PoL Physics of Life TU Dresden

CENTER FOR SYSTEMS BIOLOGY DRESDEN

Bio-Image Analysis, Programming, Biostatistics and Machine Learning for Computational Biology

Anna Poetsch, Melissa Sanabria, Allyson Quinn Ryan, Robert Haase

July 2023

@haesleinhuepf

PoL
Physics of Life
TU Dresden

CENTER FOR
SYSTEMS BIOLOGY
DRESDEN

Programming

July 2023

@haesleinhuepf

Generating arrays within for-loops

PoL Physics of Life TU Dresden

- What is the output of these programs?

```python
# we start with an empty list
numbers = []

# and add elements
for i in range(0, 5):
    numbers.append(i * 2)

print(numbers)
```

```python
numbers = [i * 2 for i in range(0, 5)]

print(numbers)
```

Subsets

```
# Arrays
numbers = [0, 1, 2, 3, 4, 5, 6, 7, 8, 9]
print(numbers)
```

```
[0, 1, 2, 3, 4, 5, 6, 7, 8, 9]
```

- What do these programs output?

Starting at (including)

Ending at (excluding)

```
subset = numbers[2:4]
print(subset)
```

Step size

```
subset_with_gaps = arr[1:8:2]
print(subset_with_gaps)
```

Image analysis Python programming

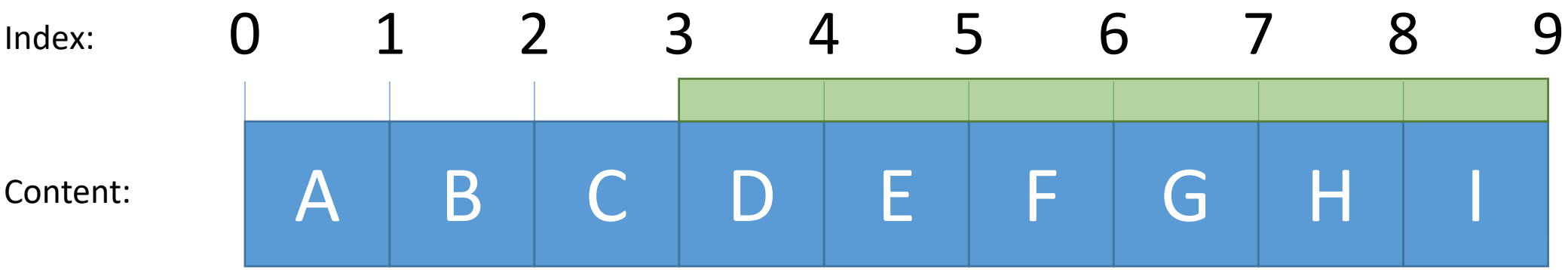
PoL Physics of Life TU Dresden

- What would be good comments in this code?

```
#
from skimage.io import imread
blobs = imread("blobs.tif")

#
from skimage.filters import threshold_otsu
threshold = threshold_otsu(blobs)
binary_blobs = blobs > threshold

#
from skimage.measure import label
labeled_blobs = label(binary_blobs)

#
import matplotlib.pyplot as plt
fig, axs = plt.subplots(1, 3, figsize=(15,15))

axs[0].imshow(blobs)
axs[1].imshow(binary_blobs)
axs[2].imshow(labeled_blobs, cmap=label_cmap)
```

Indexing, cropping, subsets

```
data = ['A', 'B', 'C', 'D', 'E', 'F', 'G', 'H', 'I']
```



What's the output of

```
data[3:]
```

?

ABC

CDEFGHI

DEFGHI

PoL
Physics of Life
TU Dresden

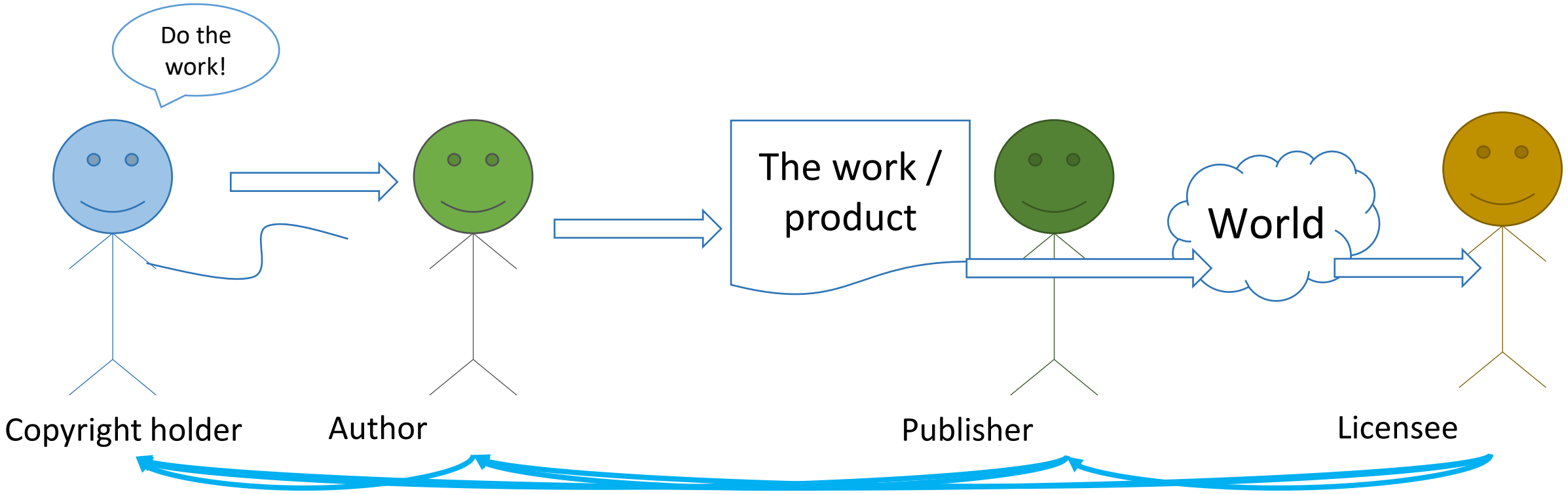
CENTER FOR
SYSTEMS BIOLOGY
DRESDEN

Data Management

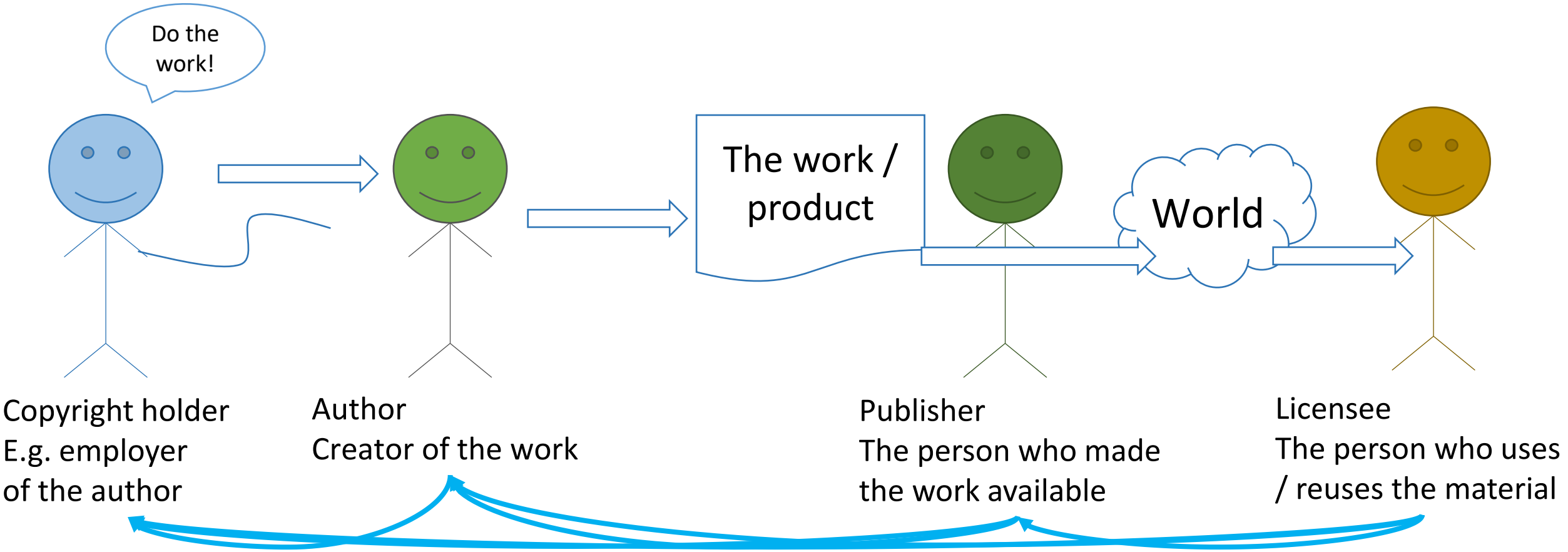
July 2023

@haesleinhuepf

Terminology



Terminology

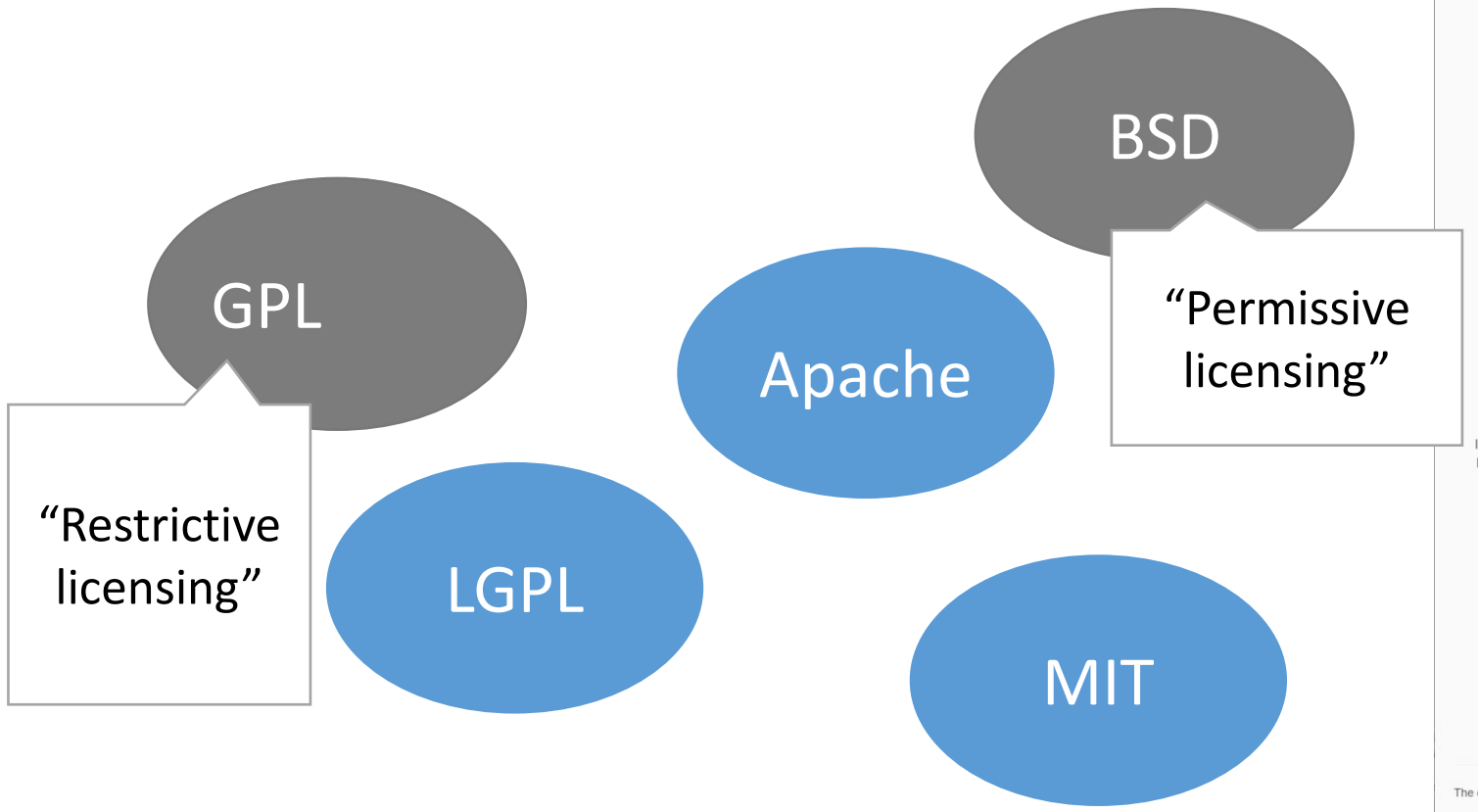


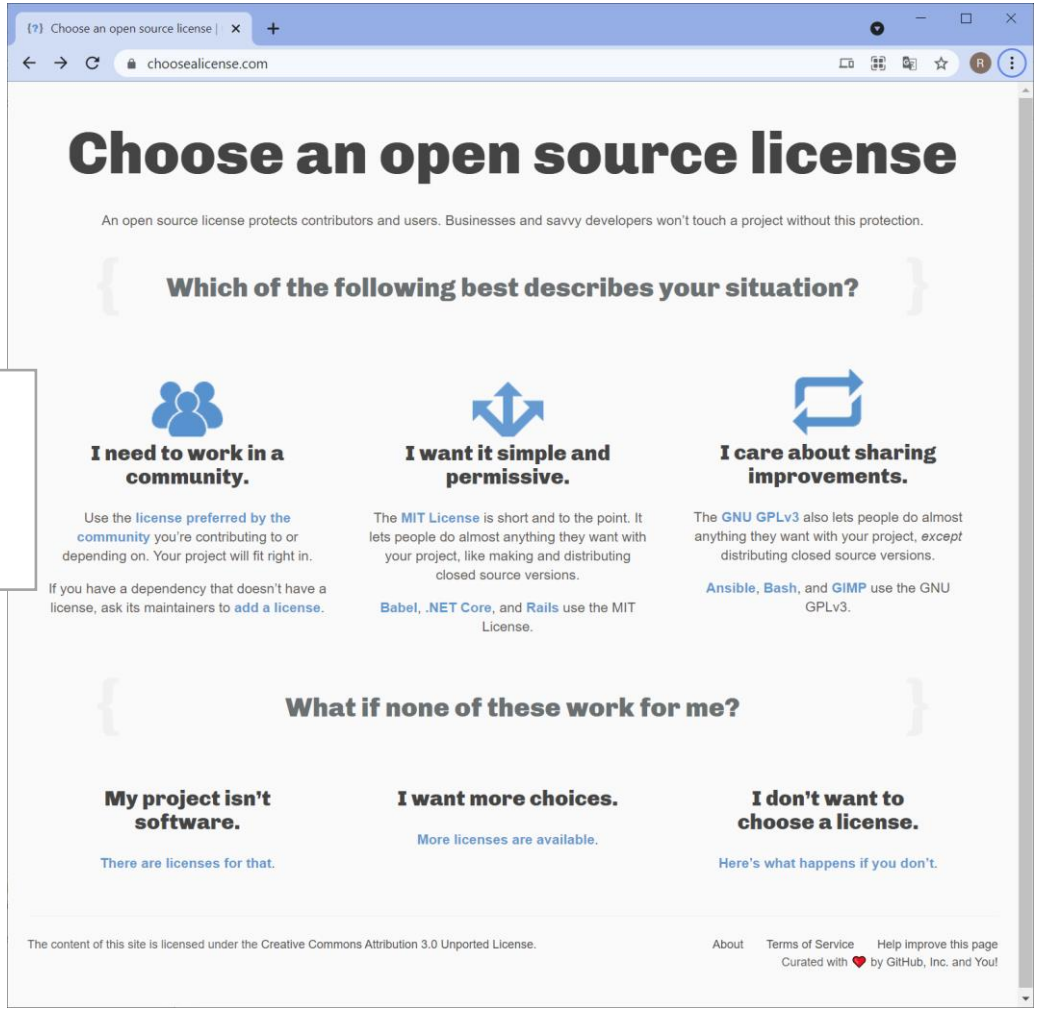
The FAIR principles

- **Findable**
  - Meta-data, unique identifiers, searchable databases
- **Accessible**
  - Open & free protocols
  - Authentication / authorization
- **Interoperable**
  - [meta] data use a formal, accessible, shared, broadly accessible language
  - References to other [meta]data
- **Reusable**
  - Properly licensed, associated relevant meta data, community-relevant community standards

Licensing software

- In the software world, other licenses are more popular, historically grown.



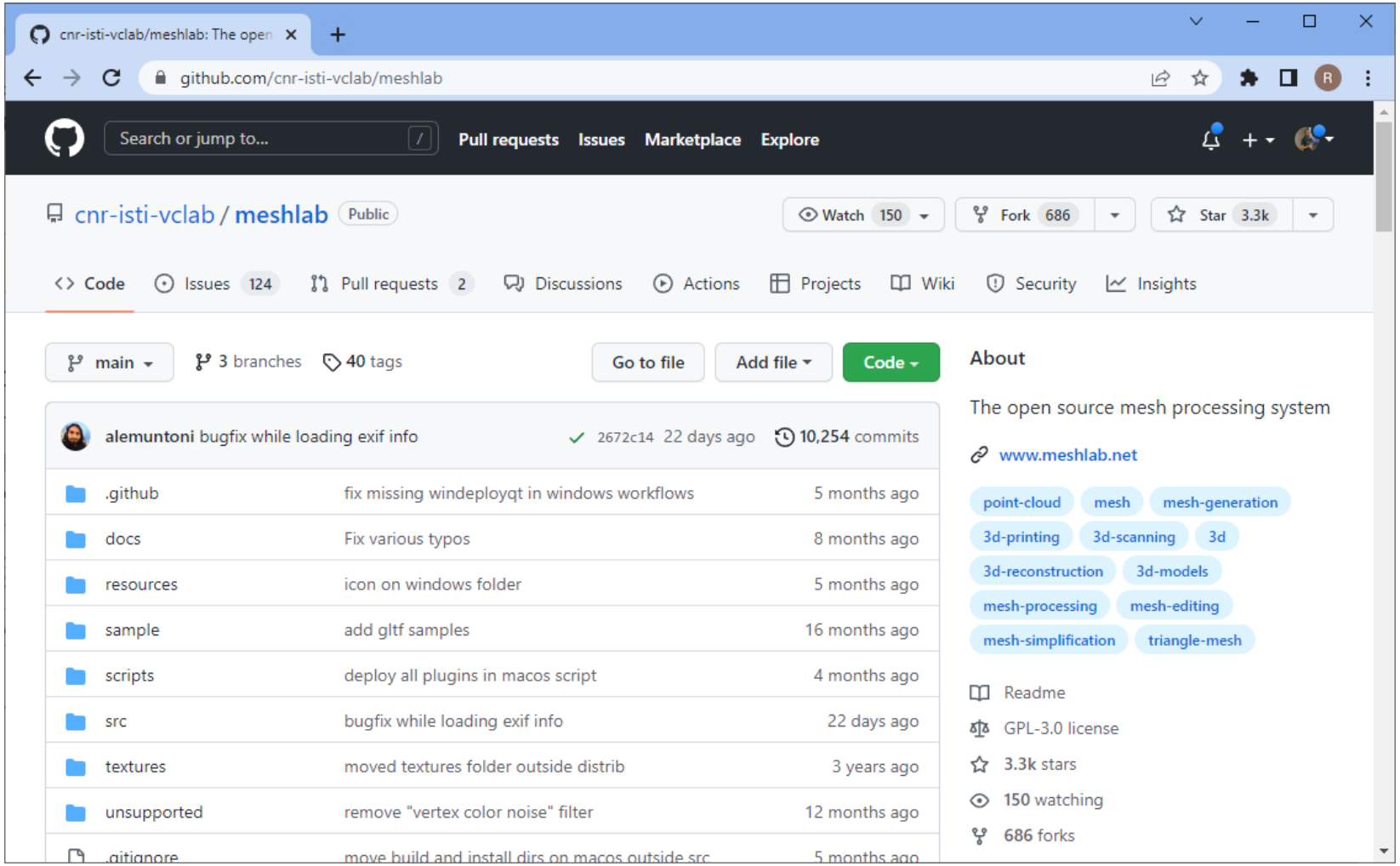


Quiz

- May I reuse code from this repository in my own BSD-licensed work?

Yes

No



Quiz: Digital Object Identifiers

PoL Physics of Life TU Dresden

- Which of these is a *unique* digital object identifier?

https://twitter.com/haesleinhuepf/status/891596662782779392

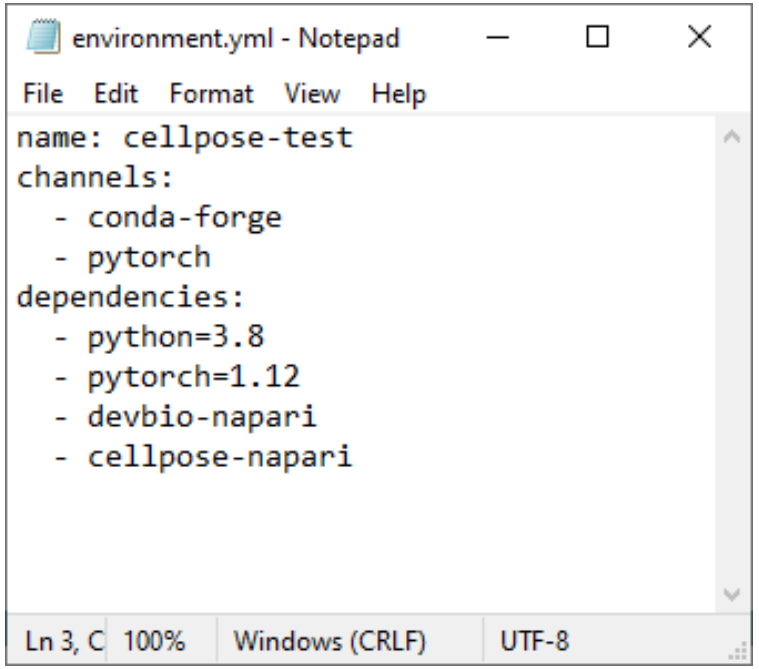
https://doi.org/10.5281/zenodo.28325

https://github.com/haesleinhuepf/devbio-napari

https://napari.org/

Documenting dependencies

- Maintain a document with the dependencies (and versions) you need in your project!
  - The conda way



In case your environment is screwed up, you can rebuild it any time.

```
conda env create -f environment.yml
```


https://conda.io/projects/conda/en/latest/user-guide/tasks/manage-environments.html#creating-an-environment-from-an-environment-yml-file
https://pip.pypa.io/en/stable/cli/pip_install/#examples

PoL
Physics of Life
TU Dresden

CENTER FOR
SYSTEMS BIOLOGY
DRESDEN

Bio-image Analysis

July 2023

@haesleinhuepf

Lecture overview: Bio-image Analysis

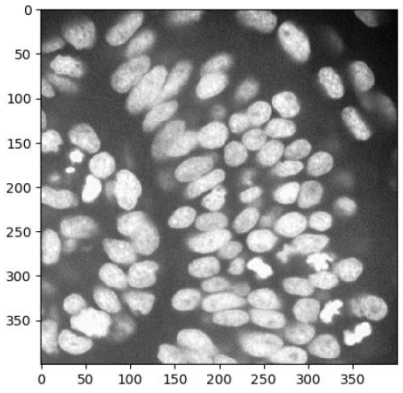
PoL Physics of Life TU Dresden

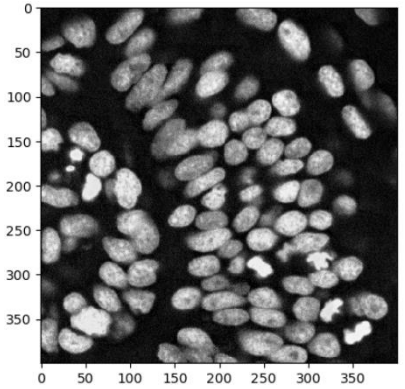
- Image Data Analysis workflows
- Goal: **Quantify observations, substantiate conclusions with numbers**

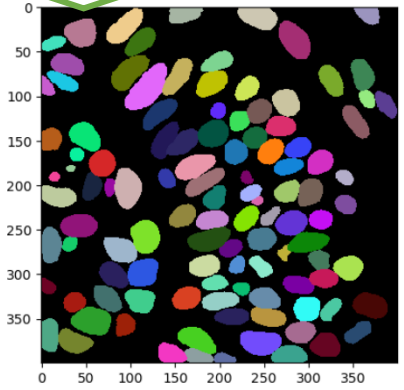
Machine learning: Pixel classification, Instance segmentation

Machine learning: Dimensionality reduction

Machine learning: Object classification

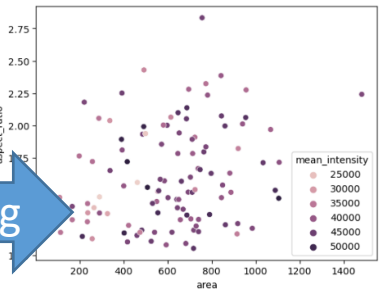






	area	mean_intensity	major_axis_length	minor_axis_length
0	594.0	40572.809764	28.611591	26.537947
1	645.0	43764.872868	33.511511	24.566916
2	1105.0	51970.561991	45.232031	31.456308
3	718.0	47015.487465	31.023274	29.5
4	791.0	49132.515803	36.382253	27.7
...	...	...	...	
105	238.0	30477.126050	20.252197	15.2
106	615.0	32886.154472	41.030760	19.8
107	110.0	33042.445455	14.366347	9.9
108	222.0	43304.180180	25.370549	11.6
109	167.0	43378.808383		

Plotting



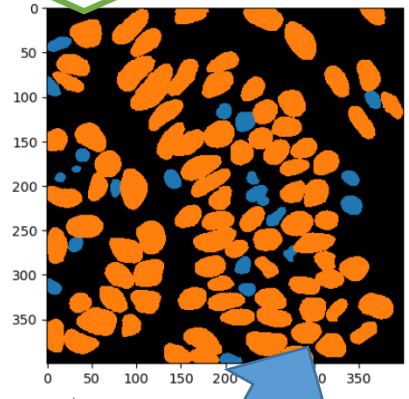


Image filtering → Image segmentation → Feature extraction → Object classification → Visualization

Bio-image analysis → Bio-statistics

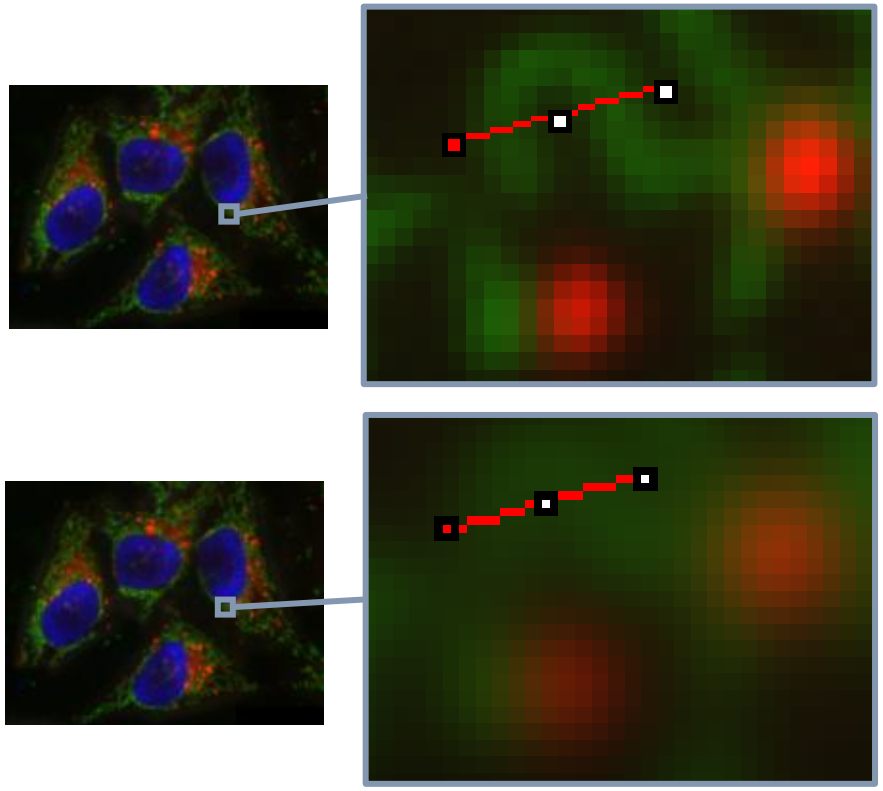
Introduction to bio-image analysis

PoL Physics of Life TU Dresden

- Bio-image analysis is supposed to be
  - **Quantitative**
    - We derive numbers from images which describe physical properties of the observed sample.
  - **Objective**
    - The derived measurement does not depend on who did the measurement. The measurement is free of interpretation.
  - **Reliable (trustworthy / validated)**
    - We are confident that the measurement is describing what it is supposed to describe.
  - **Reproducible**
    - Enabling others to re-do the experiment. For this, documentation is crucial!
  - **Replicability**
    - Others *do* execute the same analysis, potentially on other data, and see consistent results.
  - **Repeatable**
    - We can do the same experiment twice under the *same conditions* and get the same measurements.

Pixel size versus resolution

- How is the resolution of an imaging system defined?
- How is the pixel size of an image defined?



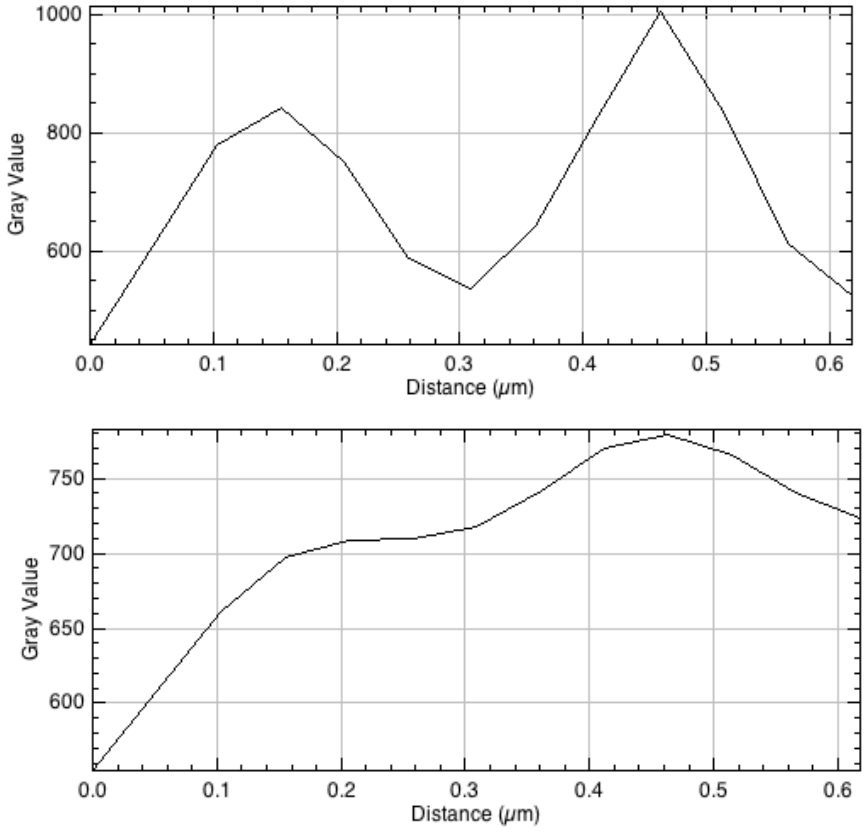
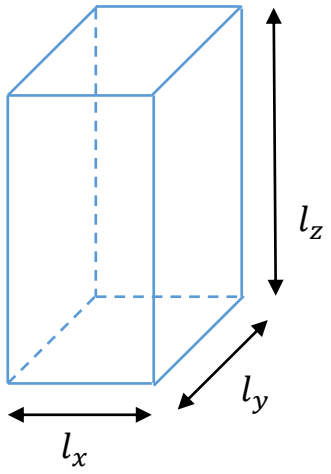
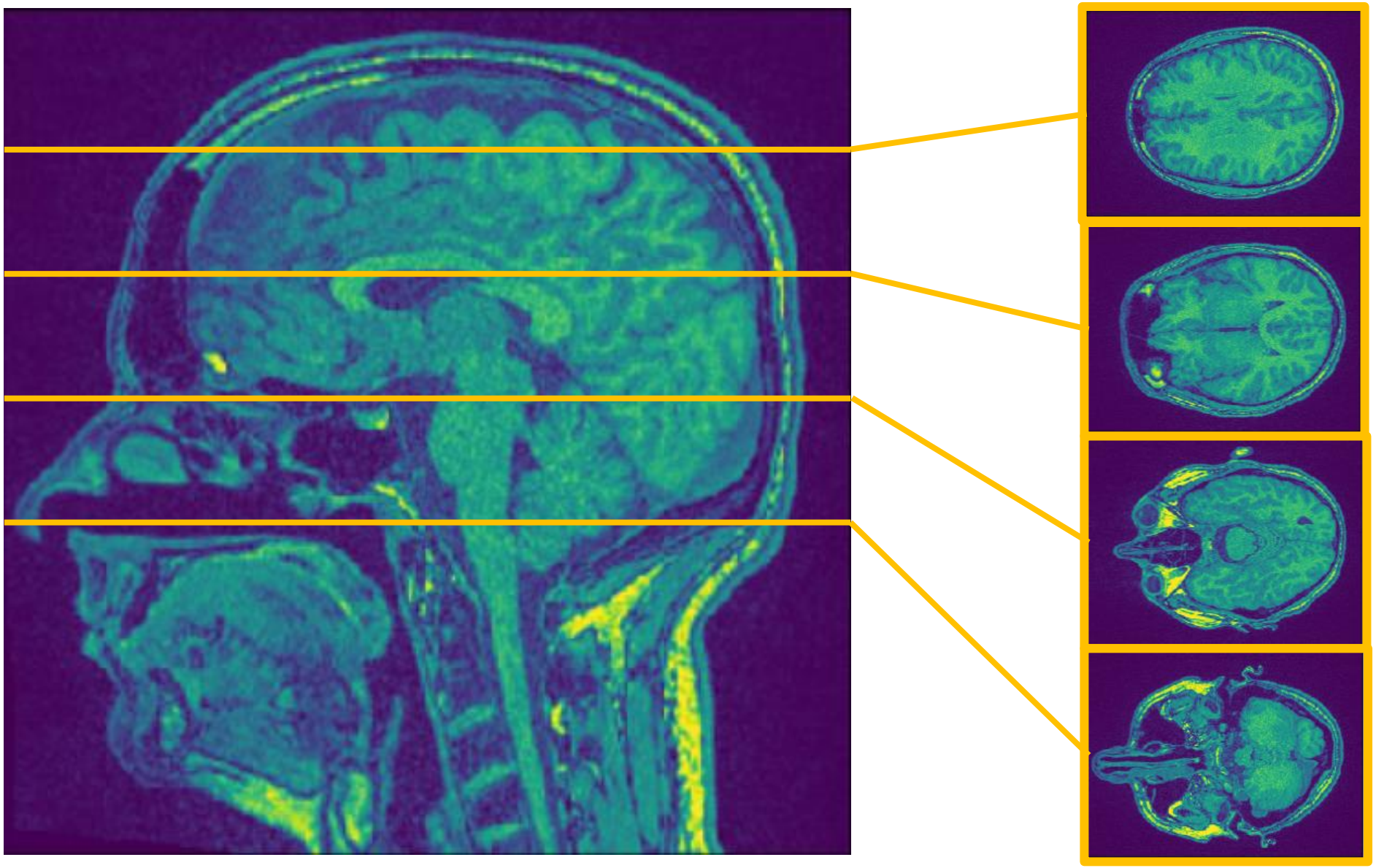


Image stacks and voxels

- 3-dimensional images consisting of voxels
- “Image stack”
- Often anisotropic



$$l_x = l_y \neq l_z$$



PoL
Physics of Life
TU Dresden

CENTER FOR
SYSTEMS BIOLOGY
DRESDEN

Image Processing Filters

July 2023

@haesleinhuepf

Filters

- An image processing filter is an operation on an image.
- It takes an image and produces a new image out of it.
- Filters change pixel values.

- There is no “best” filter. Which filter fits your needs, depends on the context.
- Filters do not do magic. They can not make things visible which are not in the image.

- Application examples
  - Noise-reduction
  - Background removal
  - Artefact-removal
  - Contrast enhancement
  - Correct uneven illumination

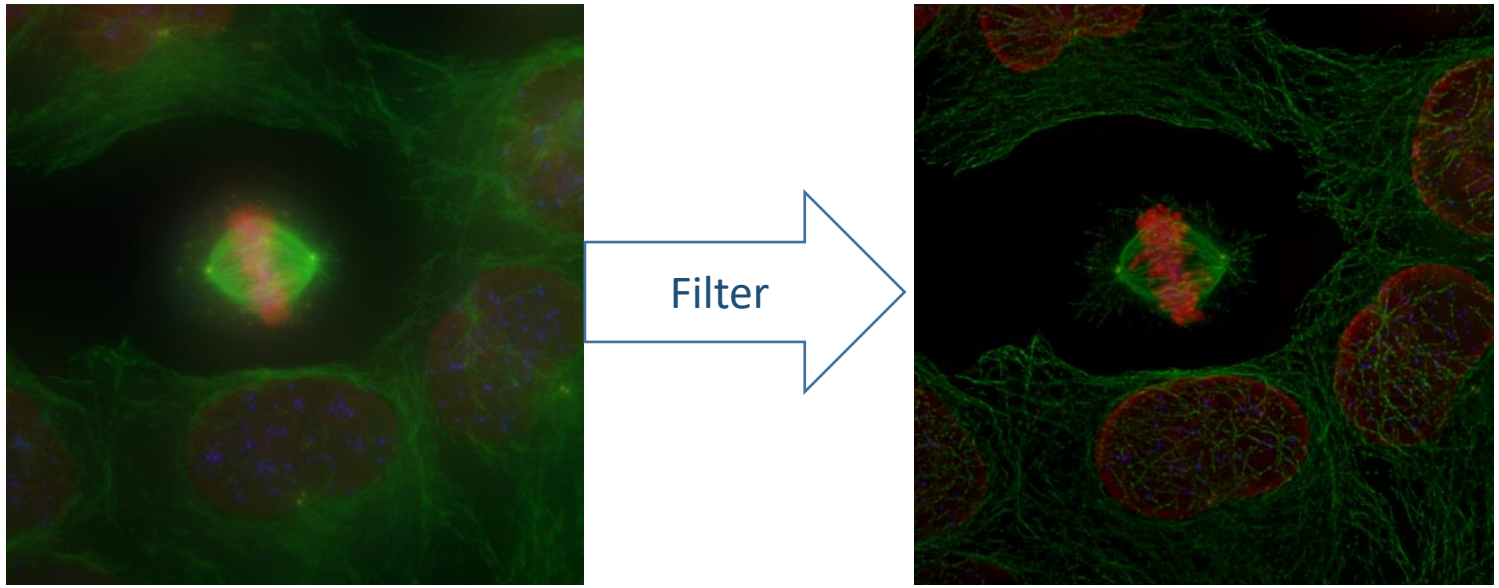
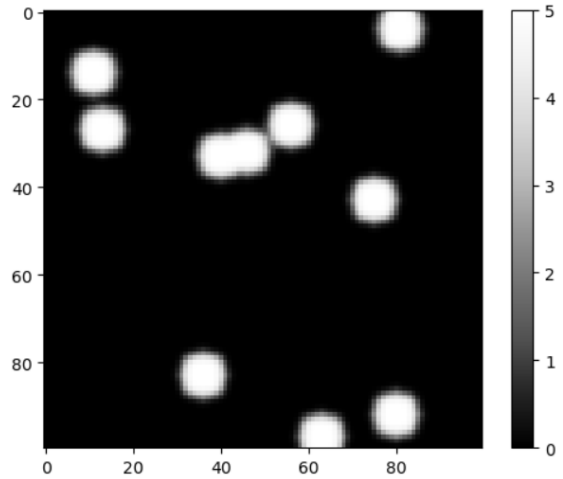


Image source: Alex Bird / Dan White MPI CBG

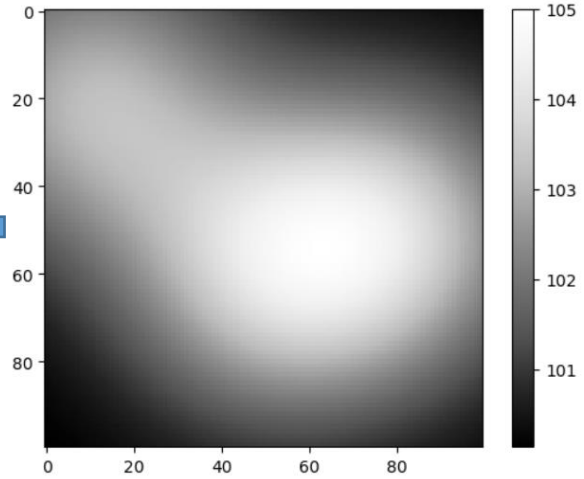
Effects harming image quality

- Image formation (simulated)

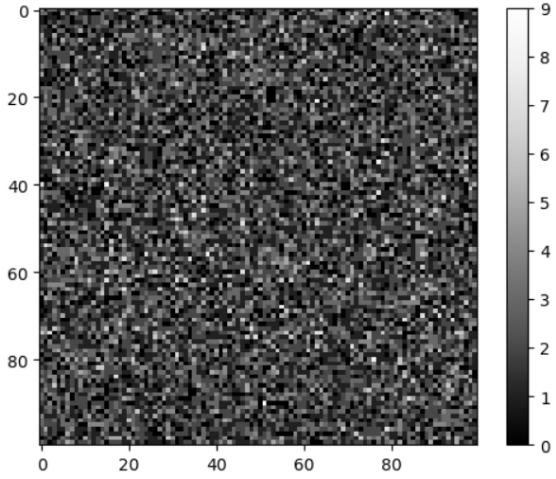
“nuclei”

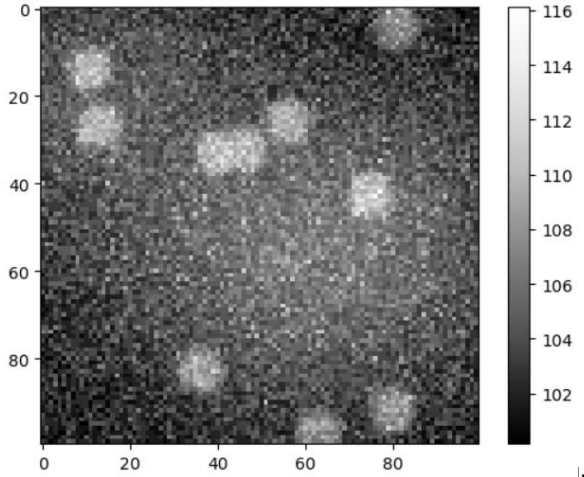


“background”

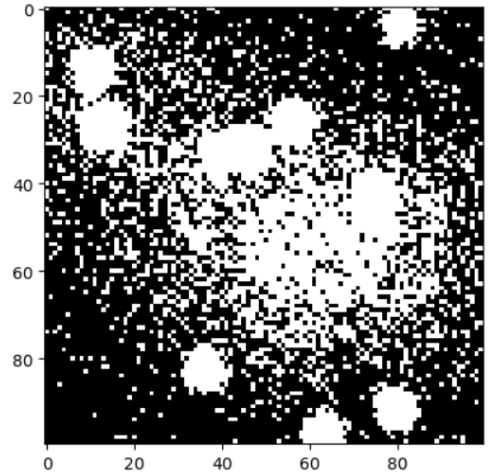


“noise”



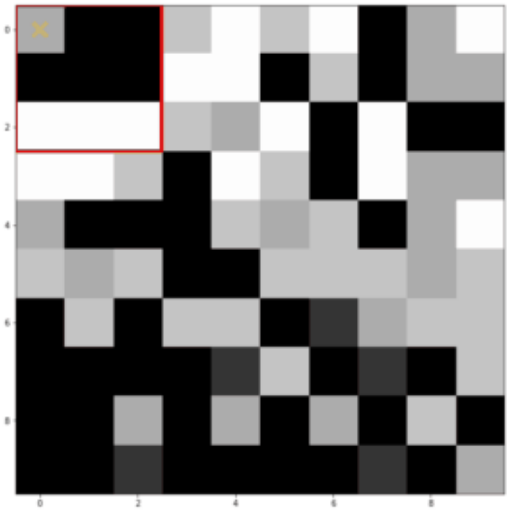


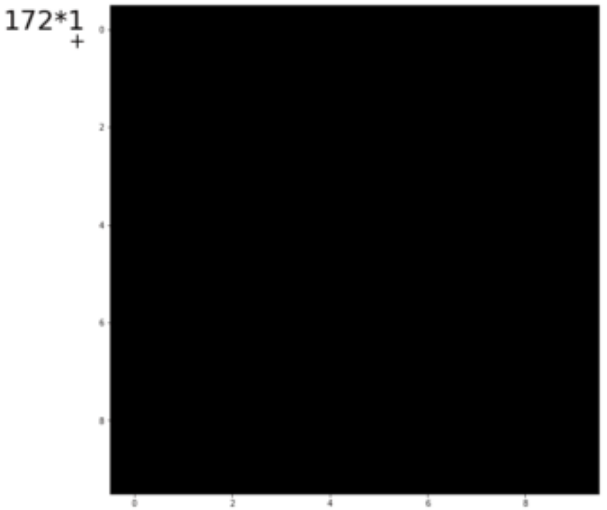
Segmentation



Linear Filters: Convolution

- What is an image processing filter?
- What is a filter kernel?
- How does convolution work?





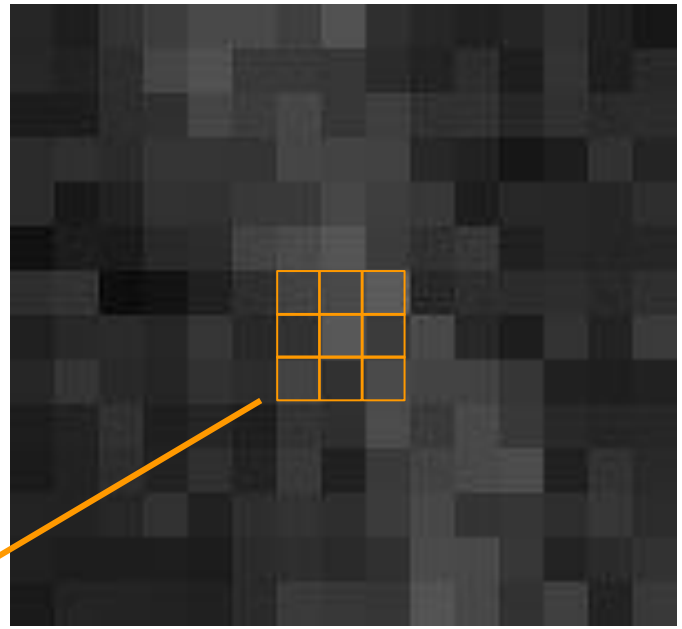
Mean filter

$$\begin{bmatrix} 1/9 & 1/9 & 1/9 \\ 1/9 & 1/9 & 1/9 \\ 1/9 & 1/9 & 1/9 \end{bmatrix}$$

Animation source: Dominic Waithe, Oxford University
https://github.com/dwaithe/generalMacros/tree/master/convolution_ani

Nonlinear Filters

- What differentiates linear and non-linear filters?



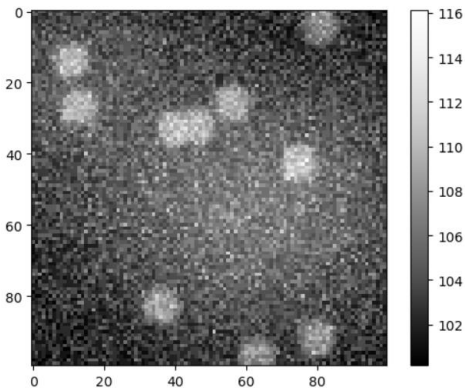
75	85	60
67	73	91
50	88	59

50 59 60 67 73 75 85 88 91

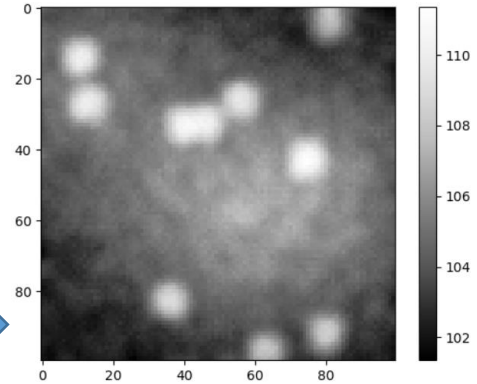
Min Median Max

Image filtering

- High-pass versus low-pass filters

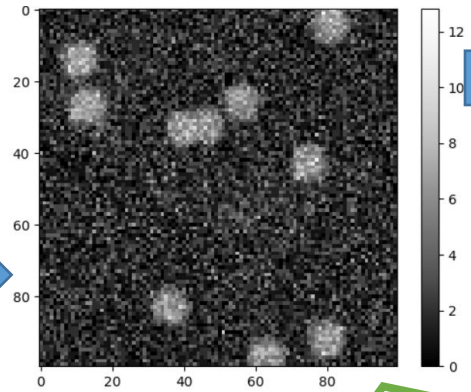


Denoise

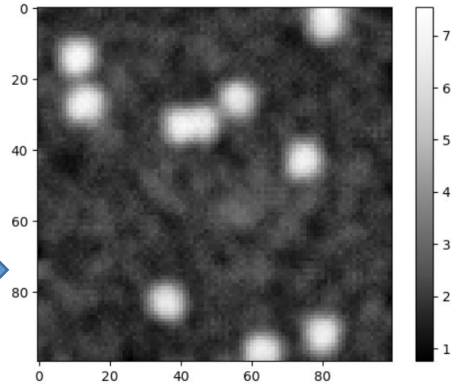


Low-pass filter: Allows low frequencies pass

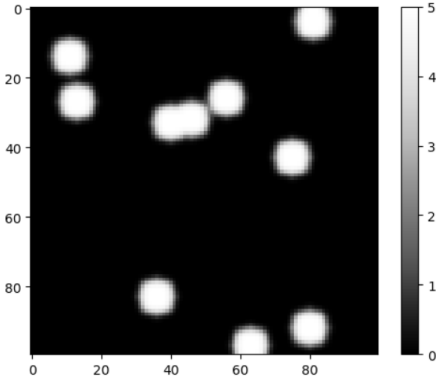
Remove background



Denoise



≈



Band-pass filter: Allows a specific range of frequencies pass

High-pass filter: Allows high frequencies pass

https://github.com/BiAPoL/Bio-image_Analysis_with_Python/blob/49a787514a367829c3e0e1832f6cc533e96d549f/03_image_processing/simulated_dataset.ipynb

PoL
Physics of Life
TU Dresden

CENTER FOR
SYSTEMS BIOLOGY
DRESDEN

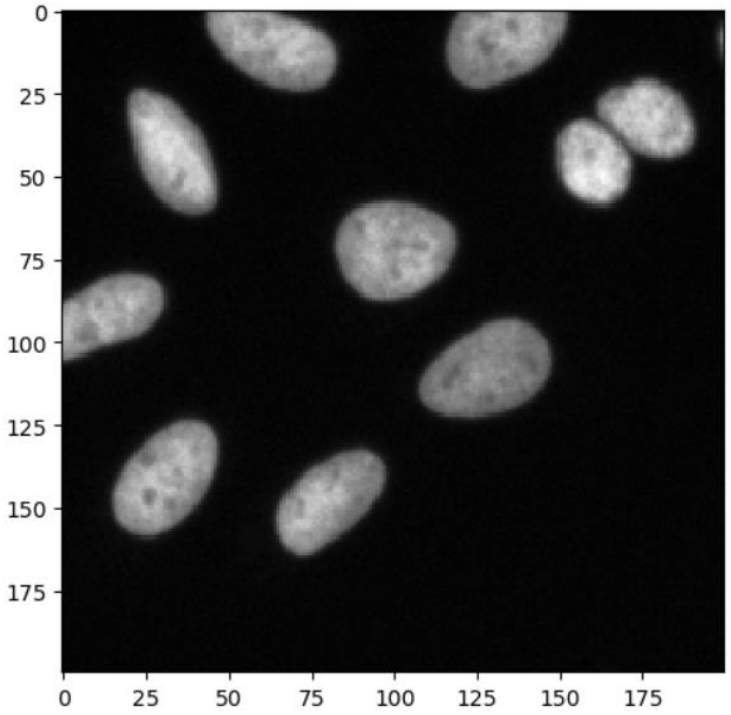
Image Segmentation

July 2023

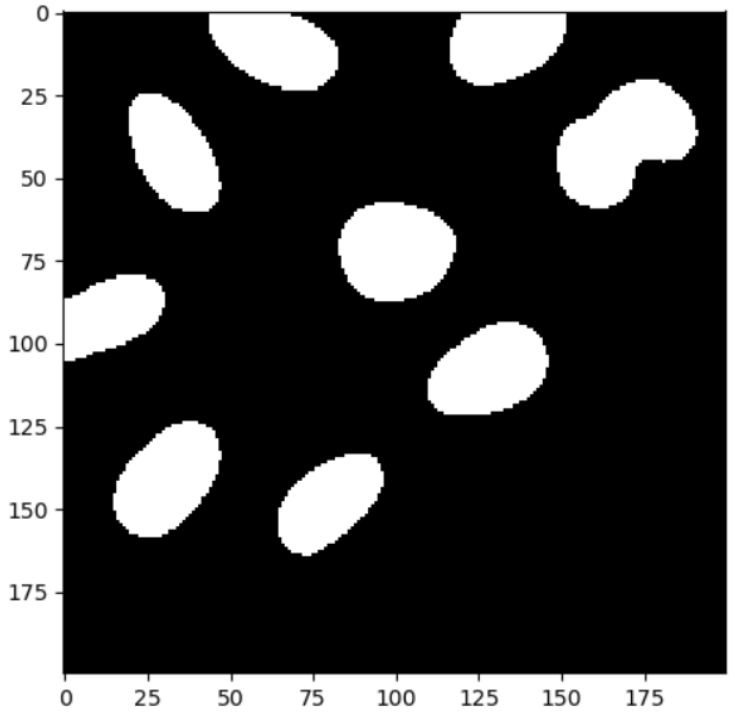
@haesleinhuepf

Terminology

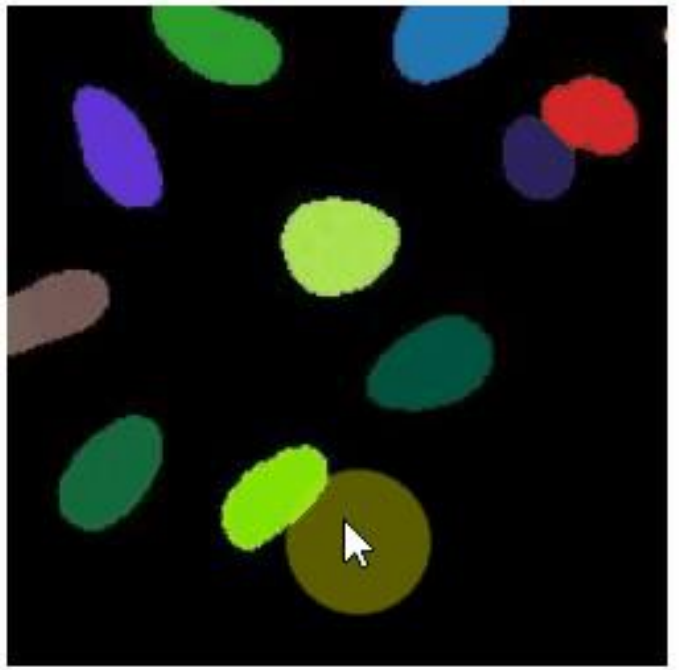
Intensity image



Binary image

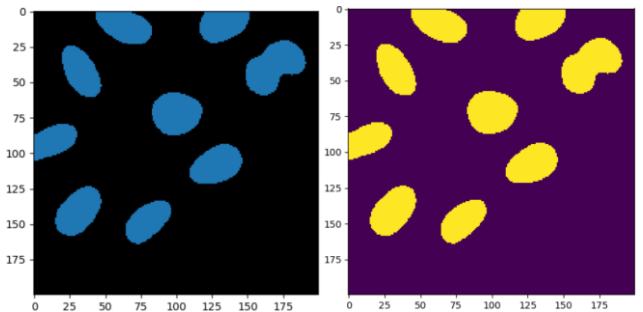


Label image



[y=152, x=92] = 0

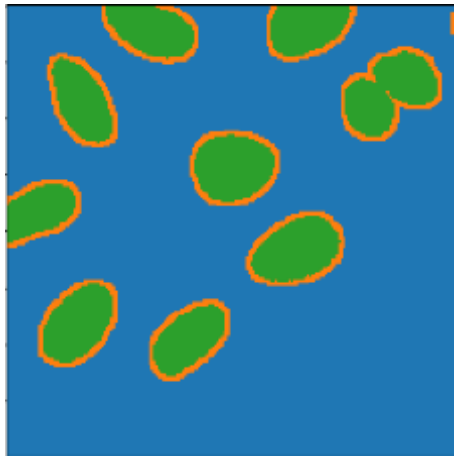
No matter how they are displayed



Terminology

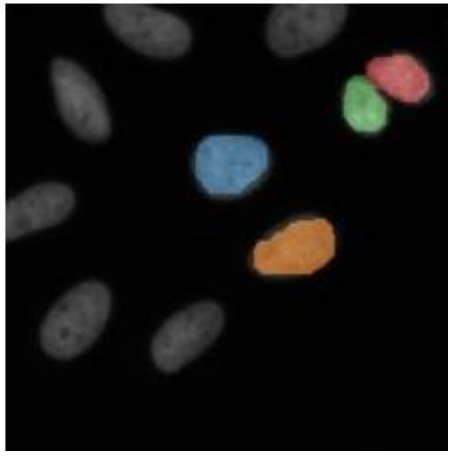
- Annotations are typically drawn by humans (e.g. to train machine learning models)

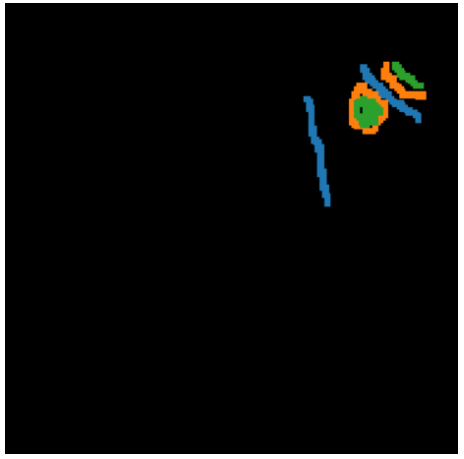
Instance segmentation



Semantic segmentation

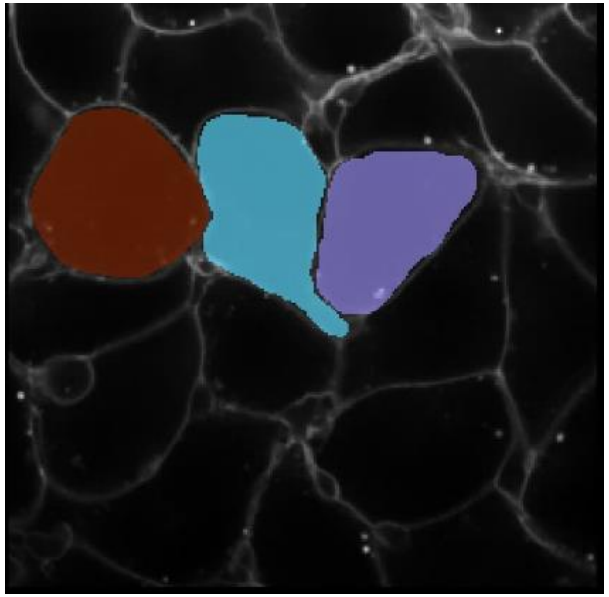
Sparse instance annotation





Sparse semantic annotation

Sparse Jaccard Index



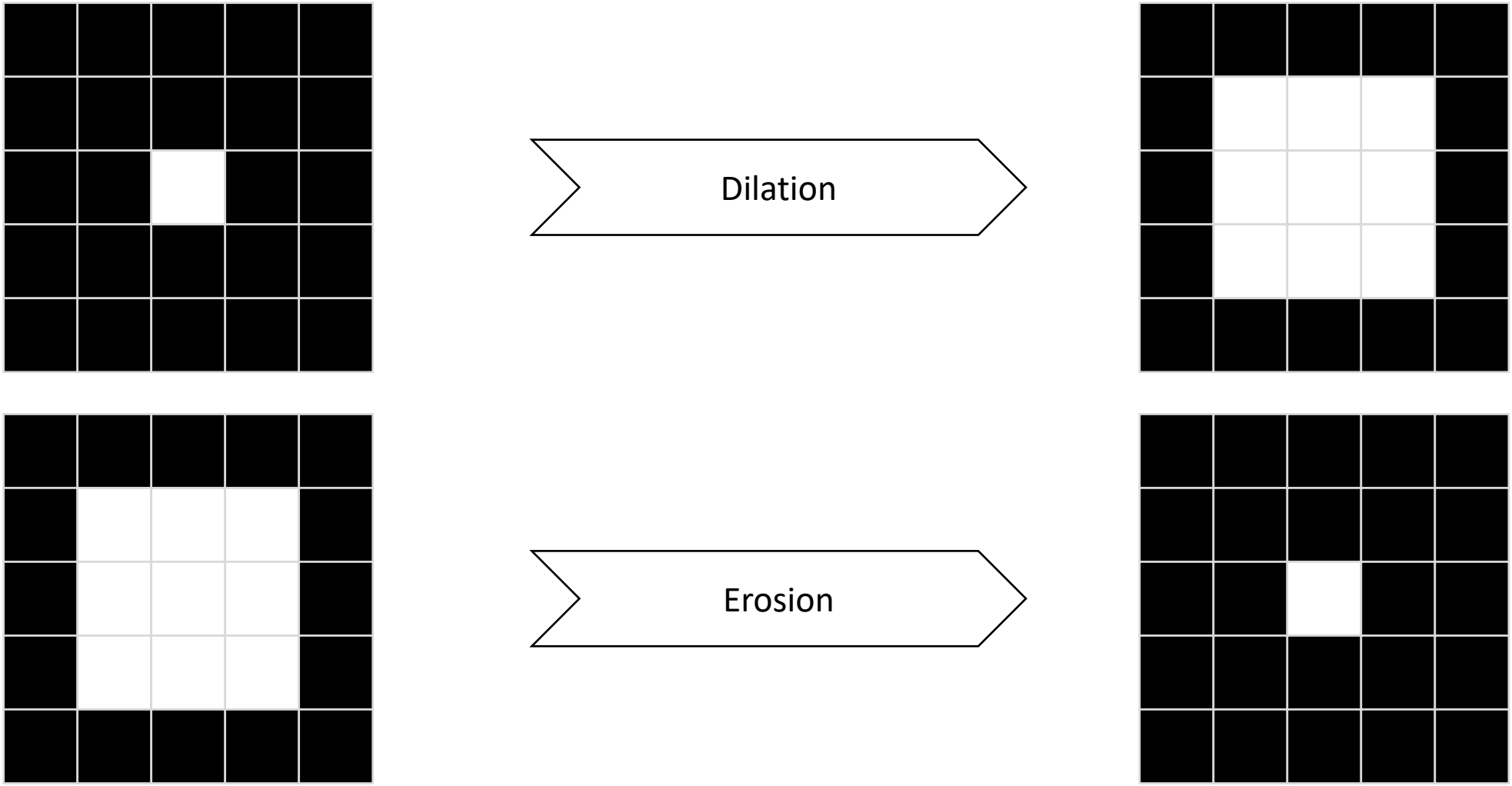
This is a …

Sparse instance segmentation

Sparse semantic segmentation

Refining masks: Dilation and Erosion

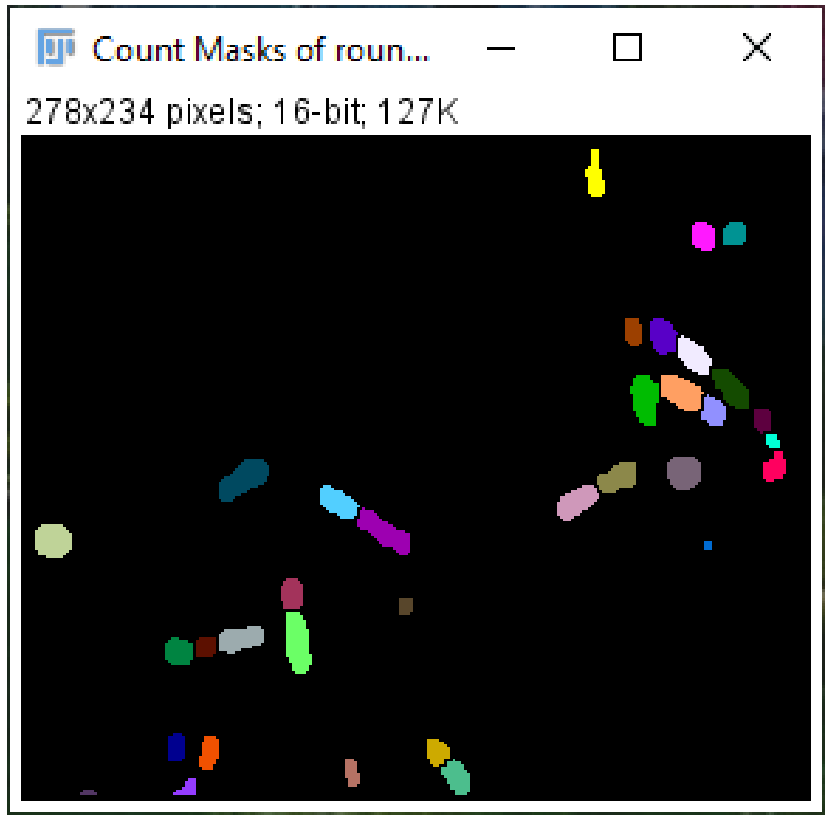
- Dilation: Every pixel with at least one white neighbor becomes white.



Quiz

- What's the name of the operation applied here?

?



Thresholding

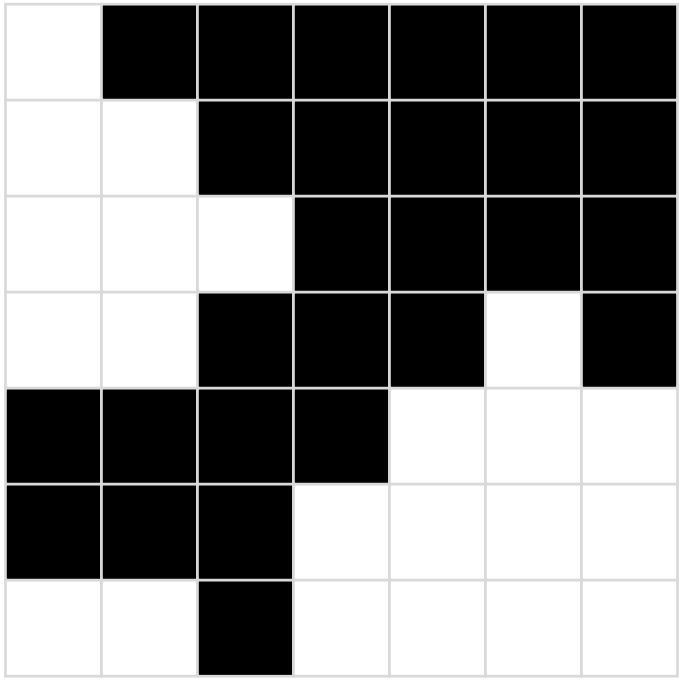
Binary closing

Connected components analysis

Watershed

Connected components labelling

- In order to allow the computer differentiating objects, connected components analysis (CCA) is used to mark pixels belonging to different objects with different numbers
- Background pixels are marked with 0.
- The maximum intensity of a labelled map corresponds to the number of objects.

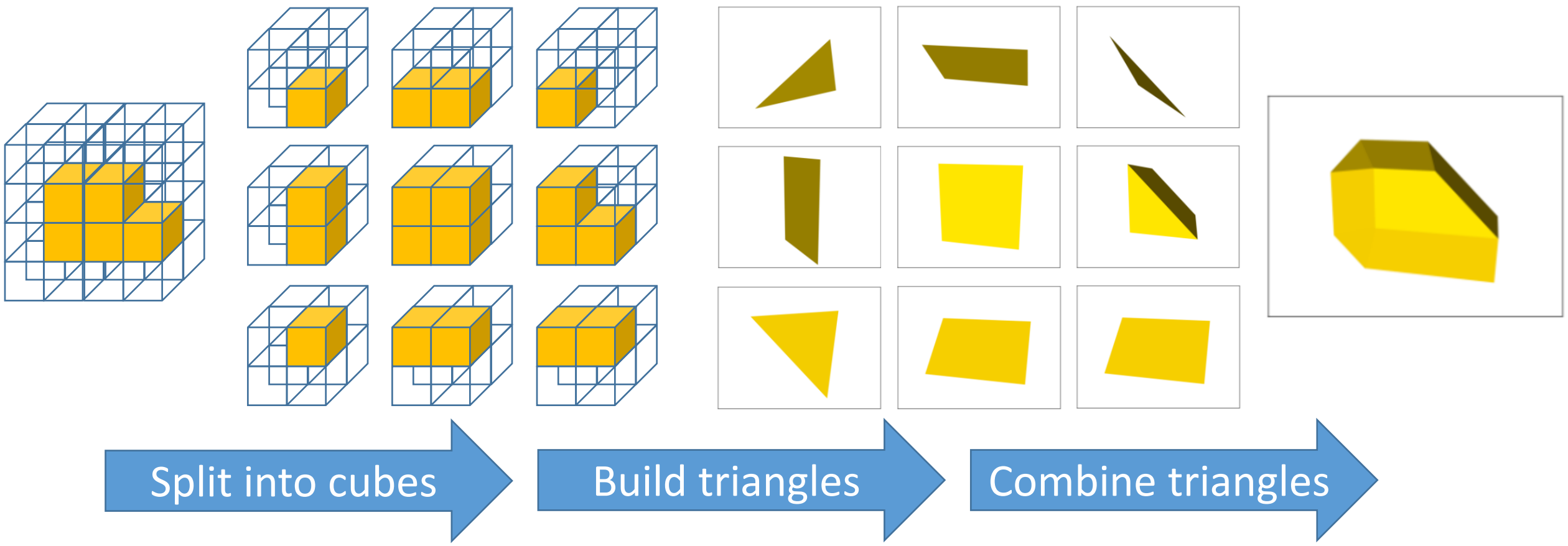


CCL

1	0	0	0	0	0	0
1	1	0	0	0	0	0
1	1	1	0	0	0	0
1	1	0	0	0	3	0
0	0	0	0	3	3	3
0	0	0	3	3	3	3
2	2	0	3	3	3	3

Marching cubes algorithm

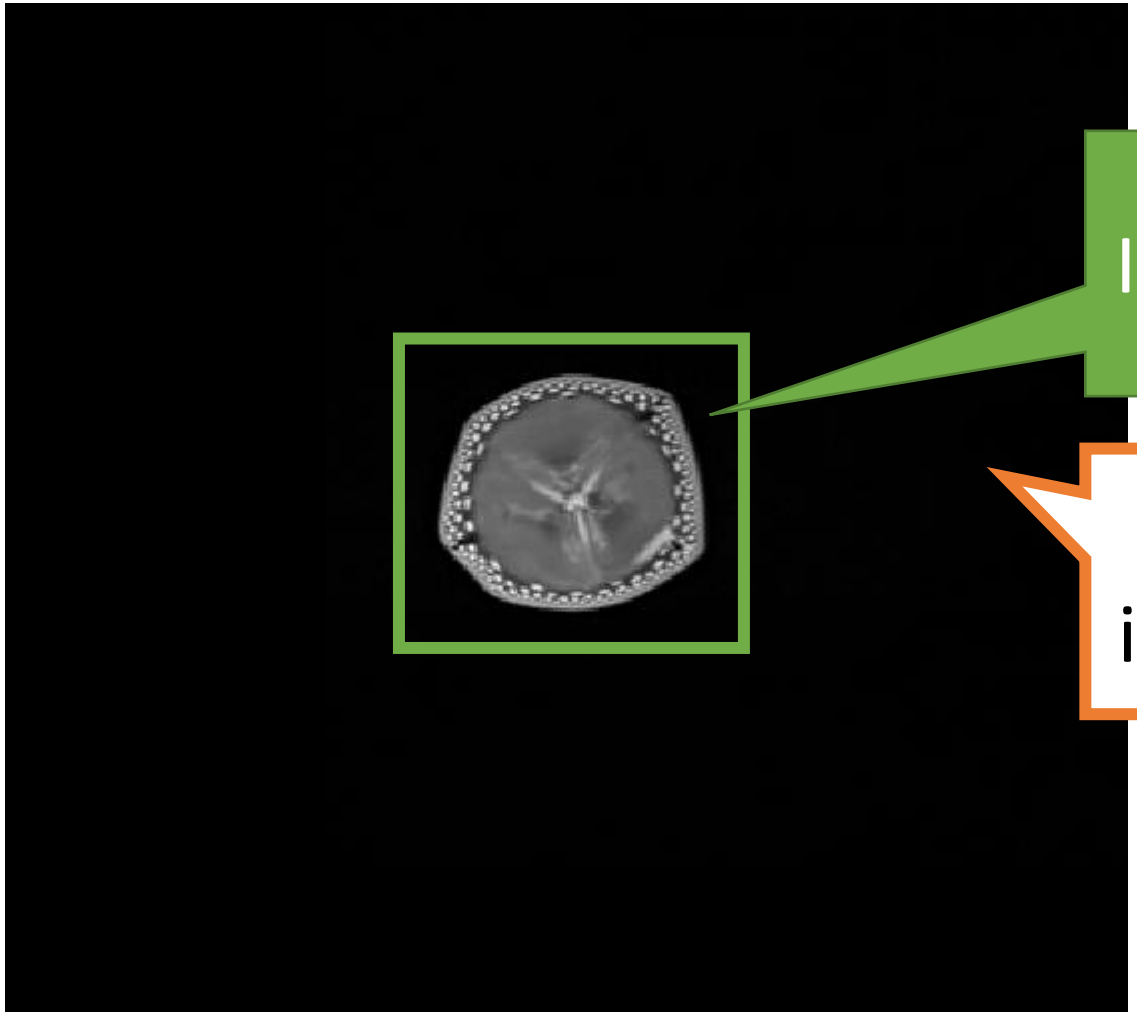
- Starting point: 3D binary image
- Cuts the image in small cubes and iterates over them



Lorensen, William E.; Cline, Harvey E. (1 August 1987). "Marching cubes: A high resolution 3D surface construction algorithm". *ACM SIGGRAPH Computer Graphics*. **21** (4): 163-169. CiteSeerX 10.1.1.545.613. doi:10.1145/37402.37422.

Cropping and resampling images

- Crop out the region you're interested in

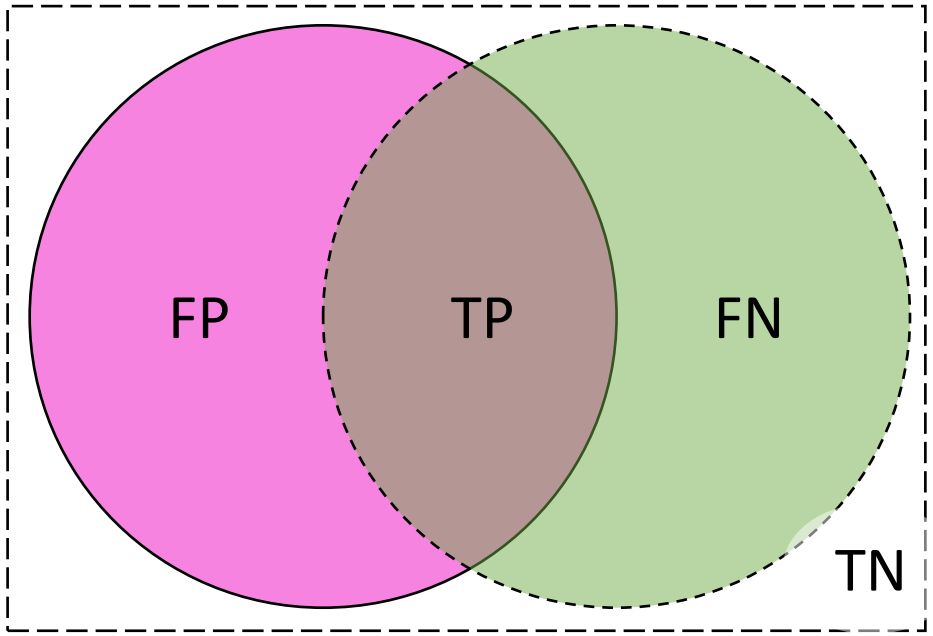


In this case
you can spare
8/9 compute time for
following processing steps

Image data source: Nasreddin Abolmaali, TU Dresden

Segmentation quality estimation

- In general
  - Define what's positive and what's negative.
  - Compare with a reference to figure out what was true and false
  - Welcome to the Theory of Sets



A — Prediction A

B — Reference B (ground truth)

ROI — Region of interest

TP — True-positive

FN — False-negative

FP — False-positive

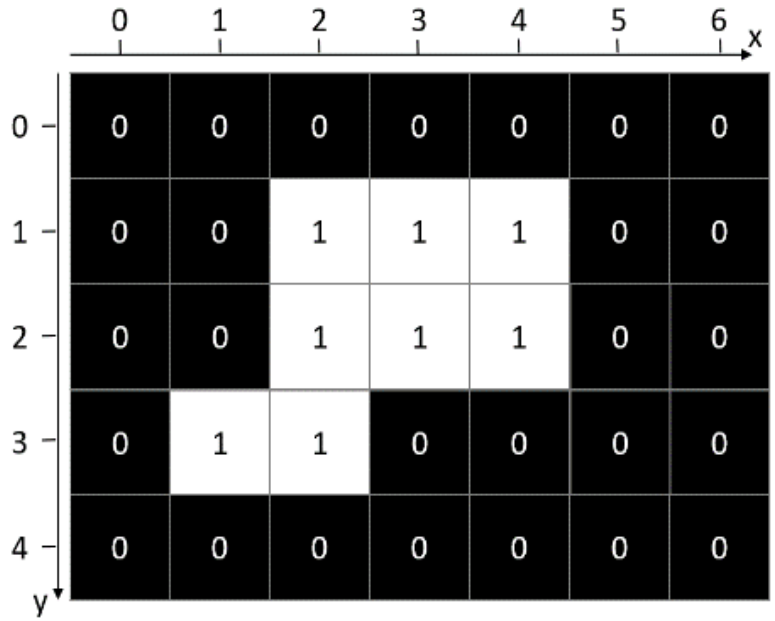
TN — True-negative

Overlap (a.k.a. Jaccard index)	$\frac{TP}{TP+FN+FP}$	How much do A and B overlap?
Precision	$\frac{TP}{TP+FP}$	What fraction of points that were predicted as positives were really positive?
Recall (a.k.a. sensitivity)	$\frac{TP}{TP+FN}$	What fraction of positives points were predicted as positives?

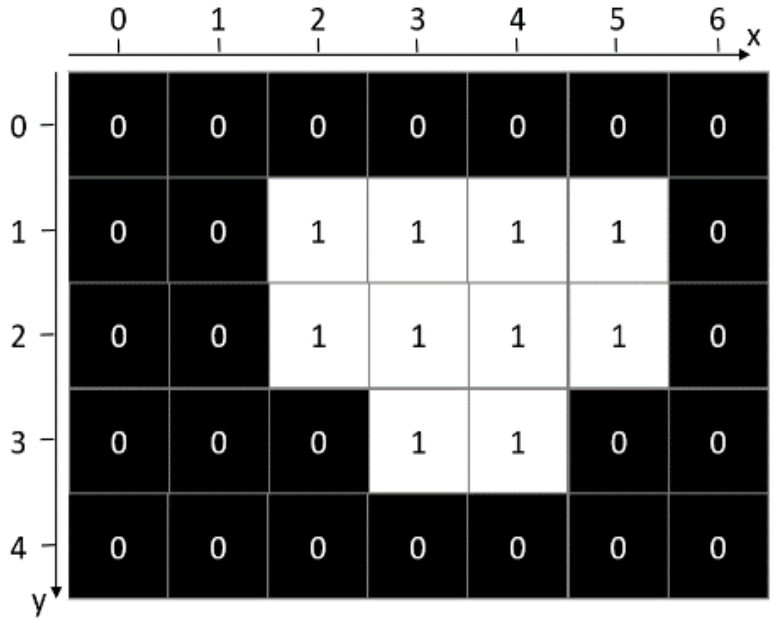
Algorithm evaluation

- Assume you are evaluating a binary segmentation algorithm by comparing its result to a given ground

Segmentation result



Ground truth



Jaccard index $\frac{TP}{TP+FN+FP}$

Precision $\frac{TP}{TP+FP}$

Recall $\frac{TP}{TP+FN}$

TP = 6
FP = 2
FN = 4
J = 6 / 12
P = 6 / 8
R = 6 / 10

PoL
Physics of Life
TU Dresden

CENTER FOR
SYSTEMS BIOLOGY
DRESDEN

Feature extraction

July 2023

@haesleinhuepf

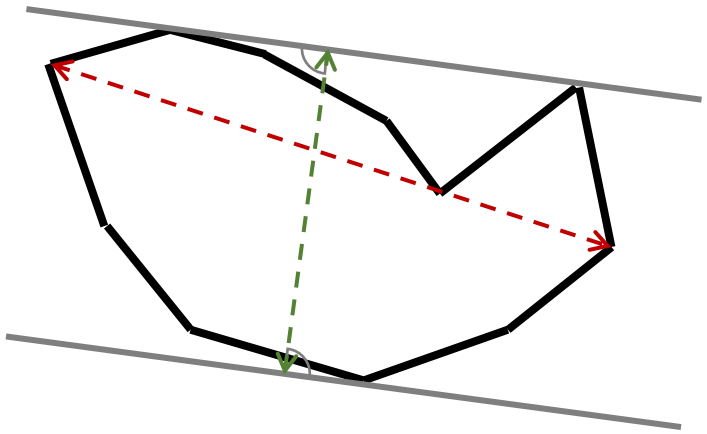
Feature extraction

- A *feature* is a countable or measurable property of an image or object.
- Goal of feature extraction is finding a minimal set of features to describe an object well enough to differentiate it from other objects.

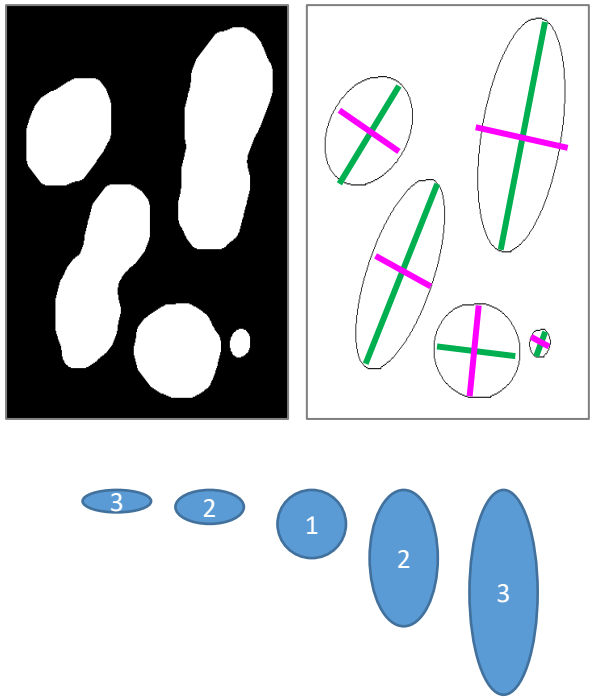
- Intensity based features
  - Mean intensity
  - Standard deviation
  - Total intensity
  - Textures
  - …
- Shape based / spatial features
  - Area / Volume
  - Roundness
  - Solidity
  - Circularity / Sphericity
  - Elongation
  - Centroid
  - Bounding box
  - …
- Spatio-temporal features
  - Displacement,
  - Speed,
  - Acceleration,
  - …
- Others
  - Overlap
  - Colocalisation
  - Network-analysis
  - …
- Mixed features
  - Center of mass
  - Local minima / maxima
  - …

Shape descriptors

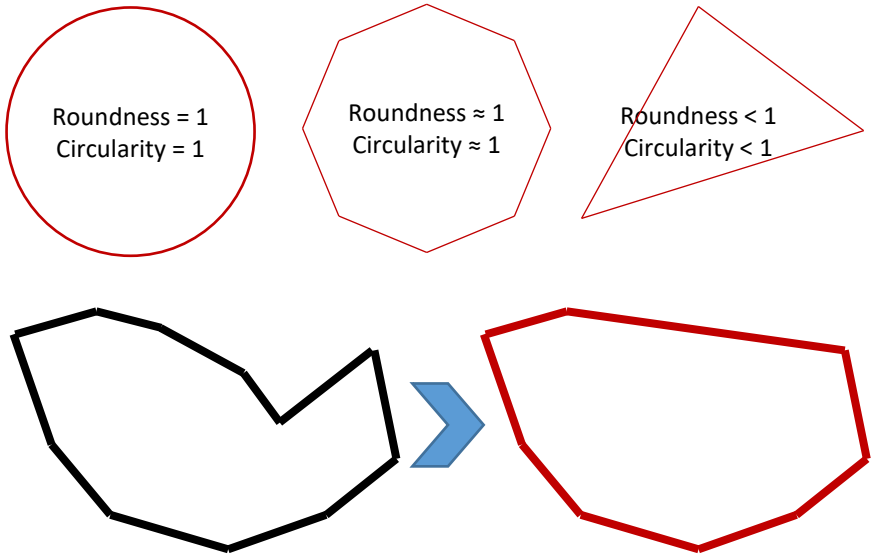
- How are objects shaped?



- Feret's diameter
- The minimum caliper ("Minimum Feret")



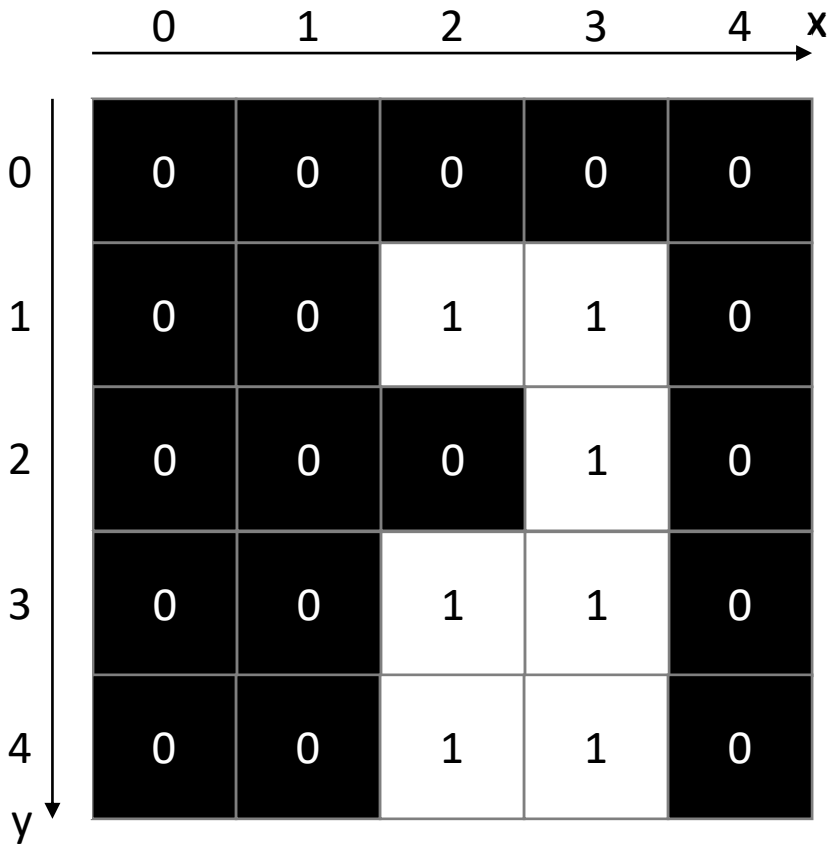
- Fit ellipse
  - Major axis … long diameter
  - Minor axis … short diameter
- Aspect ratio



- Roundness
- Circularity
- Solidity

Quiz: solidity

- What is the solidity of the white object in this image?
  - Hint: Area of the convex hull



0.875

1

5

1.2

PoL
Physics of Life
TU Dresden

CENTER FOR
SYSTEMS BIOLOGY
DRESDEN

Machine learning

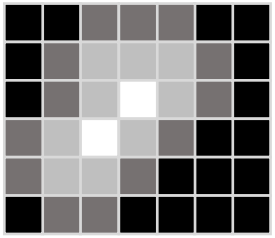
July 2022

@haesleinhuepf

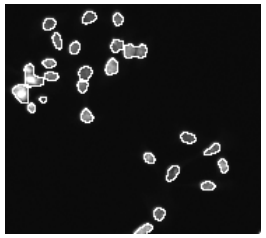
Machine learning

- Automatic construction of predictive models from given data

Pixels,

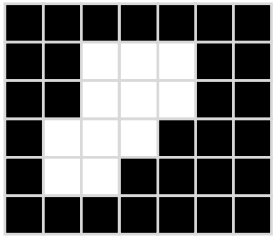


Objects,



Images, Audio, Text, Measurements, …

Dense Segmentation / Binarization



Object classification

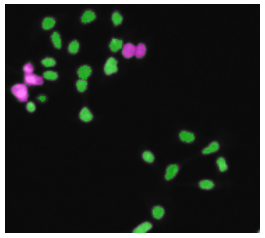
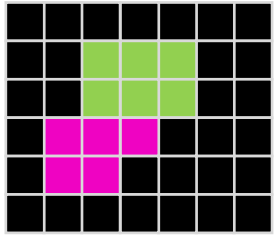


Image classification

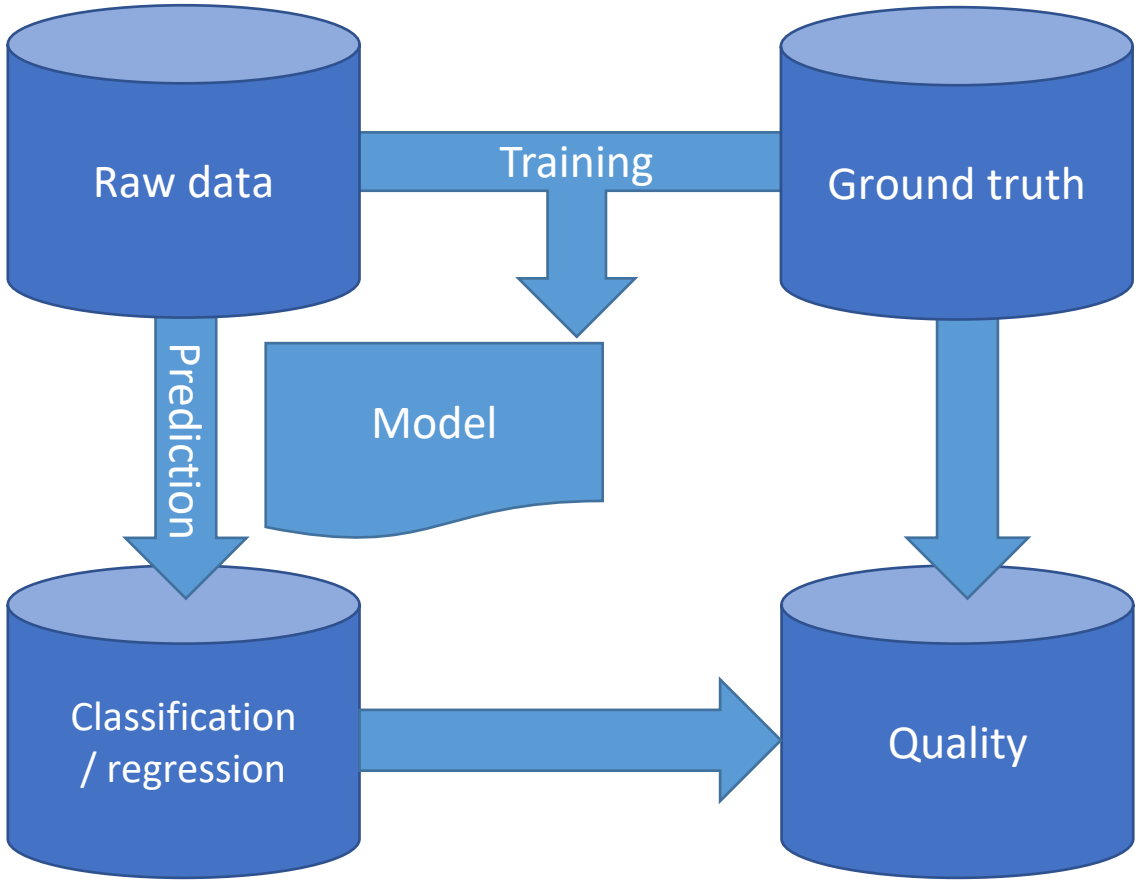
"Cat"

Instance segmentation



Cont. quantity

P_{Cat}= 0.5
$P_{Microscope}$= 0.4
Height = 80 cm

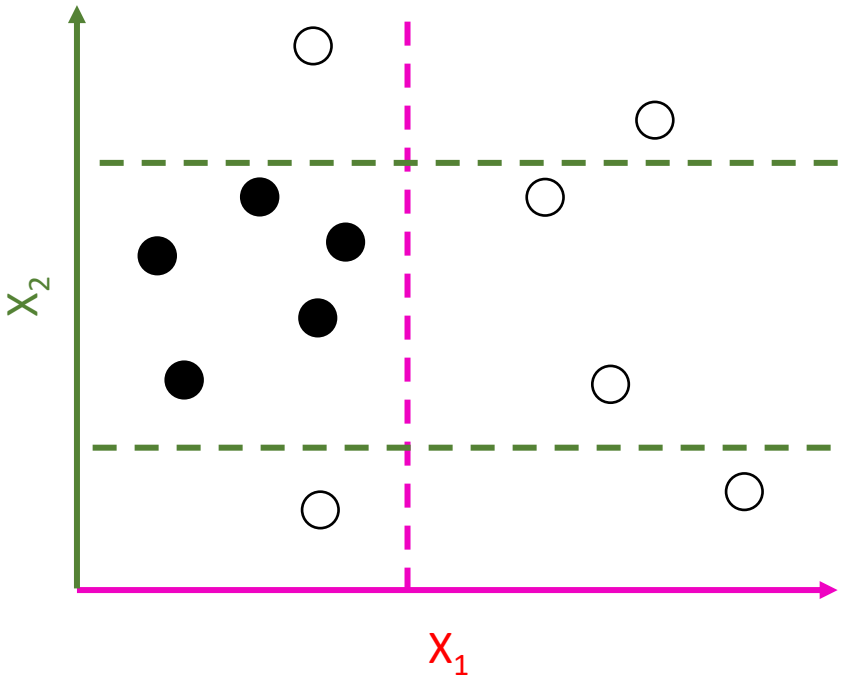


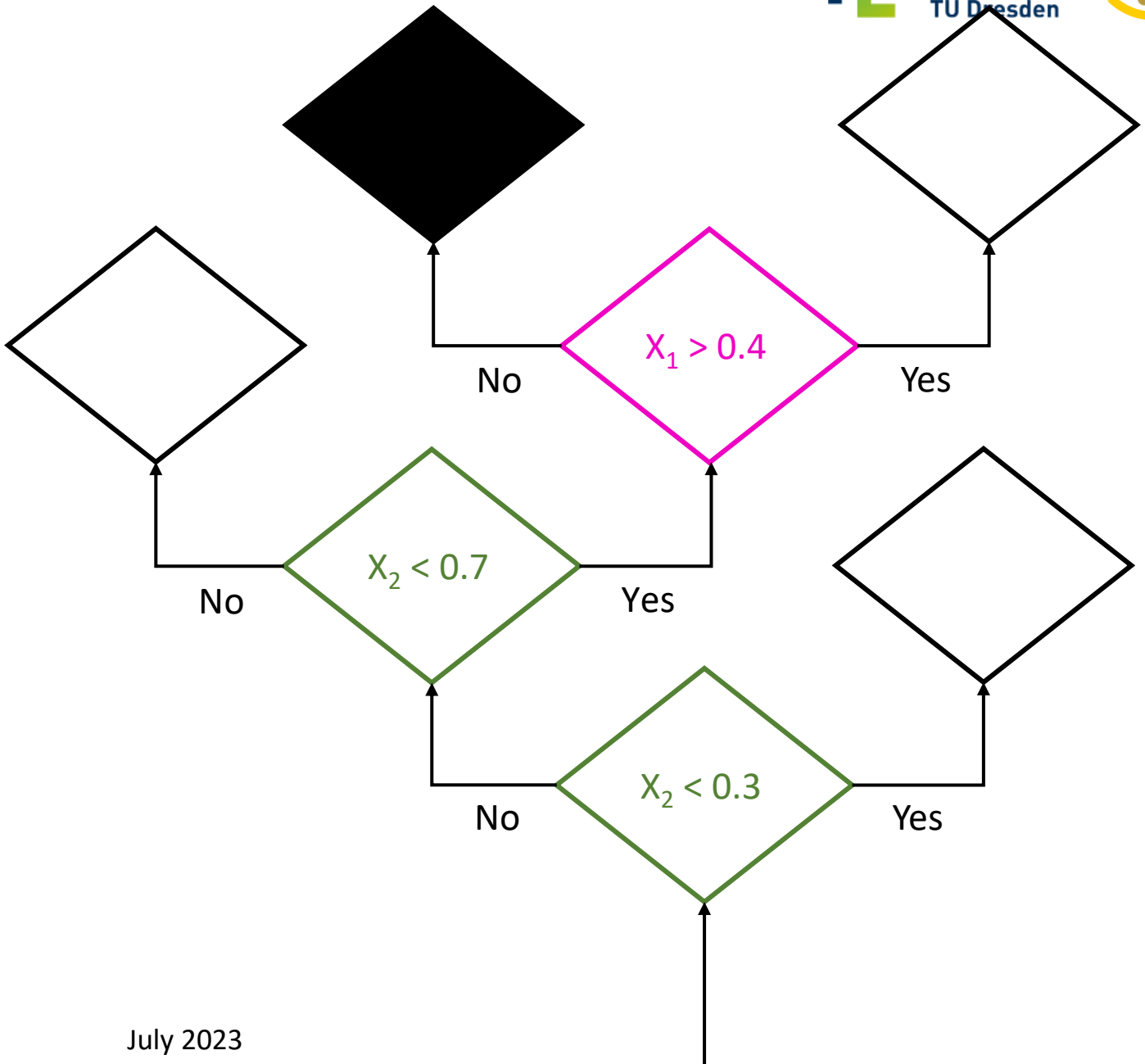
Annotated raw data, usually generated by humans

Precision, Recall

Deriving random decision trees

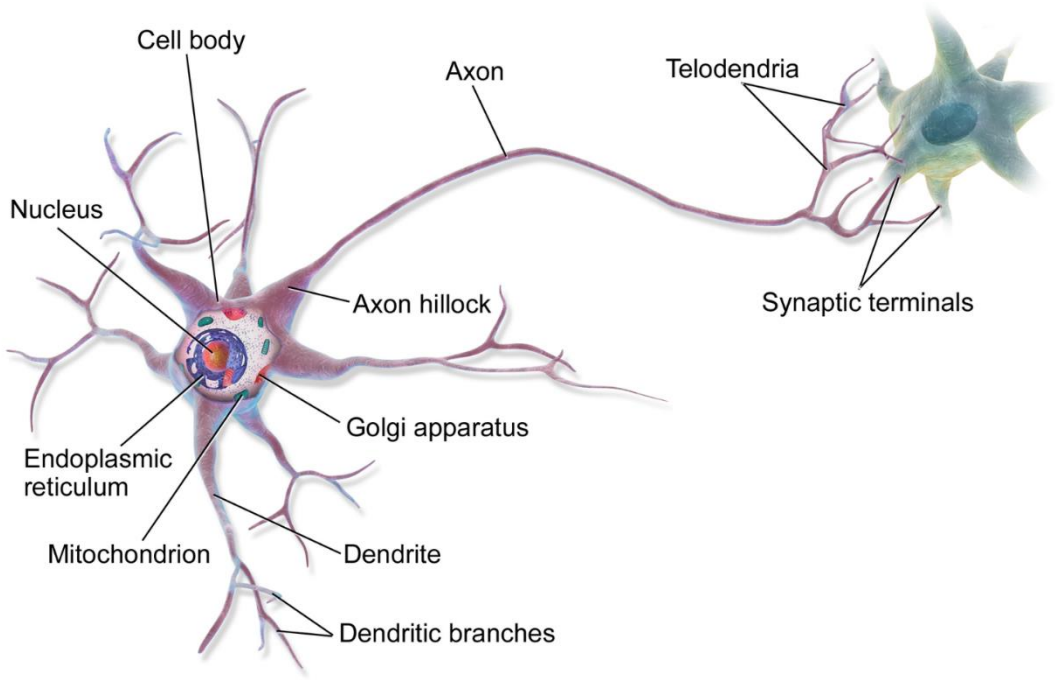
- Depending on sampling, the decision trees are different





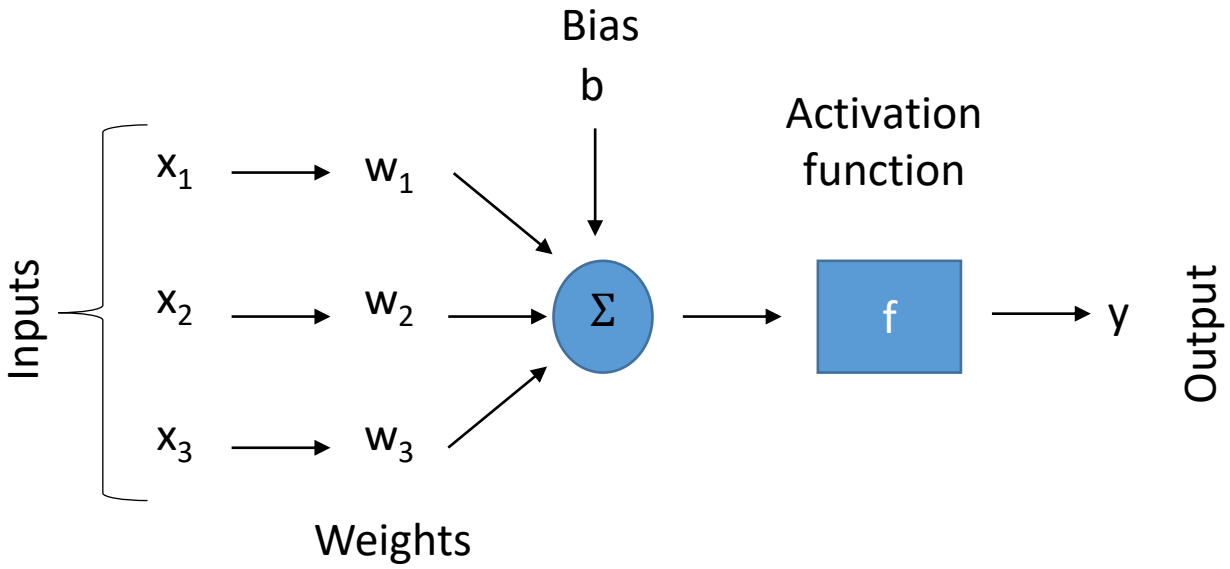
Neural networks

- How biologists see neurons



- How computer scientists see neurons

"perceptron"

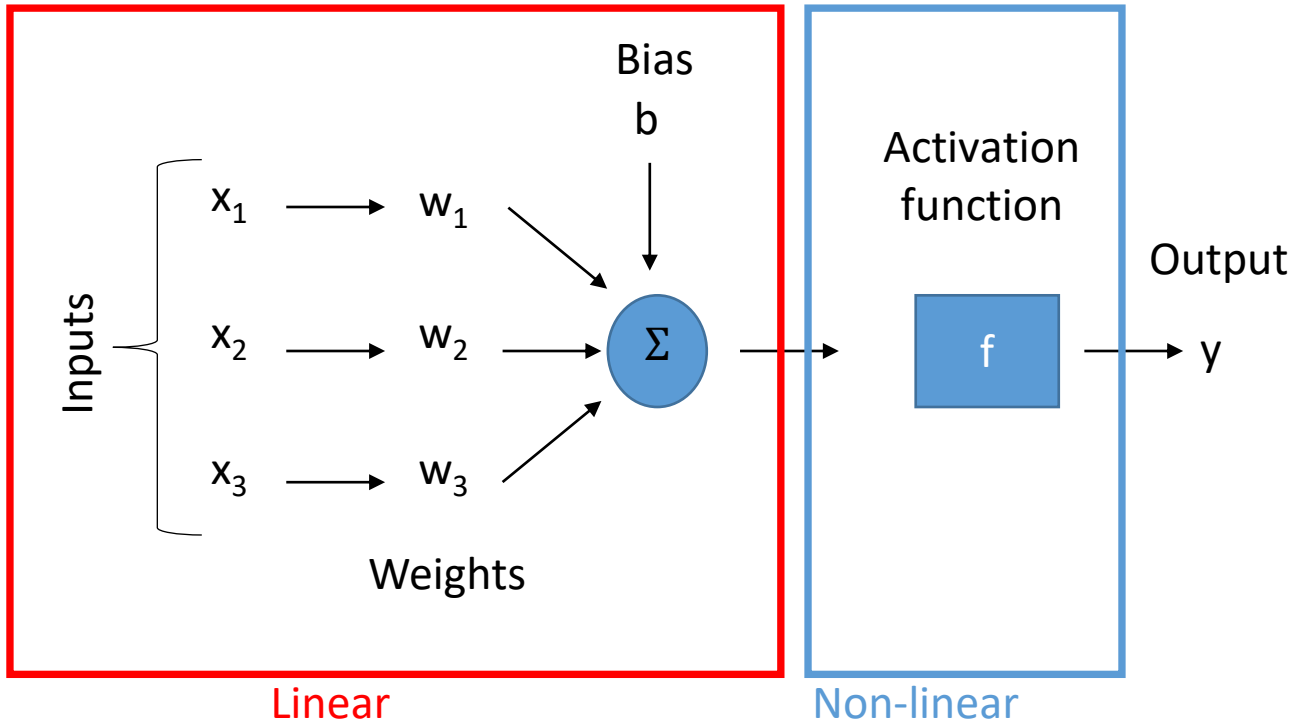


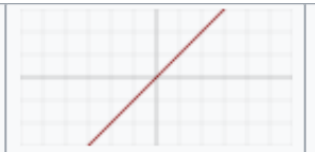
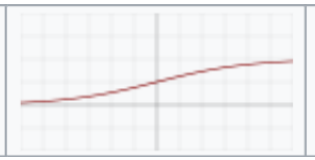
$$y = f(w_1x_1 + w_2x_2 + w_3x_3 + b)$$

Neuron image source:
https://commons.wikimedia.org/wiki/File:Blausen_0657_MultipolarNeuron.png

Activation functions

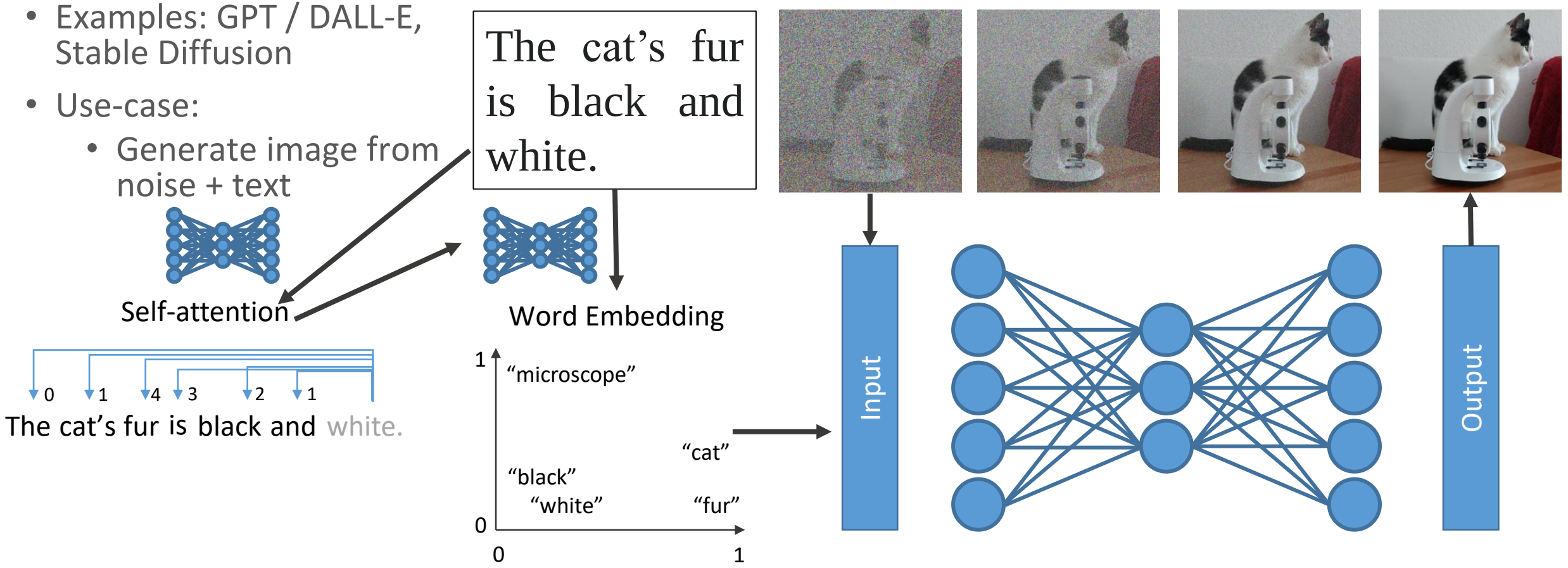
- Introduction of *non-linearity* and *activation functions* enabled what we call *deep-learning* today.



Identity		x
Binary step		$\begin{cases} 0 & \text{if } x < 0 \\ 1 & \text{if } x \geq 0 \end{cases}$
Logistic, sigmoid, or soft step		$\sigma(x) \doteq \frac{1}{1+e^{-x}}$
Rectified linear unit (ReLU)[8]		$(x)^+ \doteq \begin{cases} 0 & \text{if } x \leq 0 \\ x & \text{if } x > 0 \end{cases} = \max(0,x) = x\mathbf{1}_{x>0}$

Generative AI / Large Language Models

- Combination of neural networks + other elements + various data sources
- Examples: GPT / DALL-E, Stable Diffusion
- Use-case:
  - Generate image from noise + text



Model validation

Train dataset (e.g. 80% of the data)

- Used for training directly

Validation dataset (10% of the data)

- After every iteration see if the model overfits

Test dataset (10% of the data)

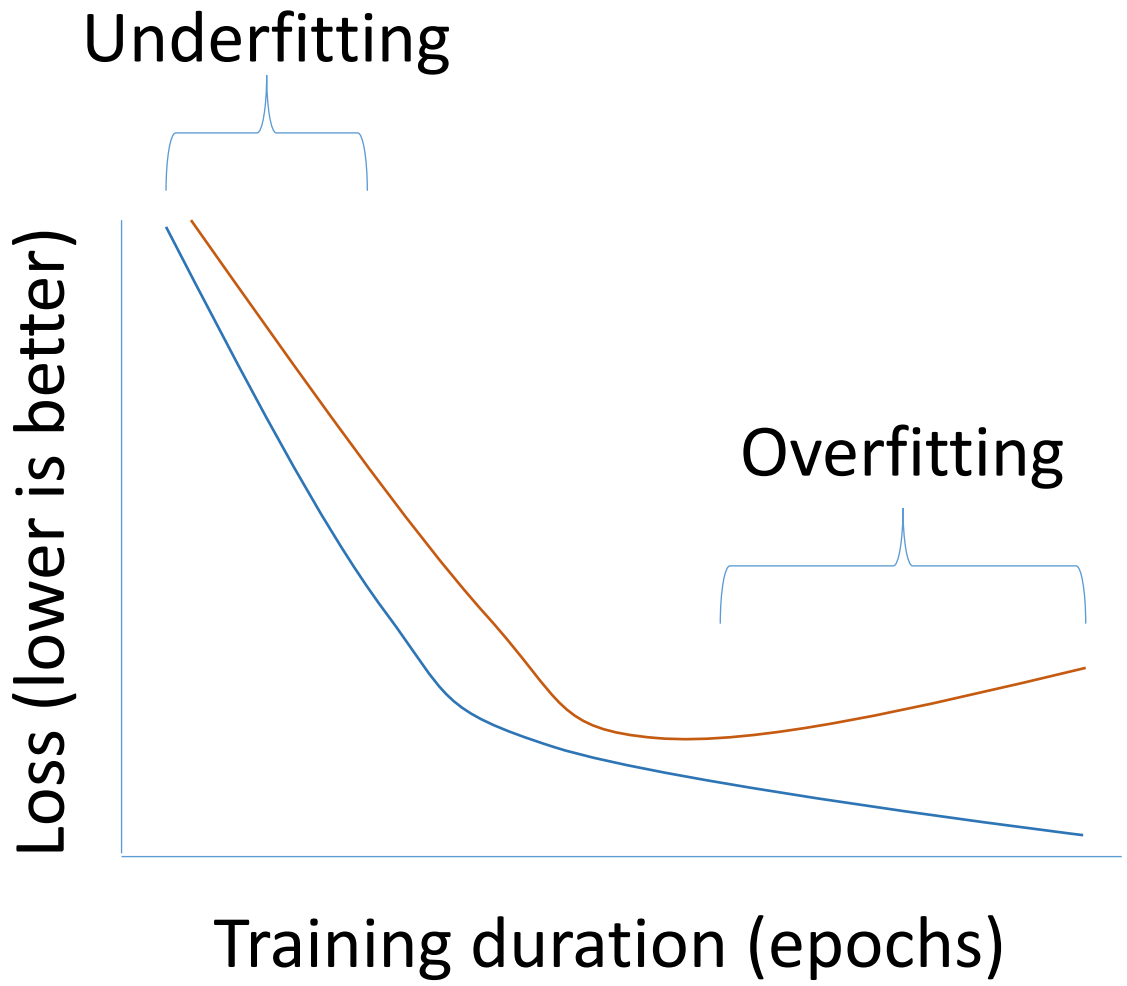
- Final evaluation after training is finished (once)

Underfitting

- A trained model that is not even able to properly process the data it was trained on

Overfitting

- A model that is able to process data it was trained on well
- It processes other data poorly



https://towardsdatascience.com/how-to-split-data-into-three-sets-train-validation-and-test-and-why-e50d22d3e54c

Generative Artificial Intelligence

- Challenges
  - Data safety / security
  - Computational cost of training neural networks
    - CO2-footprint/climate change
    - Accessibility
  - Bias: “a nice photo of a human”
  - Hallucinations
  - Glitch tokens
  - False information / fake news
  - Debugging is hard

Model: Default (GPT-3.5)

R Who developed Gradoop?

Gradoop is an open-source framework for scalable graph analytics. It was developed by the Hasso Plattner Institute (HPI) in Potsdam, Germany. The main contributors to Gradoop are researchers and developers from the HPI's Database Systems and Information Management (DIMA) group, led by Prof. Dr. Felix Naumann. The project started in 2015 as part of the HPI's research on large-scale graph data management and analysis. Since then, it has gained popularity among researchers and industry professionals working with graph data.

Sources: https://replicate.com/stability-ai/stable-diffusion
https://chat.openai.com/

PoL
Physics of Life
TU Dresden

CENTER FOR
SYSTEMS BIOLOGY
DRESDEN

Final remarks

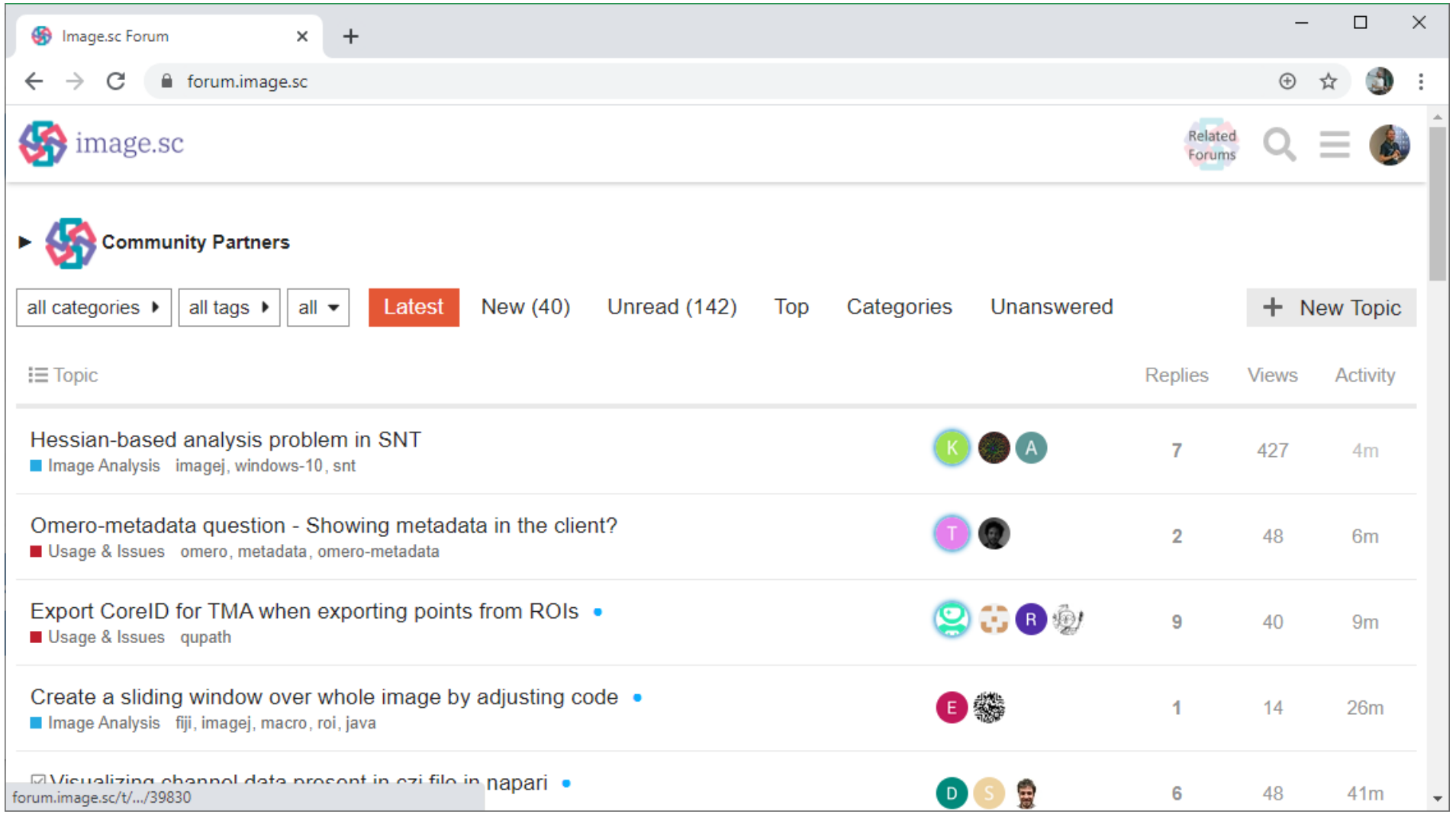
July 2023

@haesleinhuepf

ImageJ, Fiji & friends: community

PoL Physics of Life TU Dresden

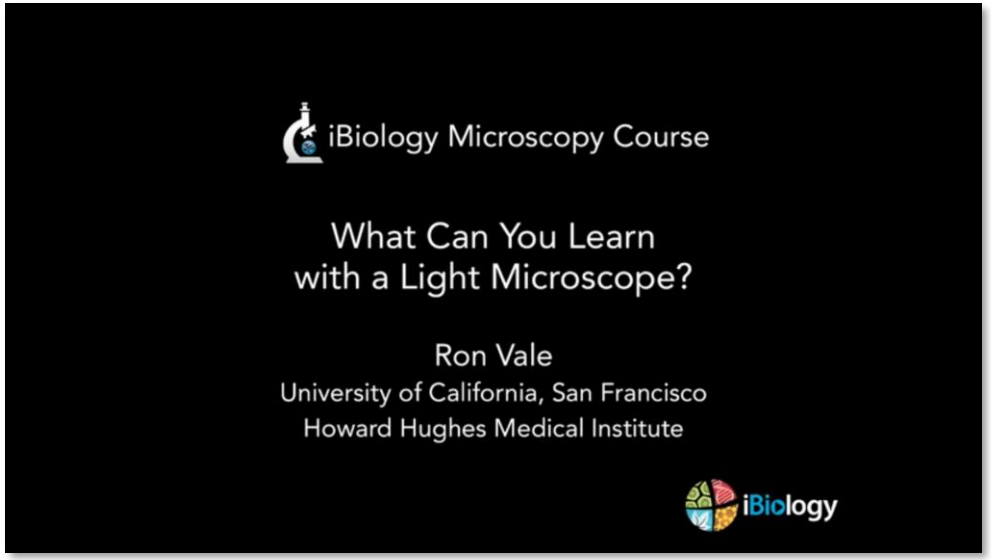
- Visit http://forum.image.sc !



What's next?

Microscopy

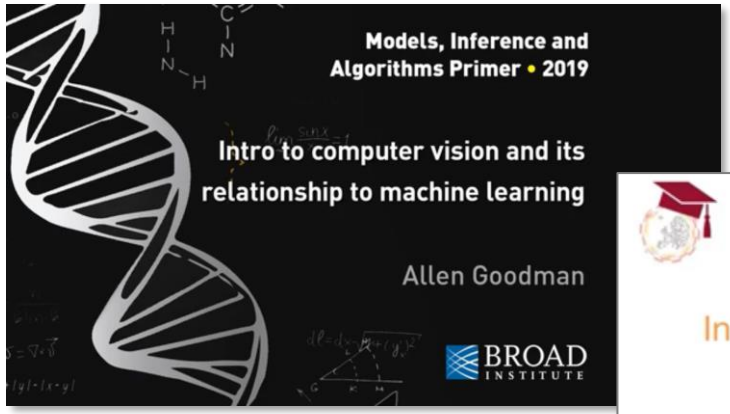
- BioDIP Dresden Light Microscopy Course: https://youtu.be/60_jgZtyR6U
- Microcourses: https://youtu.be/Tkc_GOCjx7E
- iBiology Microscopy Course: https://youtu.be/4c5ILWQmqRY

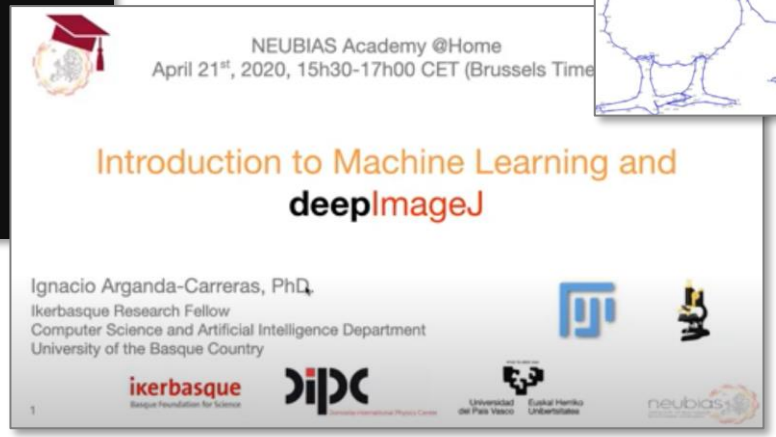


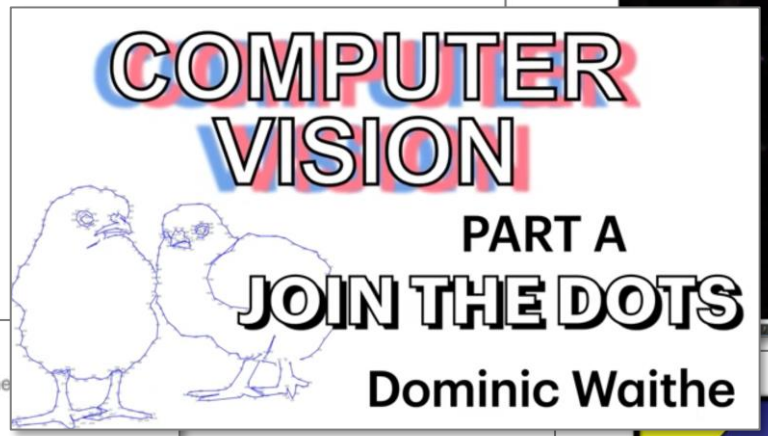
What's next?

PoL Physics of Life TU Dresden

- More machine learning for Bio-Image analysis?
  - Computer Vision / machine learning: https://youtu.be/Kzb5vTpvDBM
  - Computer vision: https://youtu.be/Smw3suzynho
  - DeepImageJ: https://youtu.be/0vTbsO8Vnuo
  - CSBDeep: https://youtu.be/ipp0mxfjhwY
  - StarDist: https://youtu.be/Amn_eHRGX5M
  - ilastik: https://www.youtube.com/ilastikTeam

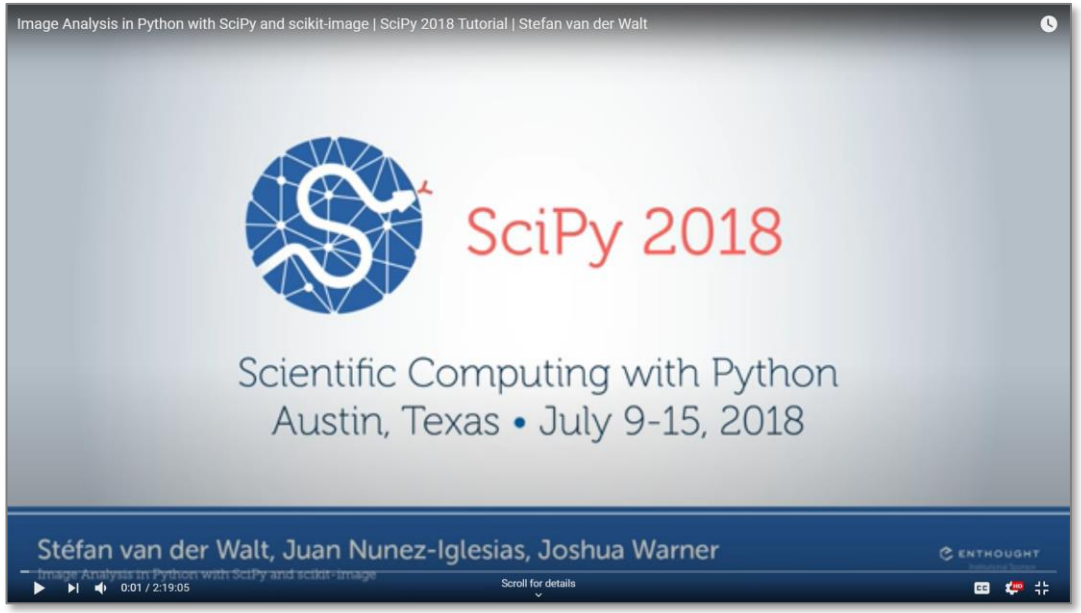


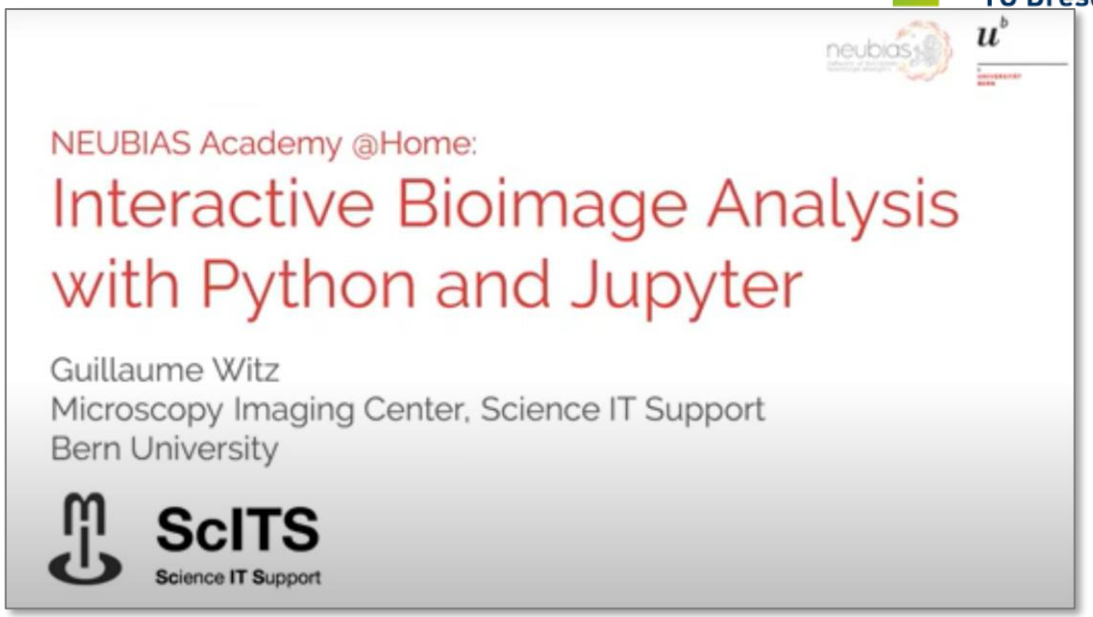




What's next?

- Image Analysis with Python
  - Python & Jupyter
    - https://youtu.be/2KF8vBrp3Zw
    - https://youtu.be/Y3pB3wnOivE
  - Scikit-image
    - https://youtu.be/pZATswy_IsQ
    - https://youtu.be/d1CIV9irQAY
  - Napari
    - https://youtu.be/VgvDSq5aCDQ



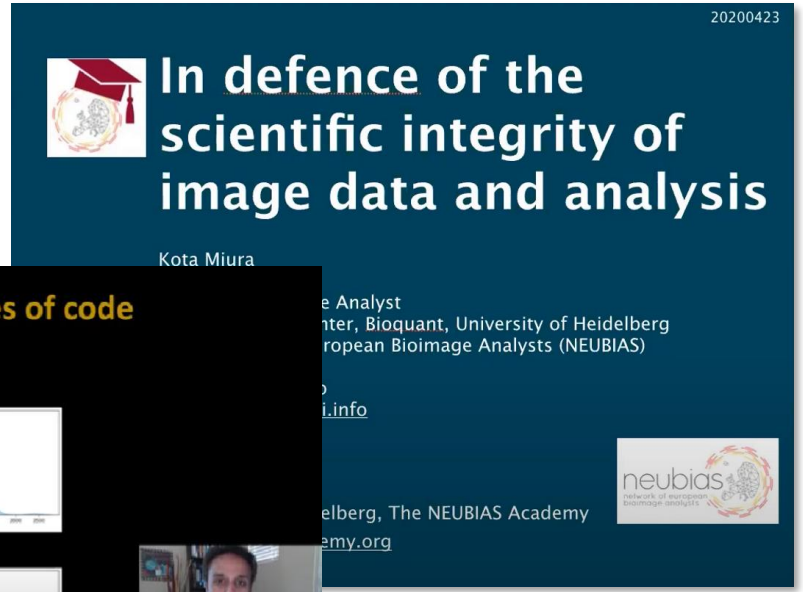


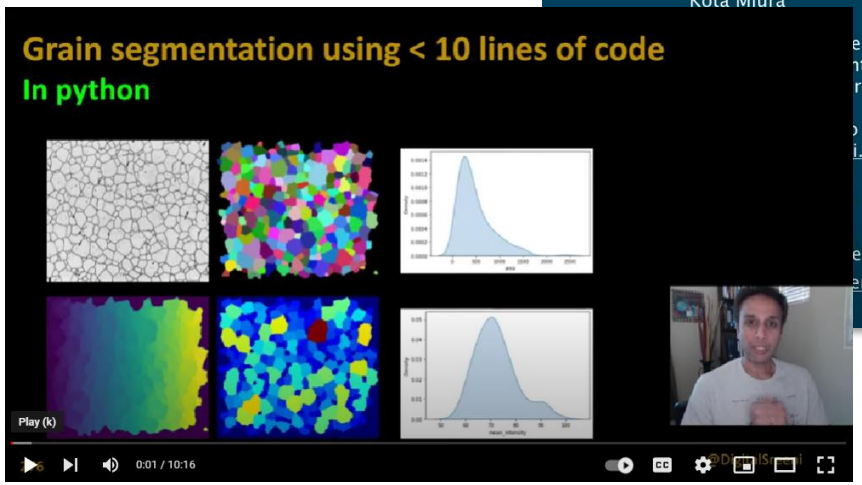
What's next?

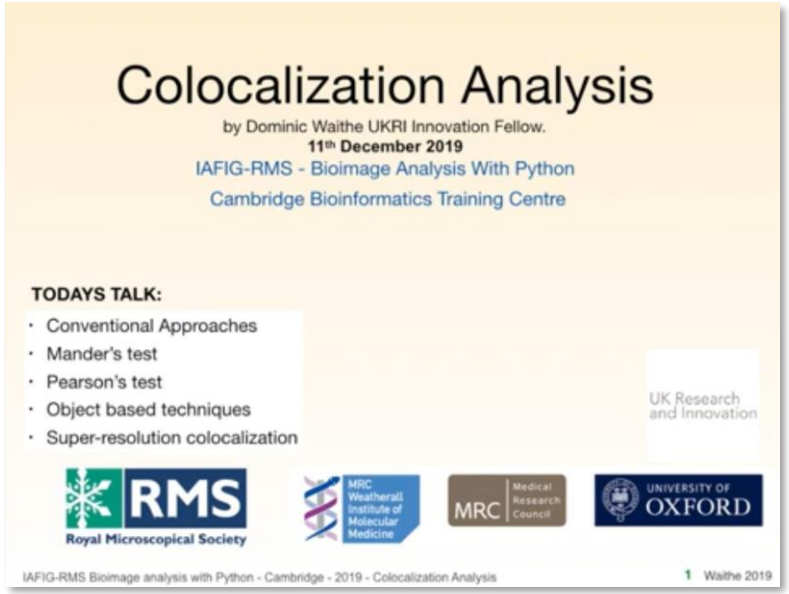
PoL Physics of Life TU Dresden

More general

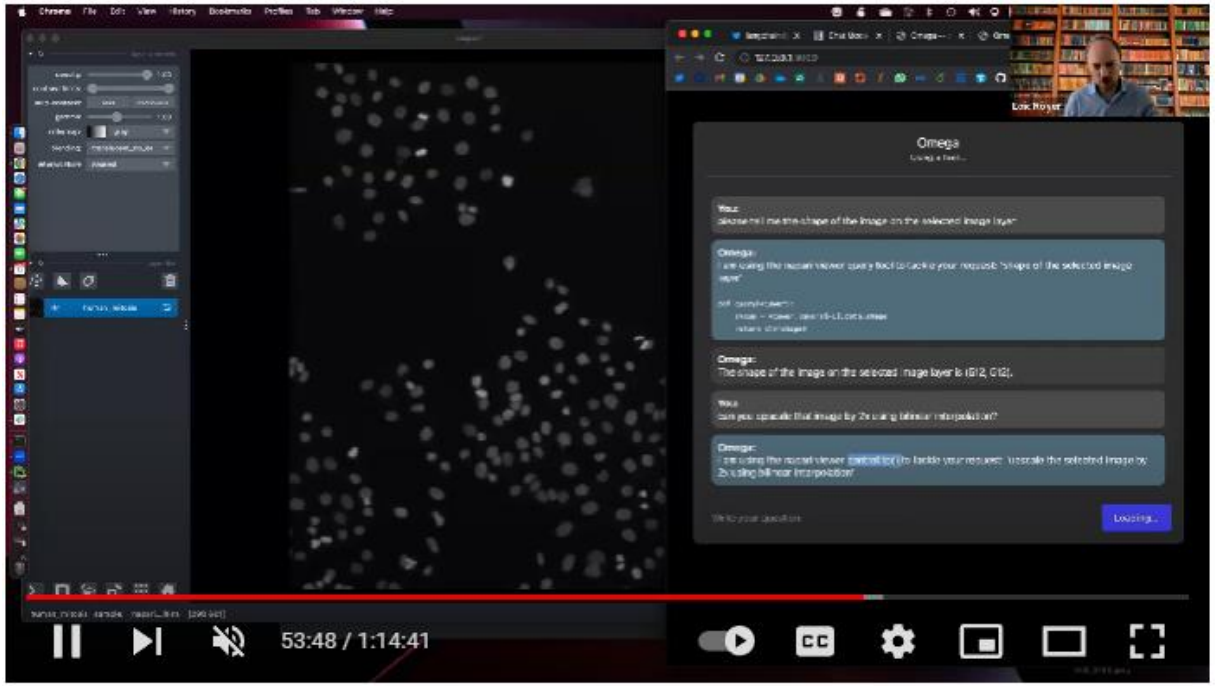
- Python for Microscopists + beyond: https://youtube.com/digitalsreeni
- Image data integrity: https://youtu.be/c_Oi2HKom_Y
- Coloc: https://youtu.be/cOrCz4qc8DI
- Automated microscopy: https://youtu.be/w0ERCrKx4gk







Further reading / watching

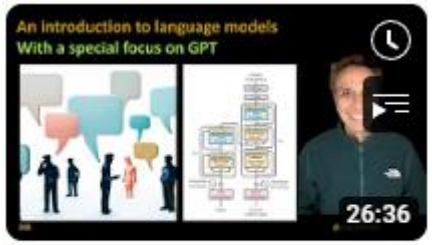


Code review of Napari-ChatGPT by Loic Royer (CZI Biohub)

Scientific Computing ...
122 subscribers
Subscribe
6
Share

https://www.youtube.com/watch?v=JMo6Sn-L_j4

DigitalSreeni
@DigitalSreeni 71.3K subscribers 380 videos
This channel walks you through the entire process of learning to code in P...



308 - An introduction to language models with focu...

311 - Fine tuning GPT2 using custom documents

309 - Training your own Chatbot using GPT

https://www.youtube.com/c/digitalsreeni

Picture source: https://twitter.com/digitalsreeni/status/1541578740584415233