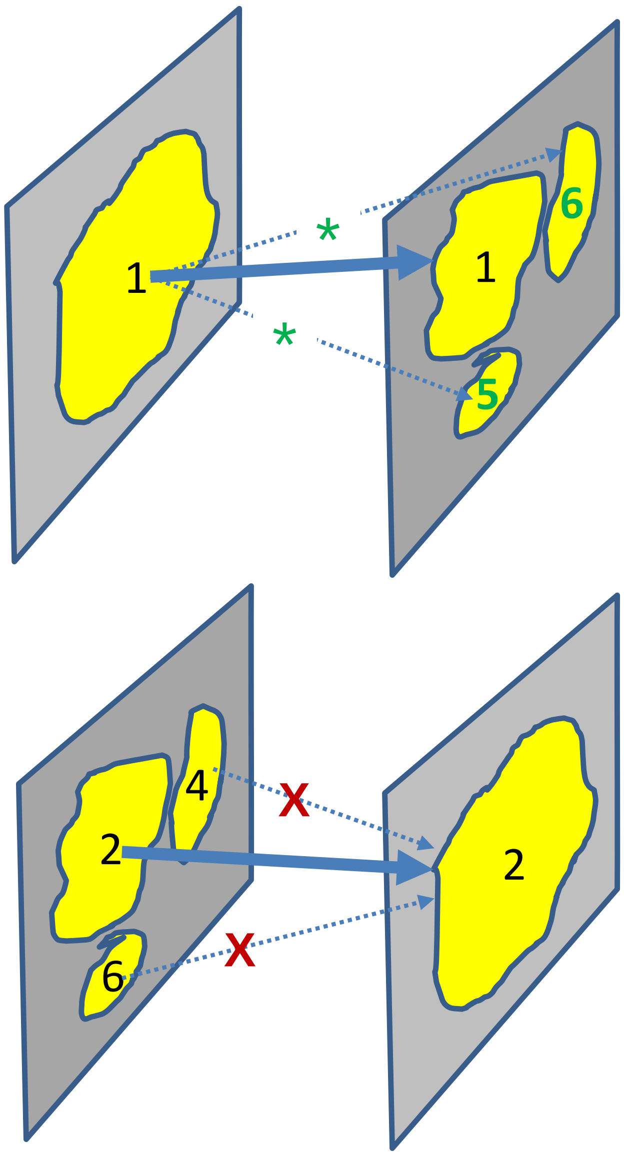
- Strong shape variations
- Weak outer borders, strong irrelevant inner borders
- Cytoplasm has same structure like background

Data provided by Dr. Sanjay Kumar. Department of Bioengineering
University of California at Berkeley. Berkeley CA (USA)

PhC-U373: Complete Training Set

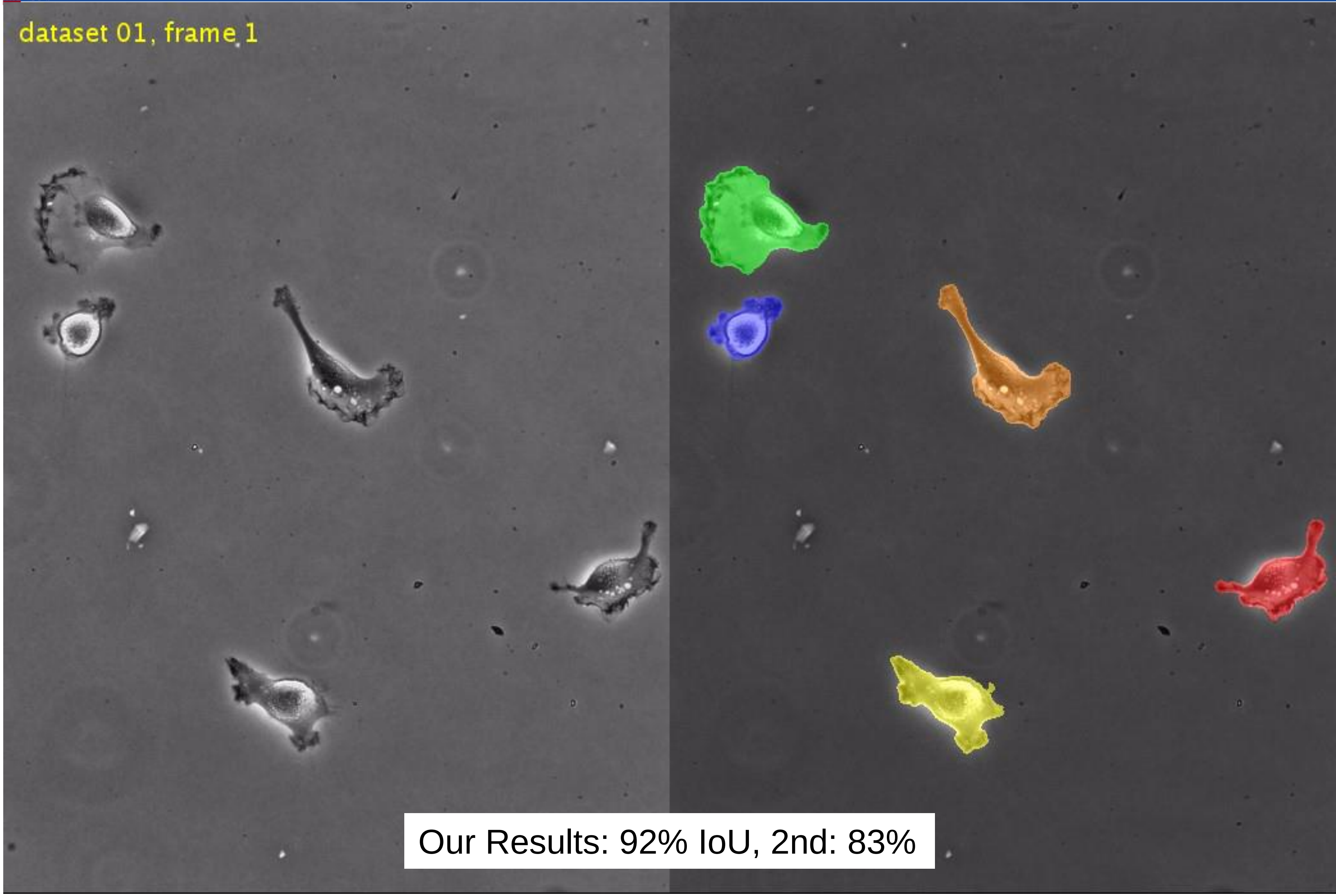


Tracking



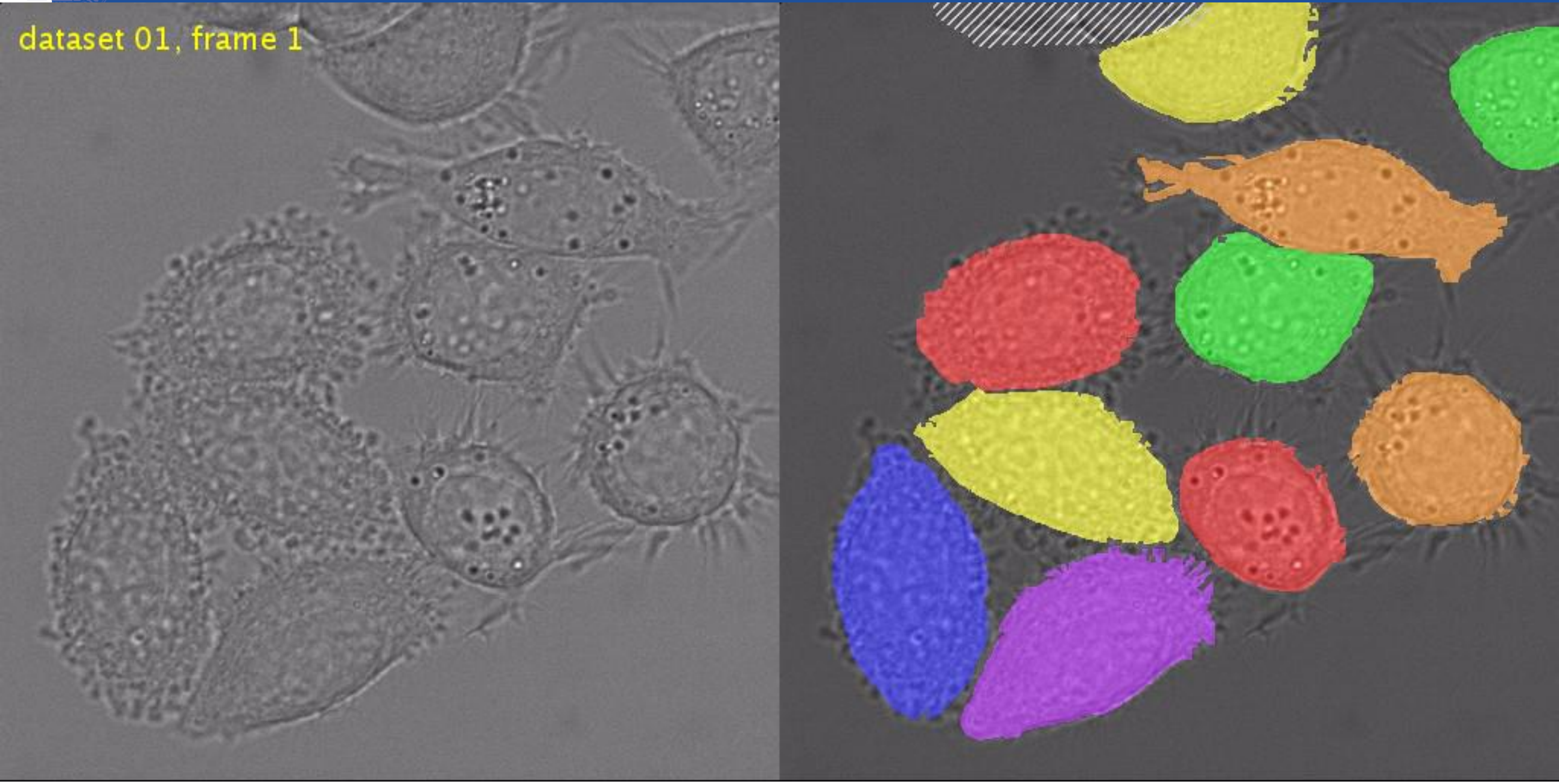
- Propagate Labels to overlapping Segments
- Resolve **one-to-many** correspondences:
  - Propagate label to max. IOU
  - Invent new labels
- Resolve **many-to-one** correspondences
  - Take label from max. IOU
  - Kill other labels

PhC-U373:



Our Results: 92% IoU, 2nd: 83%

DIC-HeLa: Complete Training Set

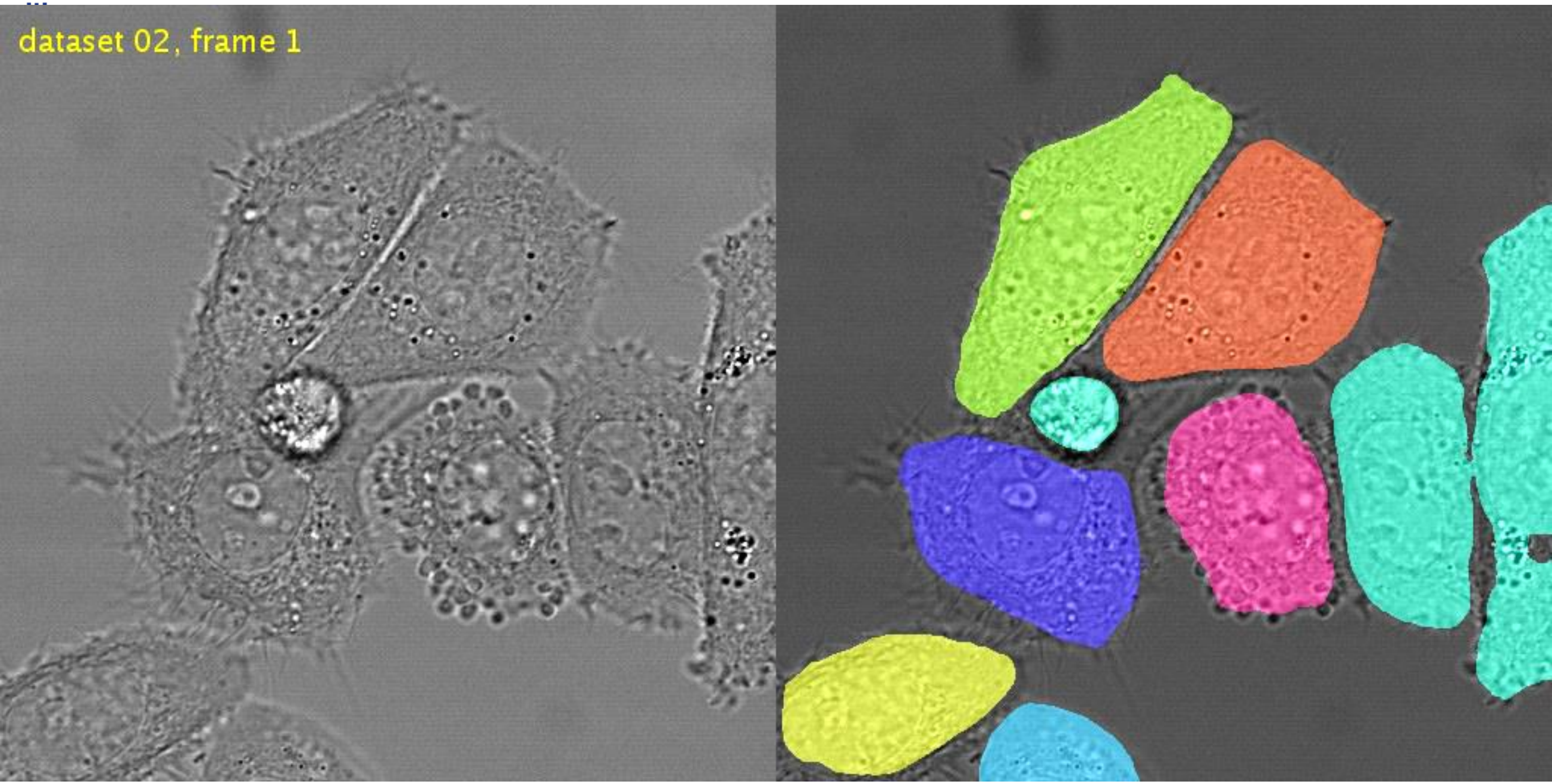


HeLa cells on a flat glass, Differential Interference Contrast (DIC) microscopy

- Touching and overlapping cells (more than one layer)
- partially invisible borders
- cells leave focal plane

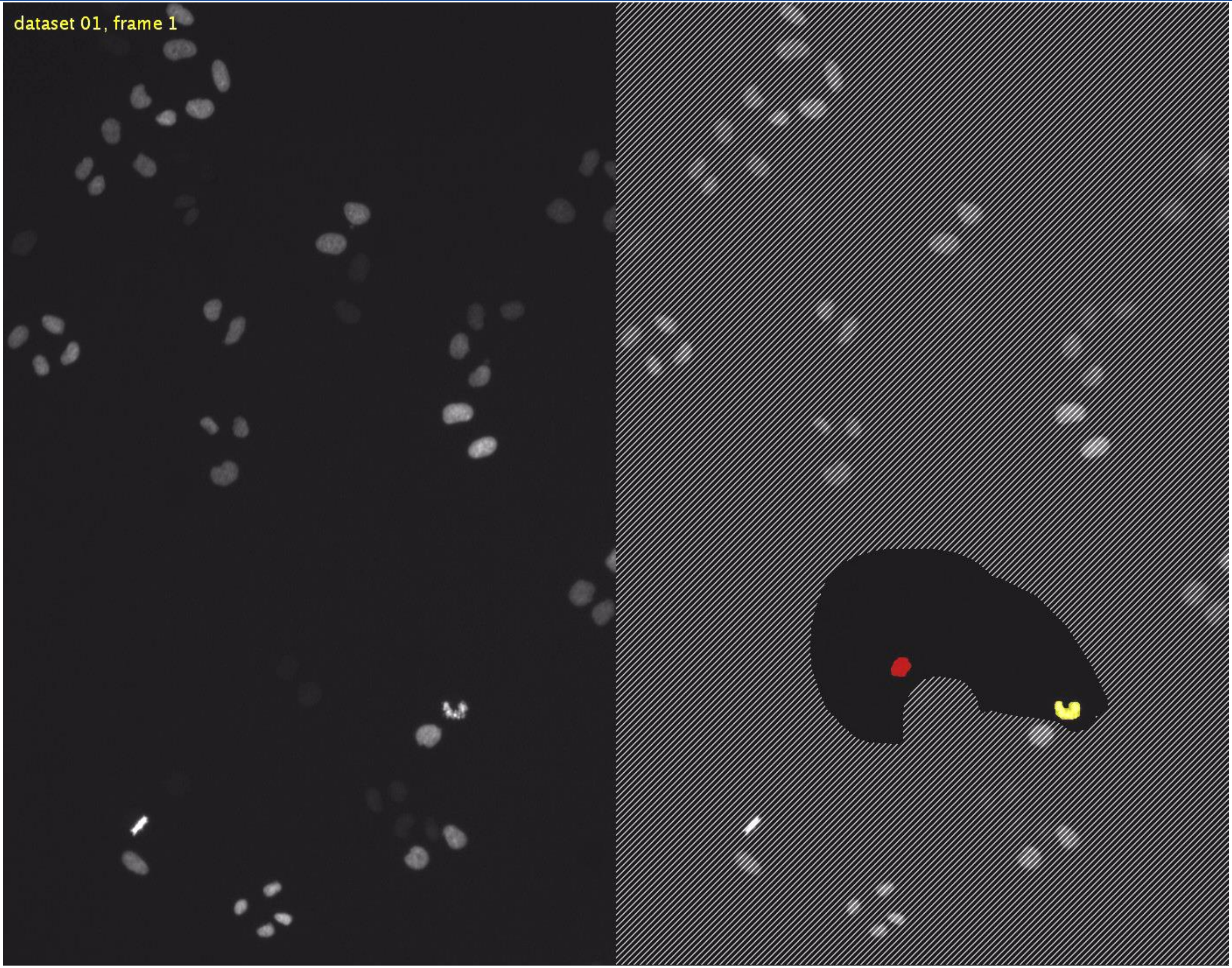
Data provided by Dr. Gert van Cappellen,
Erasmus Medical Center. Rotterdam. The Netherlands

DIC-HeLa:



Our Results: 77.6%, 2nd: 46.0%

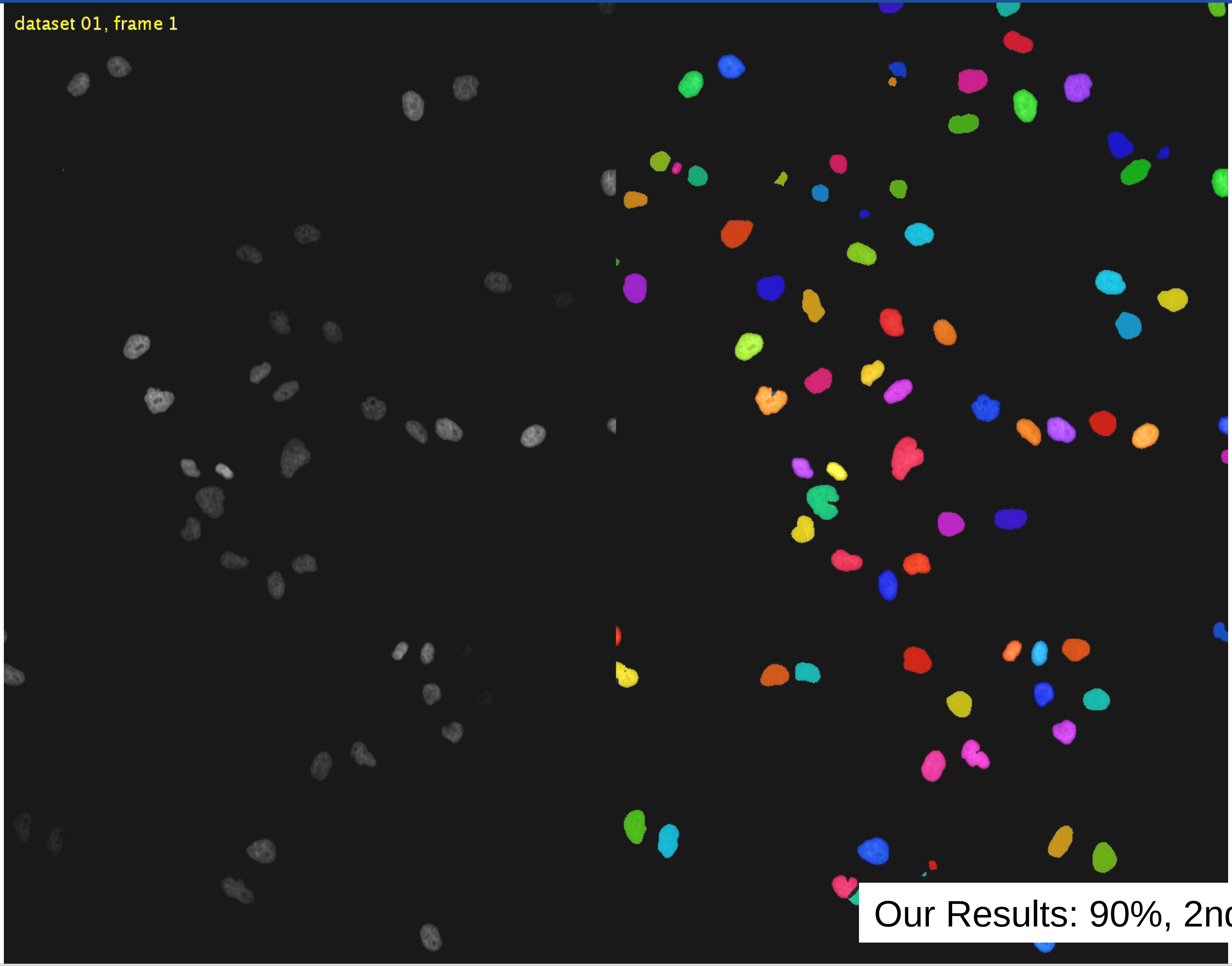
Fluo-HeLa: Training



HeLa cells stably expressing H2b-GFP

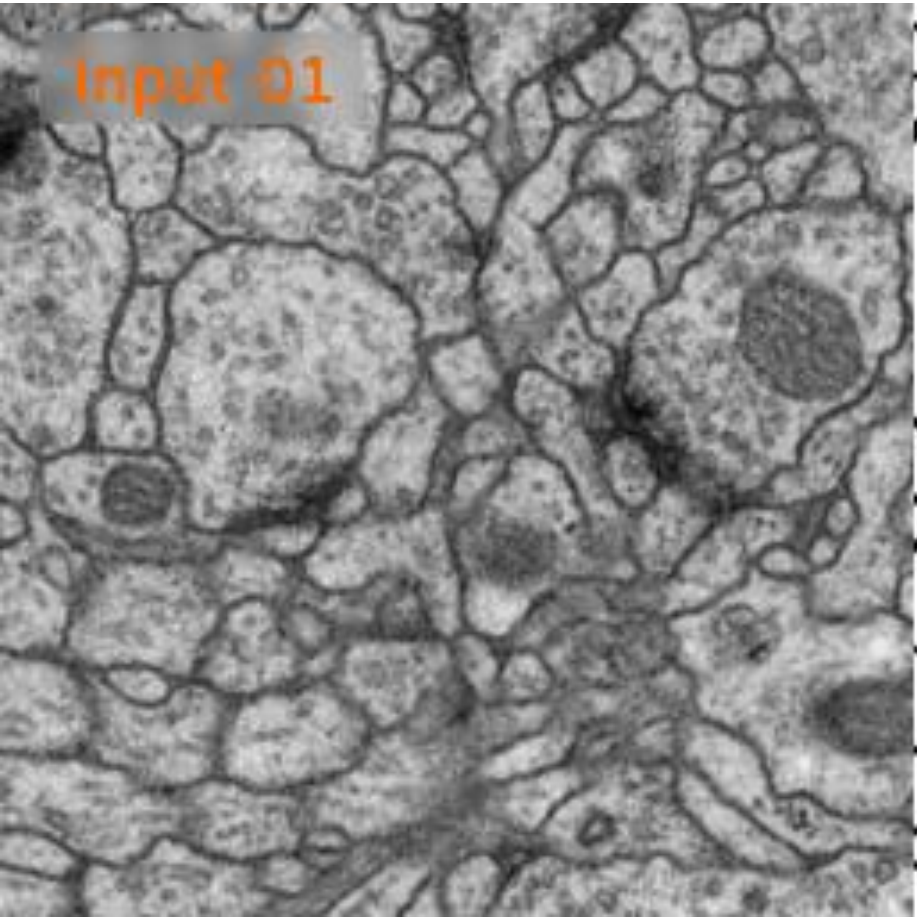
Data provided by Mitocheck Consortium

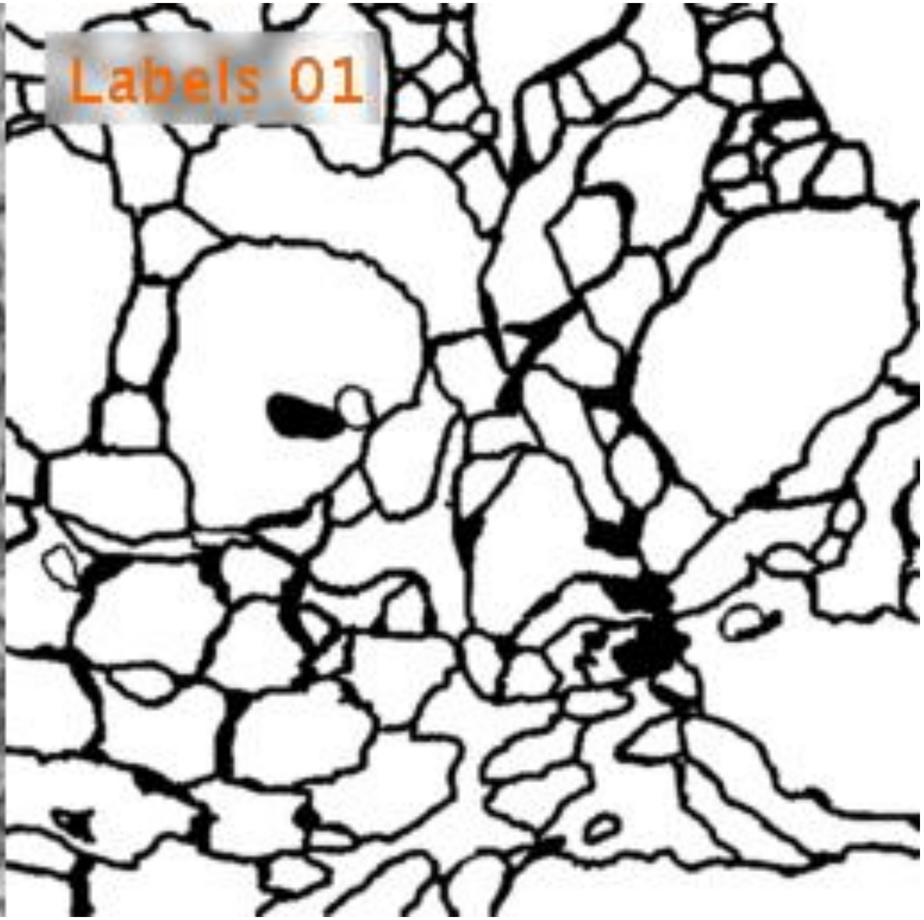
Fluo-HeLa:



Our Results: 90%, 2nd: 89%

Experiments: Neuronal structures in EM

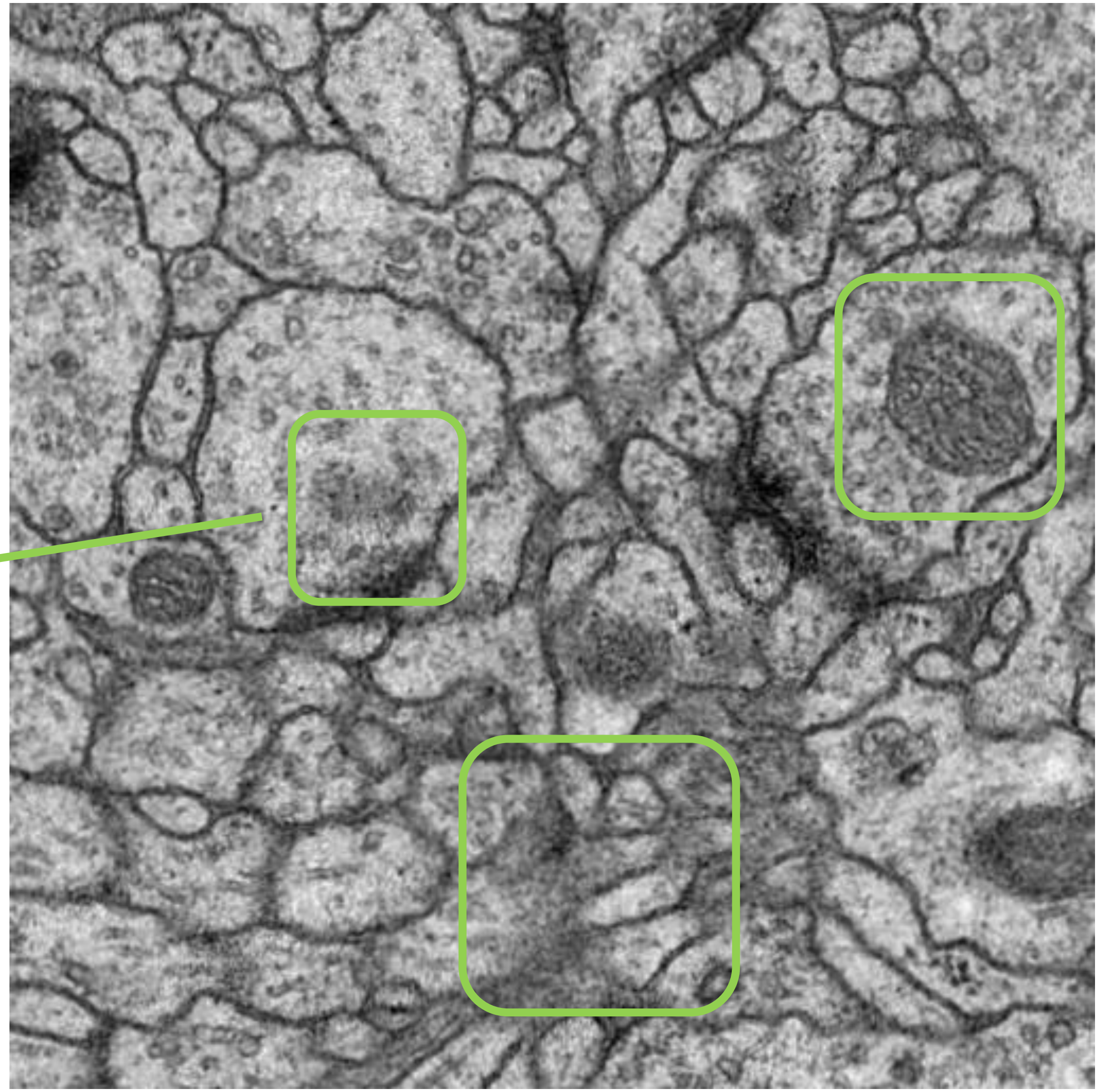




Input images

Manual labels
membranes: black,
rest (mainly cytoplasm): white

Challenges in this Data Set

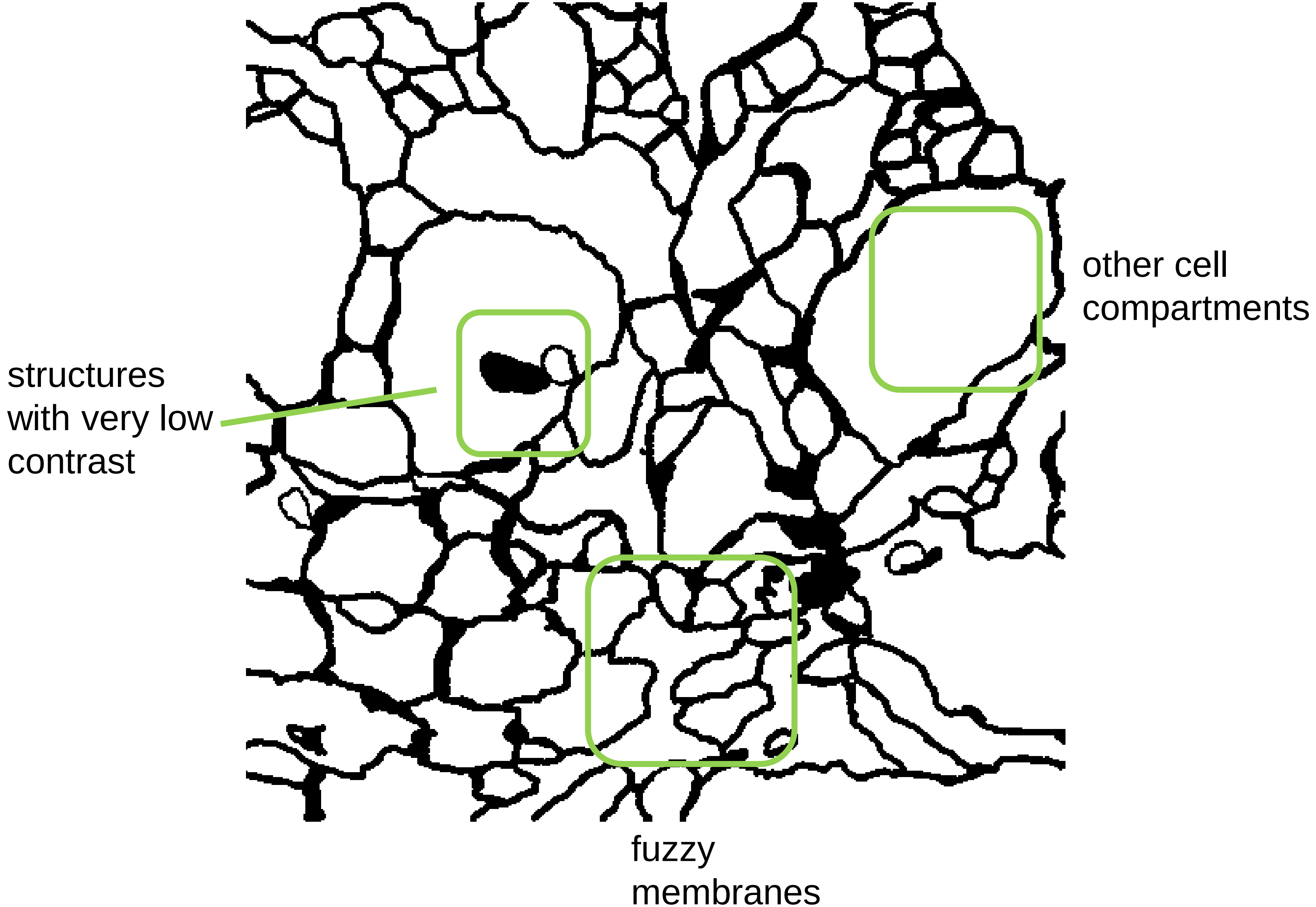


other cell compartments

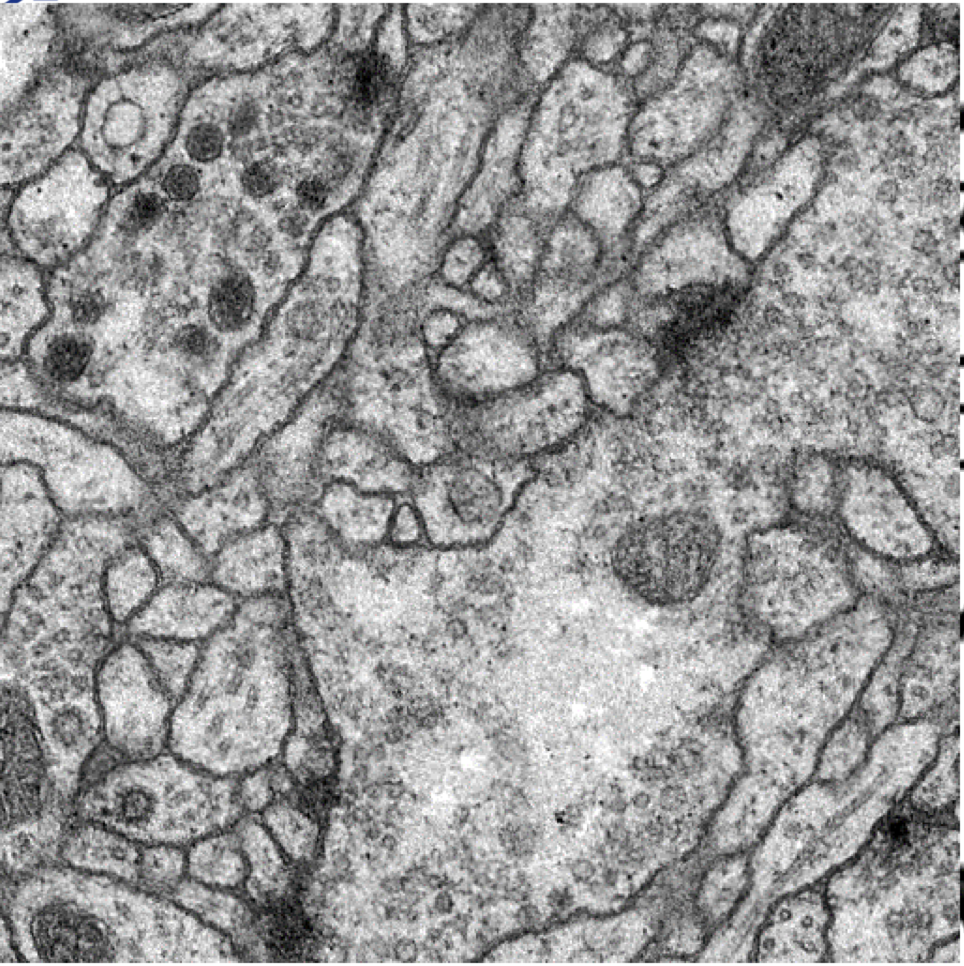
structures with very low contrast

fuzzy membranes

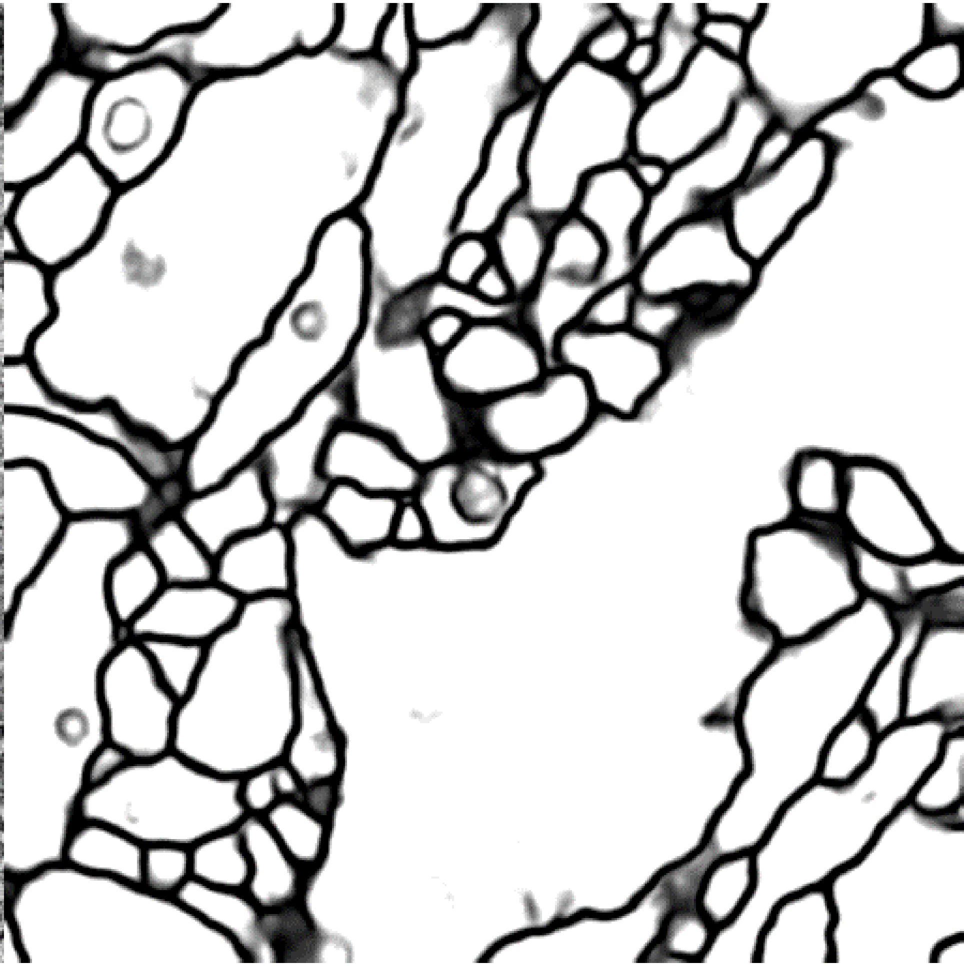
Challenges in this Data Set



Our Results



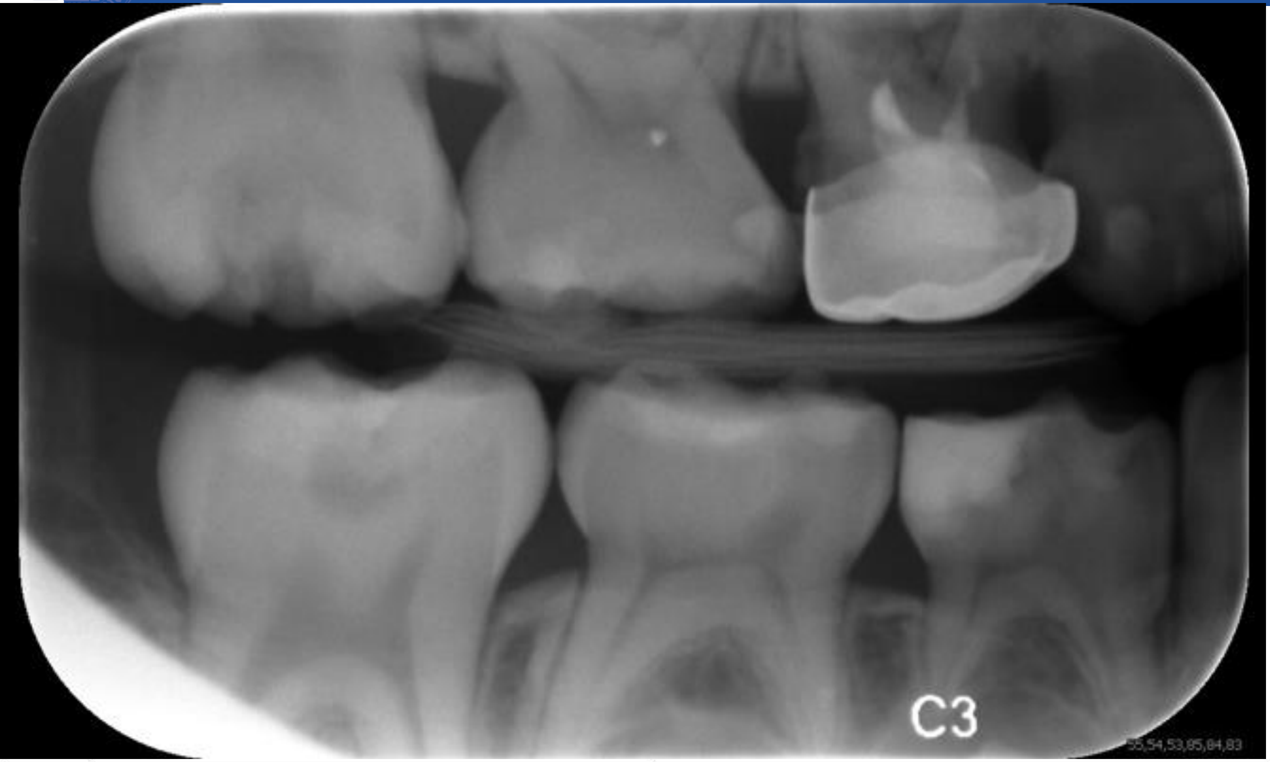
Input image



Membrane score

Our result: 0.000353 warping error (**New best score** at submission march 6th, 2015)
Sliding-window CNN: 0.000420

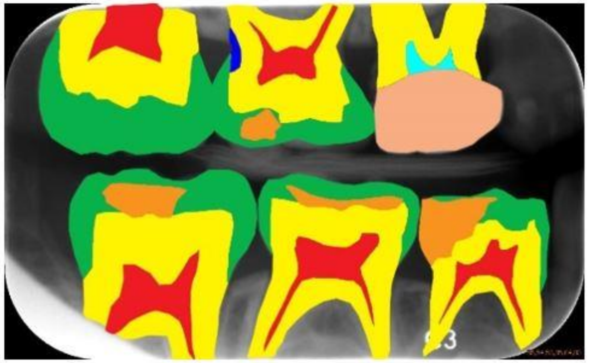
Segmentation of Dental X-Ray images



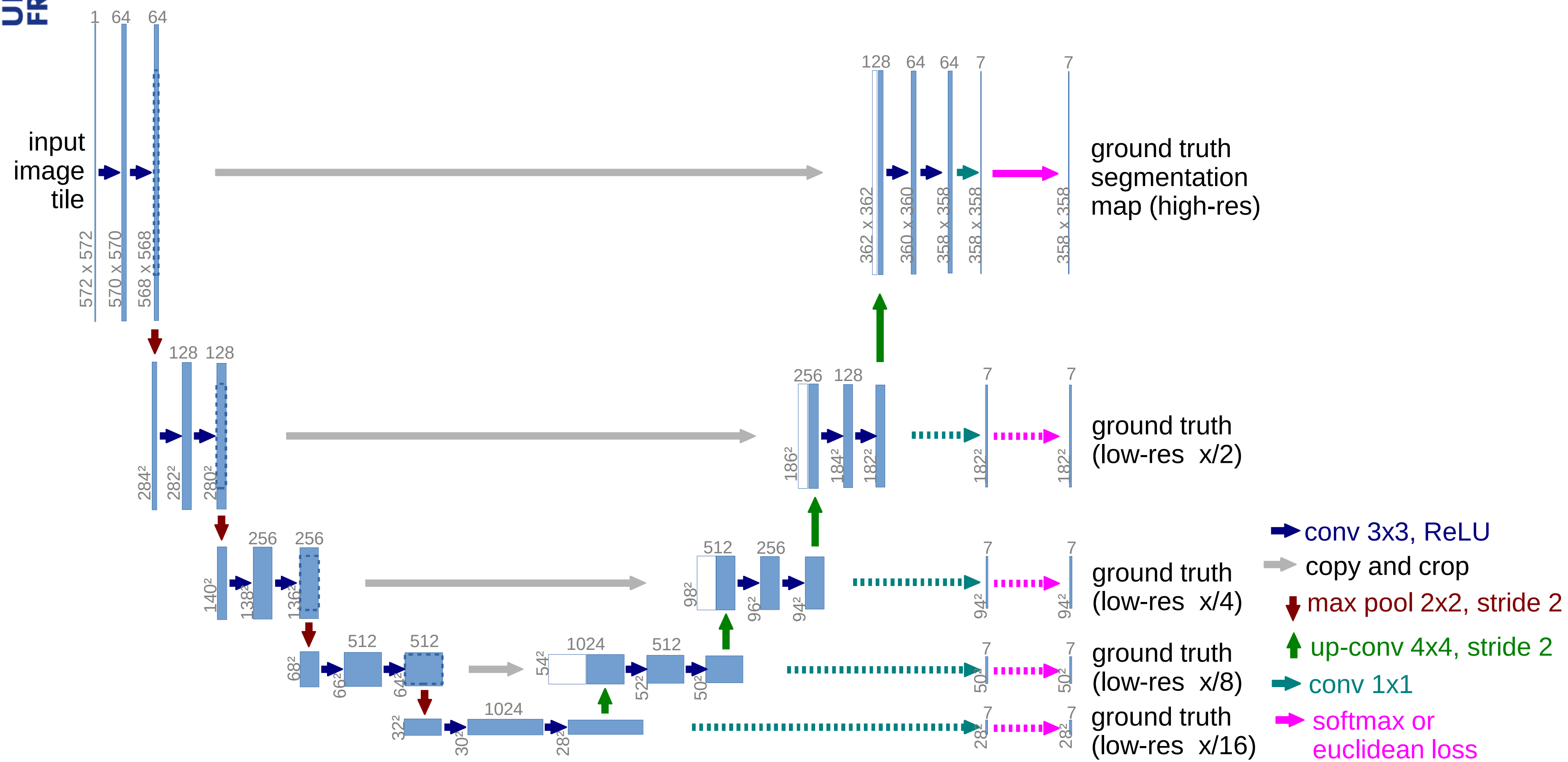
Challenges:

- subtle gray value variations at specific positions define the classes
- partial teeth not labelled

No.	Important Properties Parts
1	caries (blue color)
2	enamel (green color)
3	dentin (yellow color)
4	pulp (red color)
5	crown (skin color)
6	restoration (orange color)
7	root canal treatment (cyan color)



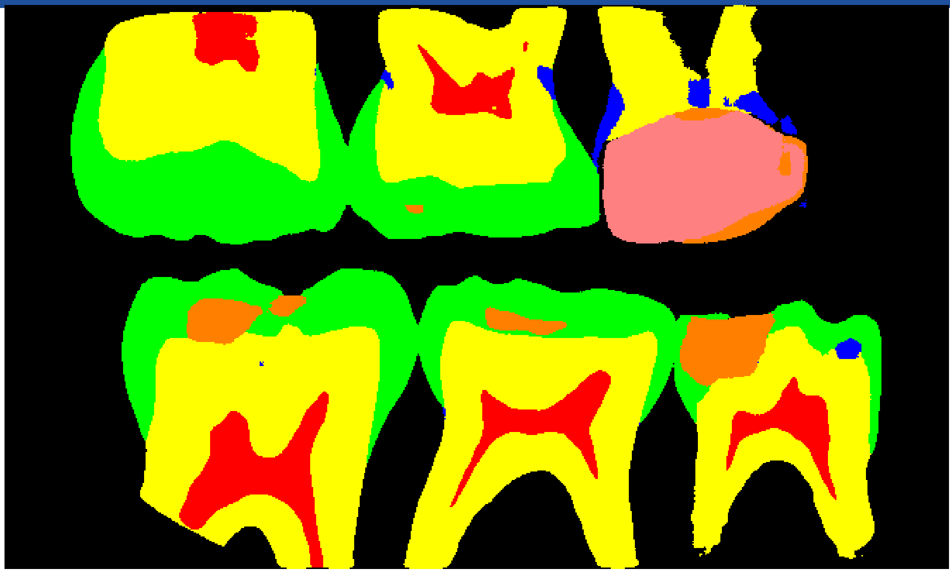
Additional Low-Resolution Loss Layers



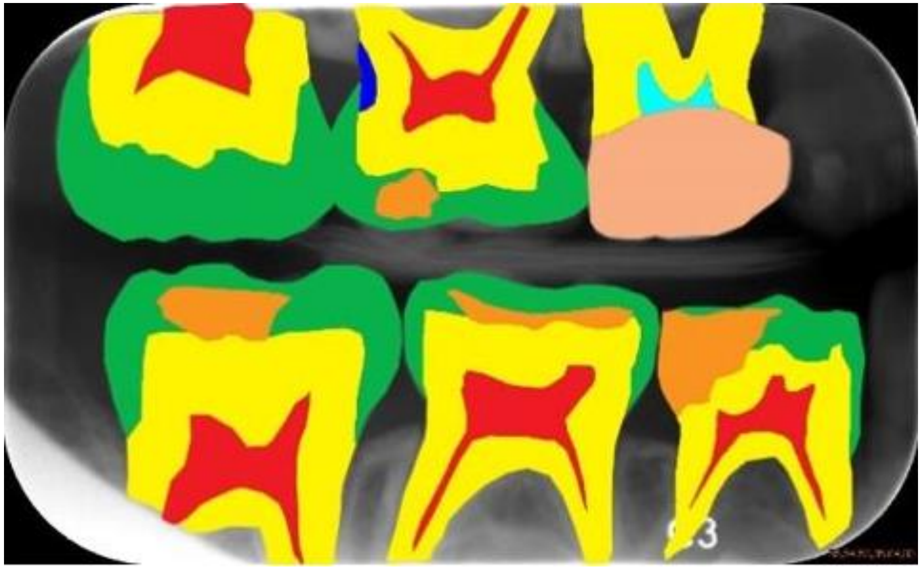
- Try to guide the training in the lower layers

Results with different Training Schemes

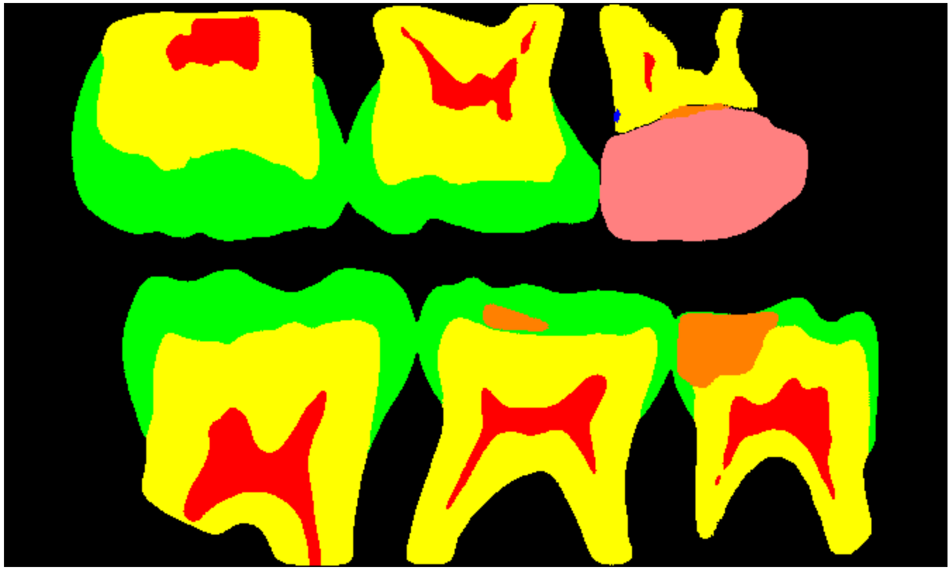
raw image



u1: only high-res loss layer

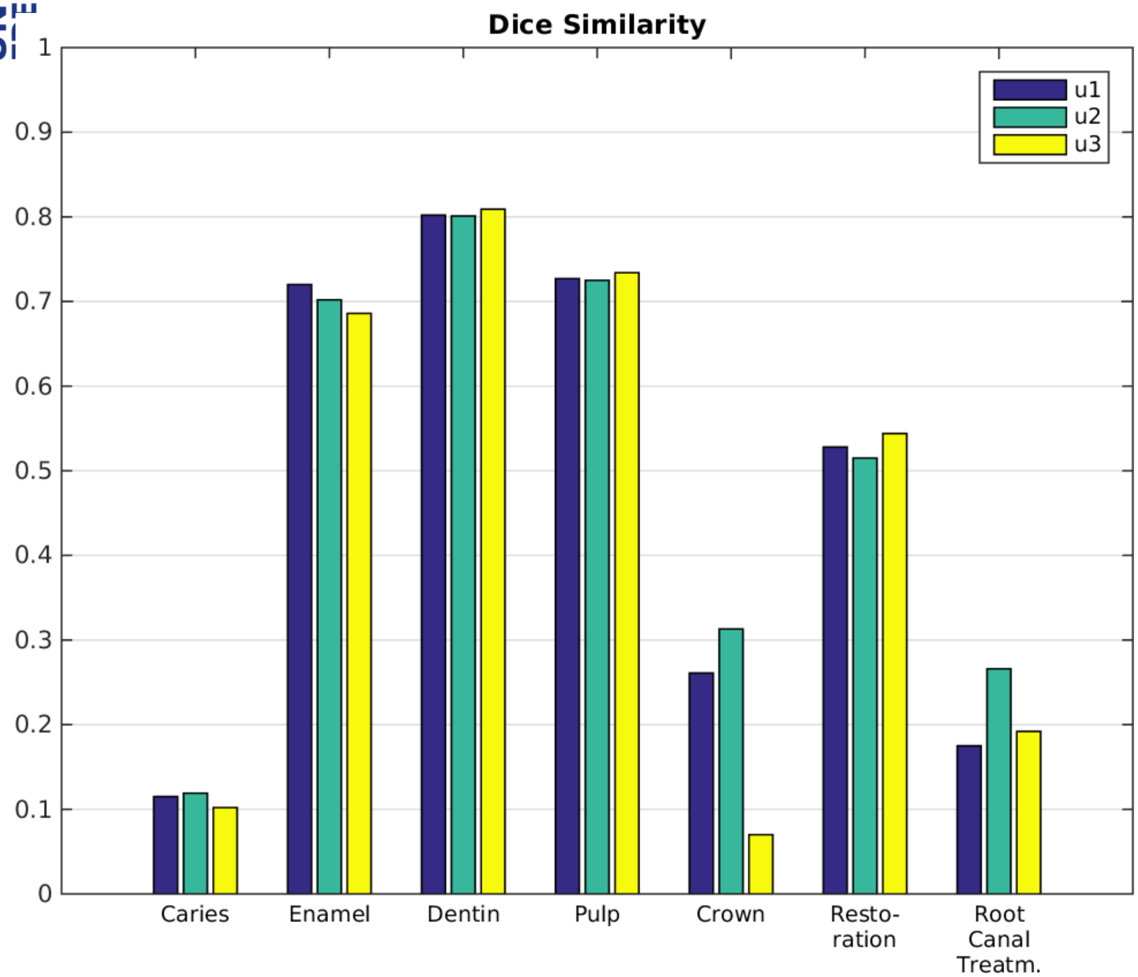


Ground truth segmentation



u2: add. low-res softmax loss layers

Quantitative Evaluation from Challenge Organizers

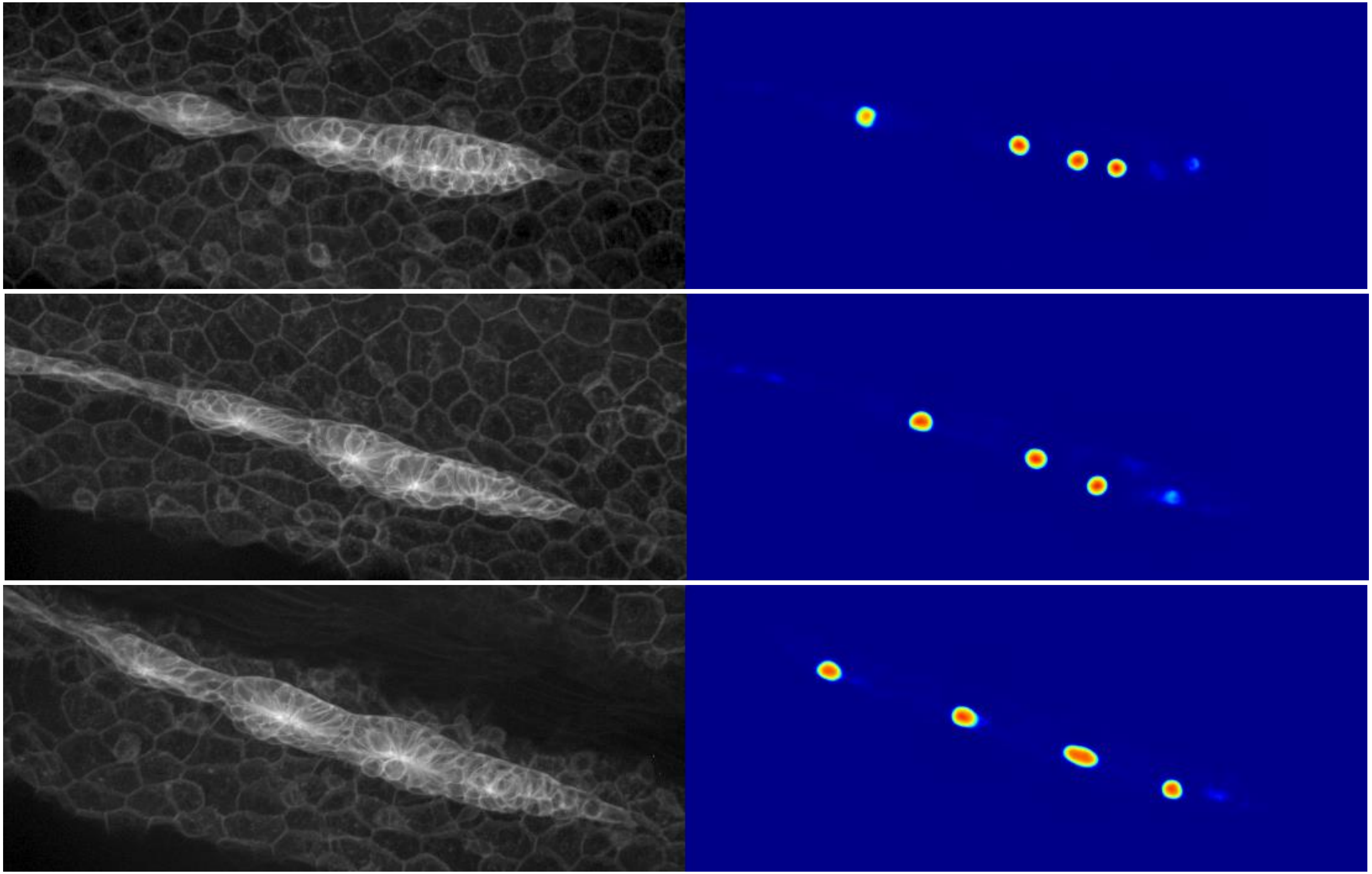


u1: high-res loss only
u2: add. low-res softmax loss
u3: add. low-res Euclidean loss

Phase 1 (pre-Conference competition):
Our result: **56.7%**
Second best: 32.2%

Phase 2 (On-site competition)
Our result: **52.5%**
Second best: 28.7%

Detection of Cell Rosettes in the Zebrafish Primordium

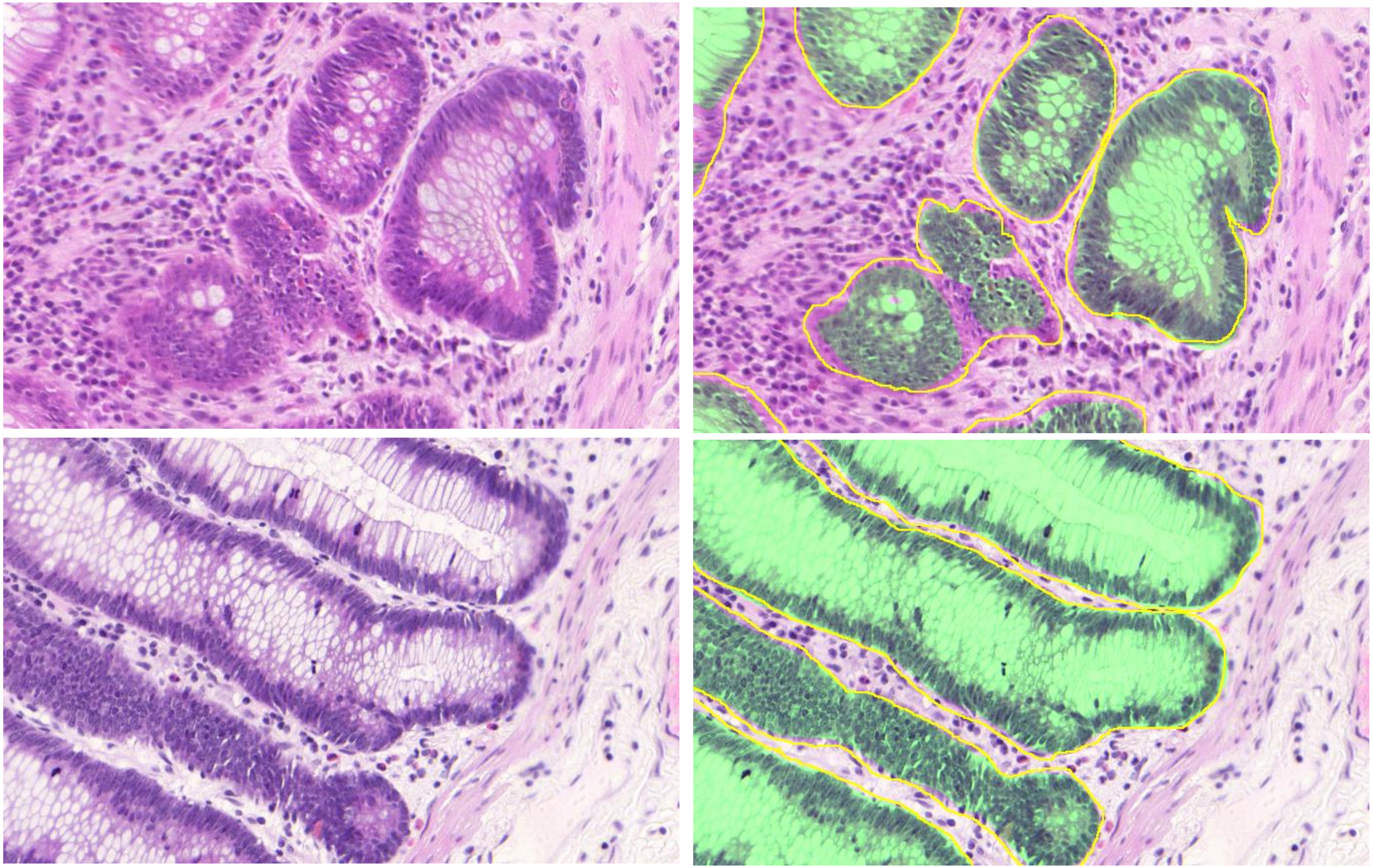


Raw images

Detection score maps

[Robert Bensch, collaboration with Virginie Lecaudey, Biology, Uni Freiburg]

Segmentation of Histopathological Images

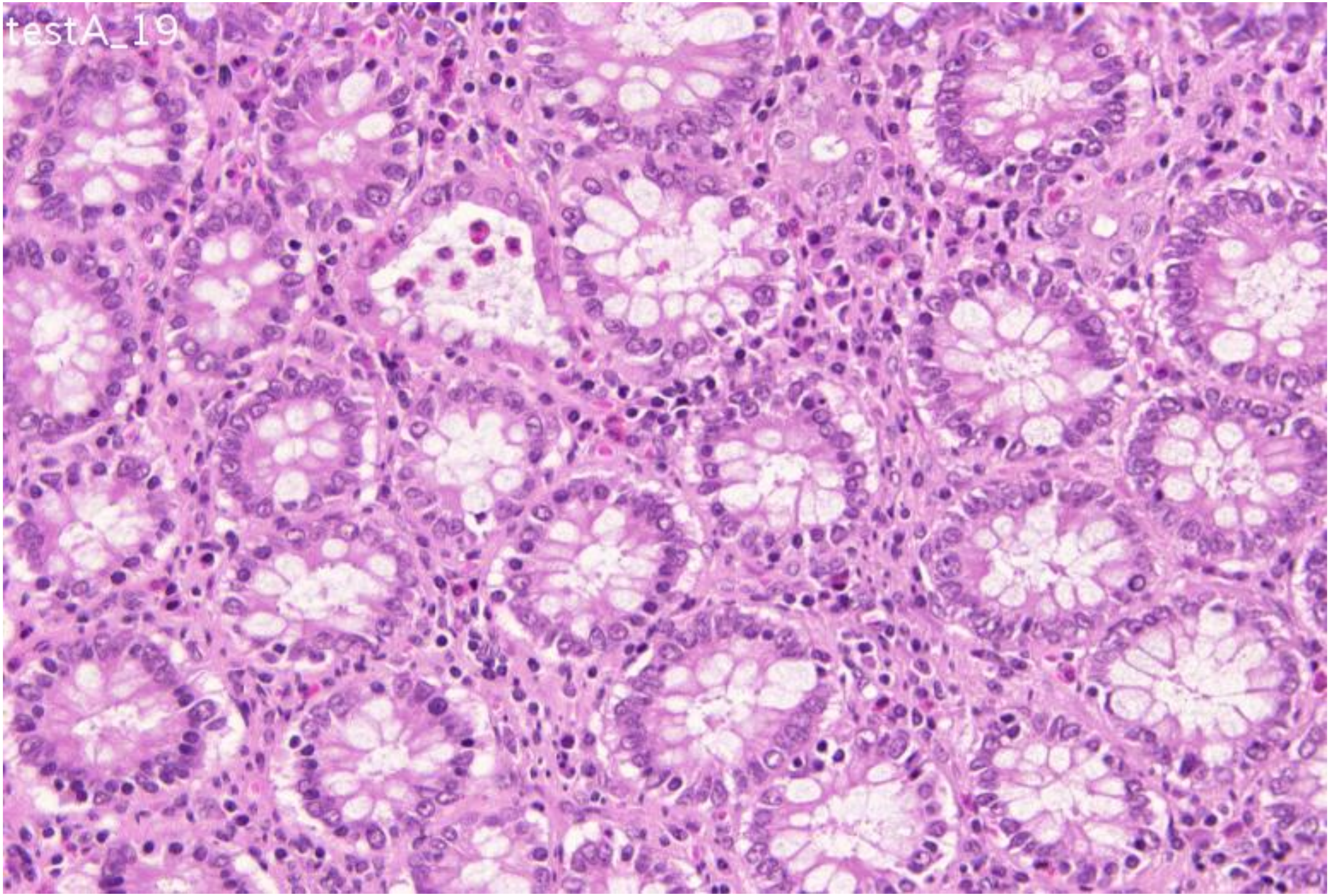


Glands in histology images

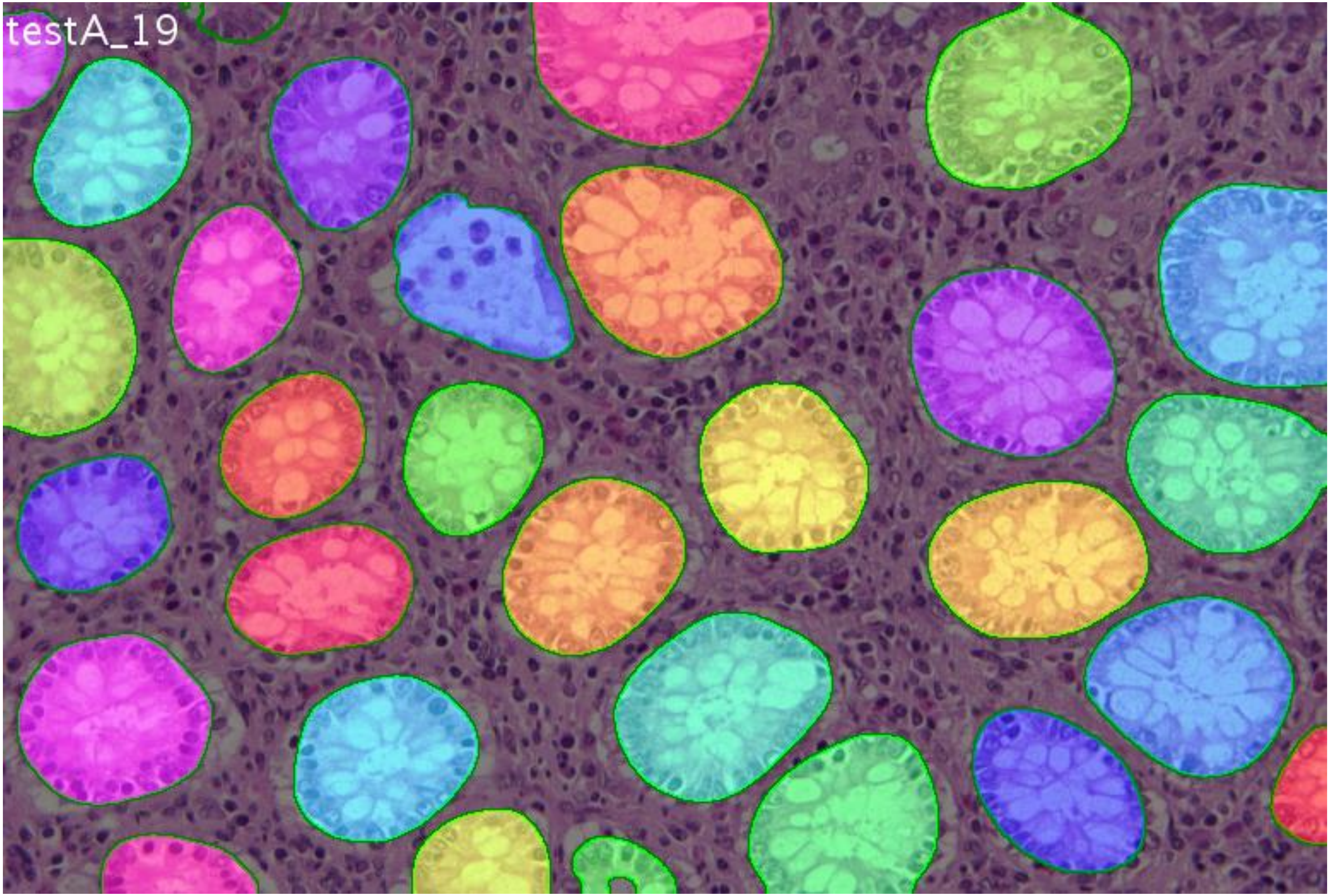
green mask: u-net result;
yellow border: ground truth annotation

[Anton Böhm, data from GlaS@MICCAI'2015: Gland Segmentation Challenge Contest]

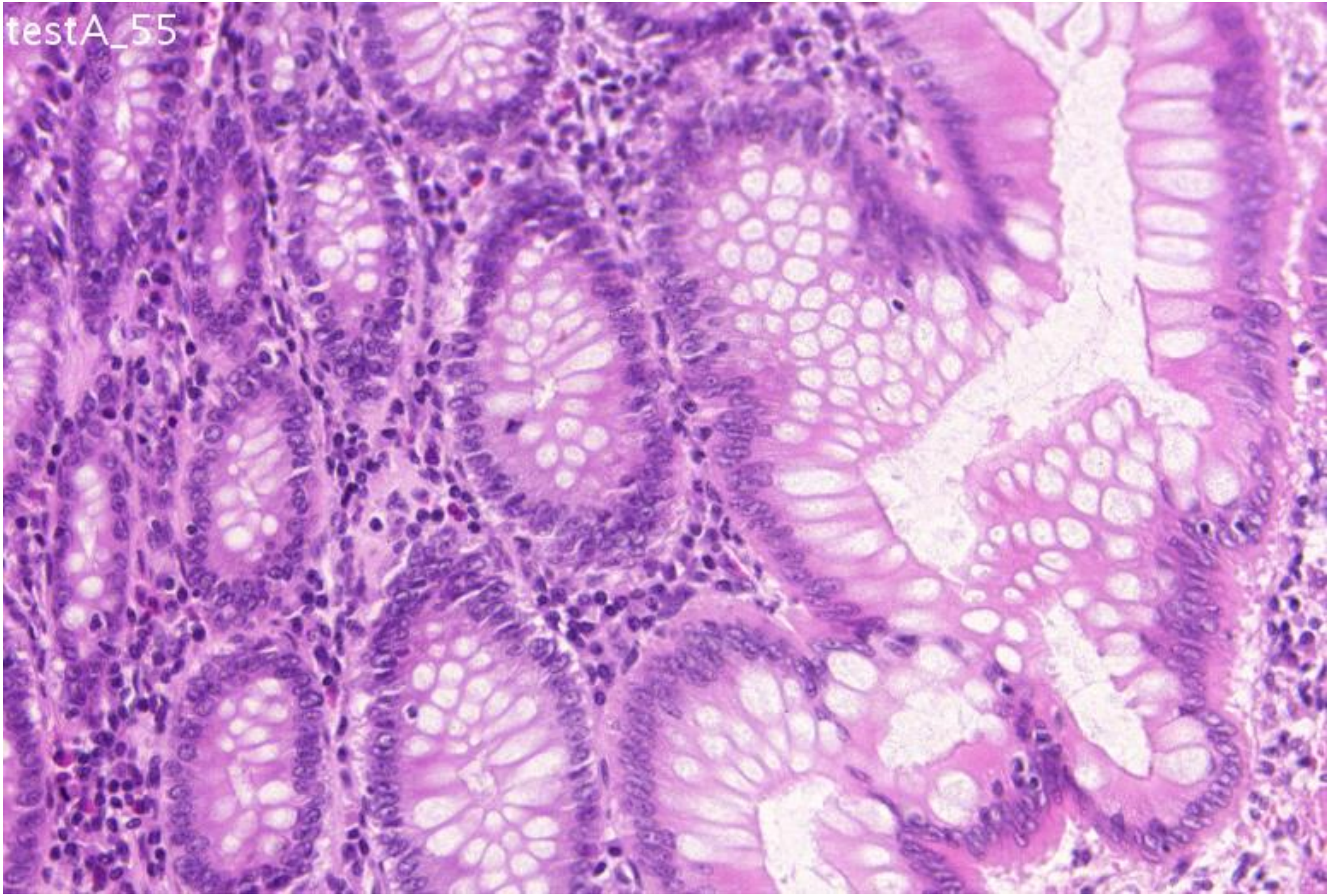
Qualitative Results: Glas Challenge Testset A



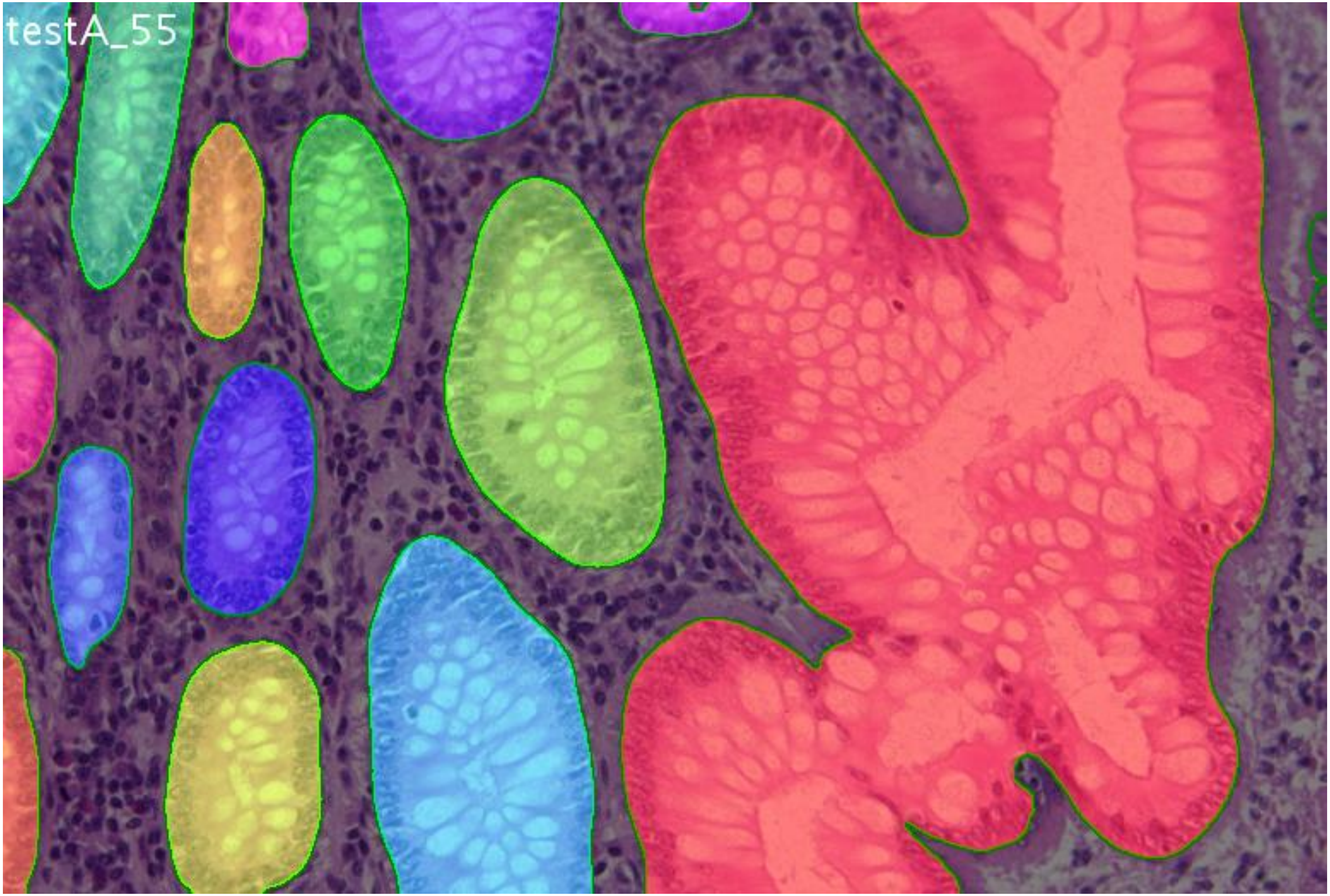
Qualitative Results: Glas Challenge Testset A



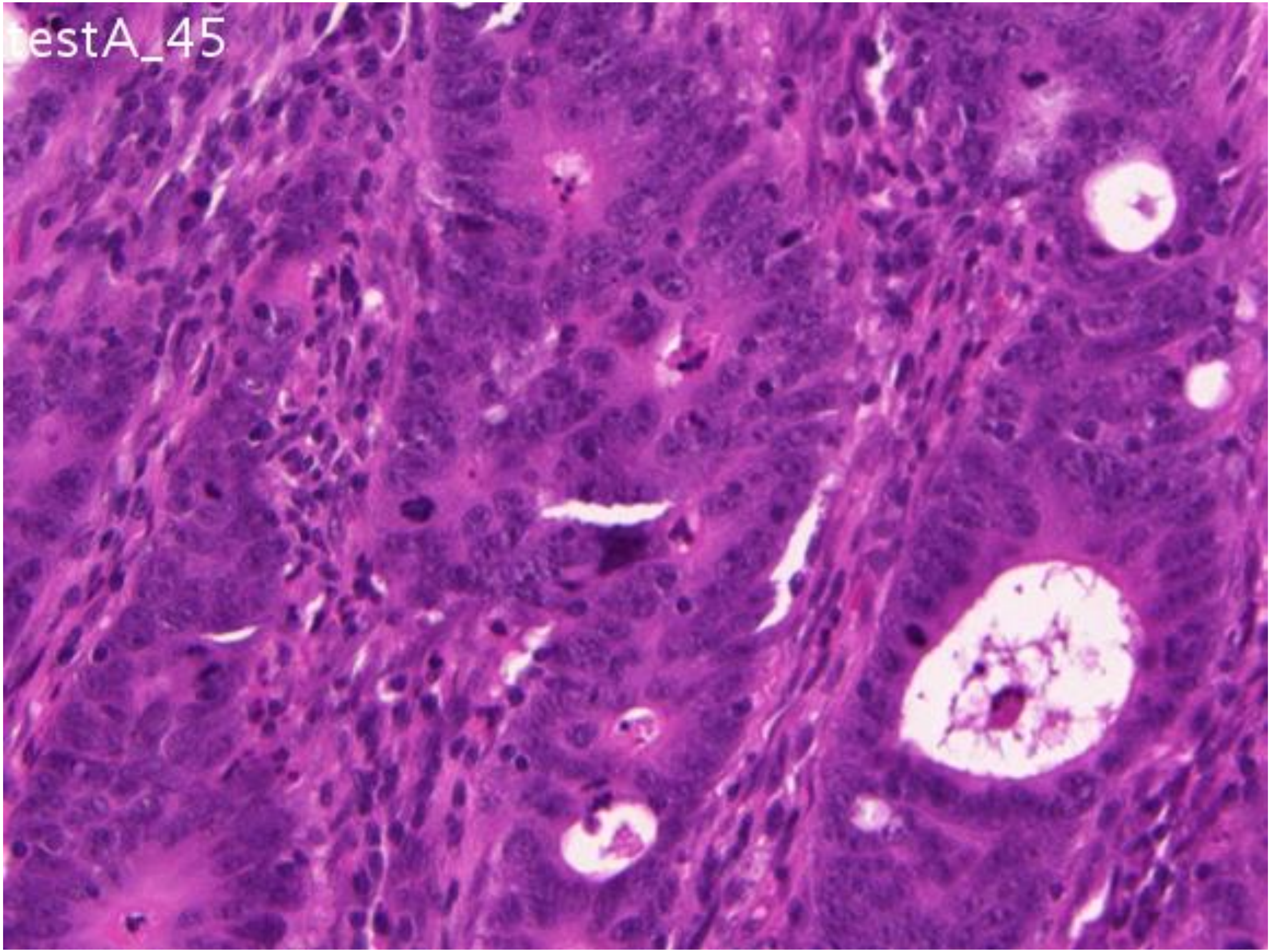
Qualitative Results: Glas Challenge Testset A



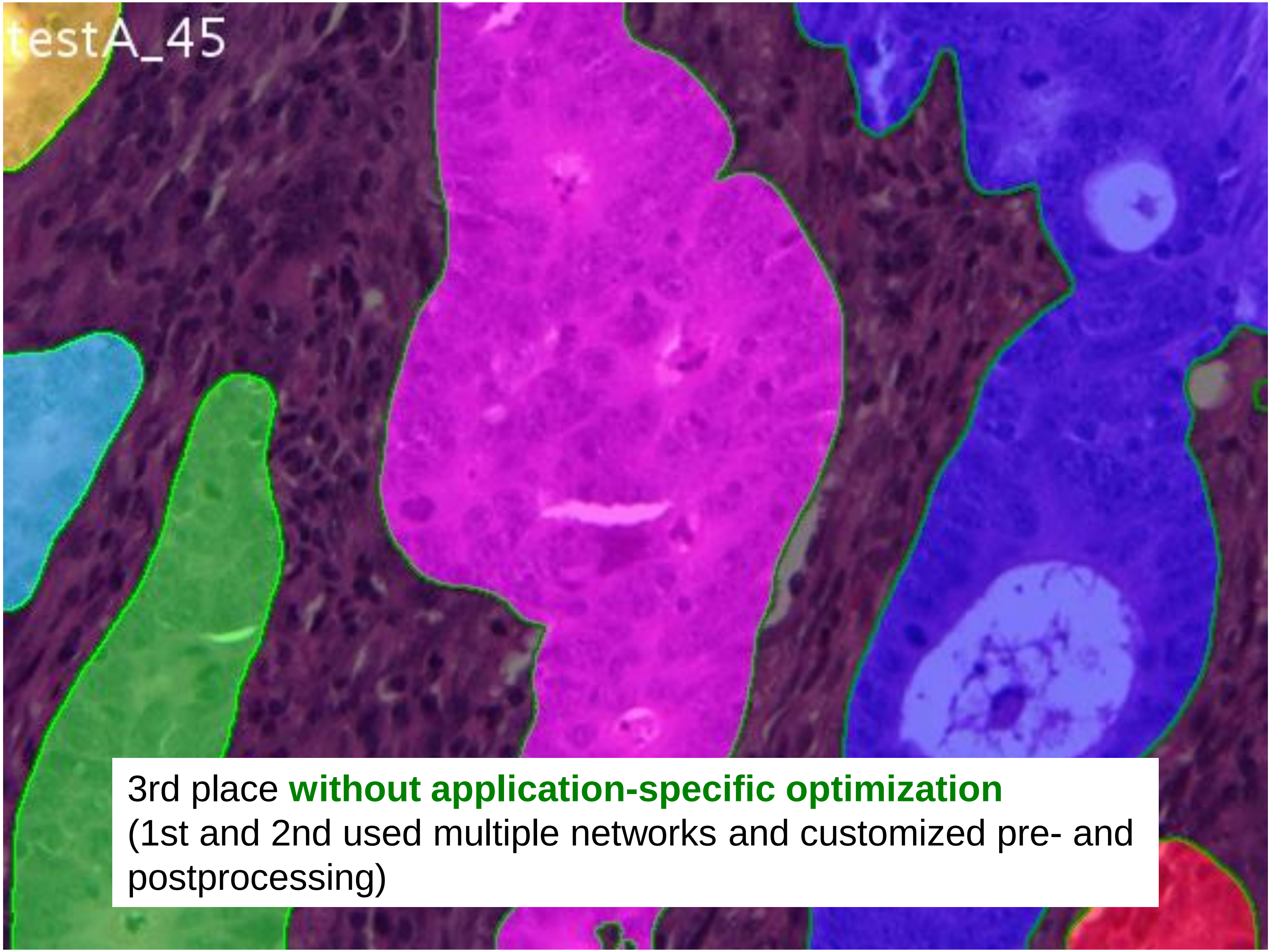
Qualitative Results: Glas Challenge Testset A



Qualitative Results: Glas Challenge Testset A

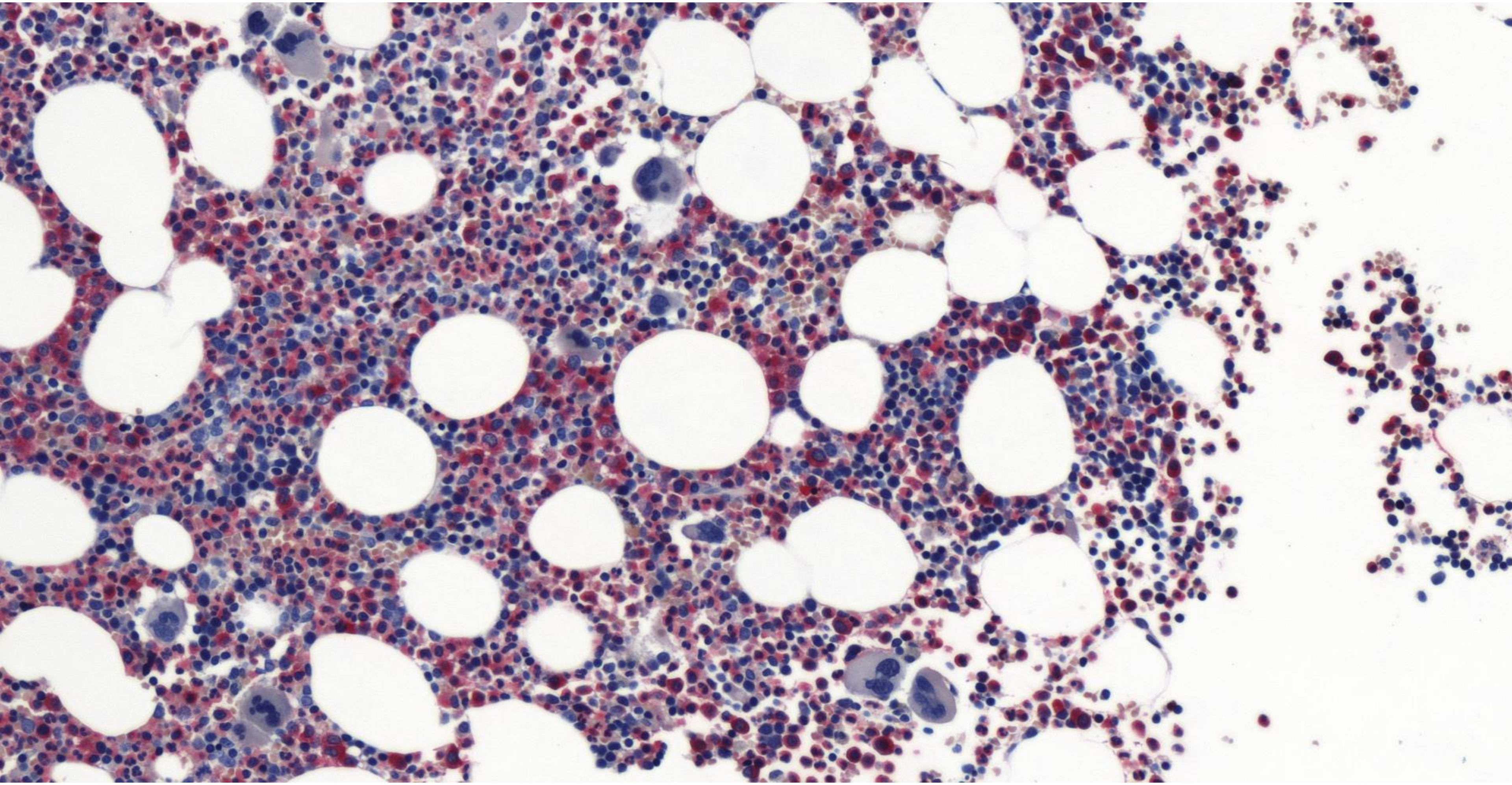


Qualitative Results: Glas Challenge Testset A



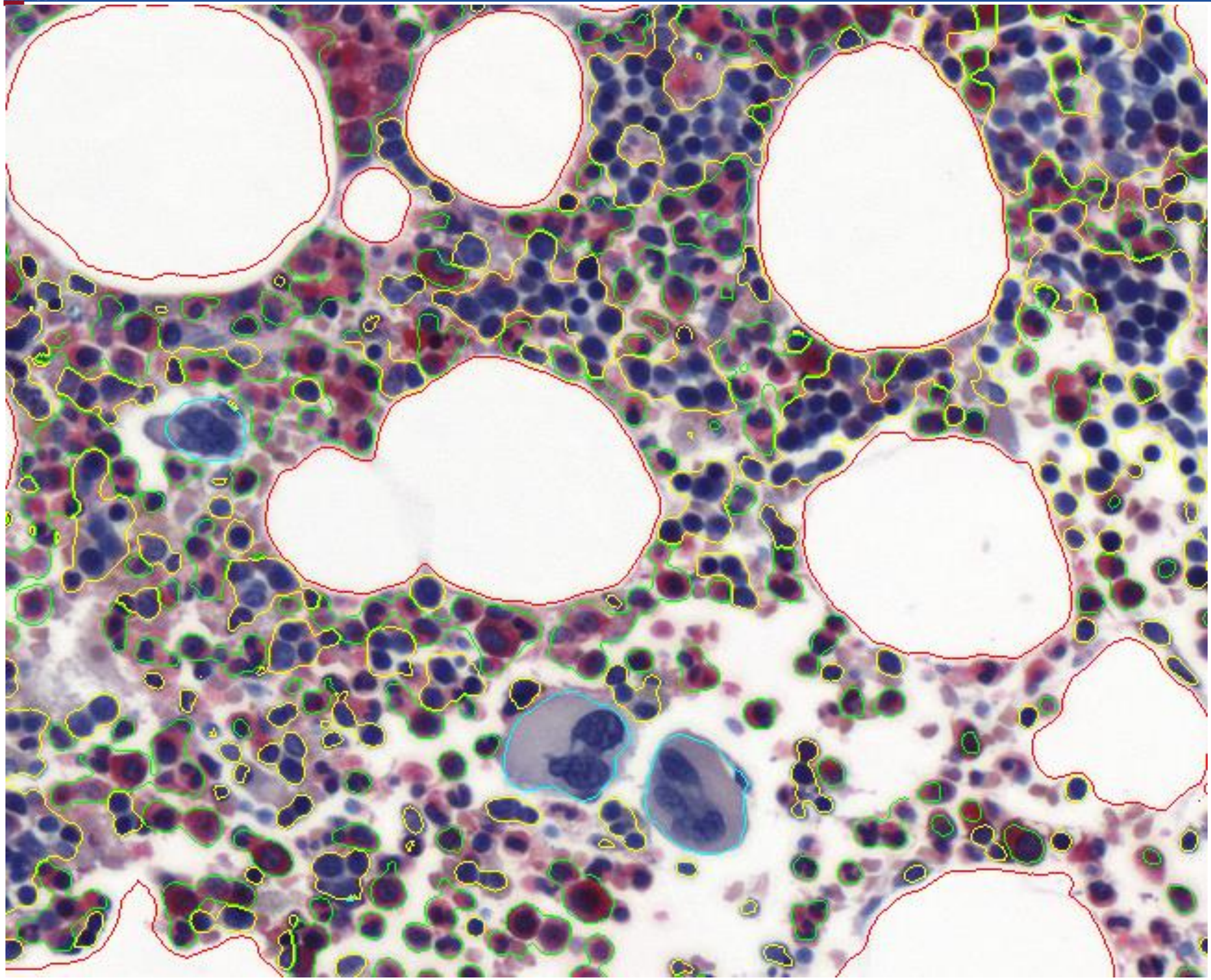
3rd place **without application-specific optimization**
(1st and 2nd used multiple networks and customized pre- and postprocessing)

Segmentation of Bone Marrow



Anton Böhm, Collaboration with K. Aumann and M. Werner (Pathology)

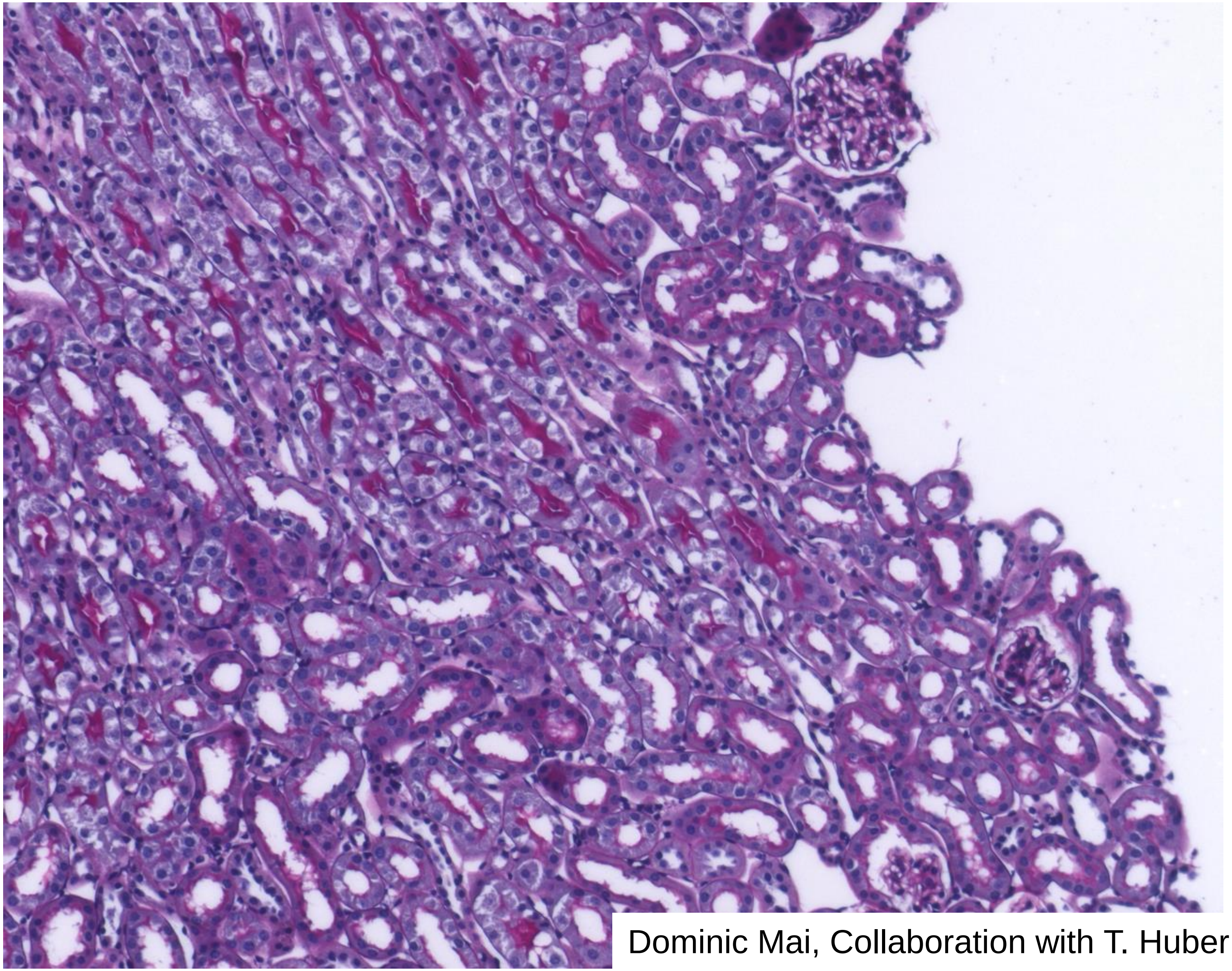
Segmentation of Bone Marrow



Erythropoesis,
Granulopoesis,
Megakaryocytes,
fat cells
trabecular bone

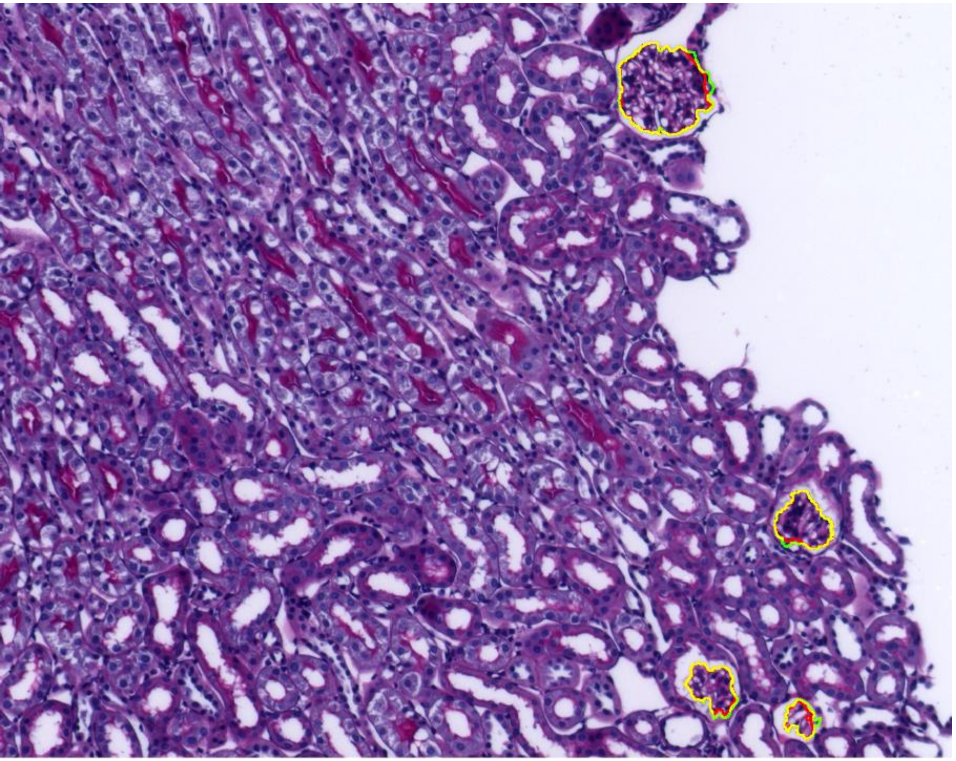
Anton Böhm, Collaboration with K. Aumann and M. Werner (Pathology)

Segmentation of Glomeruli (Kidney)

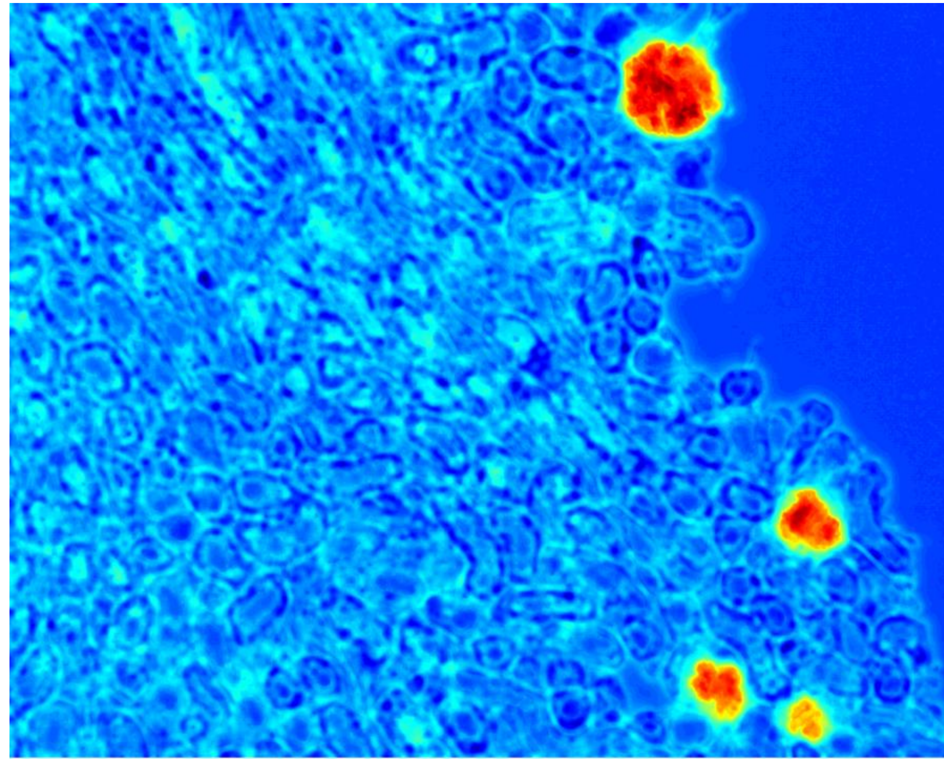


Dominic Mai, Collaboration with T. Huber (Nephrology)

Segmentation of Glomeruli (Kidney)



green: ground truth,
red: U-net

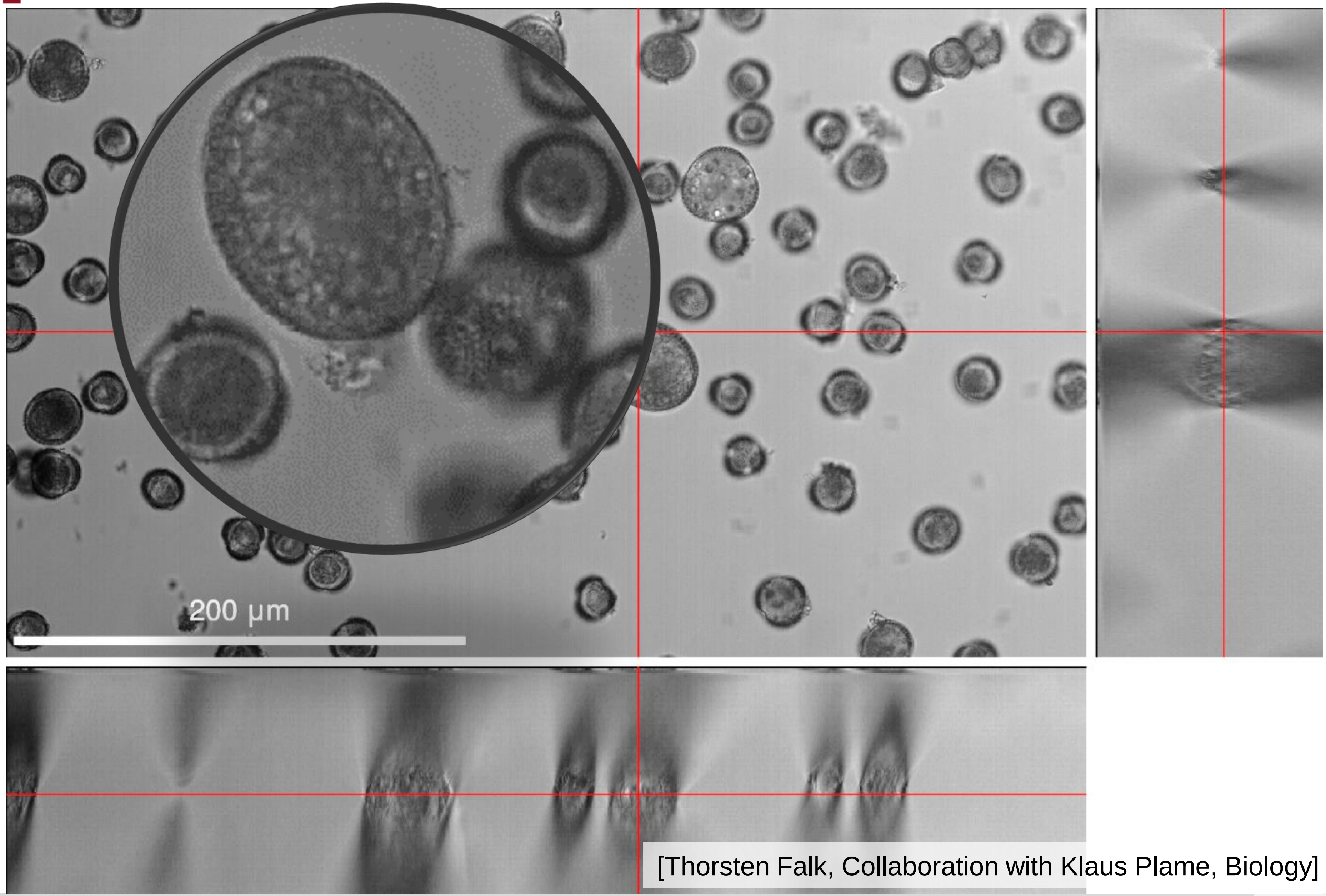


score map

Mean Intersection over Union: 82%

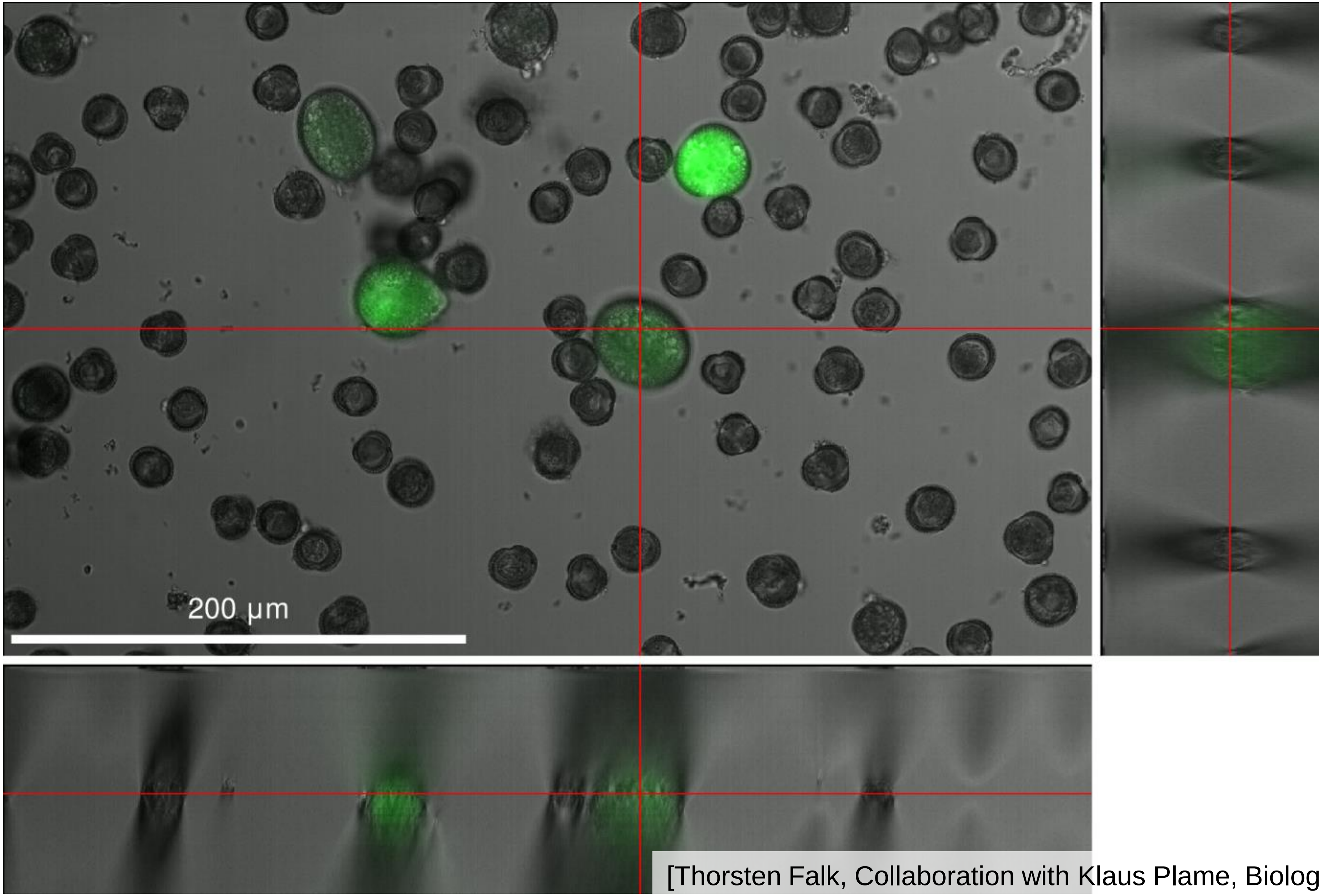
Dominic Mai, Collaboration with T. Huber (Nephrology)

Distinguish Living and Dead Cells (Short Video from Transmitted Light Microscopy)



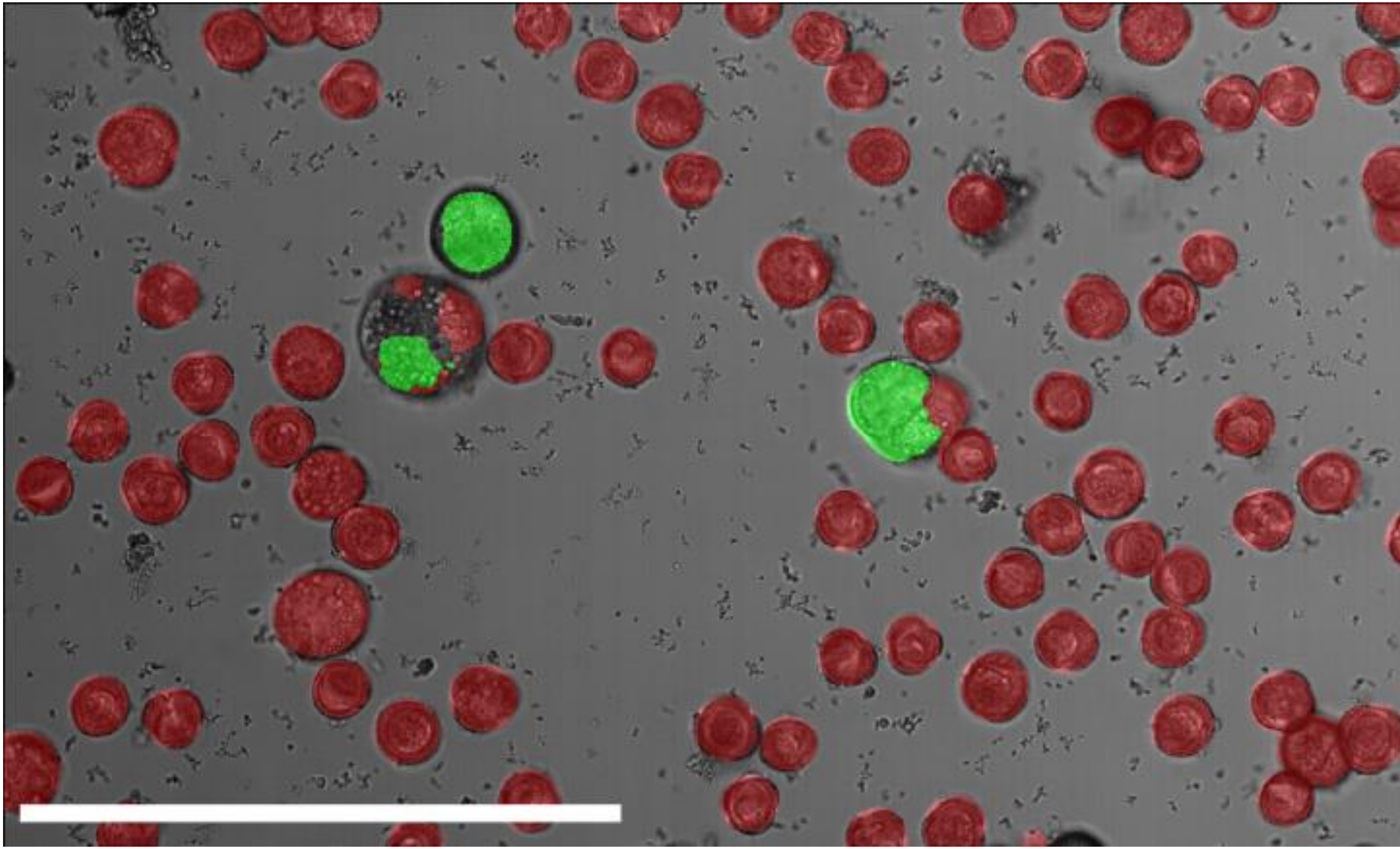
[Thorsten Falk, Collaboration with Klaus Plame, Biology]

Training Labels from CFDA Staining



[Thorsten Falk, Collaboration with Klaus Plame, Biology]

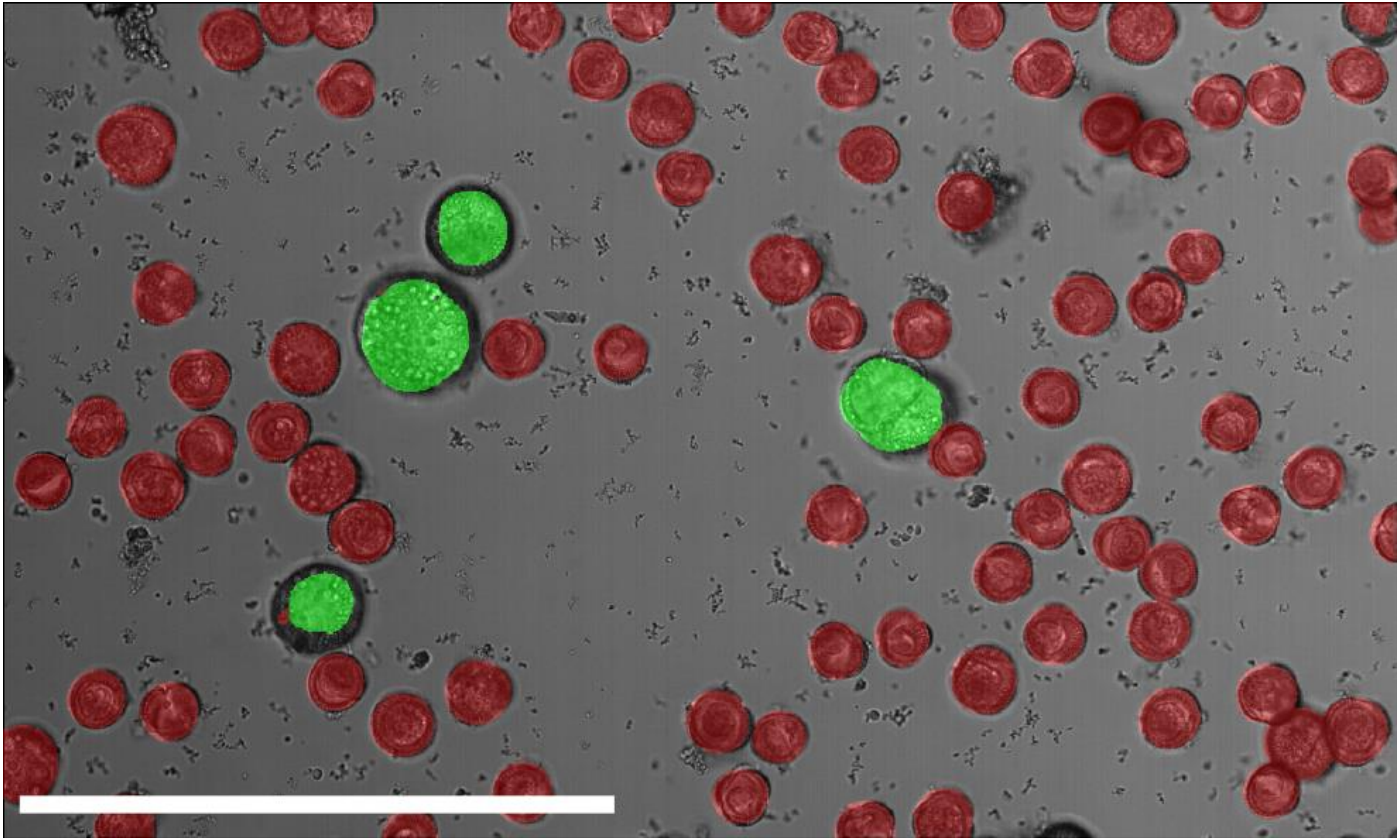
U-net Recognizes and Segments Living and Dead Cells



Result **without** time information

[Thorsten Falk, Collaboration with Klaus Plame, Biology]

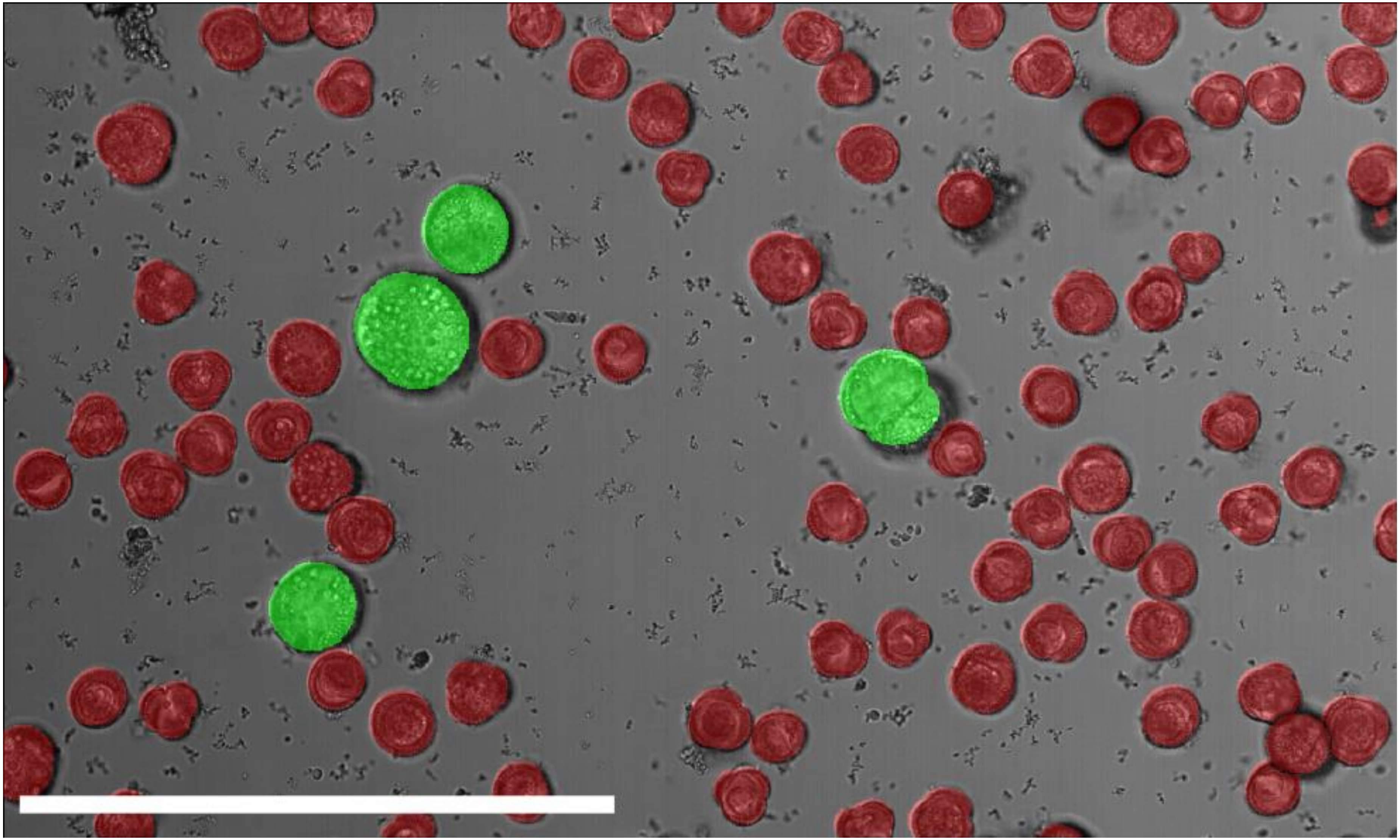
U-net Recognizes and Segments Living and Dead Cells



Result **with** time information

[Thorsten Falk, Collaboration with Klaus Plame, Biology]

U-net Recognizes and Segments Living and Dead Cells

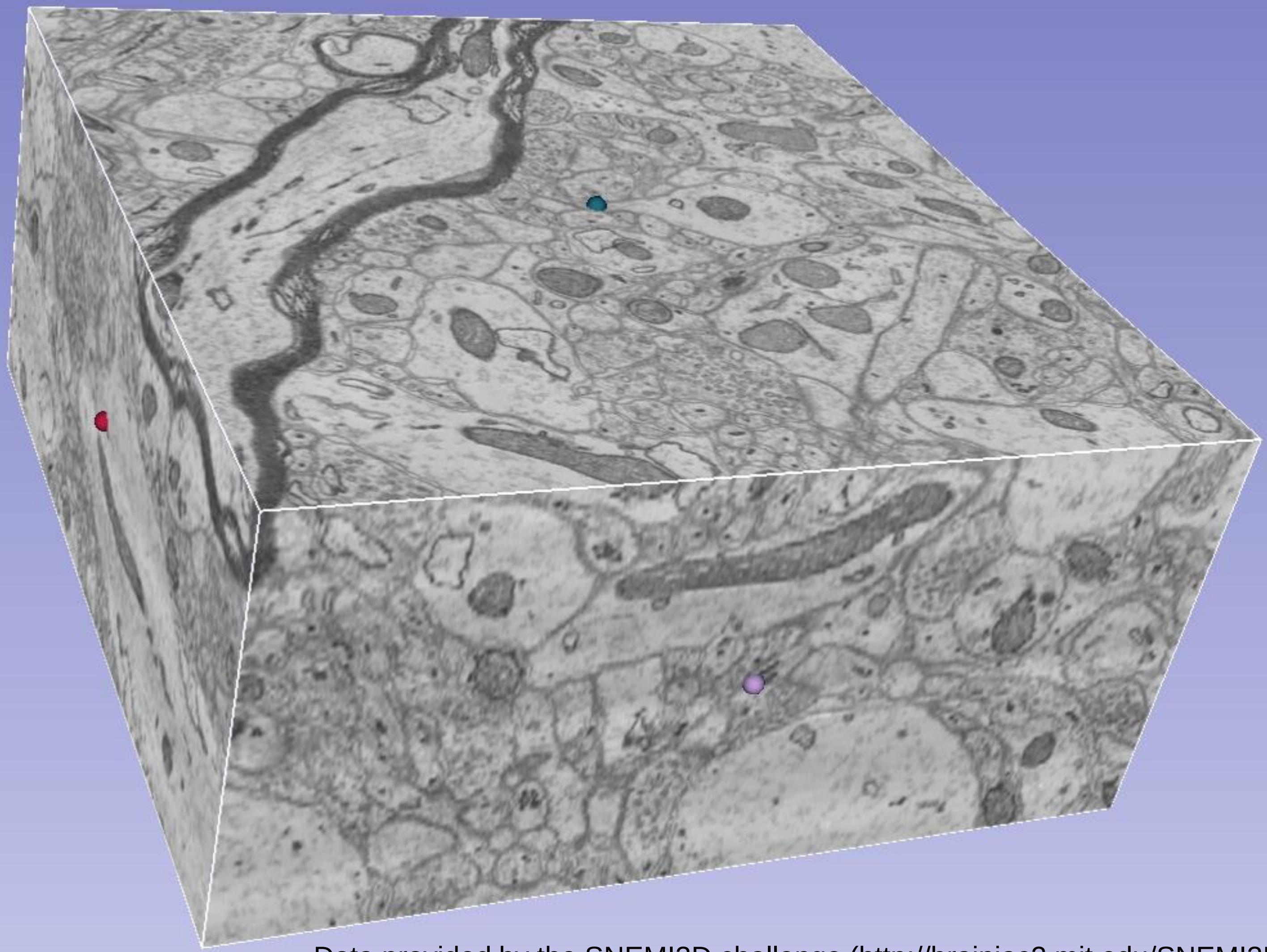


Manual ground truth annotation

[Thorsten Falk, Collaboration with Klaus Plame, Biology]

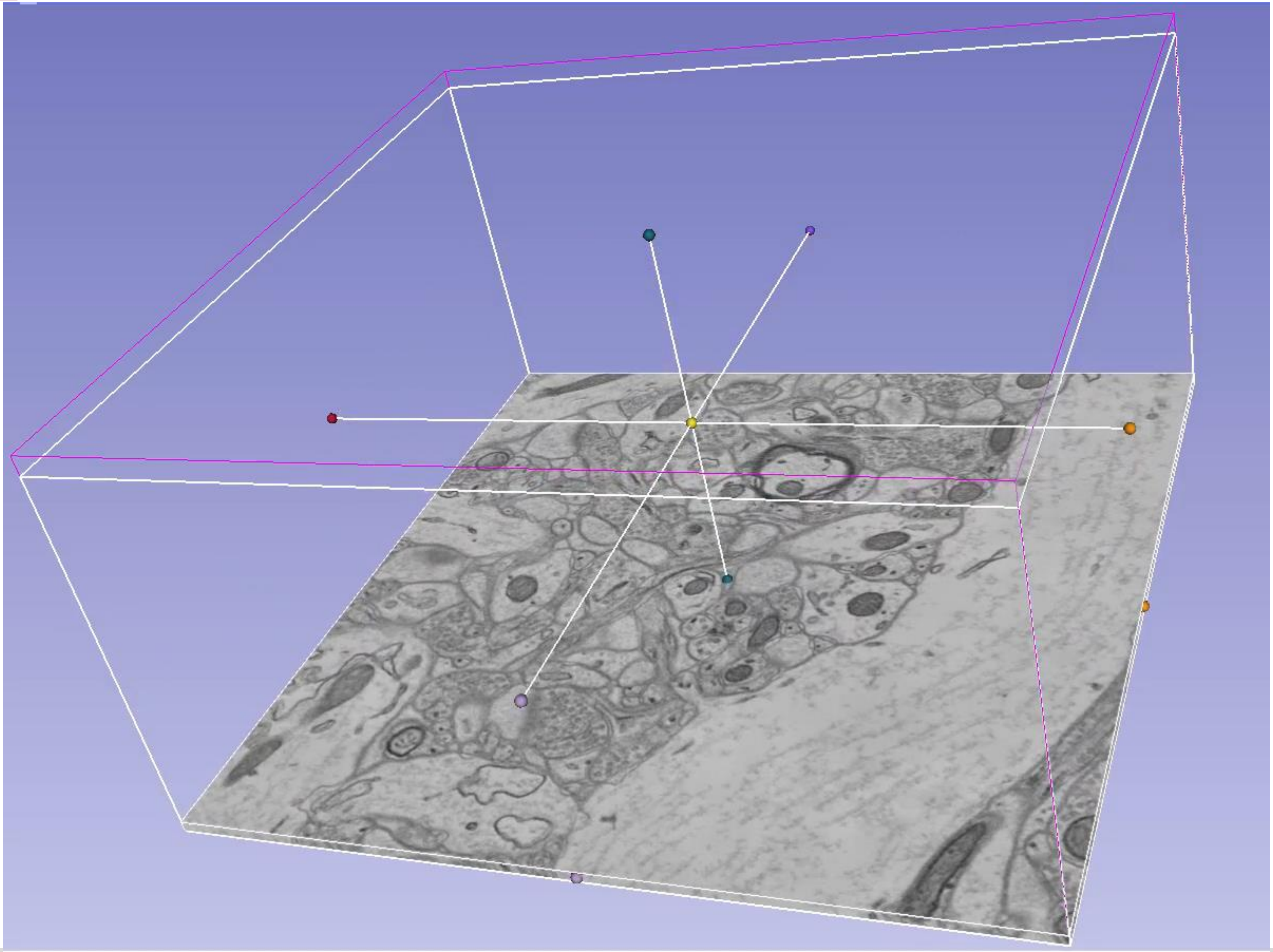
Extension to Volumetric Data

Serial section electron microscopy of mouse cortex.
1024 x 1024 x 100 voxels (element size: 6 x 6 x 30 nm³)

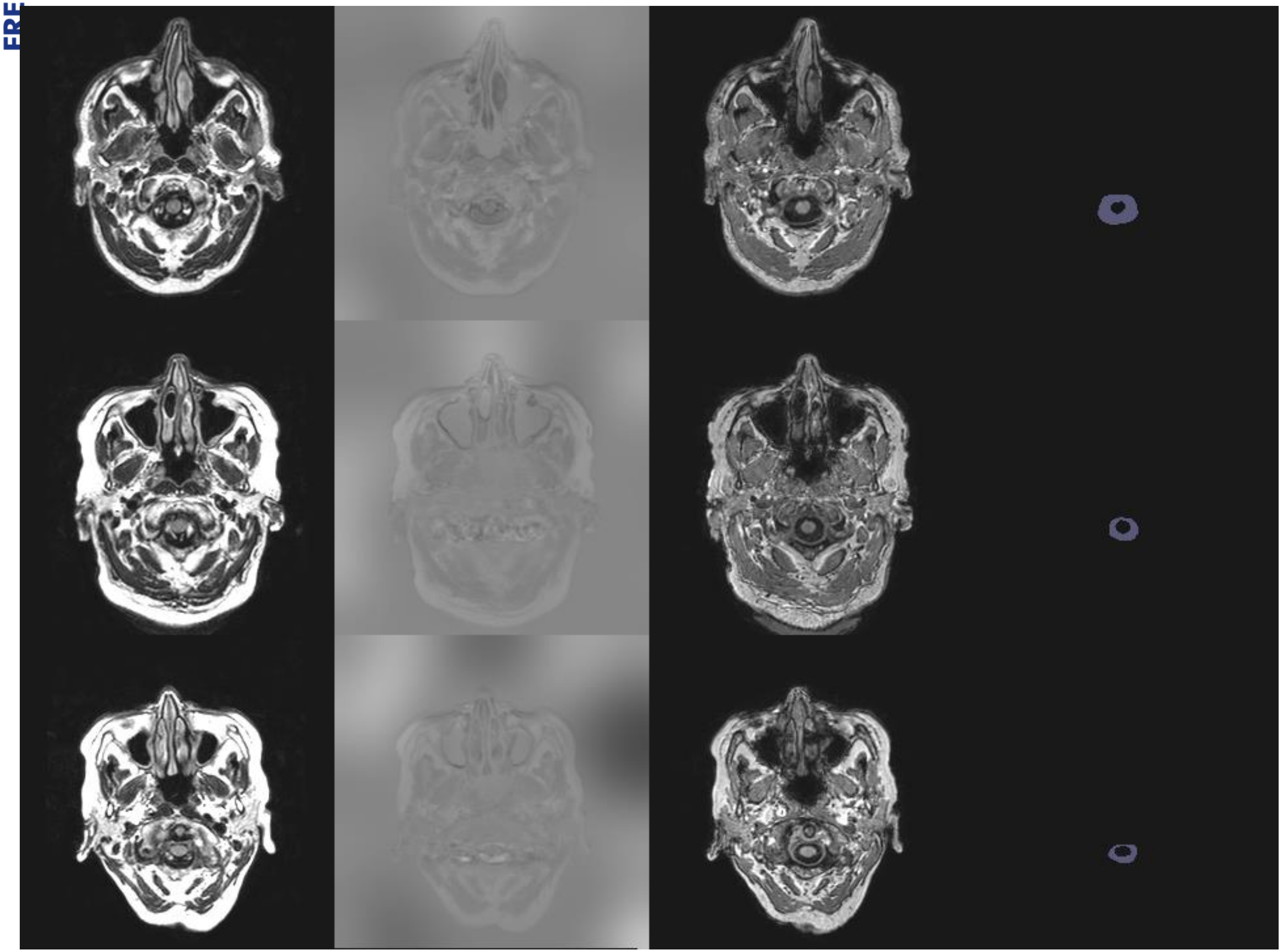


Data provided by the SNEMI3D challenge (http://brainiac2.mit.edu/SNEMI3D/home)

Results



Brain Segmentation Training Images (3 of 5)



T2-FLAIR | T1-IR | T1 | manual seg.

light blue:
gray matter,

yellow:
white matter,

dark blue:
Cerebrospinal fluid

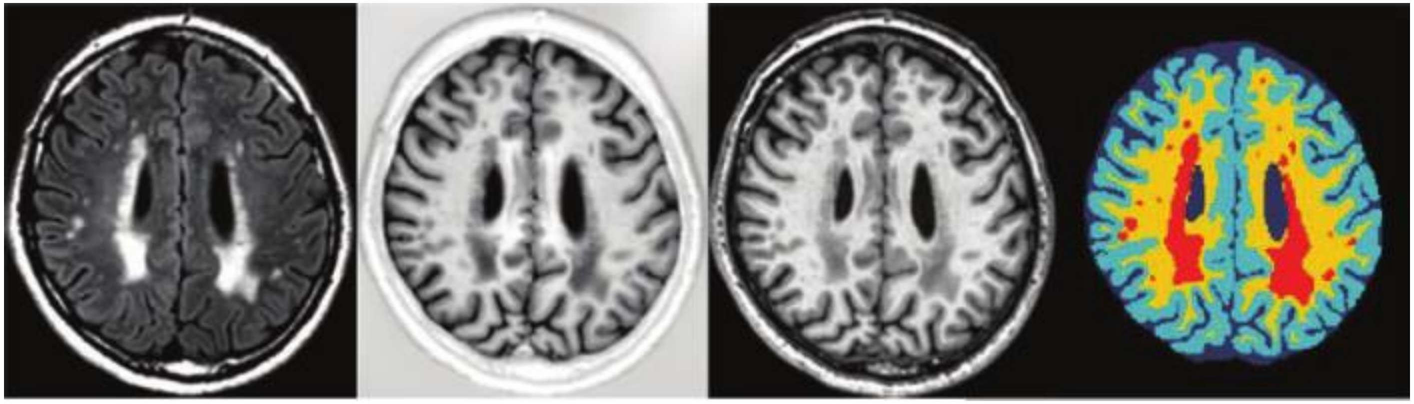
red:
WM lesions

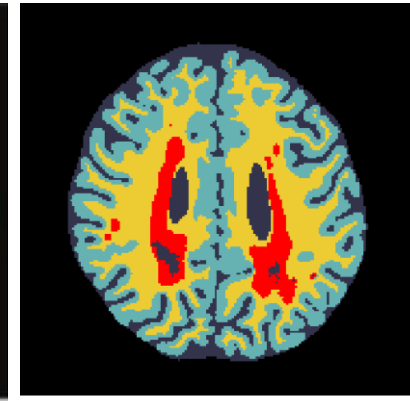
Ahmed Abdulkadir,
Collaboration with
Stefan Klöppel,
Neurology

Data provided by MRBrainS Challenge 2013 http://mrbrains13.isi.uu.nl/

Brain Segmentation Preliminary Results

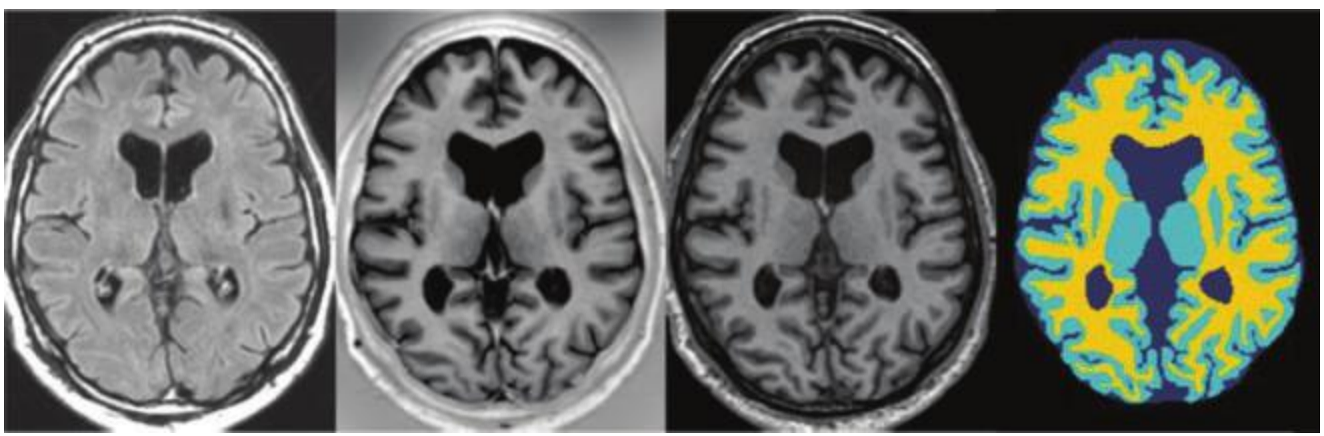
Test subject 03, slice 31

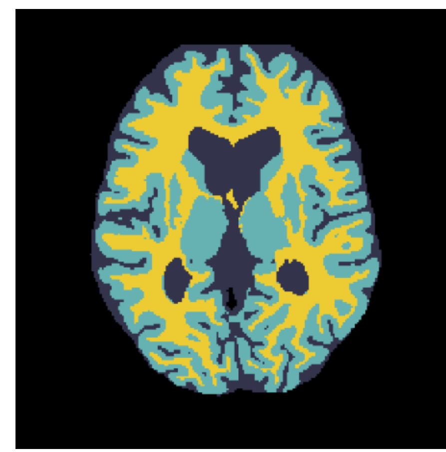




T2-FLAIR T1-IR T1 Manual seg. U-net

Test subject 09, slice 22





T2-FLAIR T1-IR T1 Manual seg. U-net

Conclusion

- Deep convolutional networks take over large parts of biomedical image analysis
- Learning segmentation (of biomedical structures) requires only few annotated training samples
- U-net architecture and elastically deformed training data do the job
- 2D implementation (based on Caffe) and ready trained networks available on our homepage
  http://lmb.informatik.uni-freiburg.de/people/ronneber/u-net

[O. Ronneberger, P. Fischer, T. Brox: U-net: Deep Convolutional Networks for Biomedical Image Segmentation. MICCAI 2015, available of ArXiv]

Programming Assignment

- Get familiar with the caffe neural network framework
  http://caffe.berkeleyvision.org/

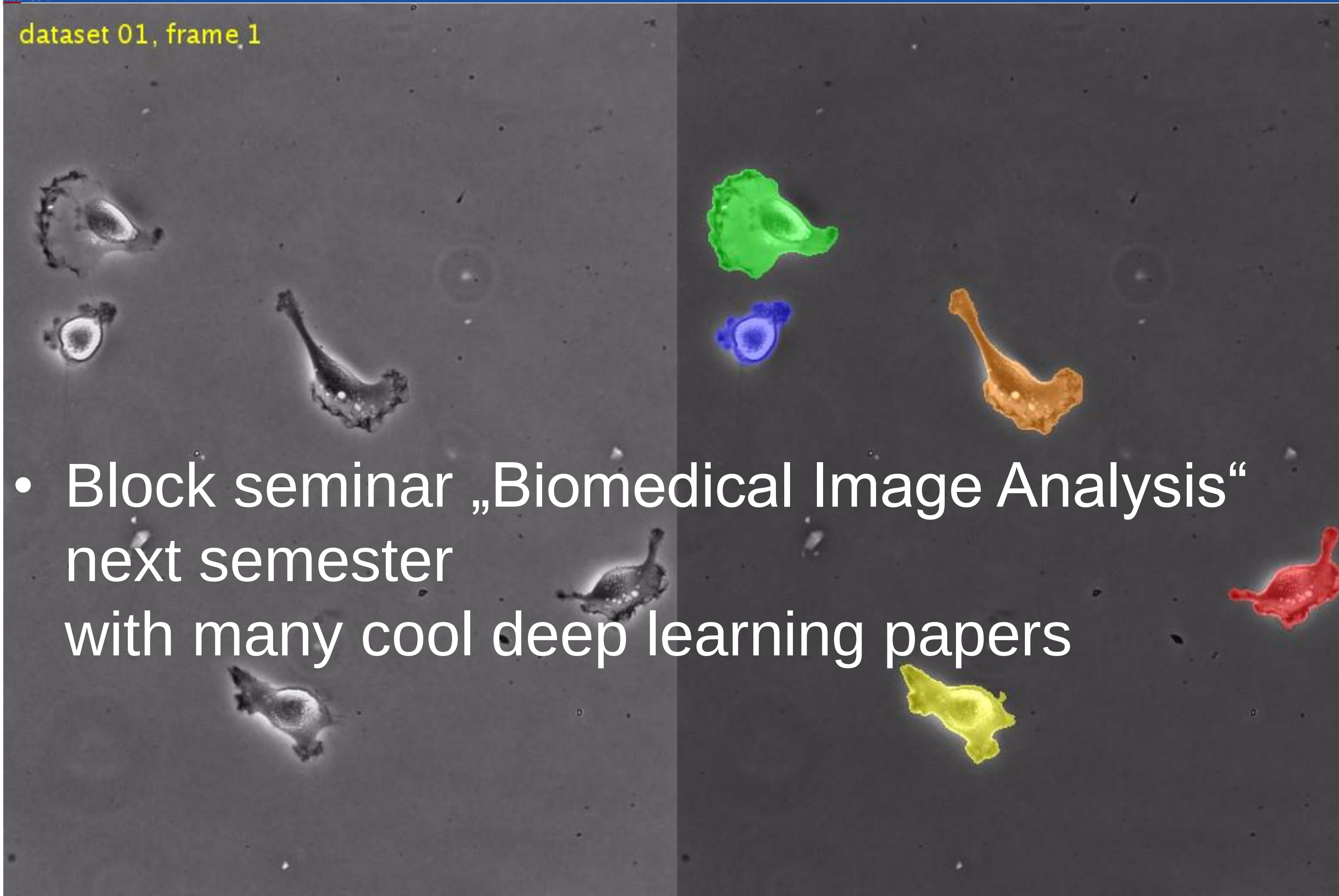
- Train a neural network for handwritten digits recognition (MNIST dataset):
  http://caffe.berkeleyvision.org/gathered/examples/mnist.html
  This can be done on CPU or GPU

- If you don‘t have a recent NVidia GPU this exercise might be a good argument why your parents should put a Titan X under the Christmas tree -- of course a much smaller GPU or the CPU alone is sufficient for MNIST. You know, I know, but maybe your parents don‘t ☺

- If you don‘t want to compile caffe yourself (quite a lot of third party libraries), we will provide a ready compiled version (Linux, 64bit) on the course homepage.

- Optional: Do some cool stuff with the ready trained networks from the “model zoo” on the caffe homepage. Here you might really need a GPU.

Advertisement



- Block seminar „Biomedical Image Analysis“ next semester
  with many cool deep learning papers