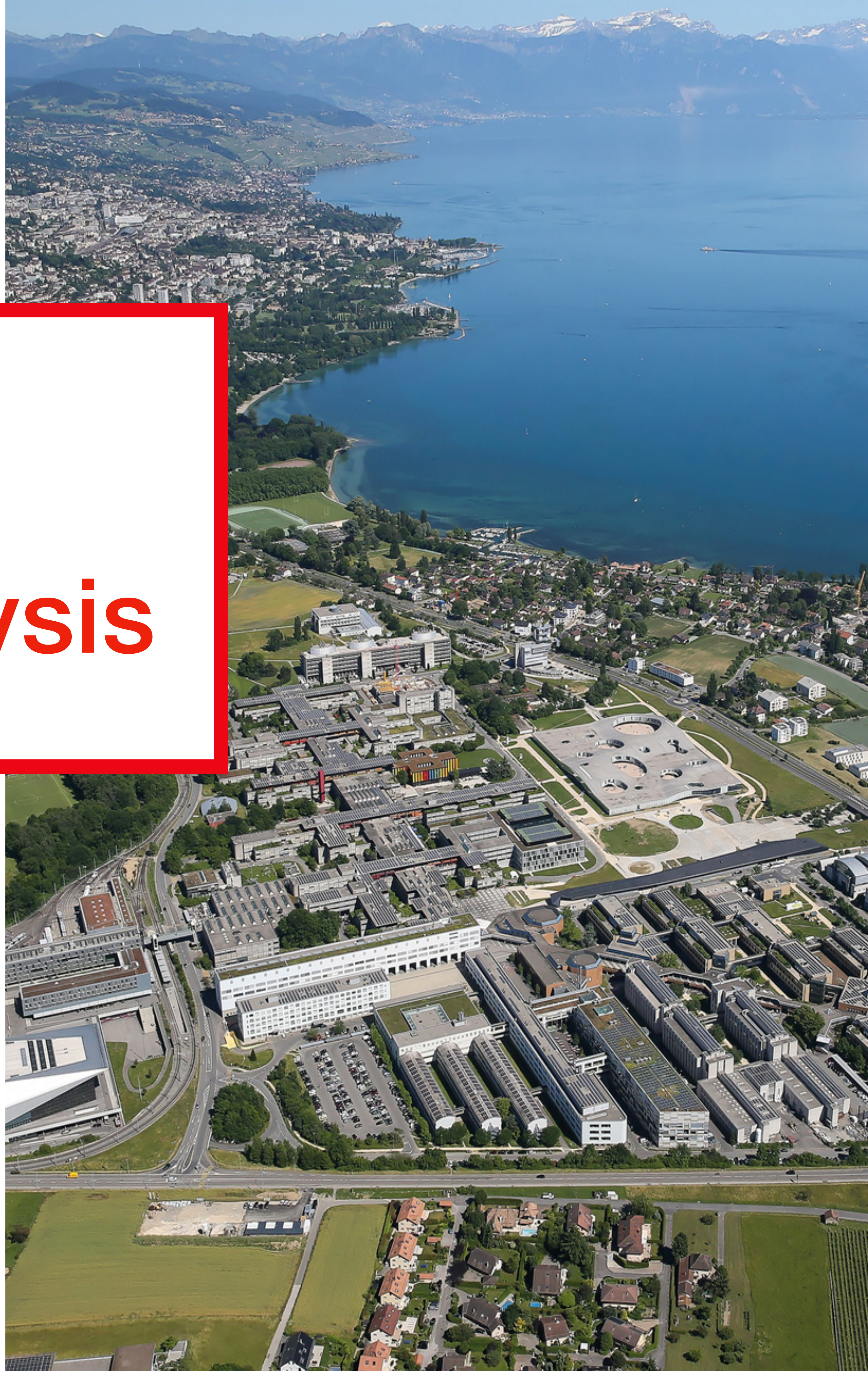
Introduction to Microscopy Image Analysis



Daniel Sage

Biomedical Imaging Group
EPFL Center for Imaging
EPFL — Switzerland
daniel.sage@epfl.ch

New microscope new device

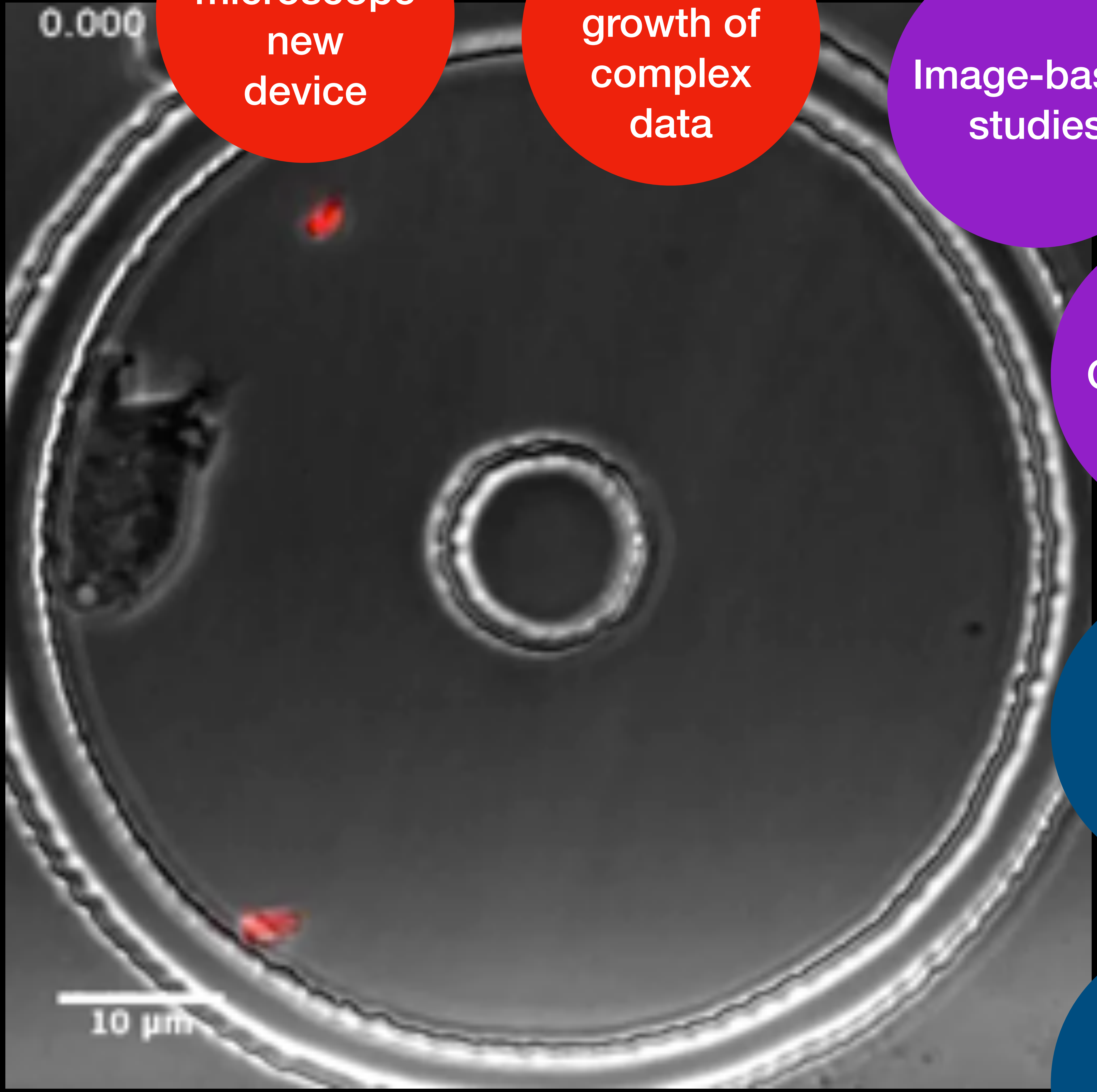
Explosion growth of complex data

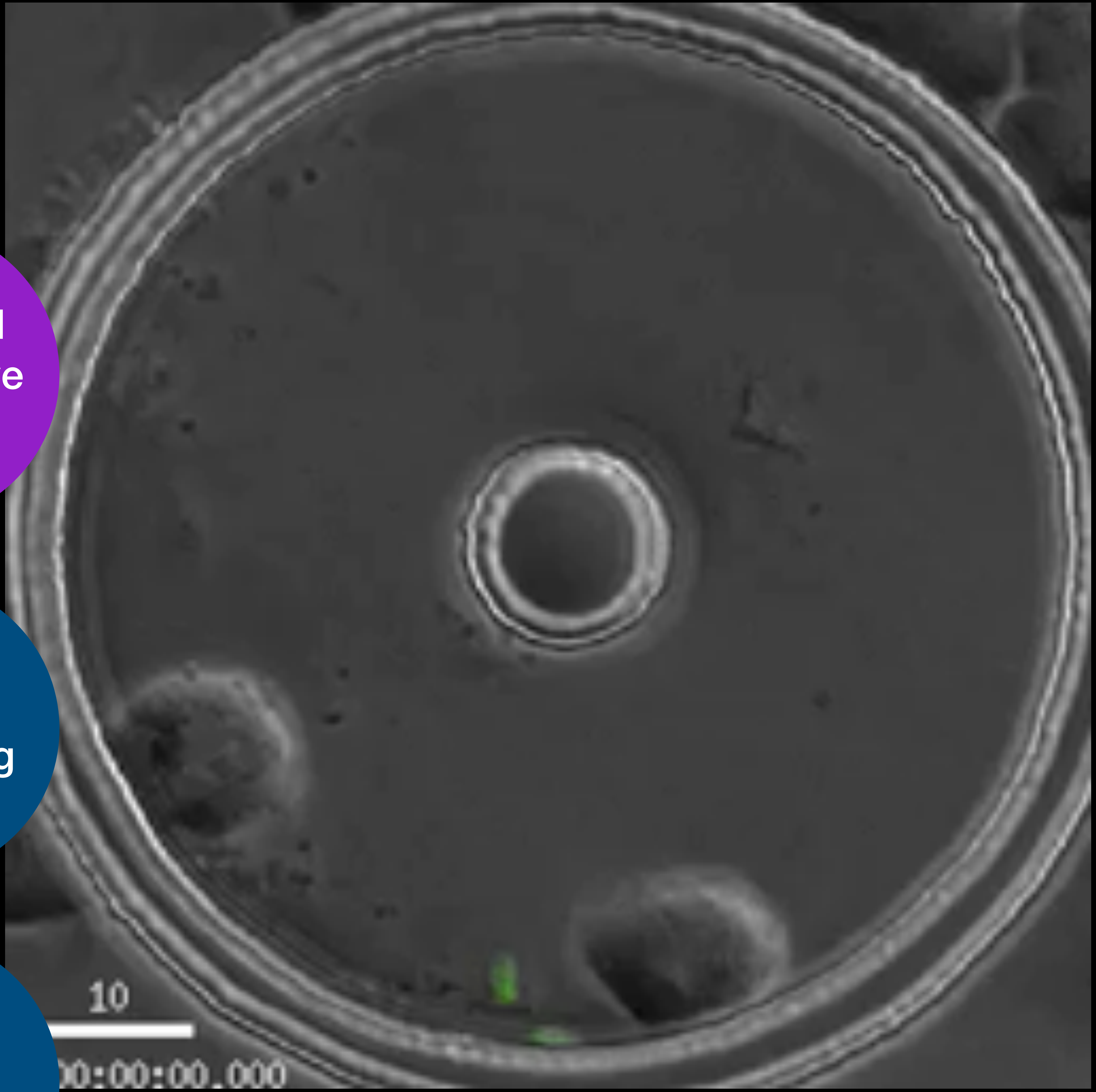
Image-based studies

Numerical Quantitative Objective

Image Processing

Machine Learning





Cell bacteria interaction, phase contrast and fluorescence channel for bacteria
Source: Matthieu Delincé and Chiara Toniolo, McKinney Lab, EPFL

Introduction to Microscopy Image Analysis

Lecture for the workshop AI4Life given by Daniel Sage, 10 June 2024

CONTEXT — BIOIMAGE INFORMATICS

METHODS — MODEL-BASED VS. DATA-DRIVEN

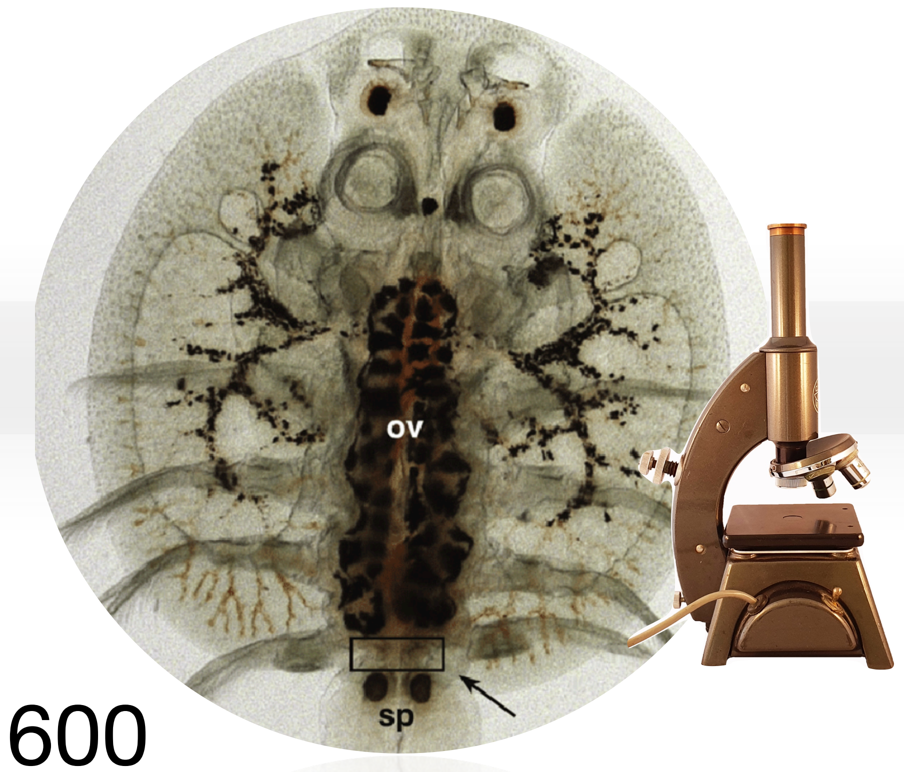
LEARNING — DATA, MODELS AND TOOLS

WRAP UP — BIOIMAGE ANALYSIS

daniel.sage@epfl.ch

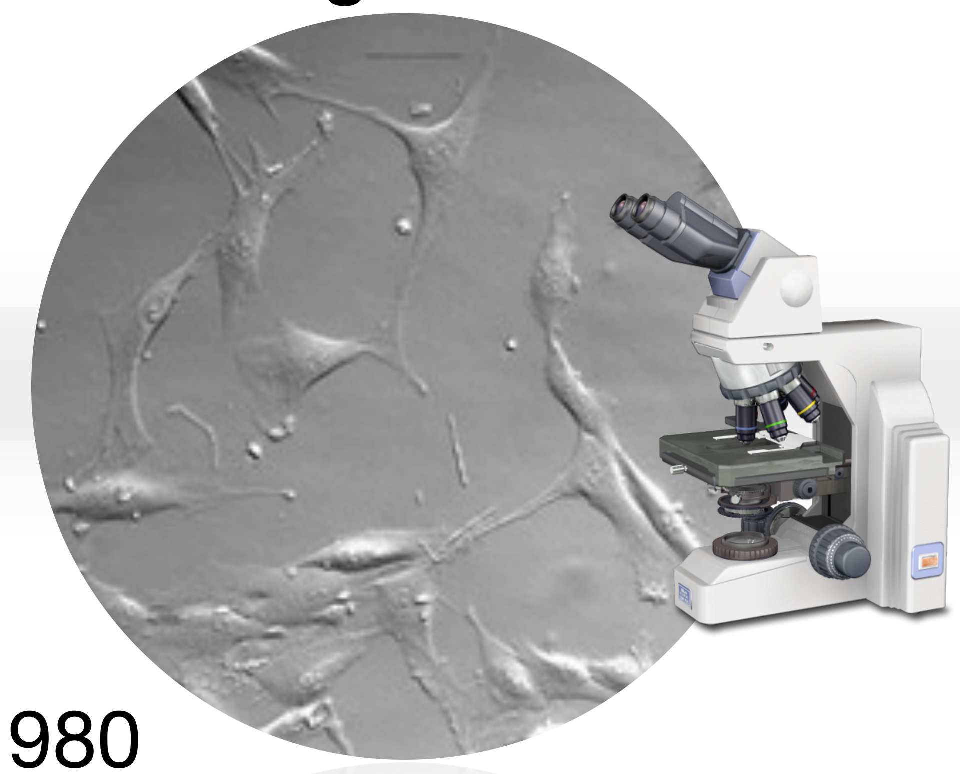
From Qualitative to Quantitative Microscopy

Observation



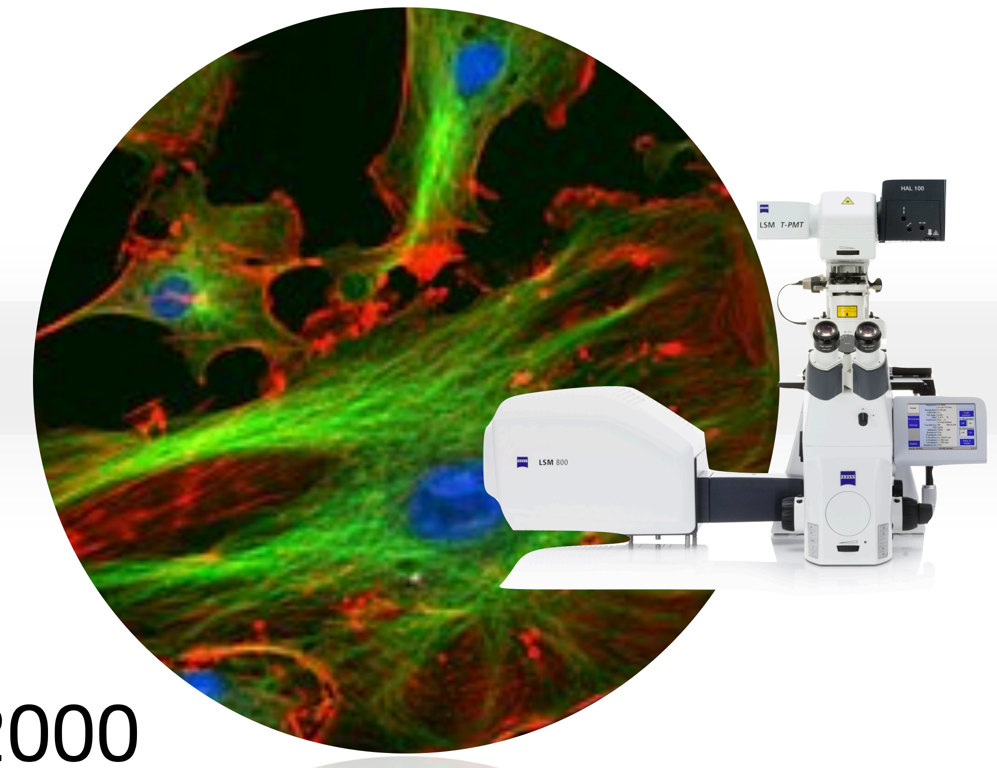
~1600

Digital era



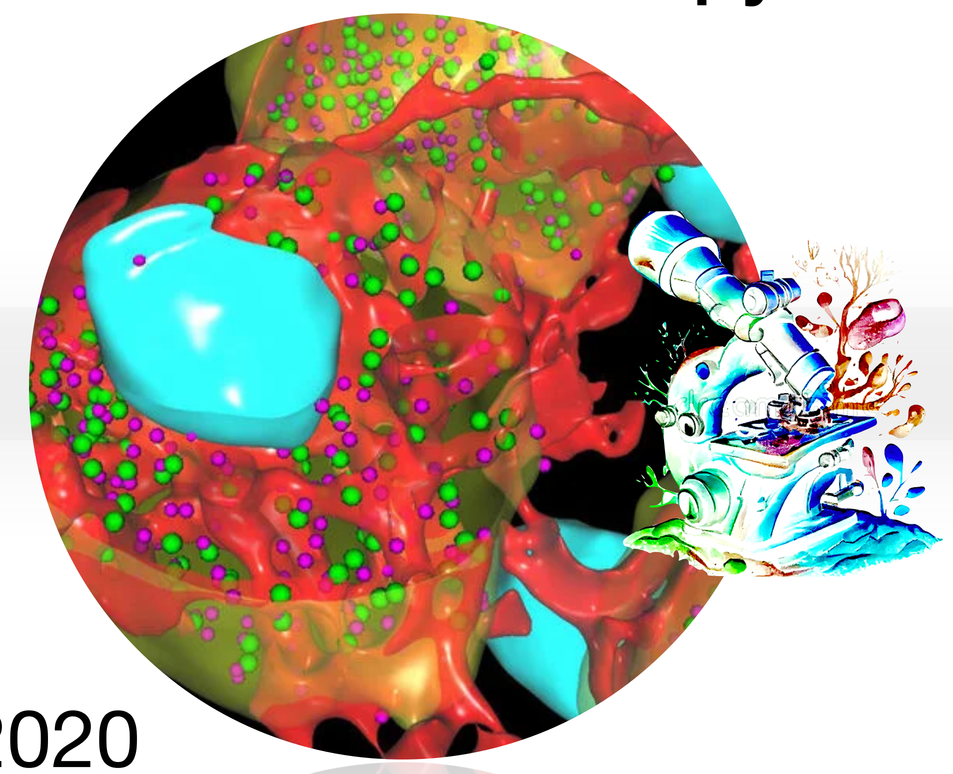
~1980

Modern time



~2000

Smart microscopy



~2020

Biochemistry
fluorescence
single-molecule

Data
data analysis
machine learning

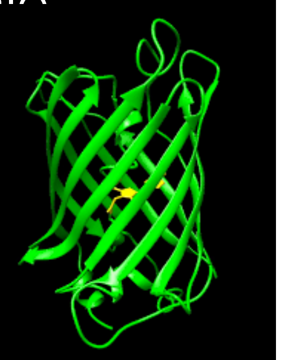
Series of (r)evolutions

Microscopy
high resolution
time-lapse

Digital
digital camera
storage

Fluorescence Microscopy

- High SNR
- Quantitative
- Specific to subcellular component
- Structural, functional and *in vivo*
- Interactions at molecular scale
- Diffraction-limited
- Autofluorescence
- Photobleaching
- Toxicity



Bioimaging has evolved from its **qualitative** origins, mainly focusing on the visualization of individual biological processes and structures, to a highly **quantitative** discipline, producing **large, multimodal and complex** datasets requiring sophisticated and robust computation and methods for analysis.

A. Mather, Building a FAIR ..., 2023

Imaging Disciplines

- **Digital Imaging**
  Image acquisition, analysis
- **Image Processing**
  Transformation of images
- **Image Analysis**
  Extraction of information
- **Computer Vision**
  High-level understanding
- **Computer Graphics**
  Synthesizing visual content

Image reconstruction
data, measure, signals ➤ image

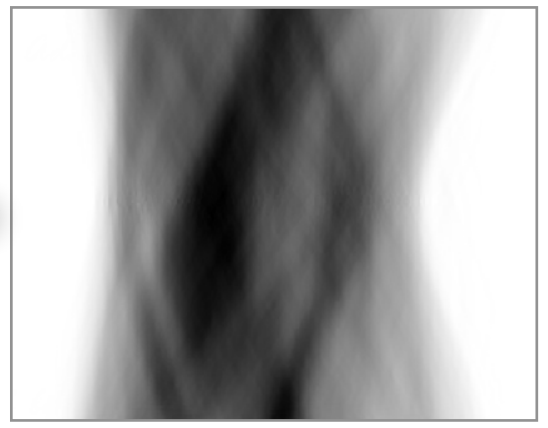



Image processing
raw image ➤ enhanced image

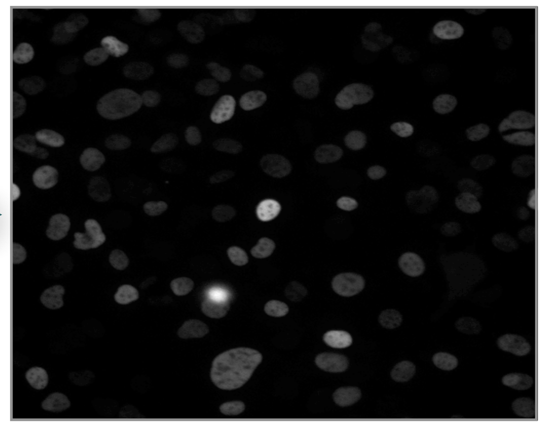
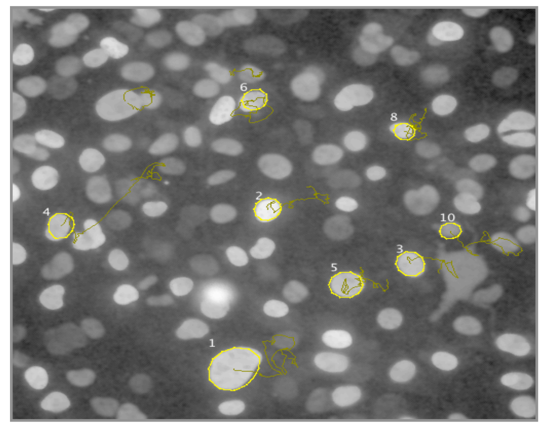
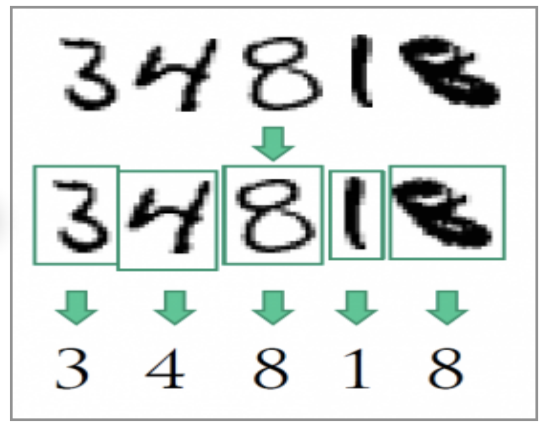
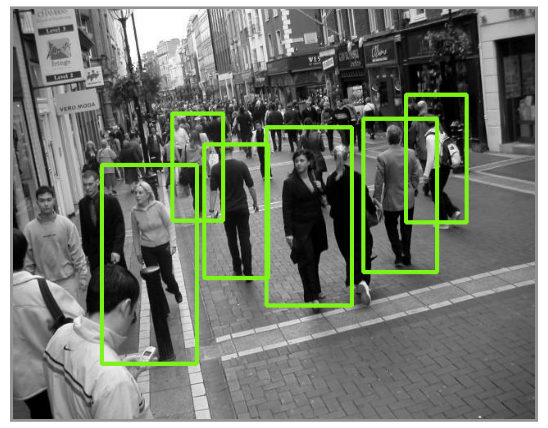



Image analysis
image ➤ number, description

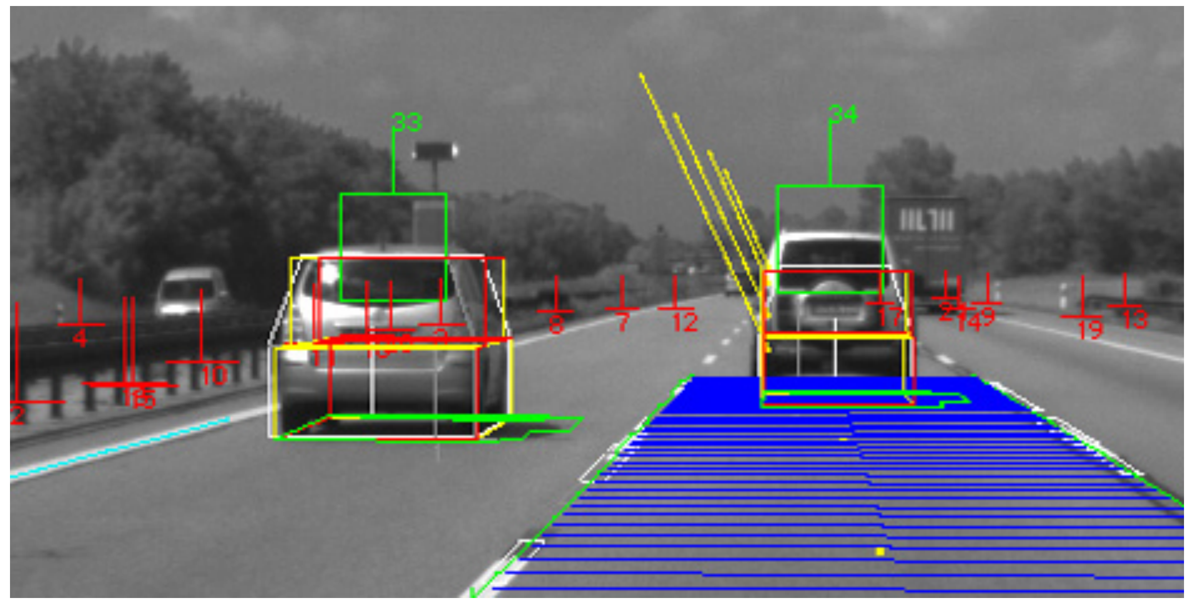




Computational Bioimaging

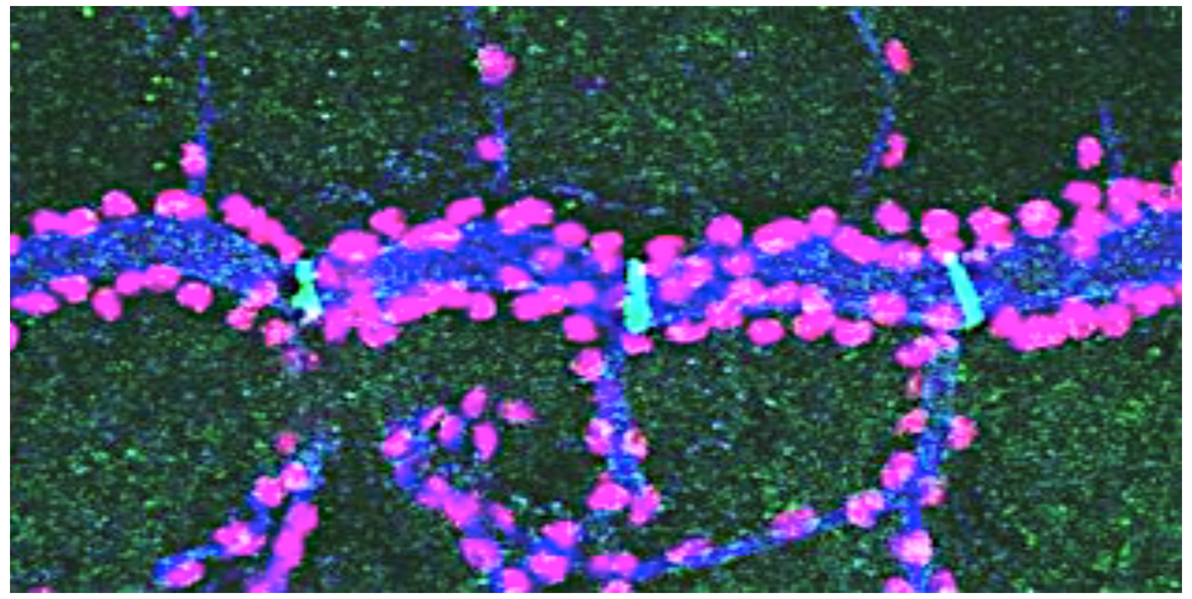
Bioimage Informatics

Computer Vision



- Human perception modeling
- Real-time processing
- Well defined shape and geometry

Bioimage Analysis



- Large number of simple particles
- Live cell / interaction / motility
- Multiple dimensions

Common Questions

Typical biological questions

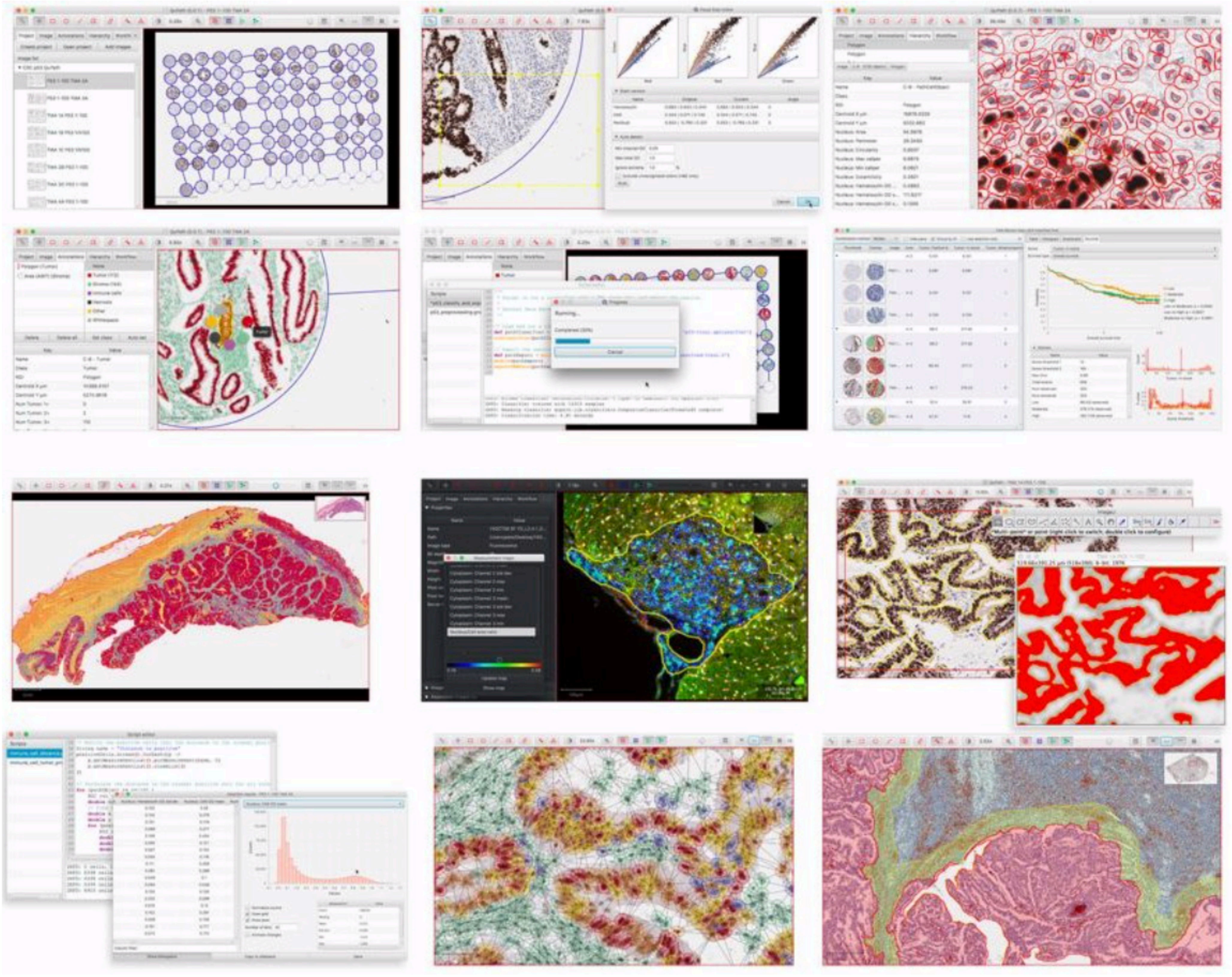
- Count cells, classify cells
- Measure the expression of protein (fluorescence)
- Mesure the granularity
- Identify the interaction
- Record the diffusion, growth rate
- Build the trace in time, cell lineage

Expectations of a bioimage analysis system

- Quantitative, objective
- Parameter-free, no human bias
- Reliable, confident, robust to input variation
- Reproducible
- Computing: reasonable fast, laptop

➡ **Do not expect a 100% correct answer in real life**

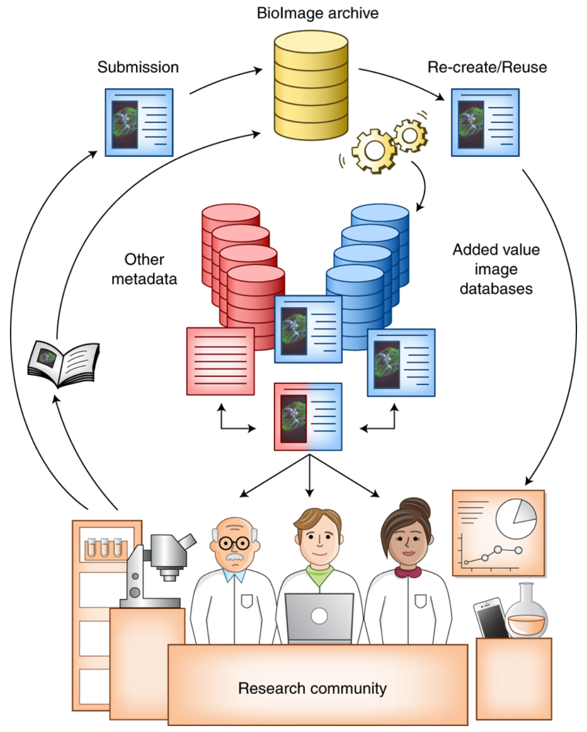
Illustration from QuPath website



Bioimage Datasets

Databases

- Data and metadata
- Annotated data for ML
- Open, re-analysable
- Centralised repositories



A call for public archives for biological image data
Jan Ellenberg et al.
Nature Methods 2018

Building a FAIR image data ecosystem for microscopy communities
Aastha Mathur et al.
J. Histoch Cell Biol 2023

Large Data

🔴 Storage

⏱ Transfer

✗ Share and annotation

⏱ Simple processing

⏱ Visualization

Public repositories of bioimages

BROAD INSTITUTE https://data.broadinstitute.org/bbbc/

CELL IMAGE LIBRARY http://www.cellimagelibrary.org/

IDR http://idr.openmicroscopy.org/
Williams, Nat Methods 2017

EMPIAR https://www.ebi.ac.uk/pdbe/emdb/empiar
Iudin, EMPIAR. Nucleic Acids Res 2023

CANCER IMAGING ARCHIVE https://www.cancerimagingarchive.net
Clark, J Digital Imaging 2013

BioImage Archive https://www.ebi.ac.uk/bioimage-archive/
Hartley, Molecular Biology 2022

Pre-trained DL models

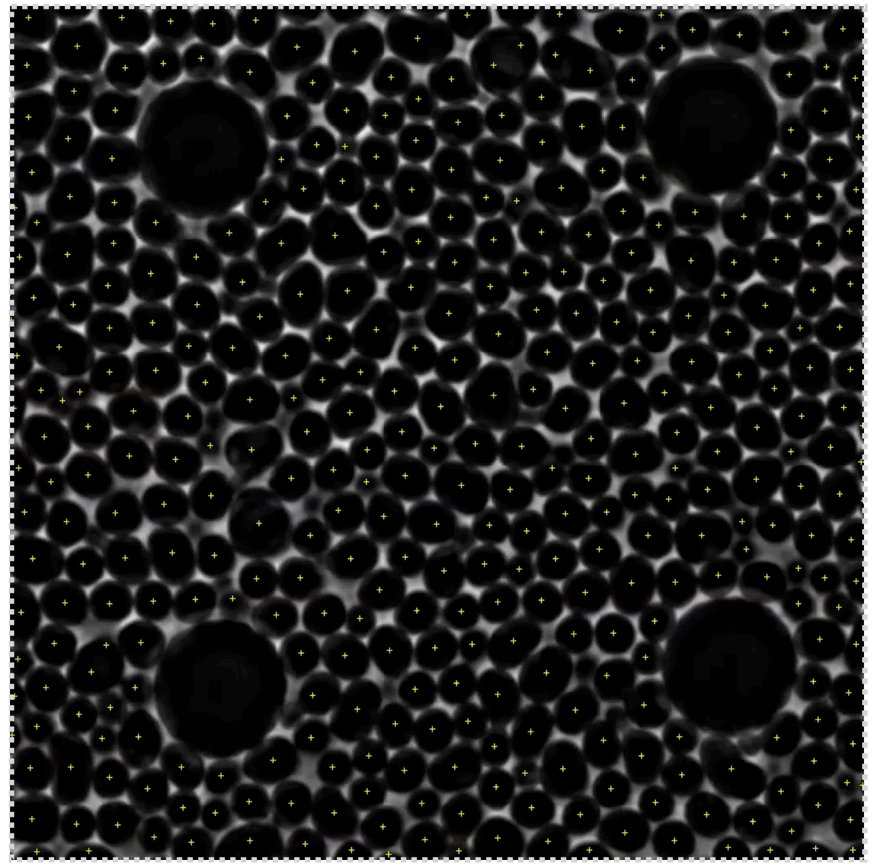
http://bioimage.org/
Trained Models Zoo

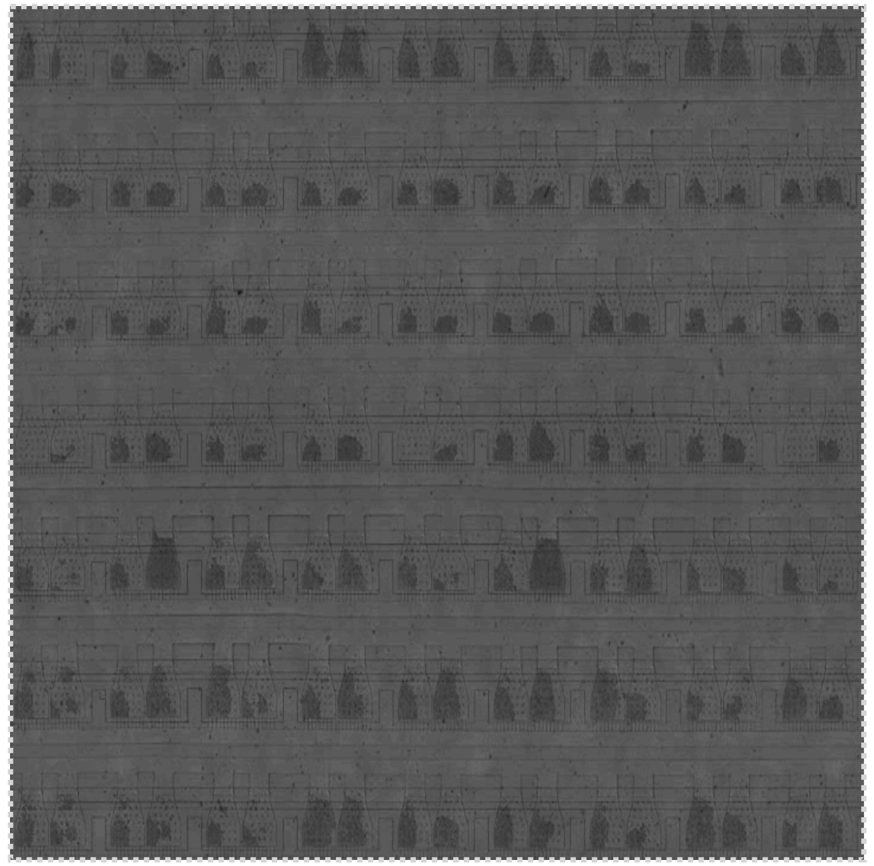
http://cellpose.org/
Cellpose for Cell Segmentation

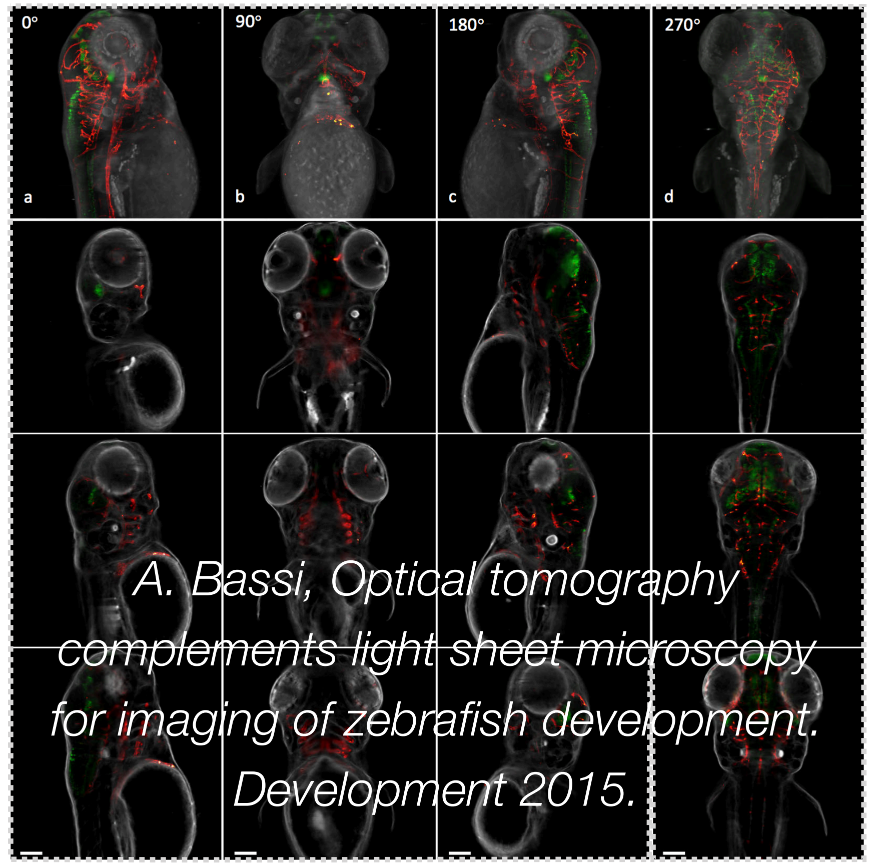
Bioimage Informatics

Bioimage informatics is a key technology for Life Science

- Quantitative discipline, unprecedented quantity and detail.
- Requiring sophisticated and robust computational and statistical methods [Kemmer 2023].
- Wide range of imaging modalities [Ouyang 2017].
- Computational advancements allow to map spatiotemporal biological processes.

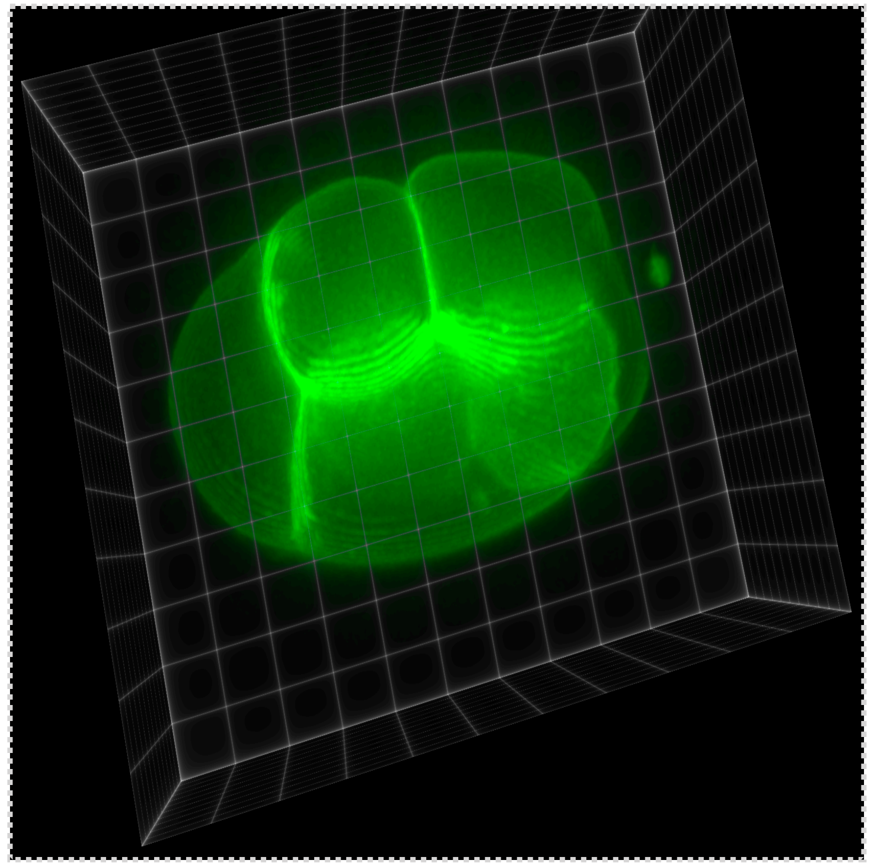


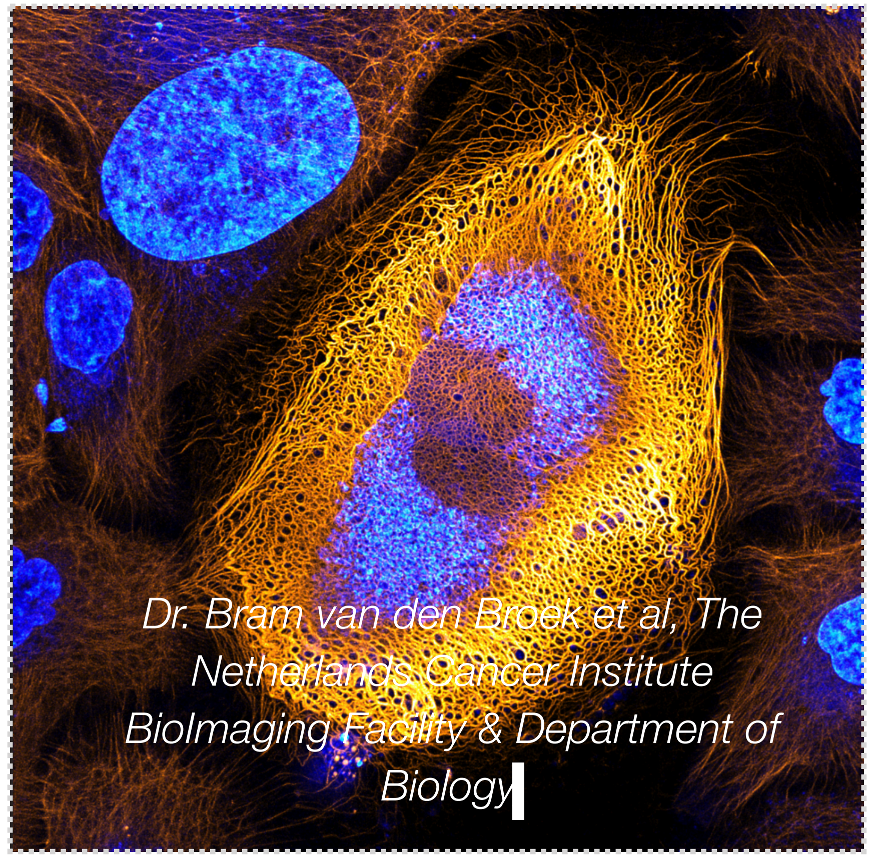


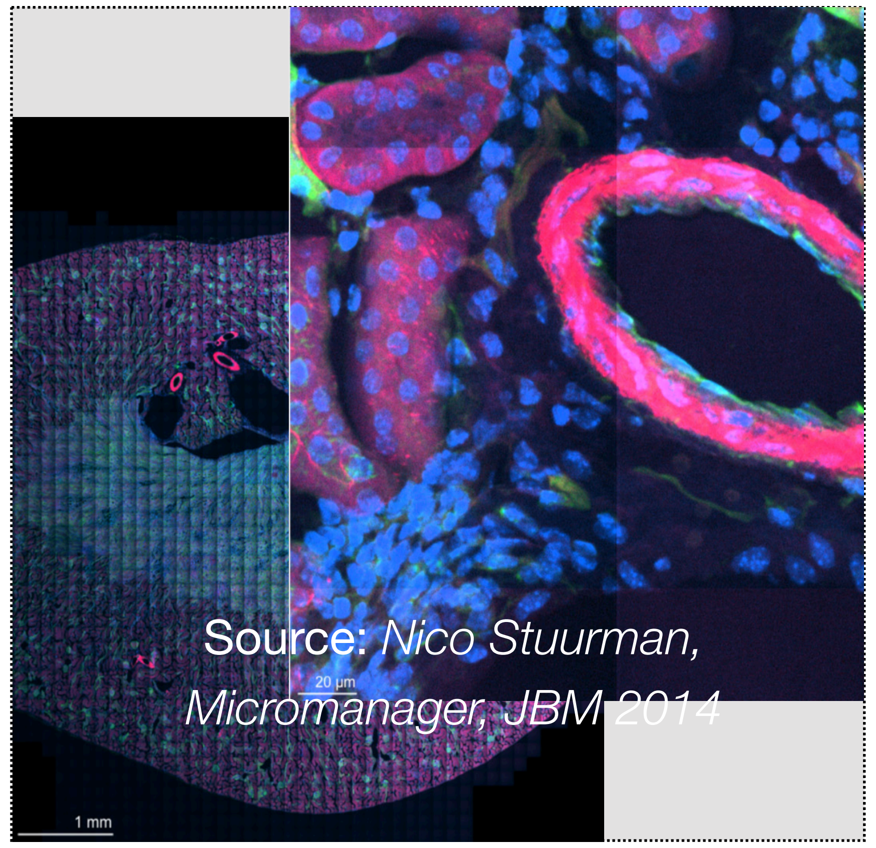


Bioimage informatics methods span

- **Spatial scales** from single molecules to cells, all the way to entire multicellular organisms [Kemmer 2023].
- **Time scales** ranging from fractions of ms to days.
- **Multiplexing** visualization of a multitude of sample characteristics in parallel [Ellenberg et al. 2018].







Software

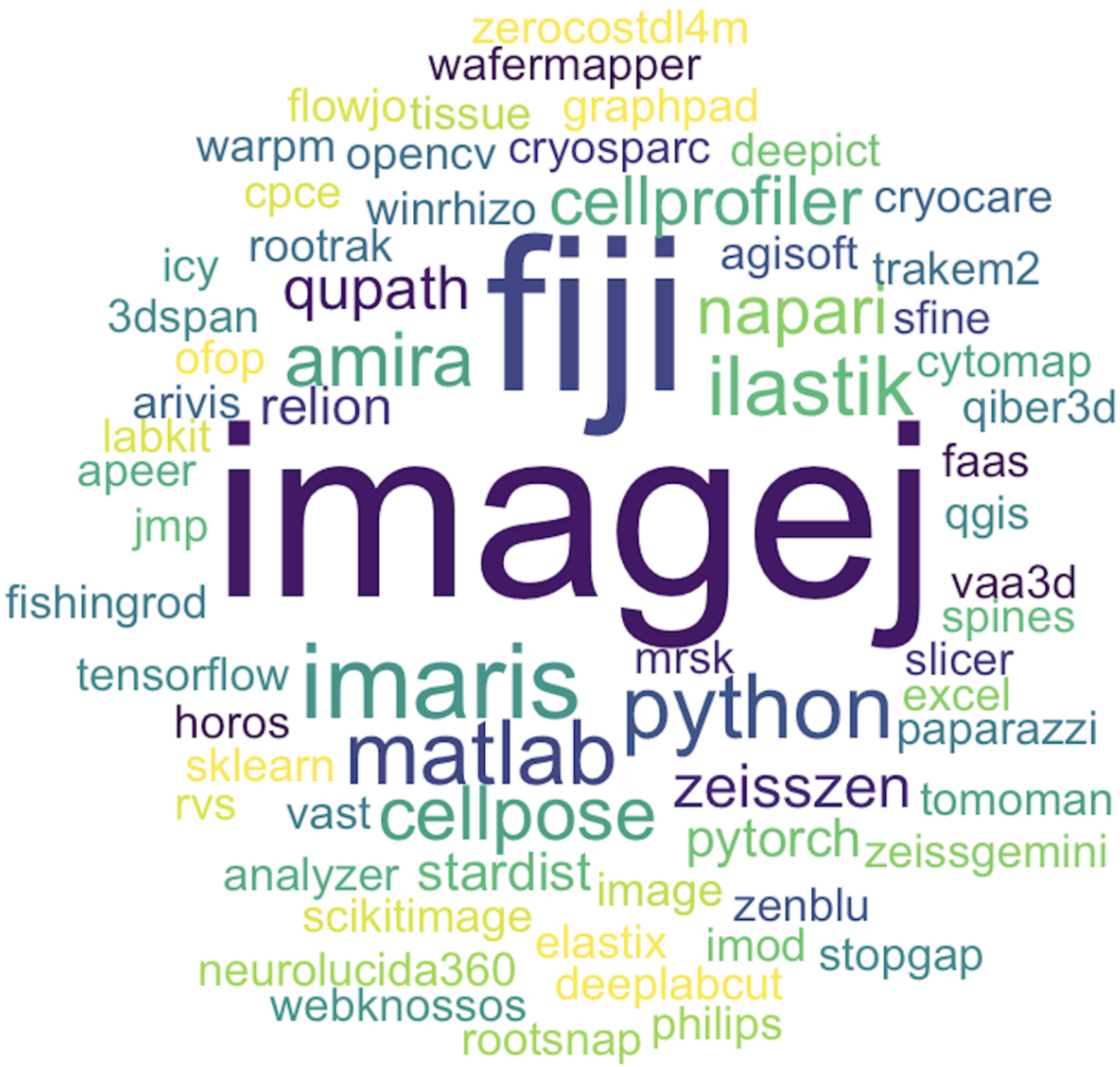
Software Packages in Java

FIJI
ImageJ2
Loci

ImageJ
NIH
Vanilla version

QuPath
Pete Blankhead

ICY
Pasteur Institute



Automation of image analysis problem

Automation, speed up, many images
Non-human bias
Reproducibility, documentation

Software Packages in Python

CellProfiler
Broad Institute

NAPARI
Biohub

scikit-image
image processing in python

OpenCV

PyTorch

TensorFlow

jupyter

Community

Conferences

I2K

european light microscopy initiative elmi

neubias network of european bioimage analysts

Resource for training

image.sc

Forum for getting help

Bioimage Analysis Societies

GLOBAL BIOIMAGING growing collaboration

globalbioimaging.org

BioImaging North America

www.bioimagingnorthamerica.org/

AI4Life

ai4life.eurobioimaging.eu/

and more ...

EURO BIOIMAGING

www.eurobioimaging.eu

FRANCE-BIOIMAGING

france-bioimaging.org/

Swiss BIAS

swissbias.github.io/

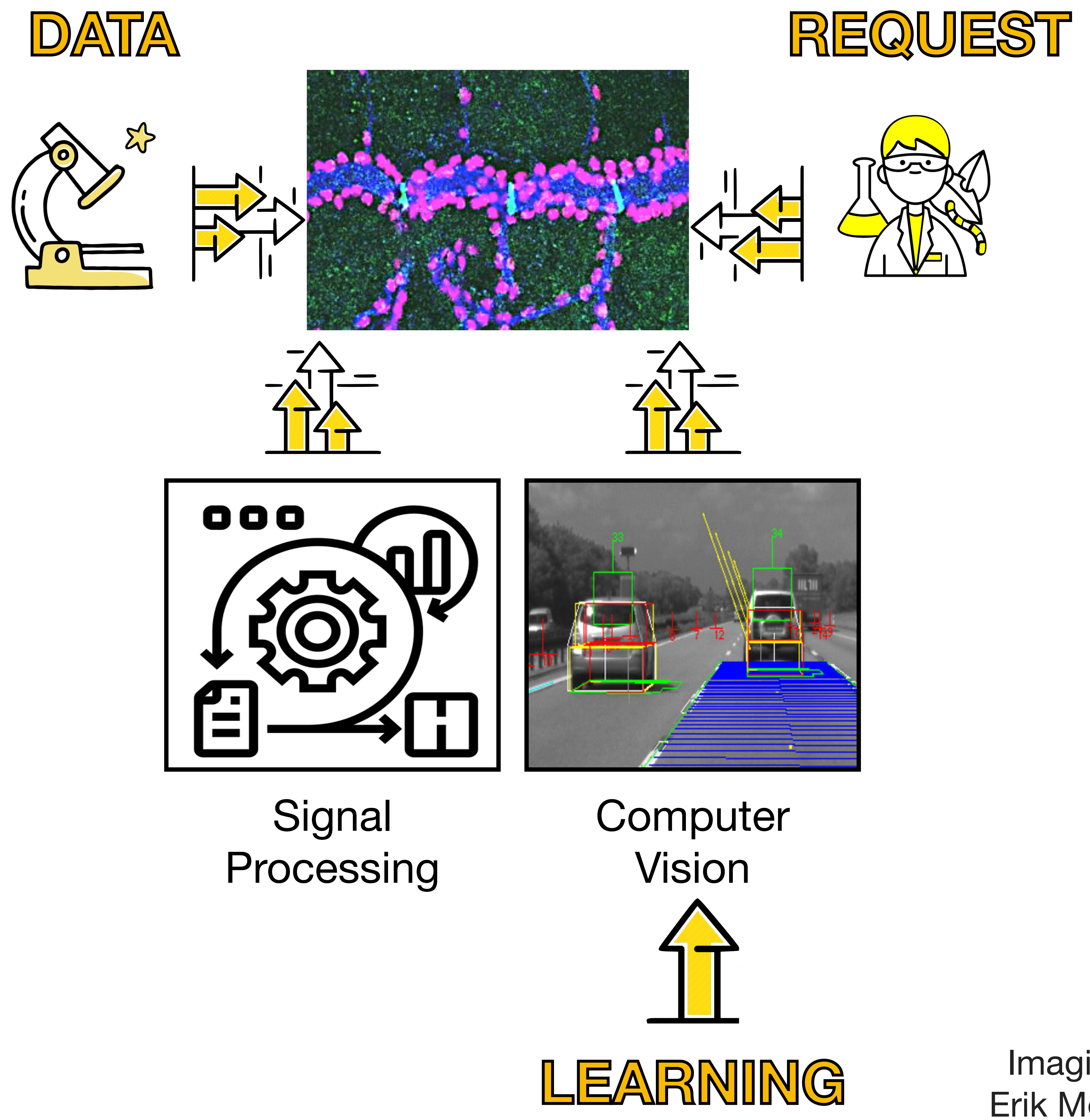
QUAREP-LiMi

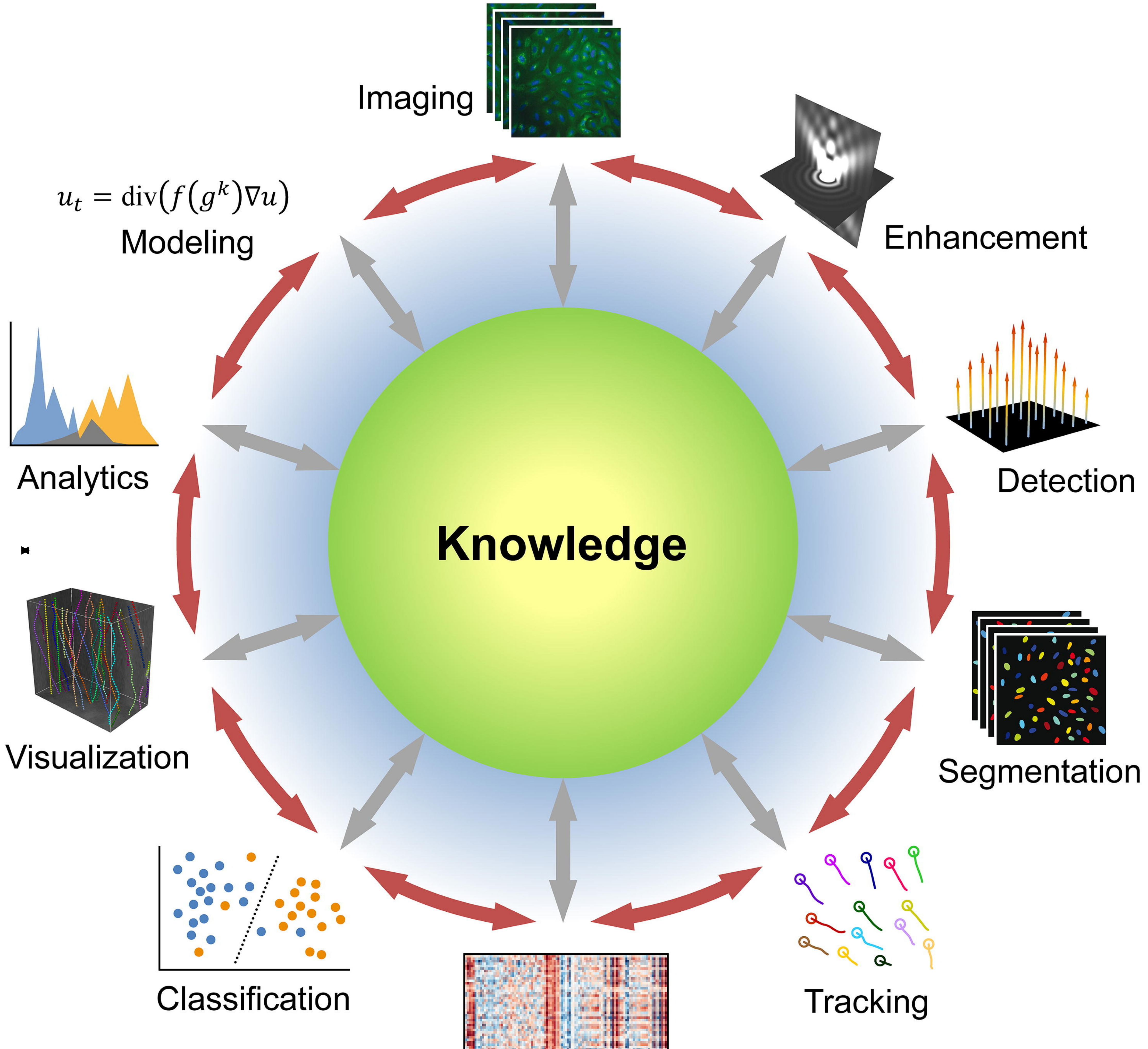
quarep.org/

German BioImaging Gesellschaft für Mikroskopie und Bildanalyse

gerbi-gmb.de/

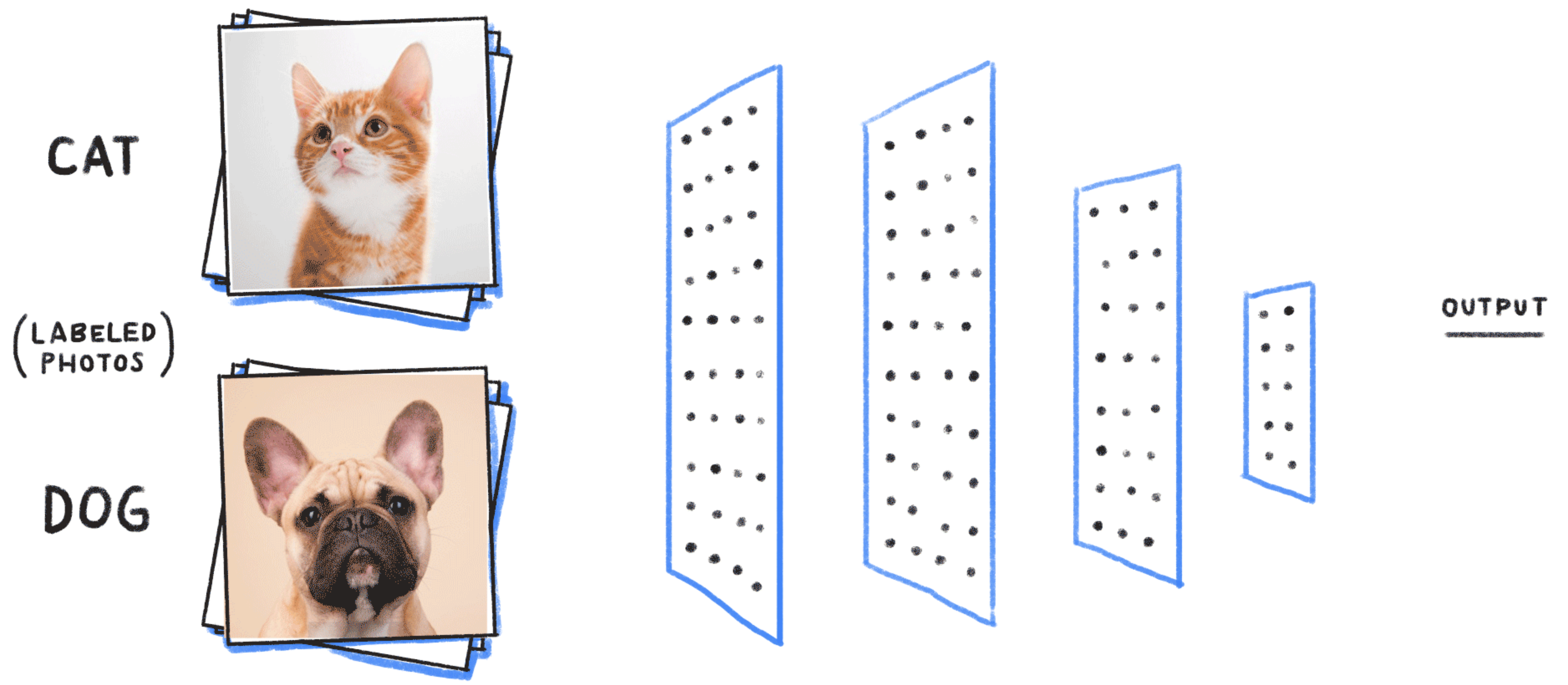
Bioimage Analysis





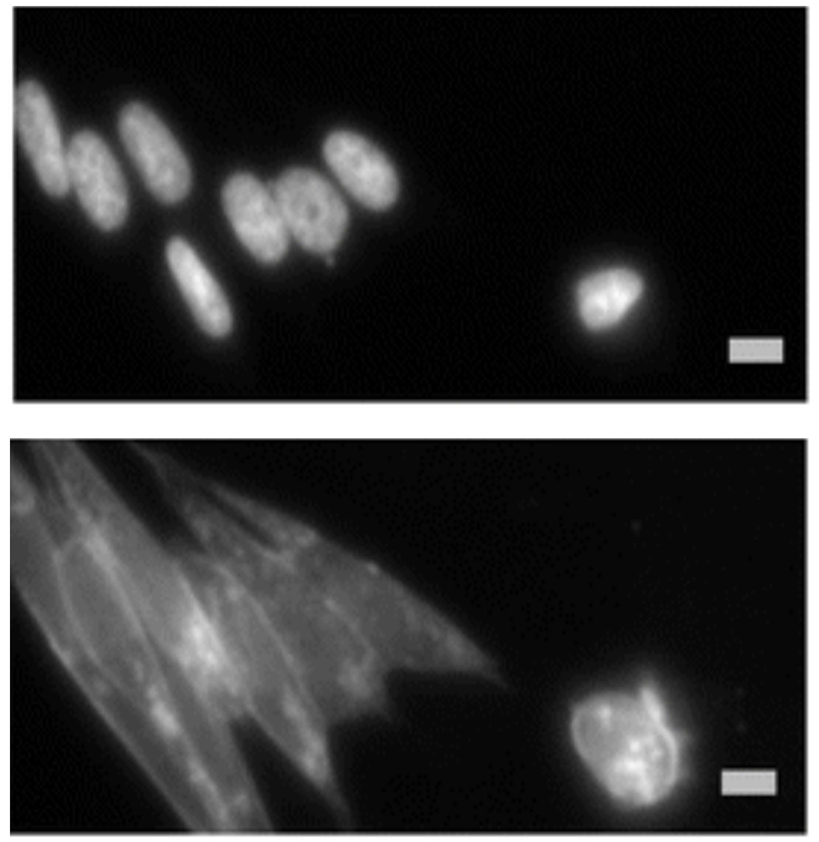
Imagining the future of bioimage analysis
Erik Meijering, Nature Biotechnology, 2016.

Image Classification | Style Transfer | Image Restoration | In-painting | Object Detection | Segmentation



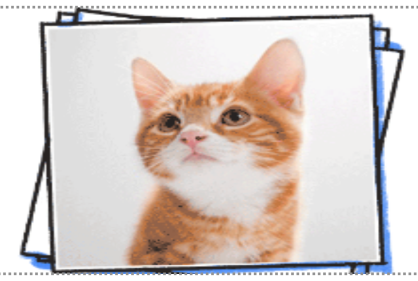
Source: https://becominghuman.ai/ Venkatesh Tata

Esteva et al., Nature, 2017



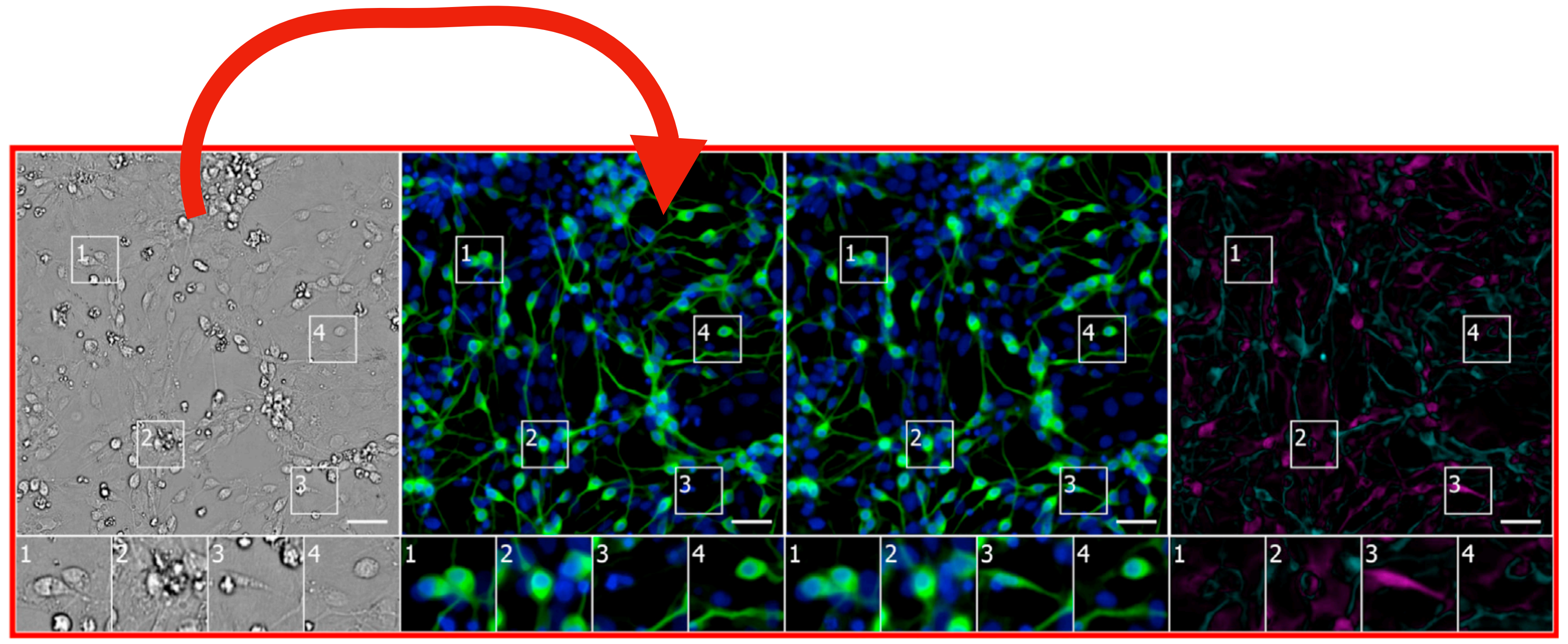
Assessing microscope image focus quality with deep learning
Yang et al., BMC Bioinformatics 2018

Image Classification | **Style Transfer** | Image Restoration | In-painting | Object Detection | Segmentation

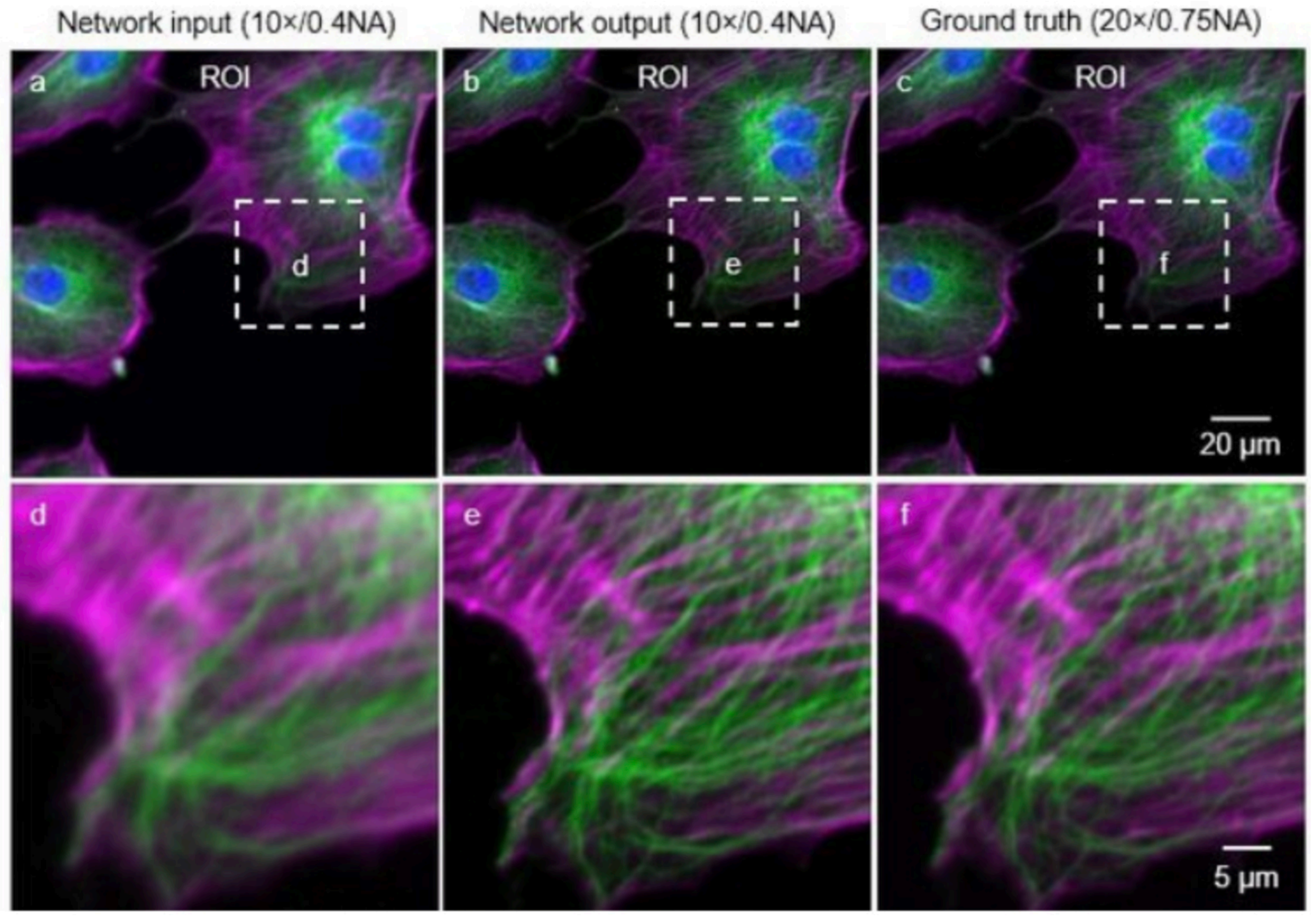


Virtual Acquisition

From Brightfield to Fluorescence



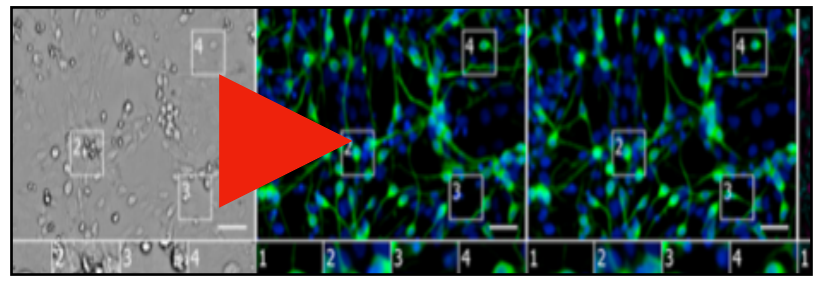
E. M.Christiansen, *In Silico* Labeling, Cells 2018

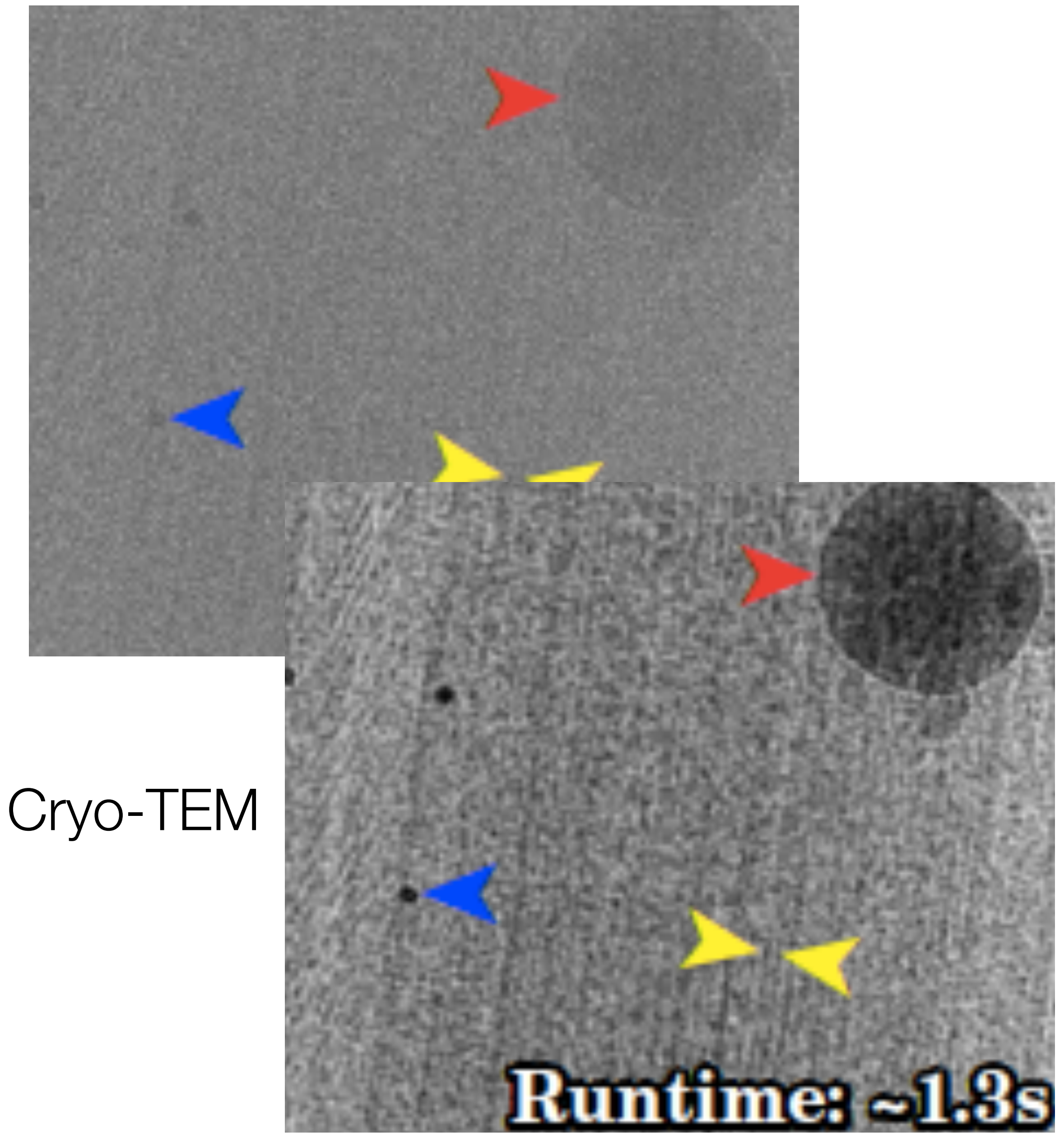


H. Wang, Nature Methods, 2019,

Image Classification | Style Transfer | Image Restoration | In-painting | Object Detection | Segmentation

CAT

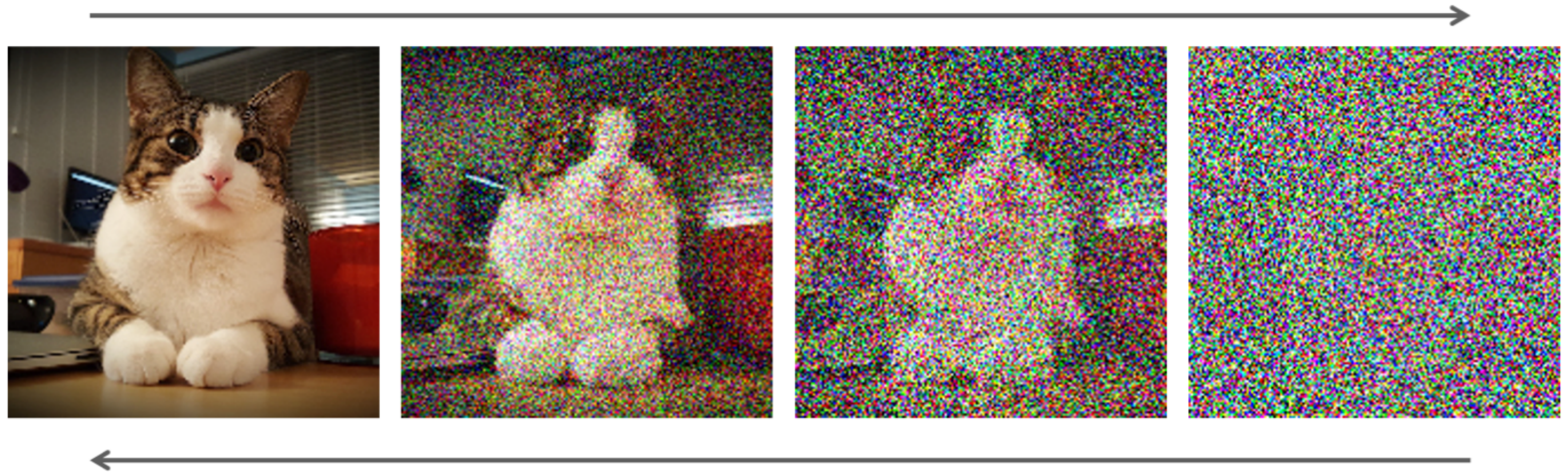




Cryo-TEM

Krull et al., Noise2Void -CVPR, 2019.

Denoising



Generative Denoising Diffusion Models

Image Classification

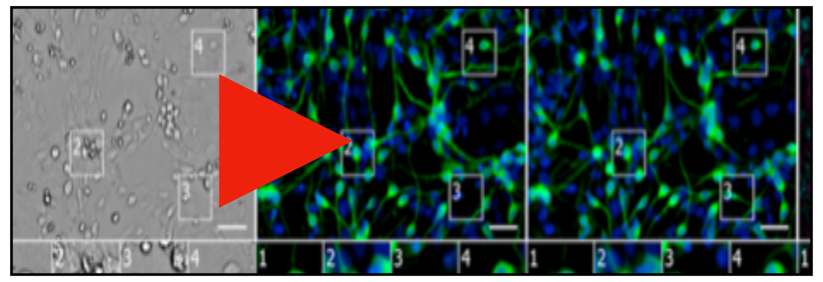
Style Transfer

Image Restoration

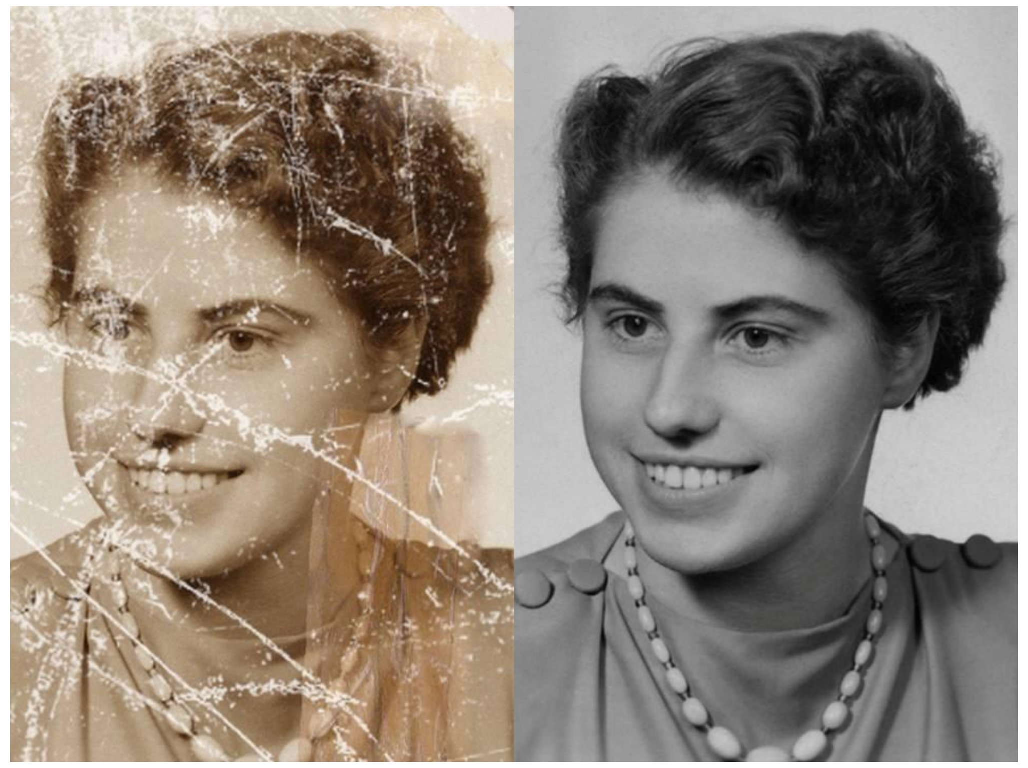
In-painting

Object Detection

Segmentation

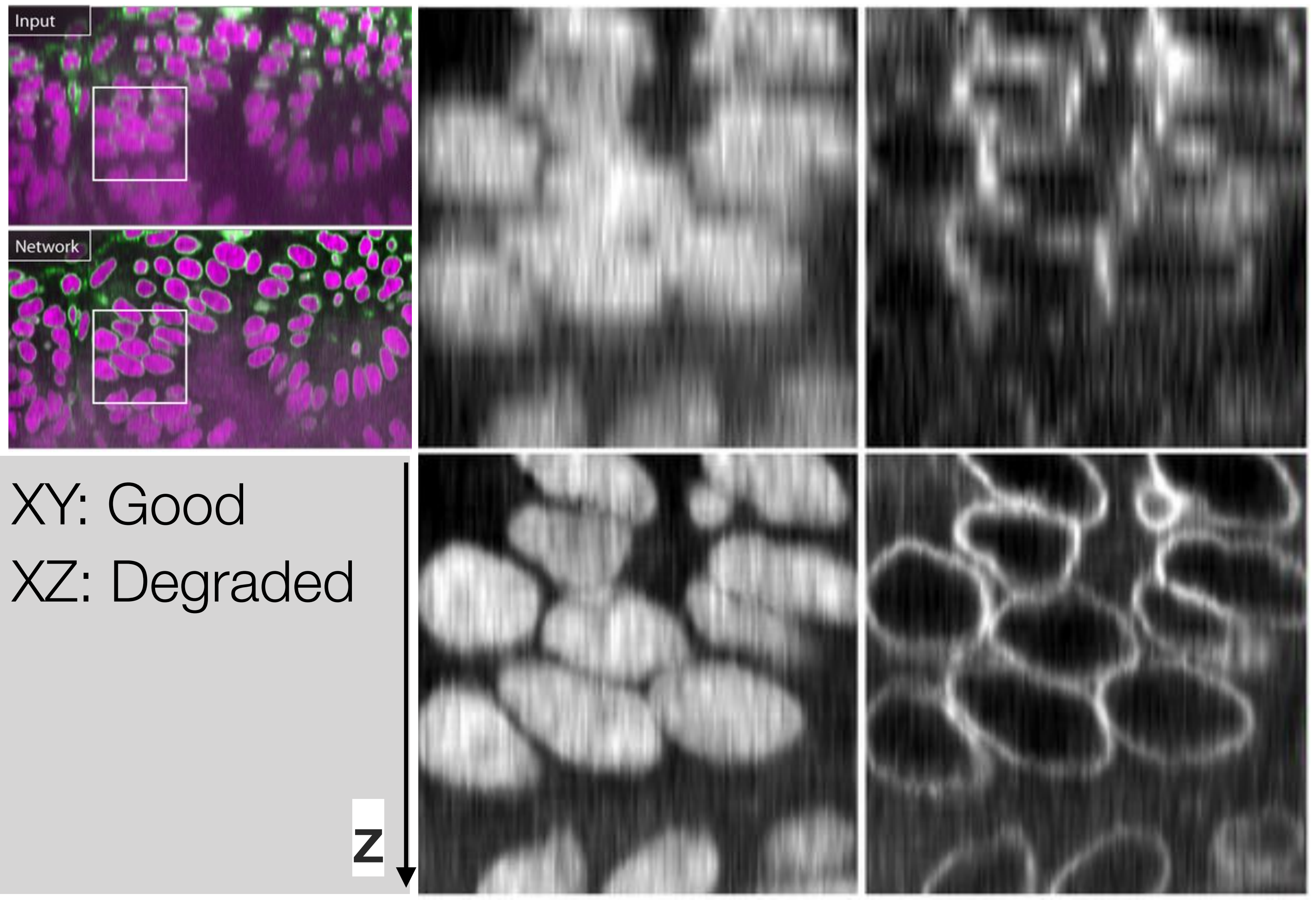


Fill in missing parts



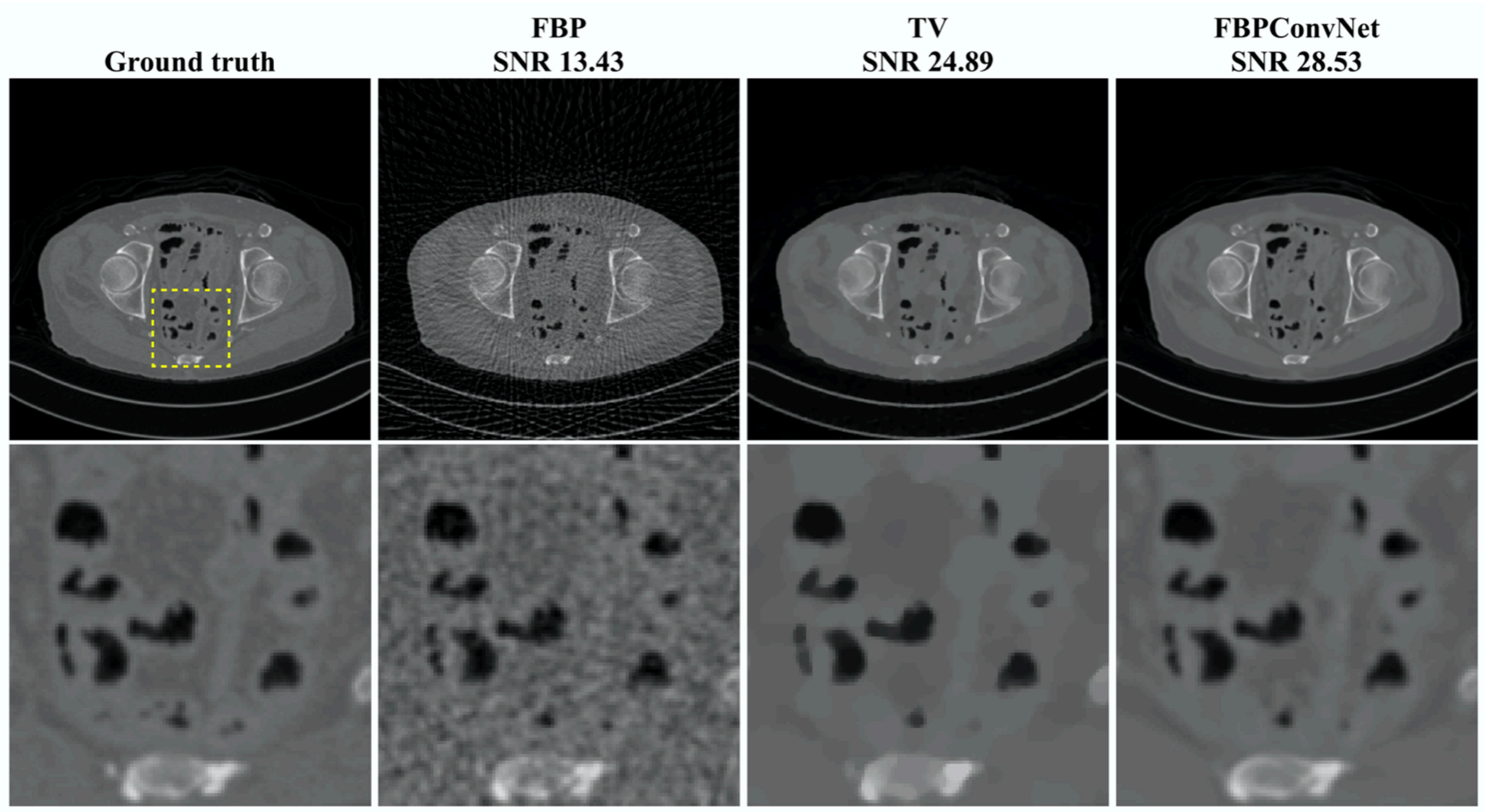
Source: Medium.com, Tarun Bonu

Axial restoration



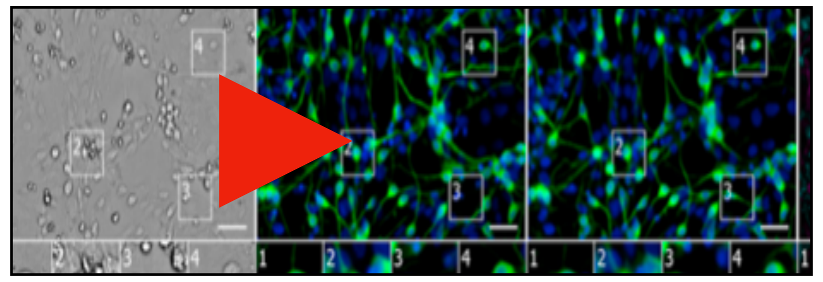
CARE, M. Weigert, 2019

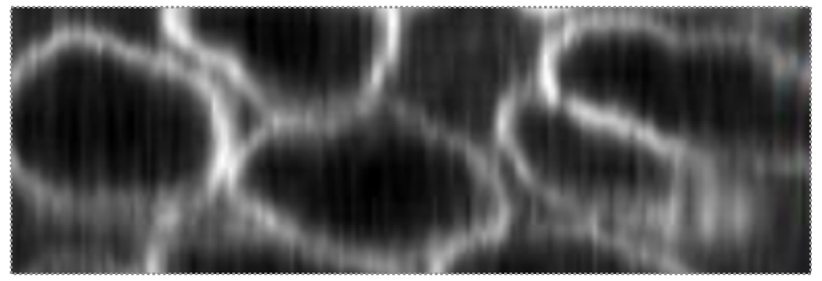
Artefact Correction

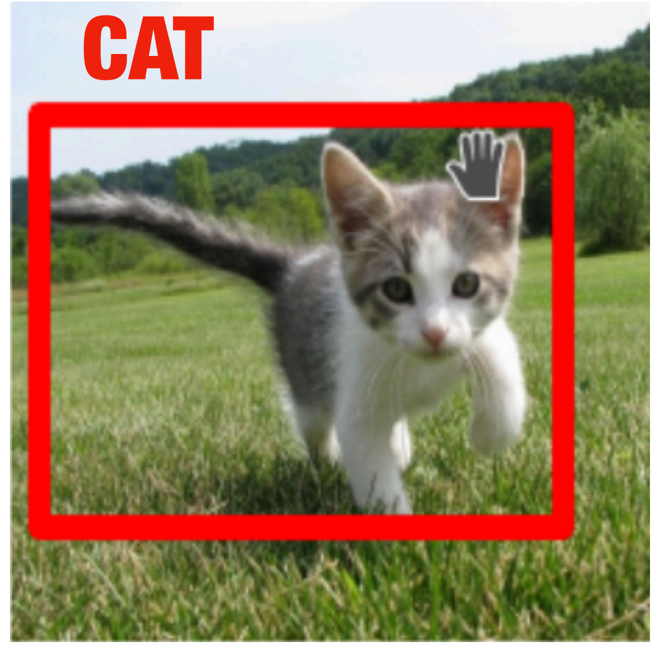


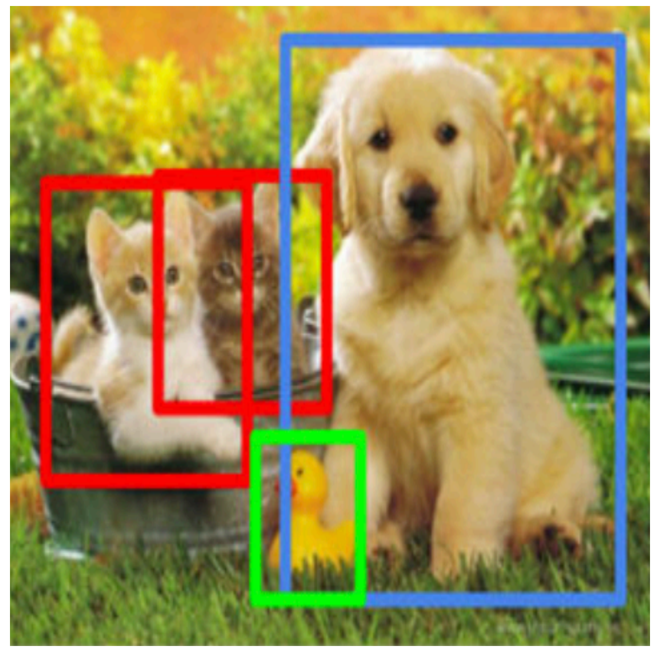
Jin et al., IEEE Trans. Im Proc., 2017

Image Classification | Style Transfer | Image Restoration | In-painting | **Object Detection** | Segmentation

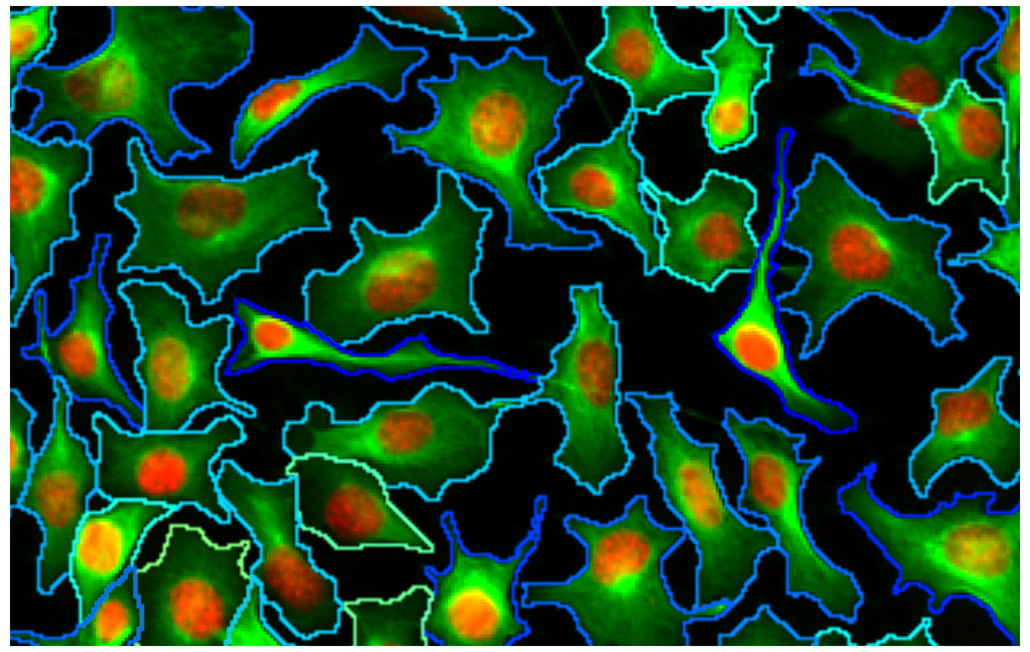




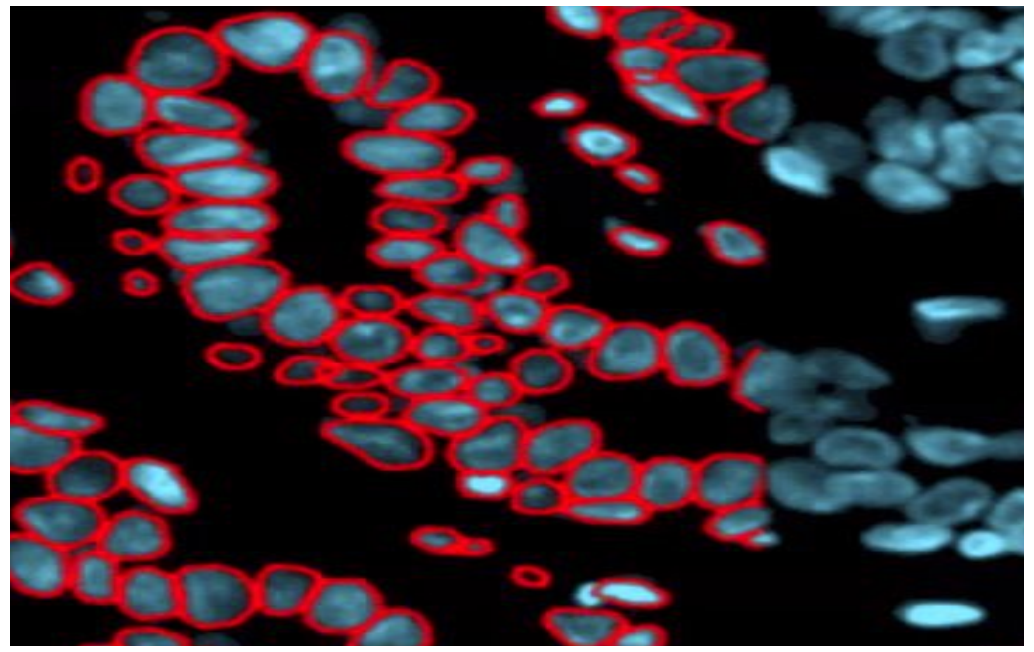




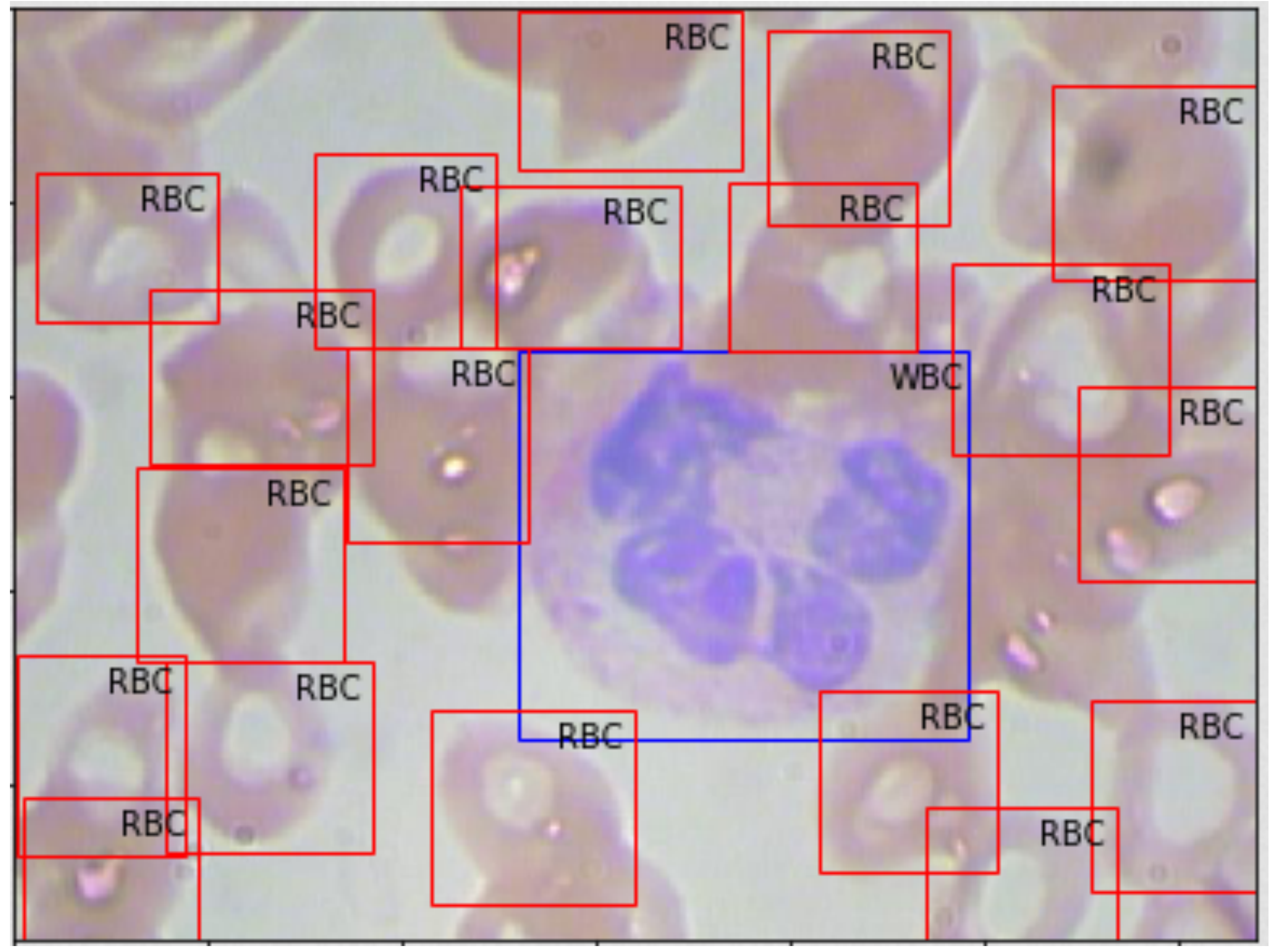
CAT CAT DUCK DOG



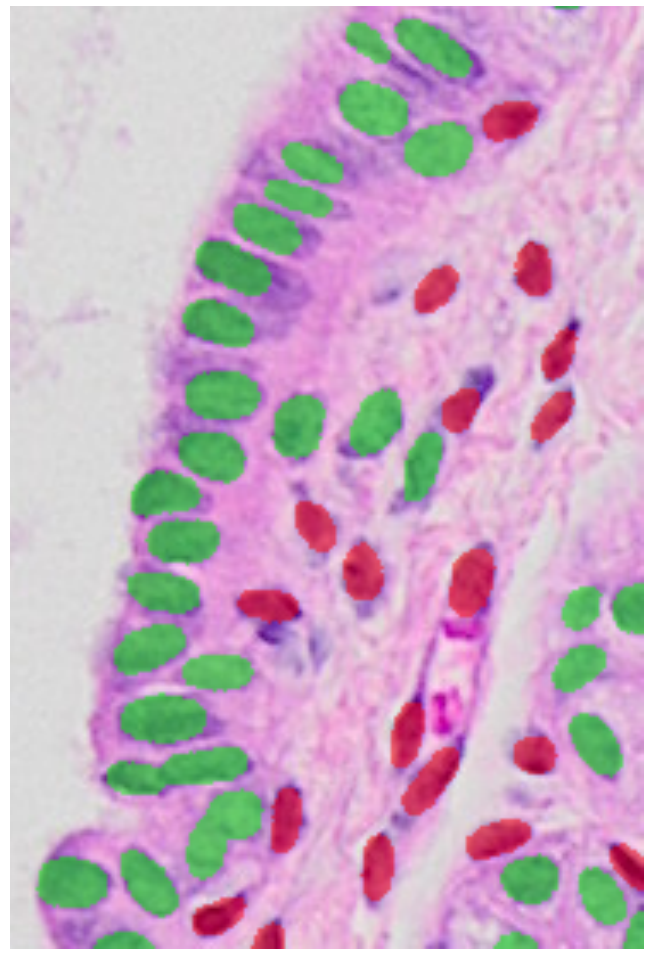
Cellpose
C. Stringer
2021



Stardist
M. Weigert
2018



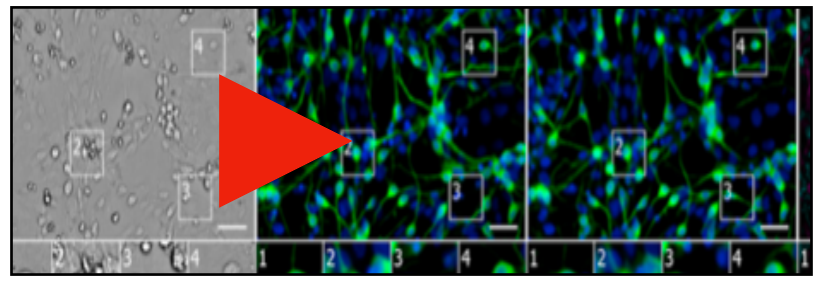
Blood cell detection Bounding Box
YOLOv3

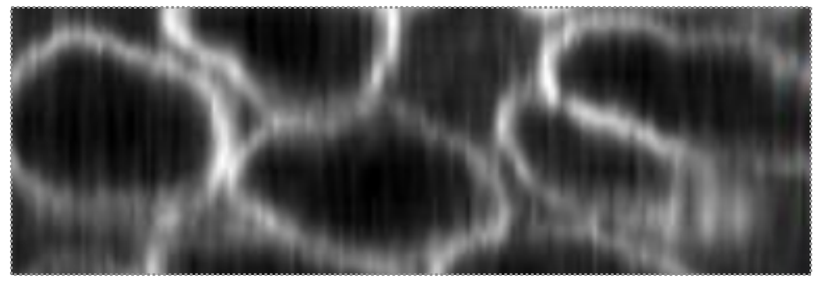


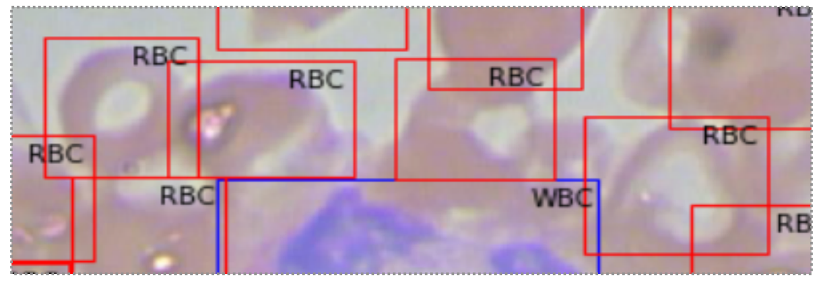
Mouse/human cells

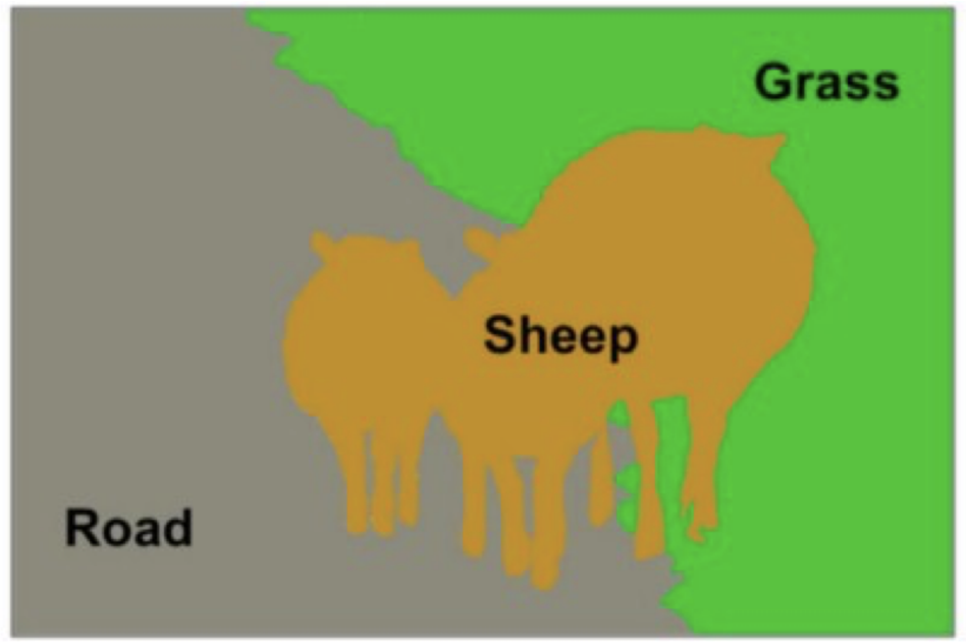
Juppet et al., 2021.

Image Classification | Style Transfer | Image Restoration | In-painting | Object Detection | **Segmentation**

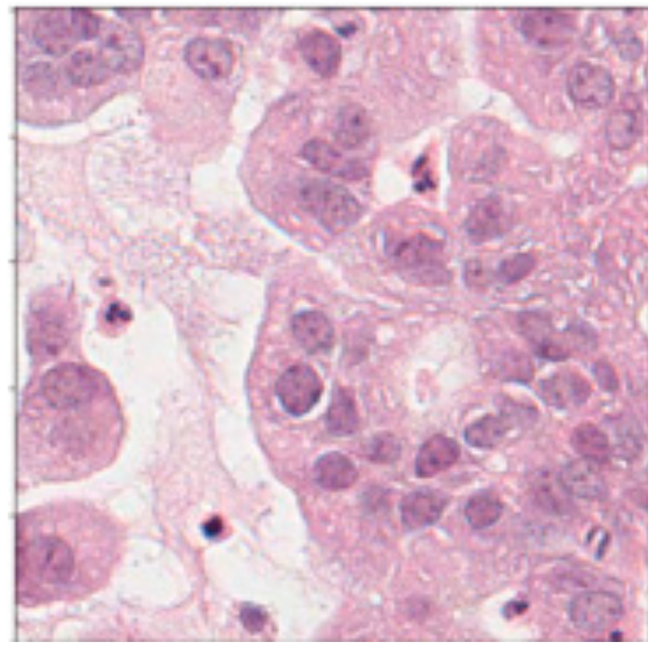


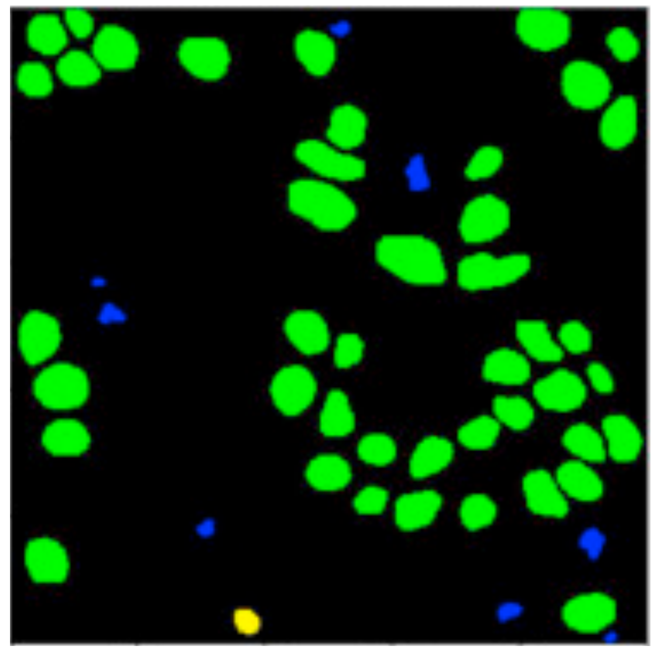






Semantic segmentation



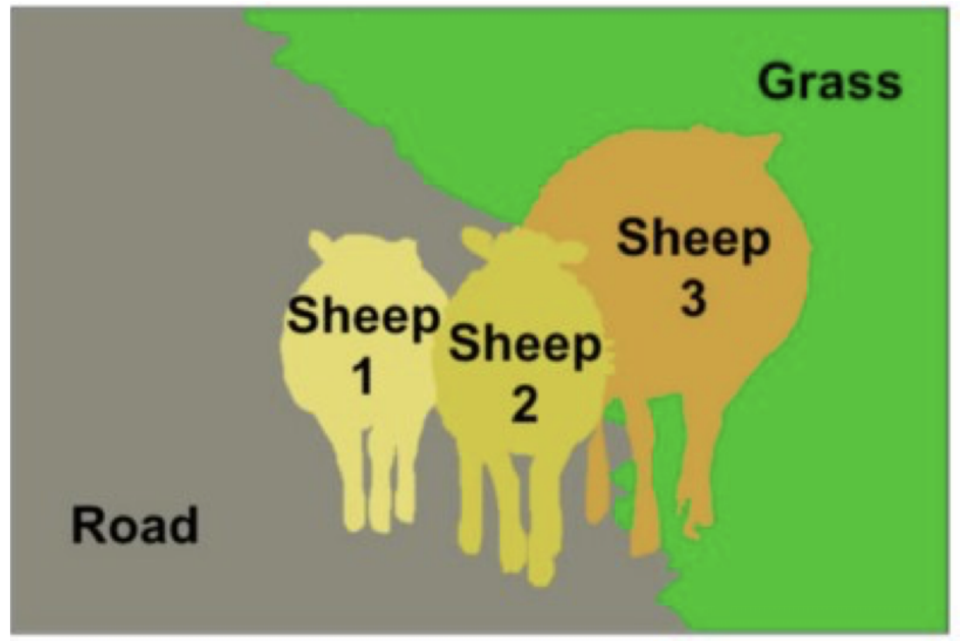


Semantic Segmentation

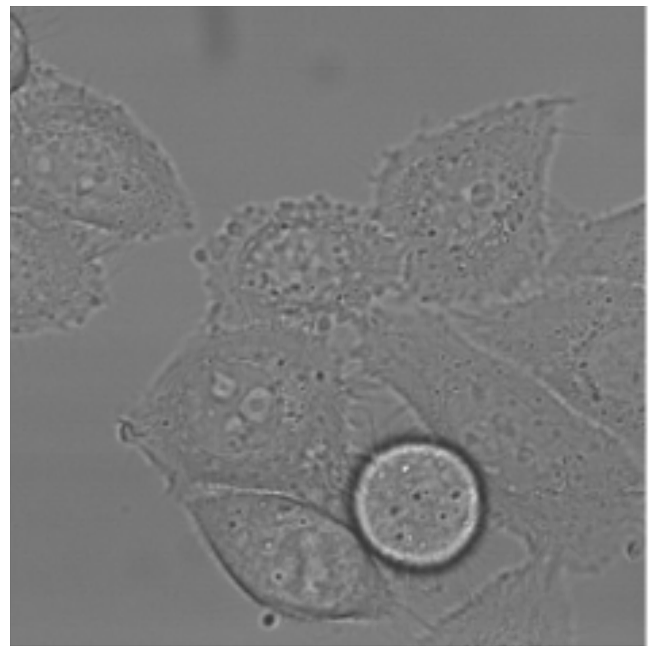
Pixel classification

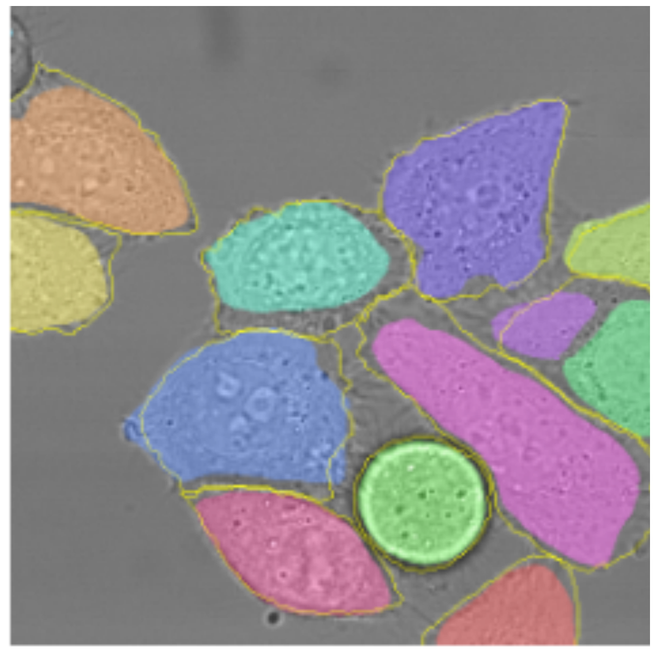
Pixel classification

- Binary
- Multiple classes



Instance segmentation



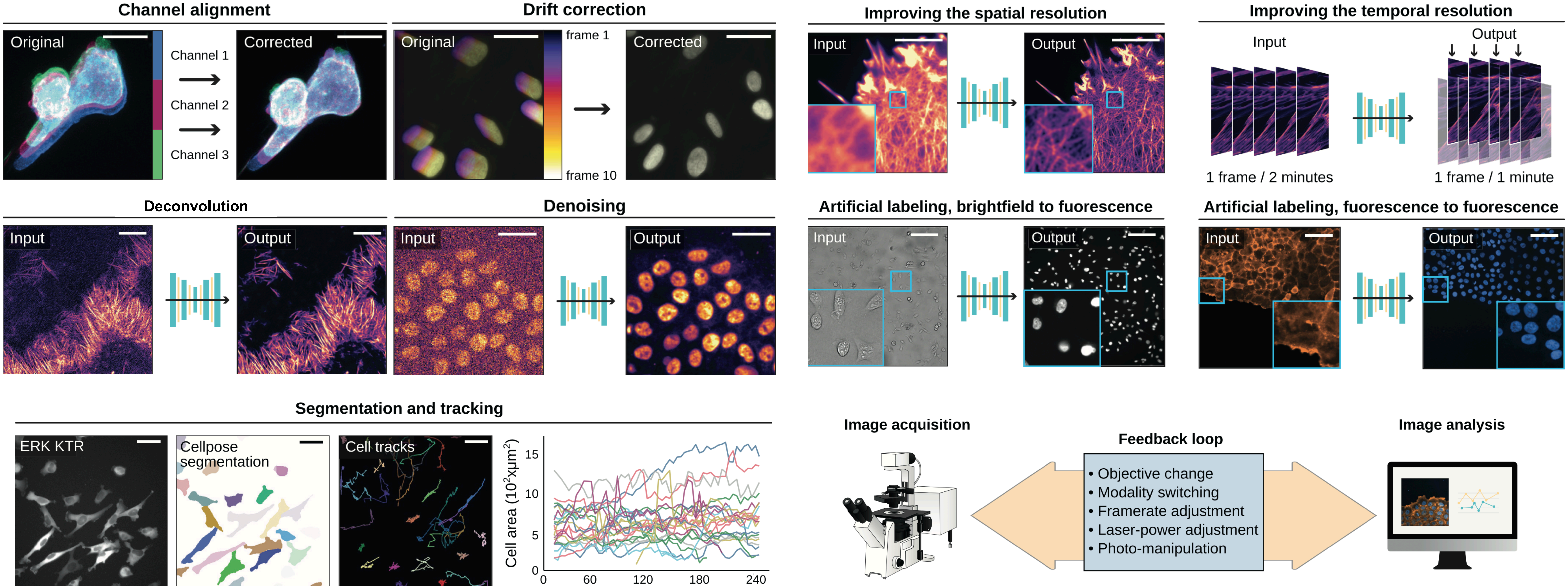


Instance Segmentation

HeLa cells

Ronneberger et al., U-Net MICAI, 2015.

Live Cell Imaging in Deep Learning Era



smart microscopy

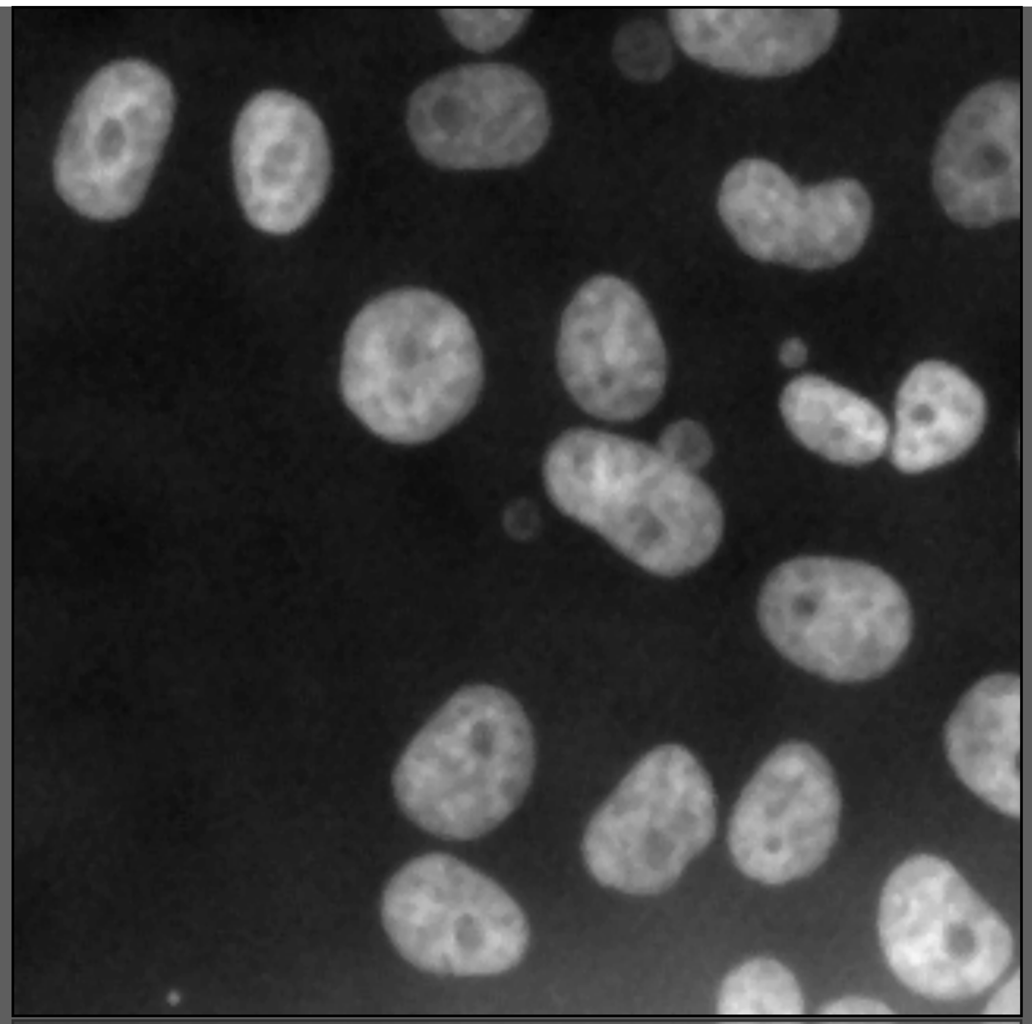
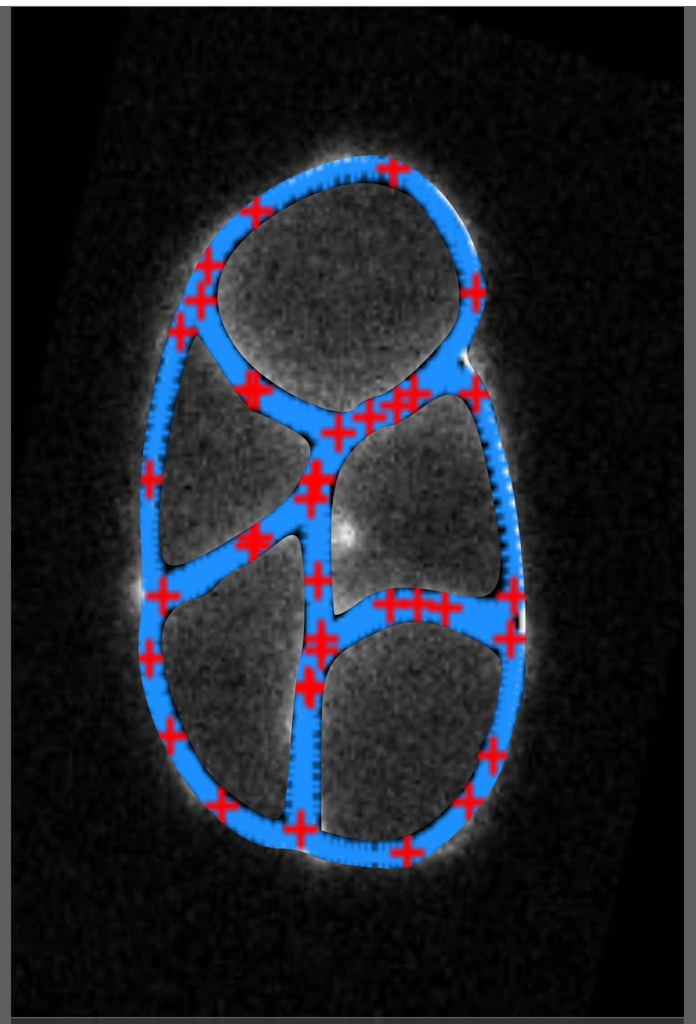
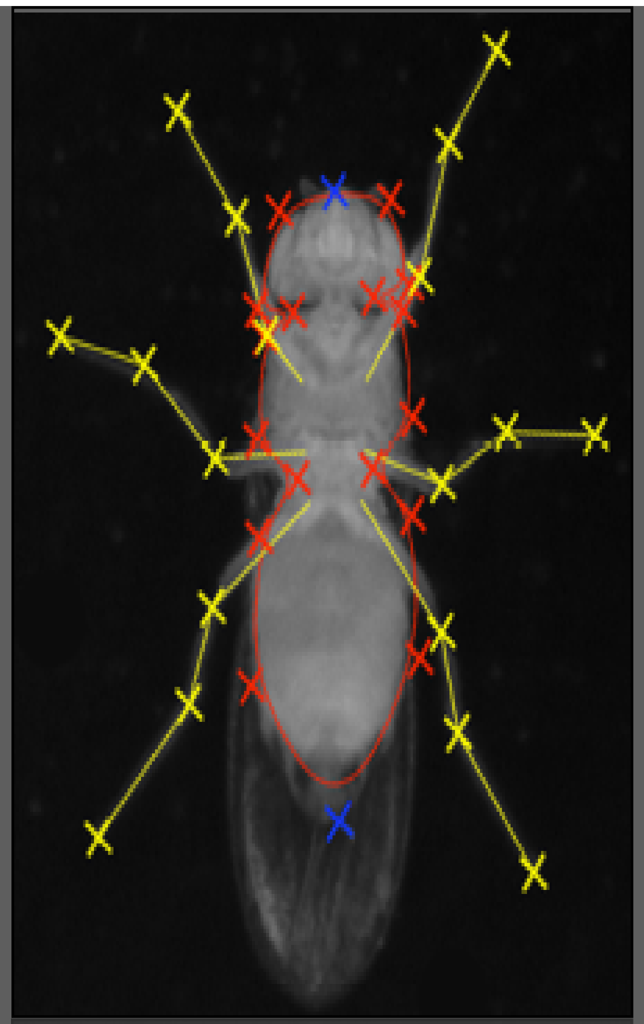
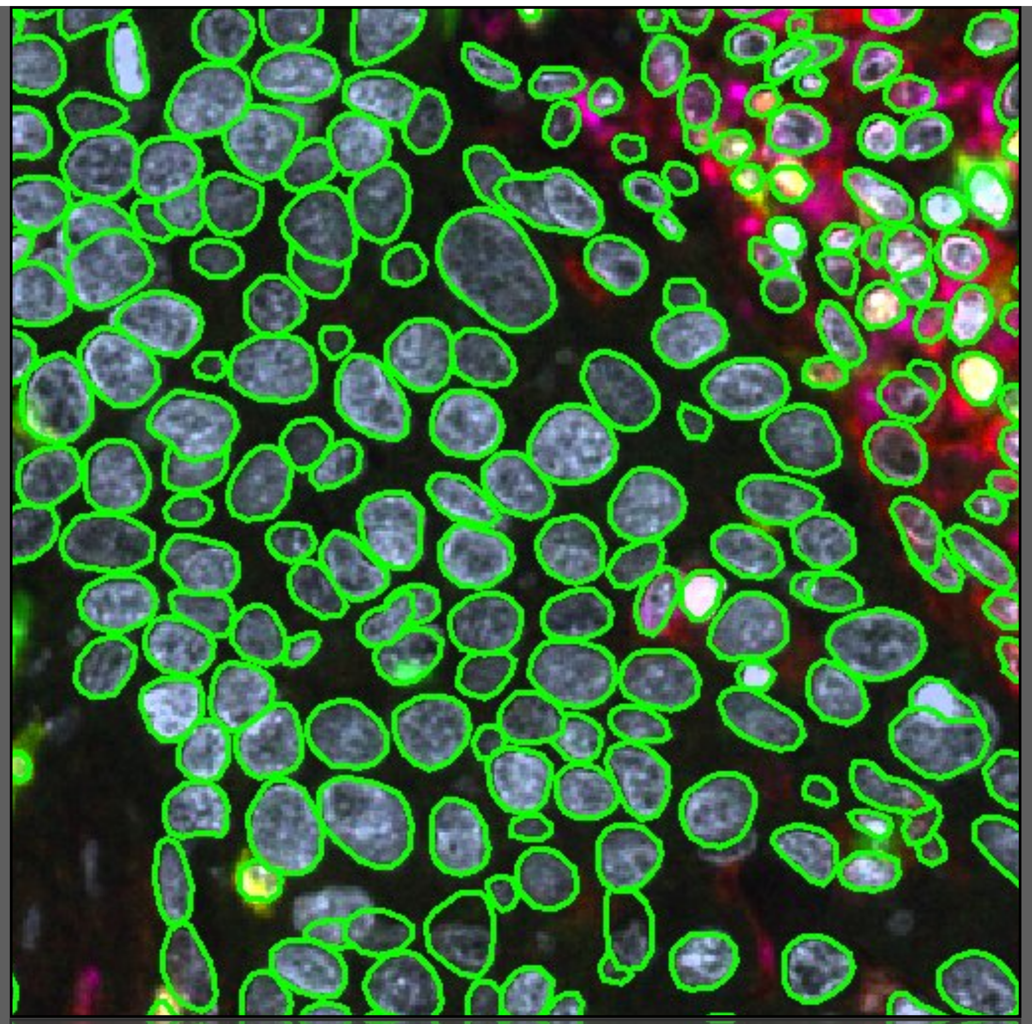
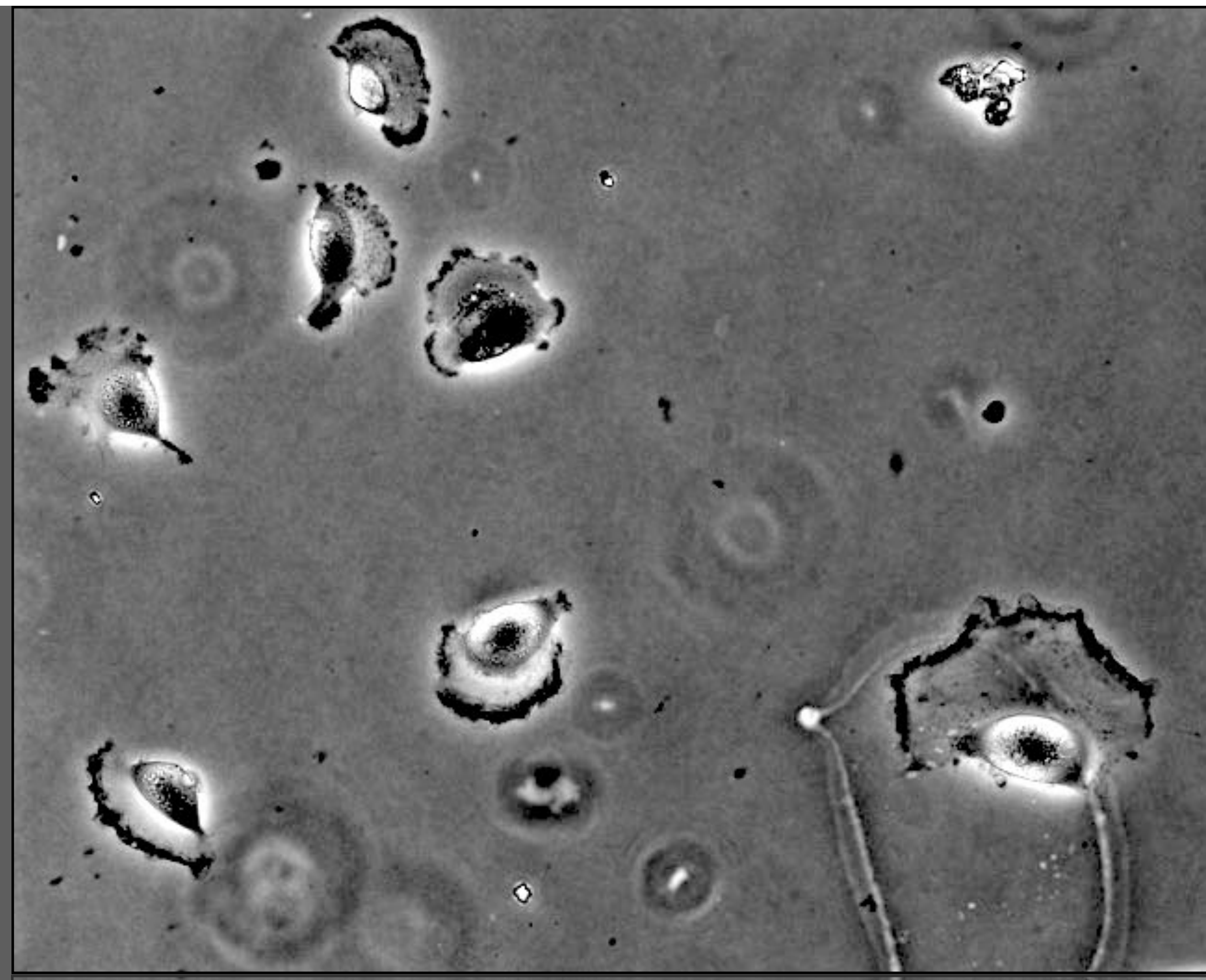
Live Cell Imaging in Deep Learning Era
J. Pylvänäinen et al., Current Opinion in Cell Biology, 2023

Image Segmentation

- ➡ Grouping pixels into regions
- ➡ Segment image into objects
- ➡ Classify objects of the image

- ➡ Dimension
- ➡ Large number of objects, dense
- ➡ Highly variability: shape, color, ...
- ➡ Rare phenotype of interest

Science or Art ?

Introduction to Microscopy Image Analysis

Lecture for the workshop AI4Life given by Daniel Sage, 10 June 2024

CONTEXT — BIOIMAGE INFORMATICS

METHODS — MODEL-BASED VS. DATA-DRIVEN

LEARNING — DATA, MODELS AND TOOLS

WRAP UP — BIOIMAGE ANALYSIS

daniel.sage@epfl.ch


CCD
CPU
IMAGEJ
Fluo
RAM
WEB
ILASTIK
GPU
DATA
NAPARI
SMART
Rule-based
Model-based
Machine Learning
Deep Learning
Trained Models
Foundation Models
morpho
wavelet
level-set
snake
Inverse Problem
RF
UNET
stardist
sam
1985
2000
2005
2010
2015
2020
2025

Rule-based

Model-based

Machine Learning

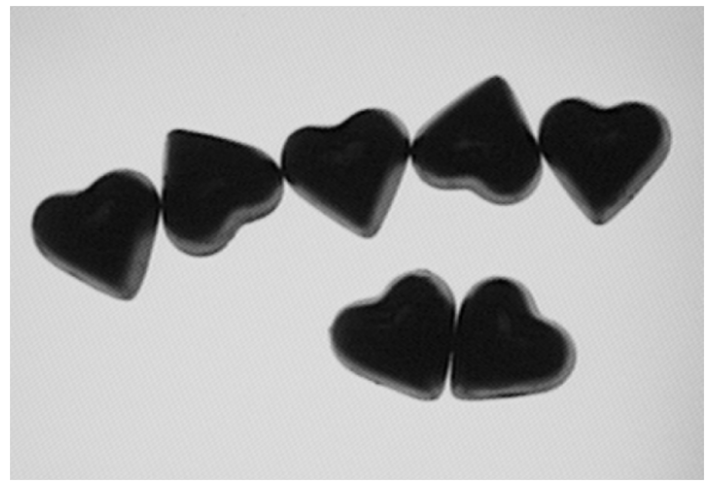
Deep Learning

Trained Models

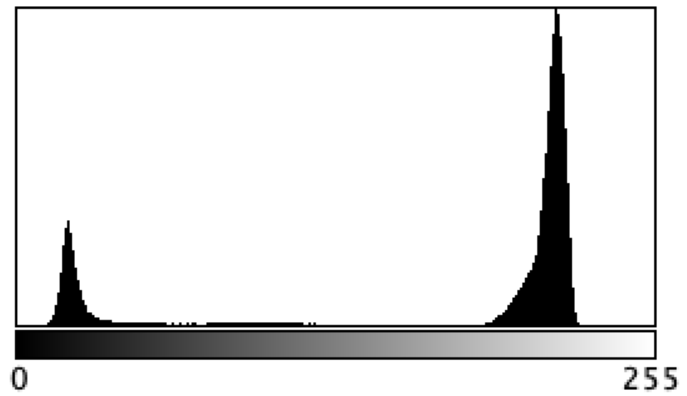
Foun-dation Models

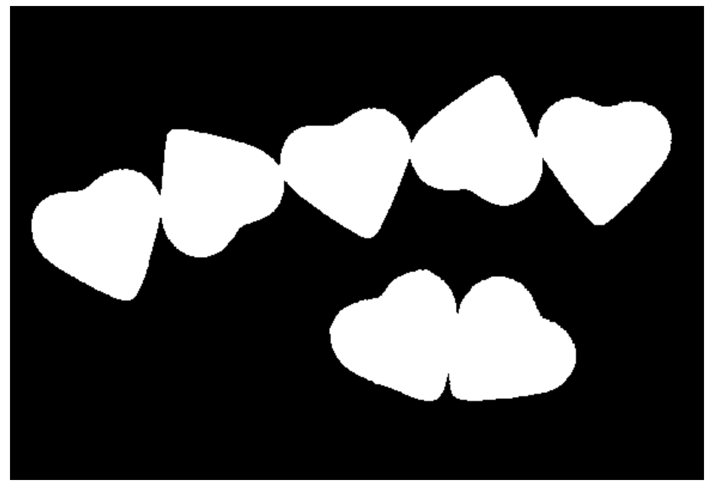
Rule-based Approach

- Intensity-based adhoc routines
- Preprocessing, filters



Packaging of Swiss chocolates



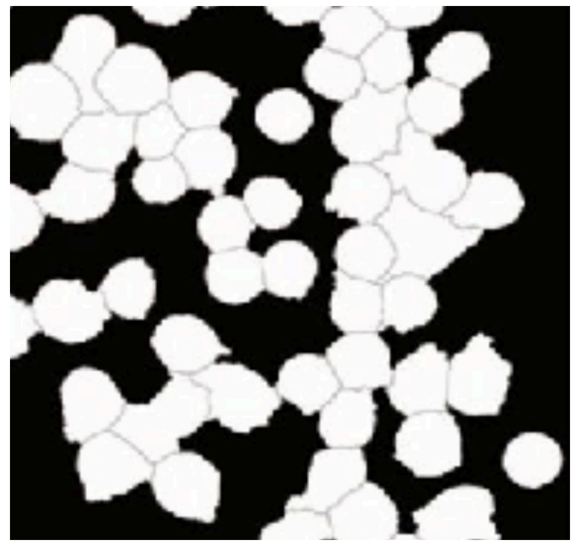


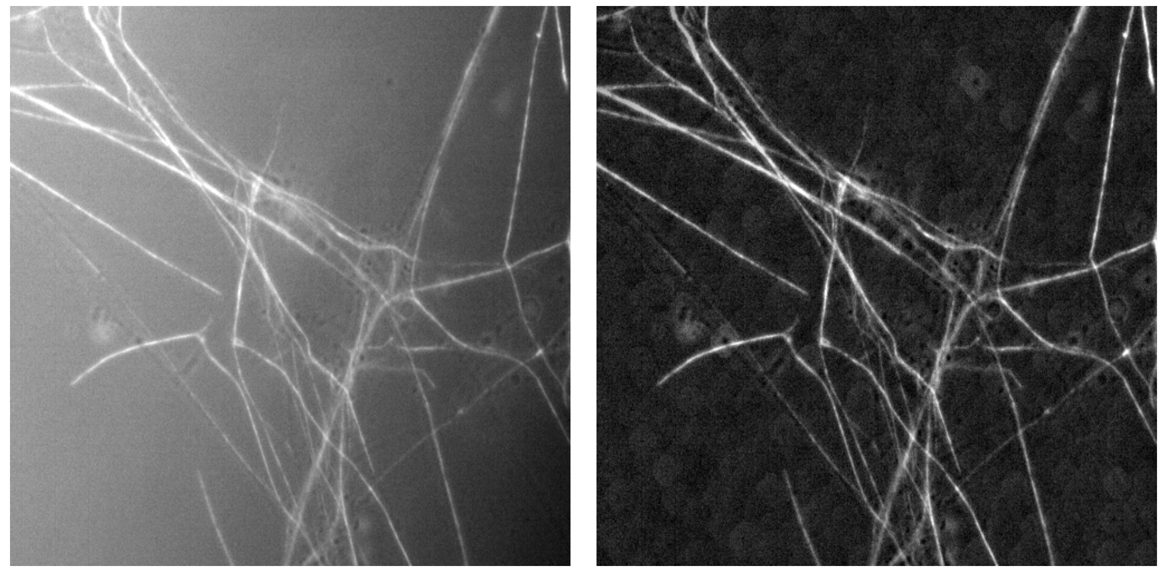
Aggregation

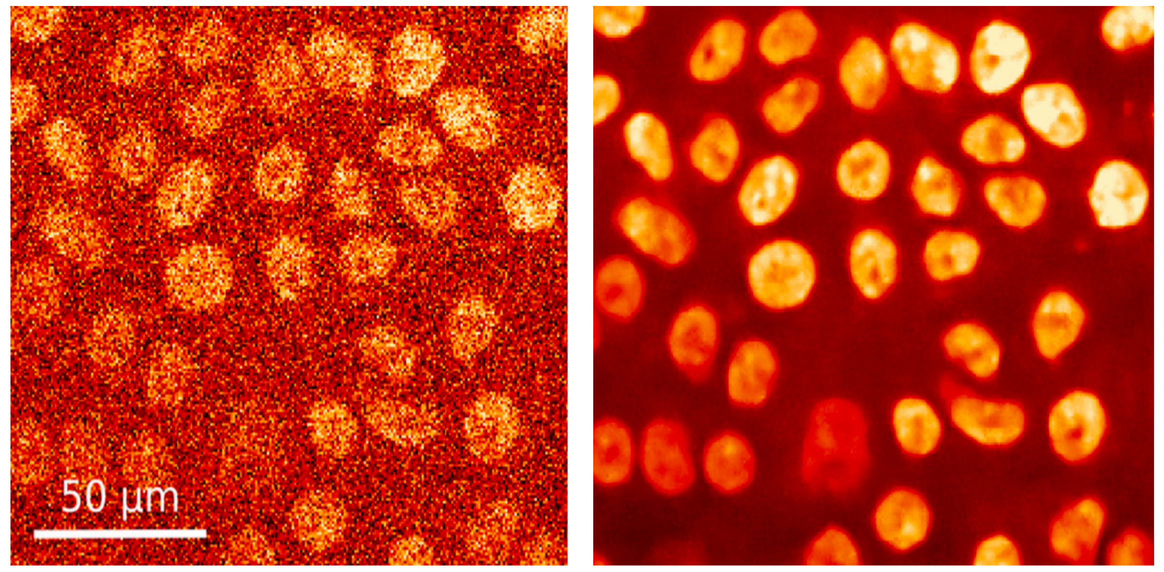
- Super-pixel SLIC
- Region-growing/ Watershed

Preprocessing / simplification

- Structure enhancing
- Denoising
- Projection over time
- Flatten background







Rule-based

Model-based

Machine Learning

Deep Learning

Trained Models

Foundation Models

Classic Image-Analysis Workflow

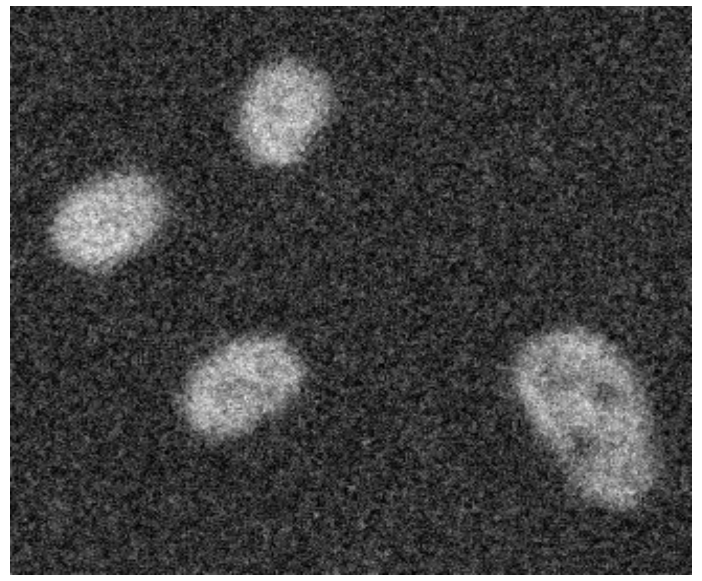
low-level information
many pixels (~Gb)

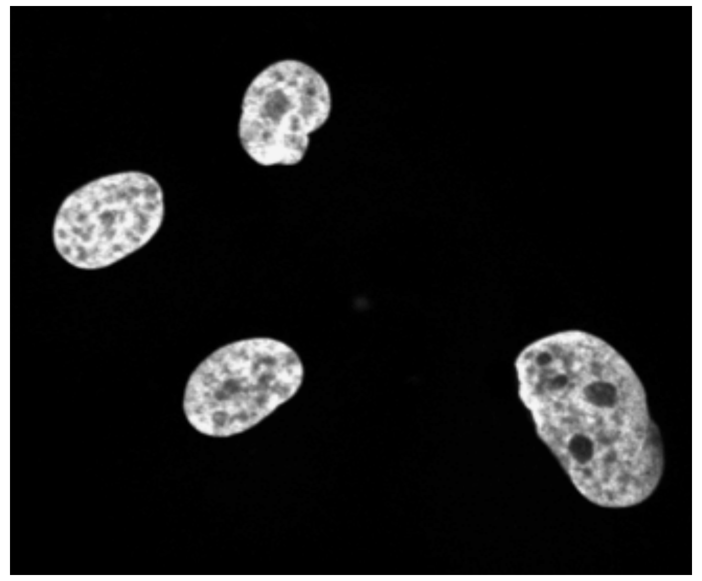
image formation model

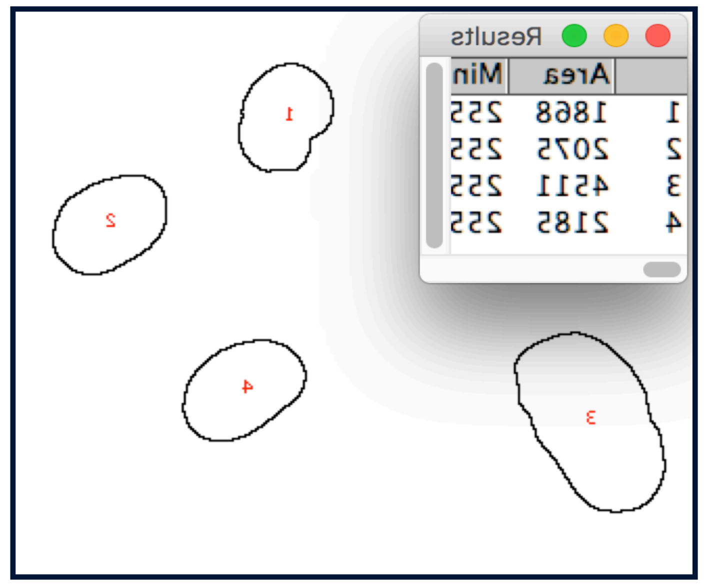
Priori knowledge

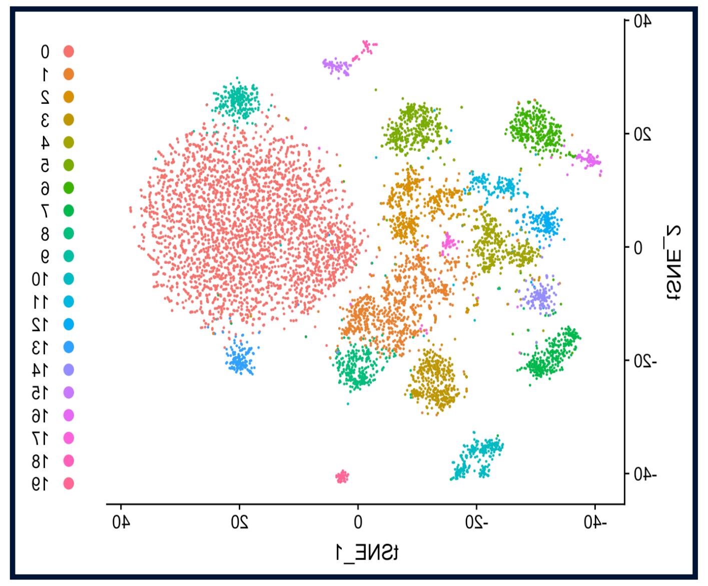
Data analysis
bio-statistics

high-level information
few objects (~Kb)









Preprocessing → Segmentation → Detection → Measurement → Analysis

Raw data

Pixel

Classified Pixel

Object

Numbers, Stats

Decision

Parameters, step-by-step error propagation

Rule-based

Model-based

Machine Learning

Deep Learning

Trained Models

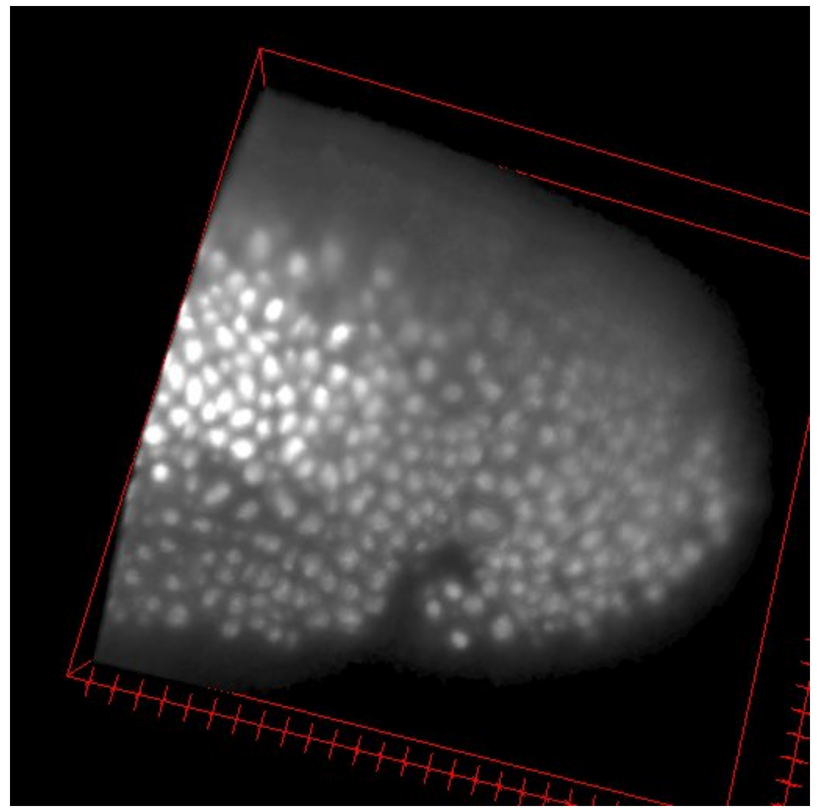
Foun-dation Models

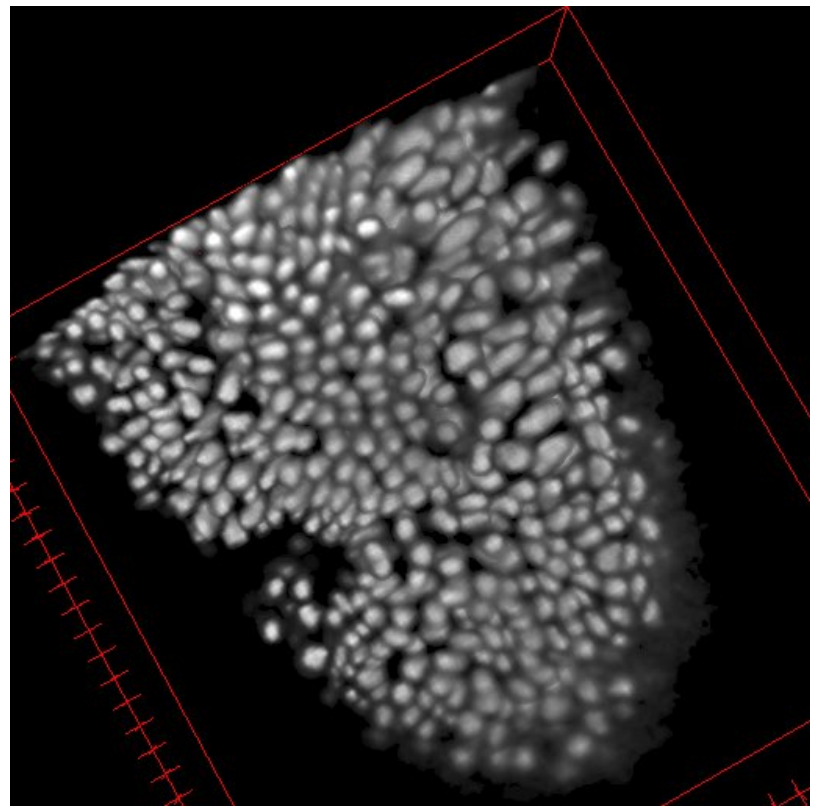
Model-based Approach

PHYSICS-DRIVEN RECONSTRUCTION

Mathematical formalism: forward model

- Inverse problems: Deconvolution
- Correlative imaging
- Image reconstruction
- Super-resolution microscopy



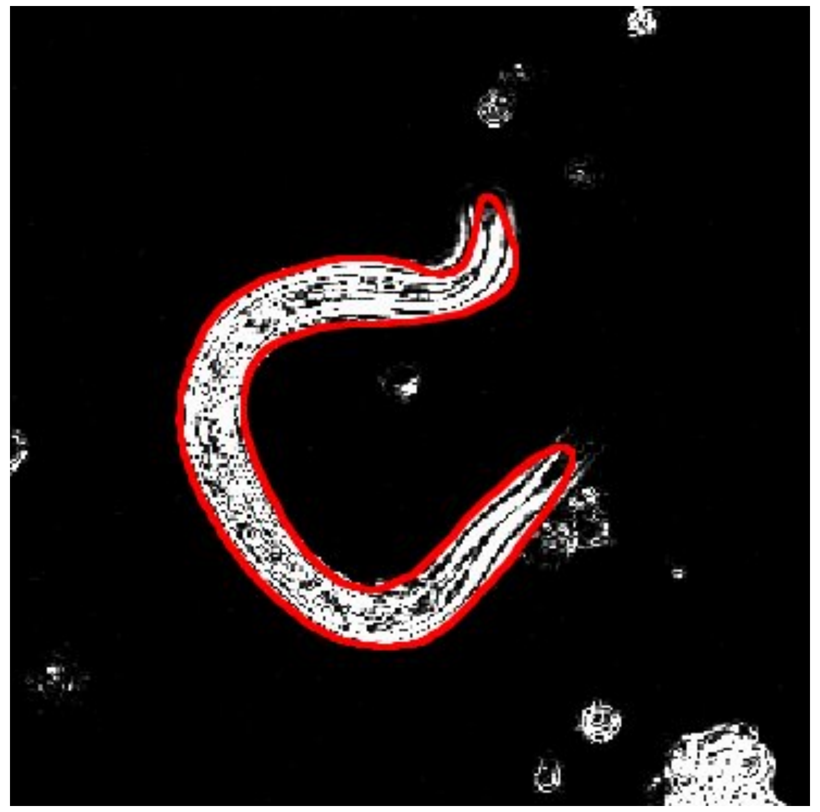


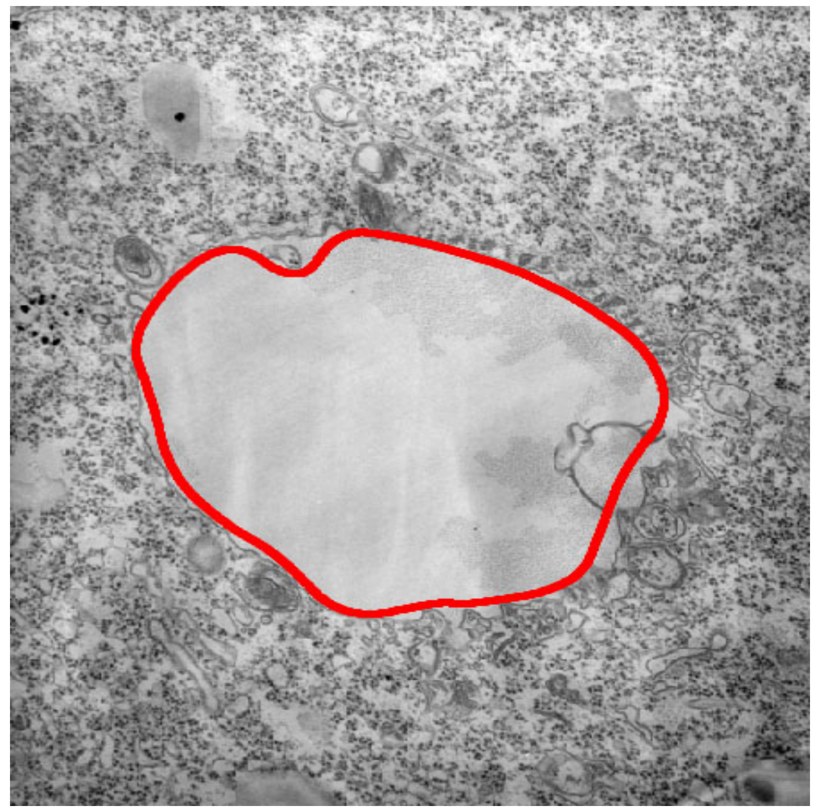
Sage et al. DeconvolutionLab2 Methods 2017.

SHAPE MODELING FOR SEGMENTATION

Mathematical formalism: shape representation

- Level-set
- Active contours
- Graph-cut
- Mathematical morphology





Badoual et al. Active Contours 2017.

Rule-based

Model-based

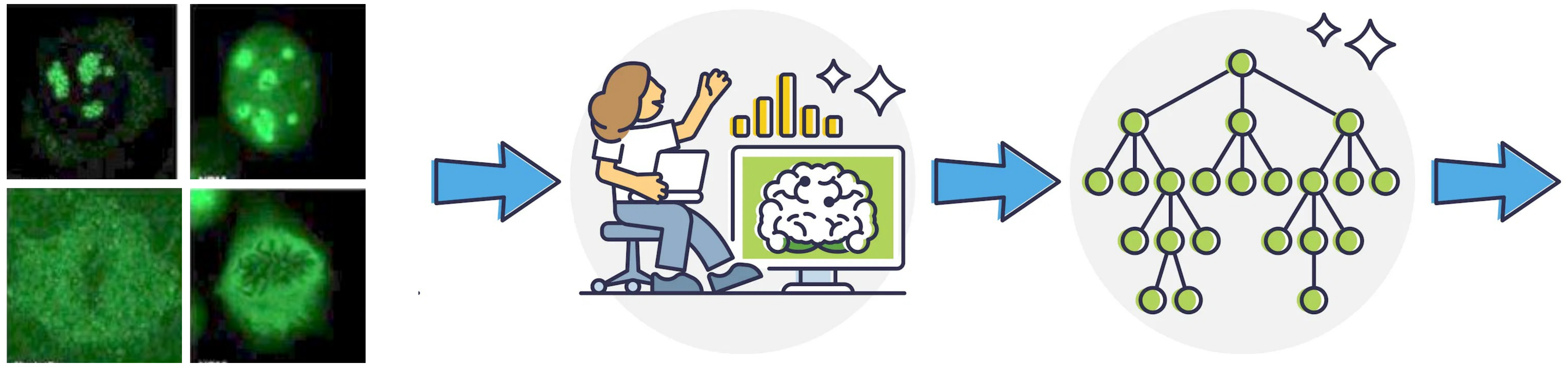
Machine Learning

Deep Learning

Trained Models

Foundation Models

Machine Learning



Handcrafted features: filterbank

Classification: supervised learning

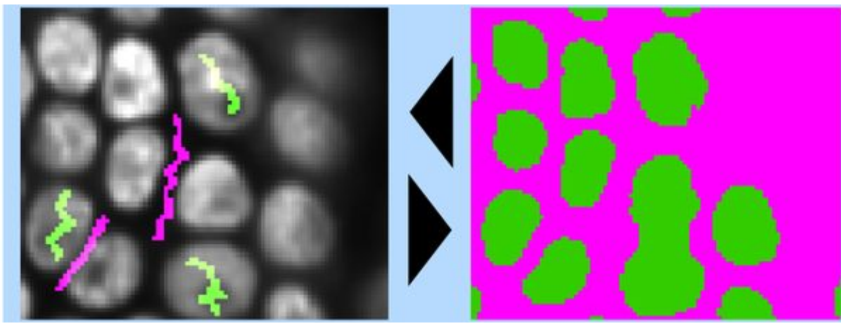
Anaphase
Metaphase
Interphase

Pixel Classification Annotation by user interface and feature classifier

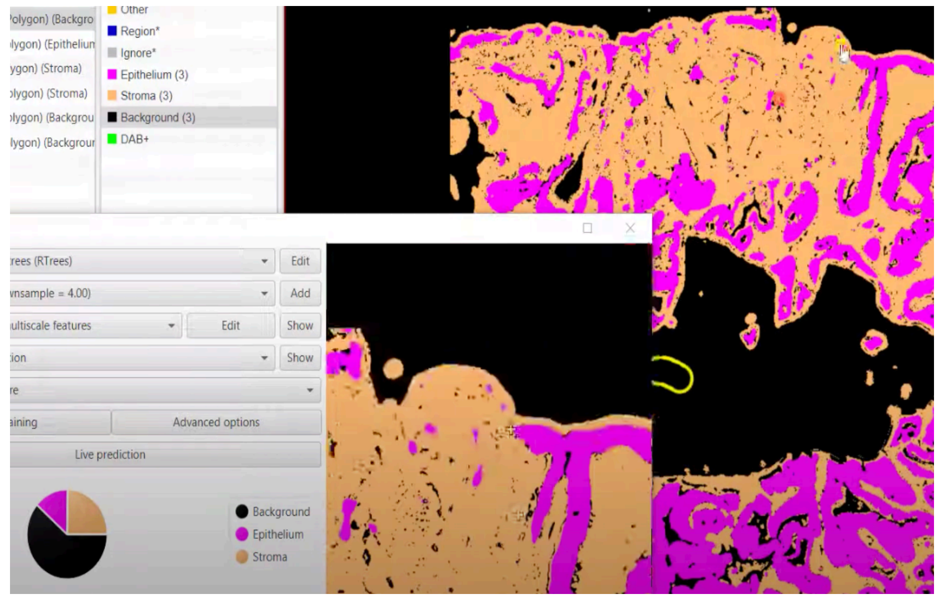
Fiji

WEKA The University of Waikato

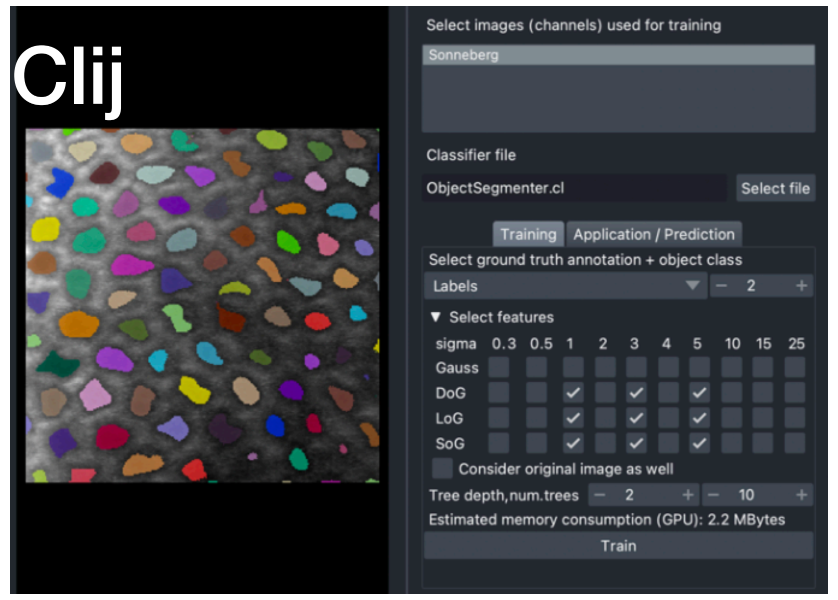
LabKit



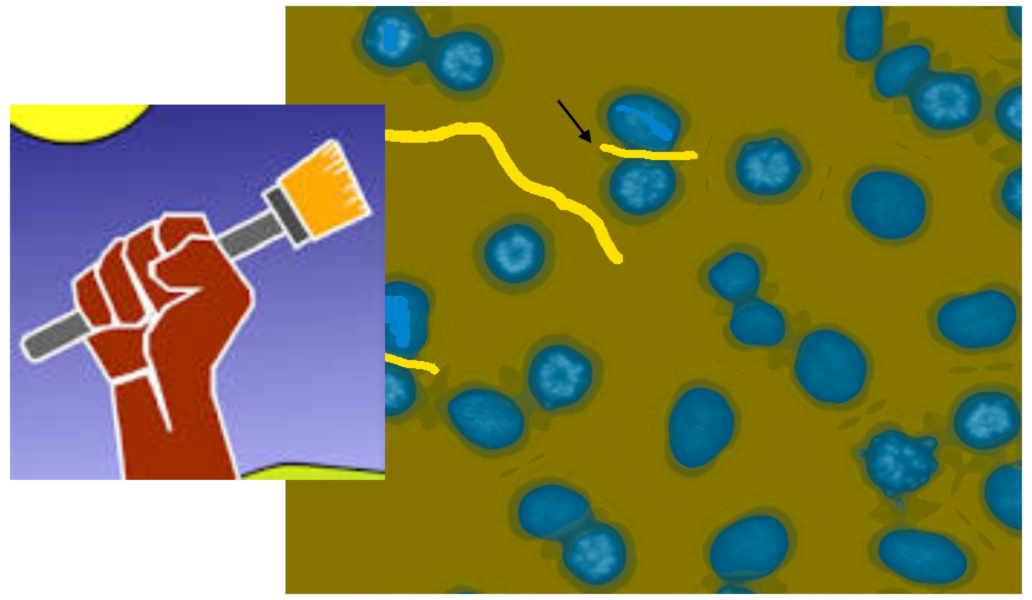
QuPath



Napari



Ilastik



Rule-based

Model-based

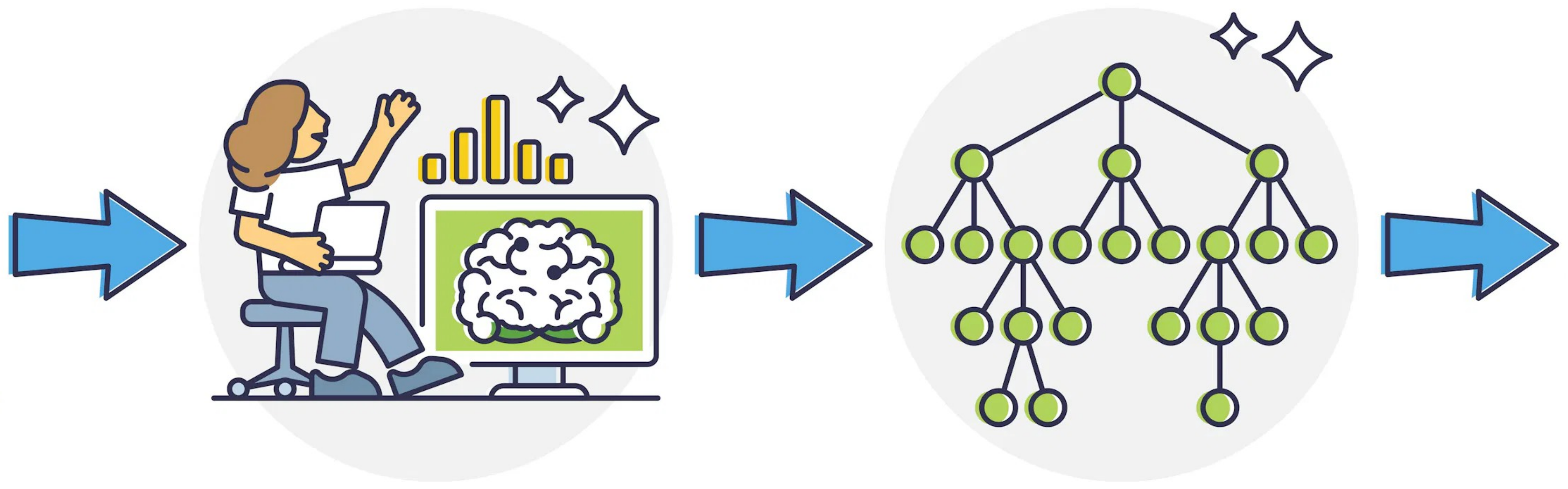
Machine Learning

Deep Learning

Trained Models

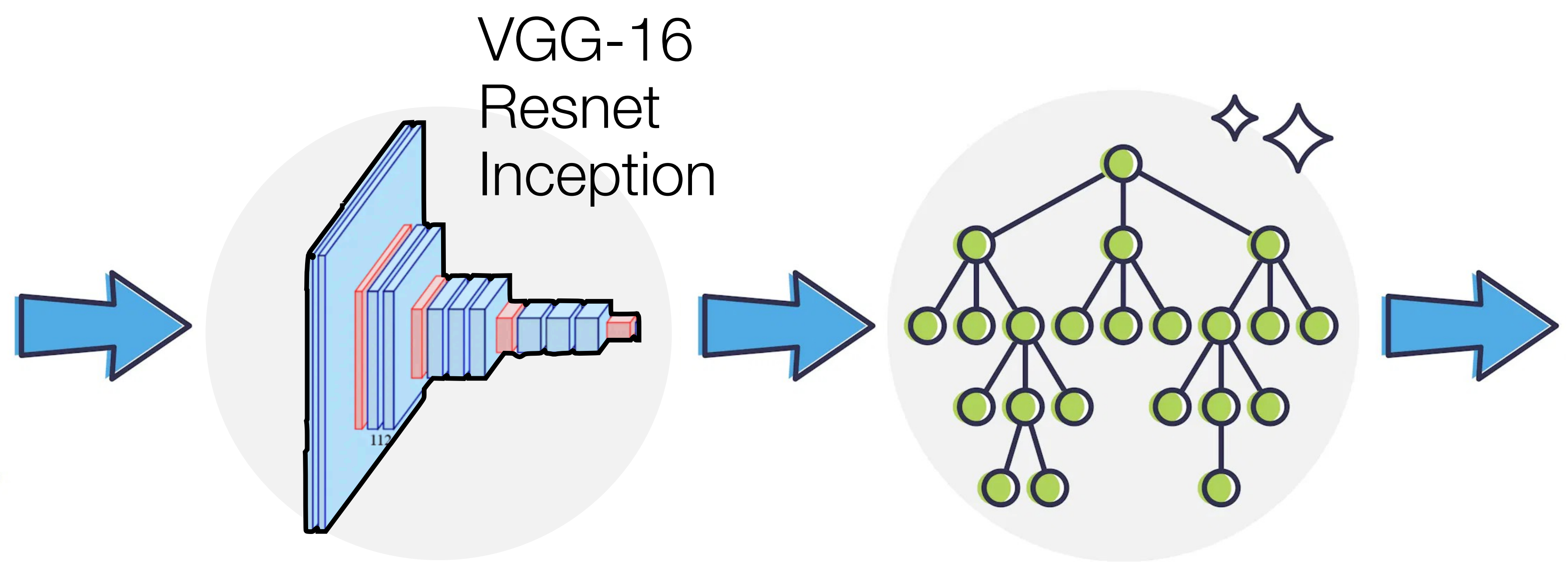
Foun-dation Models

Machine Learning



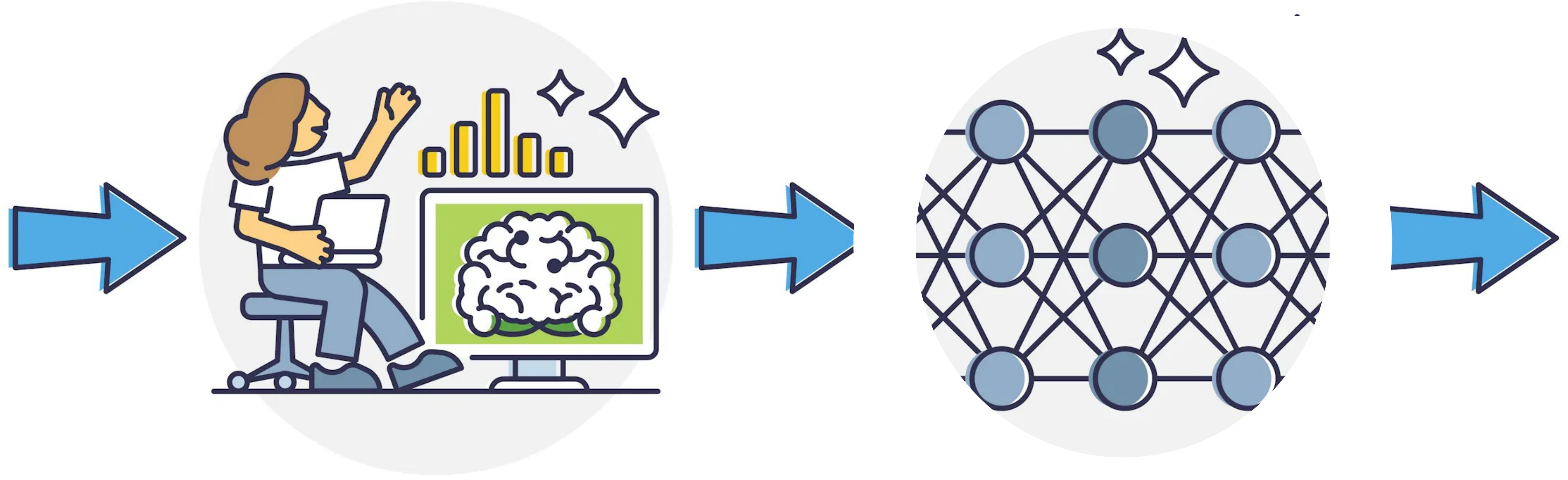
Hand-crafted features

Random-Forest



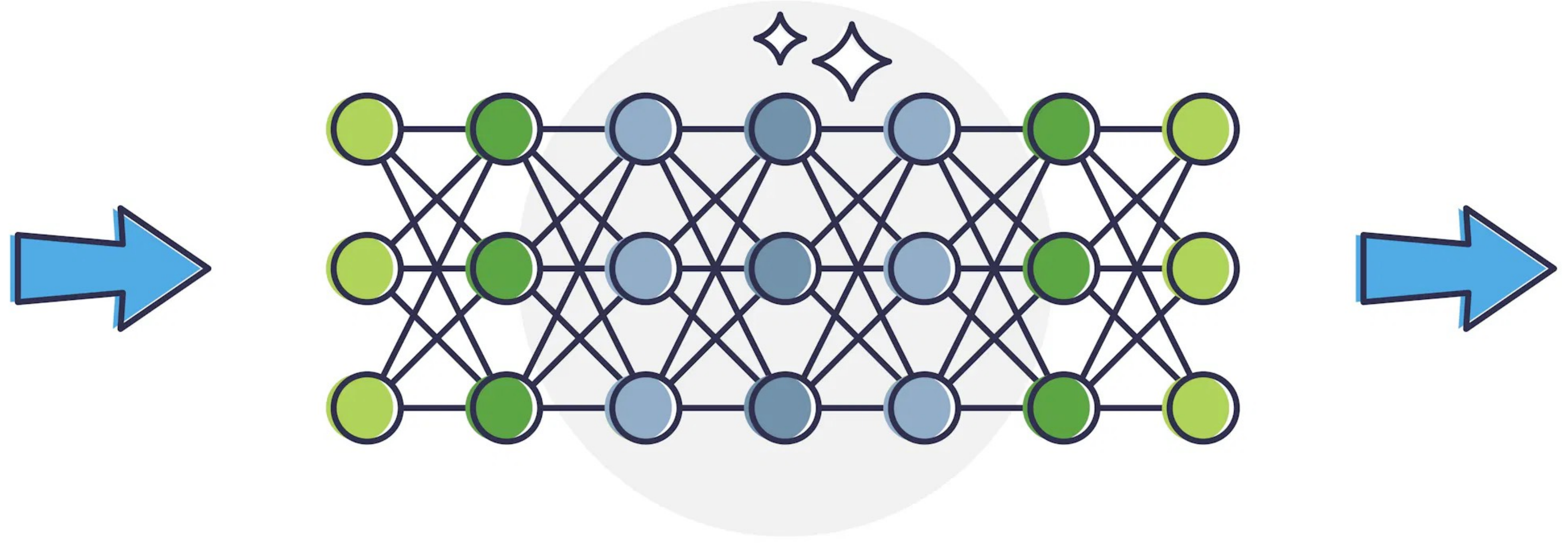
Feature extraction

Classifier



Hand-crafted features

Fully connected network



End-to-end learning -> Deep architecture

Rule-based

Model-based

Machine Learning

Deep Learning

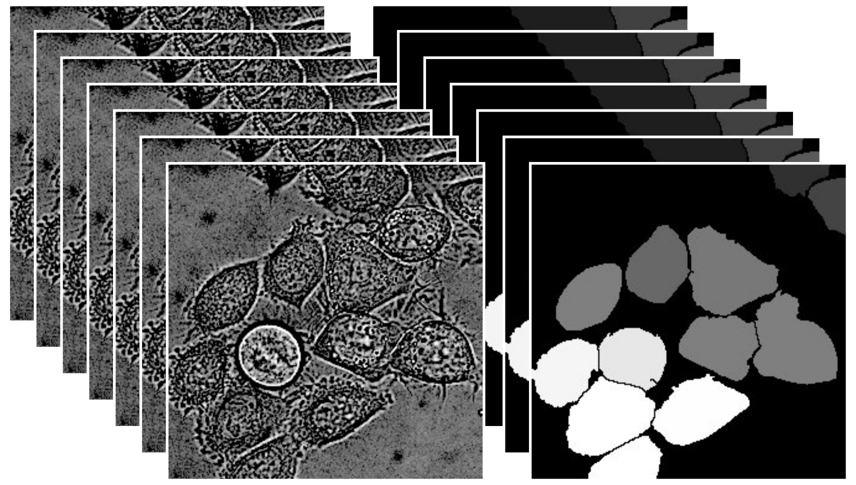
Trained Models

Foundation Models

Deep Learning — Supervised Training

Training

Pairs of images
Training dataset



raw annotated

Trained model

Inference

Trained model

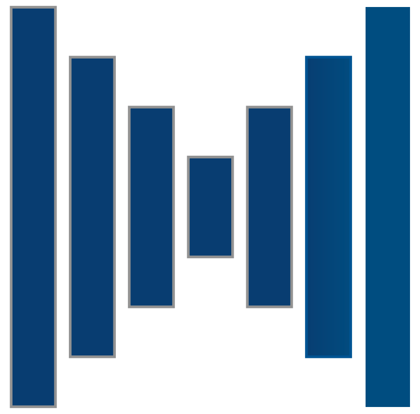
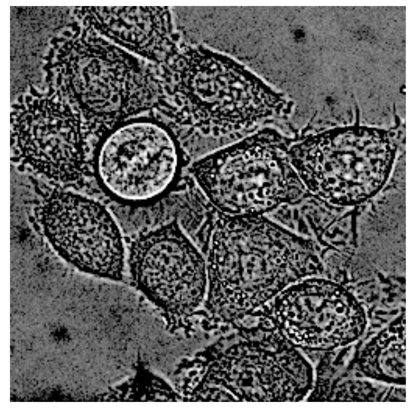
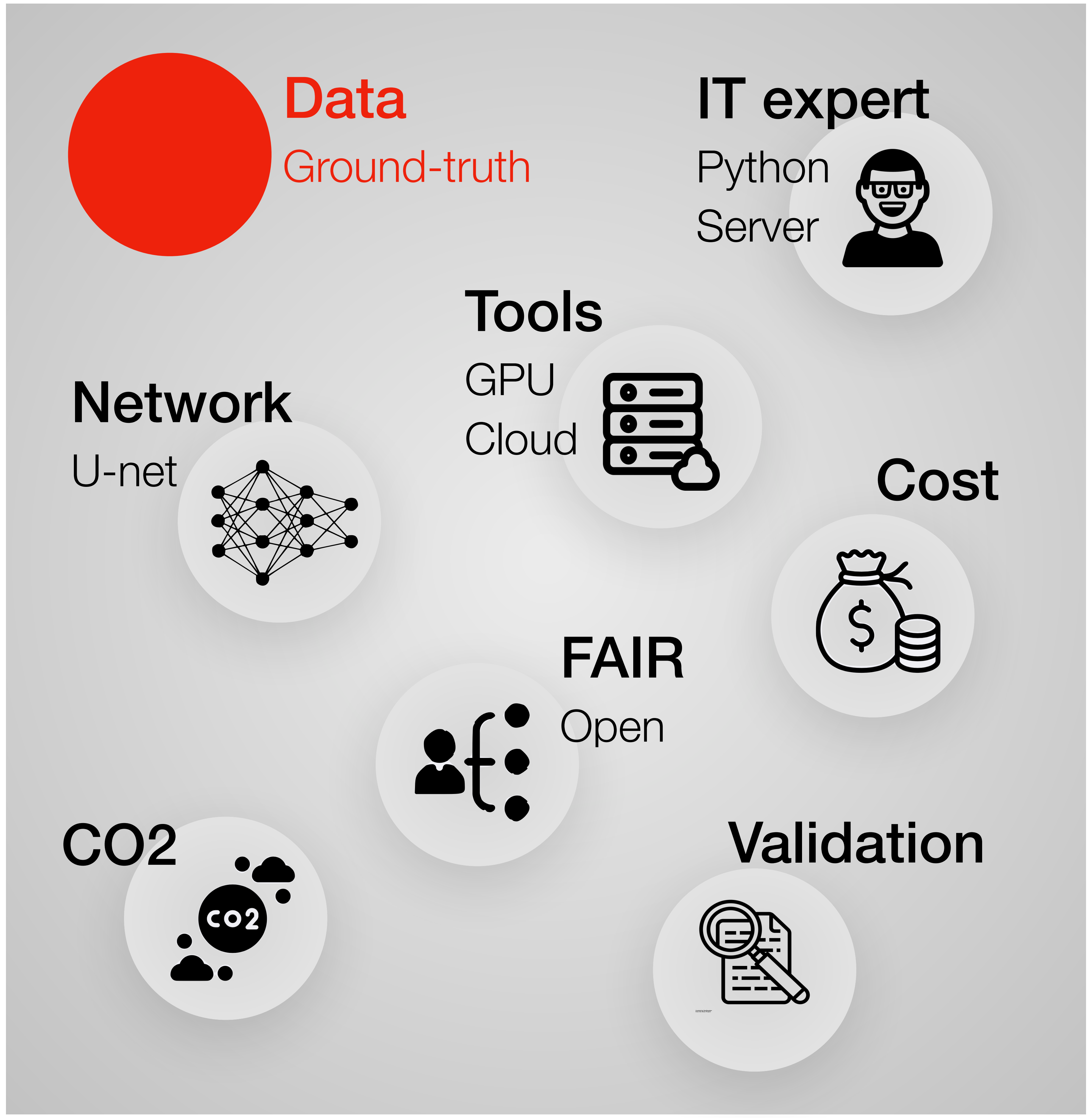


Image to analyse



raw

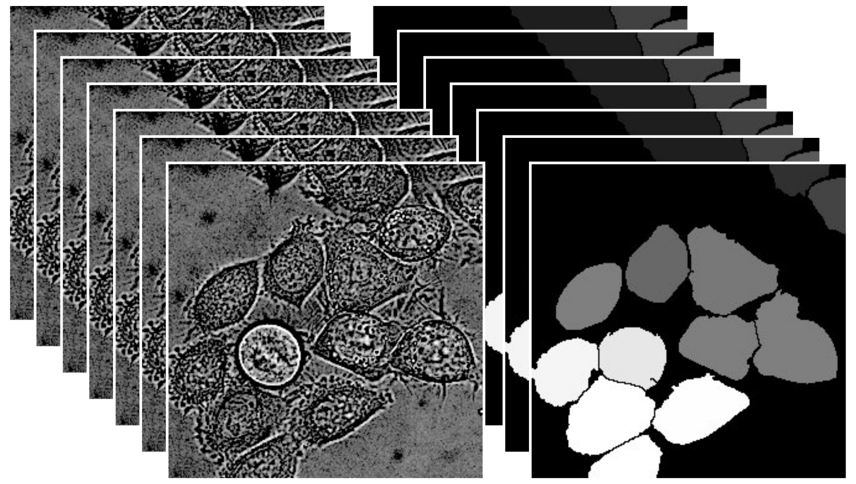
Prediction



Deep Learning — Supervised Training

Training

Pairs of images
Training dataset



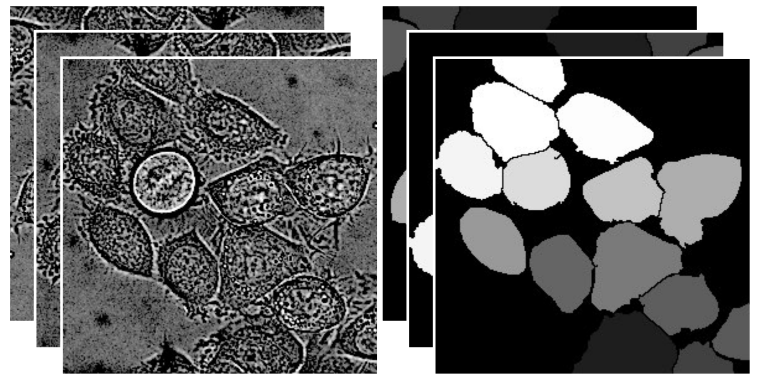
raw annotated

Trained model

Fine Tuning

Trained model

Few pairs of application-dependent images



raw annotated

Specific trained model

Inference

Trained model

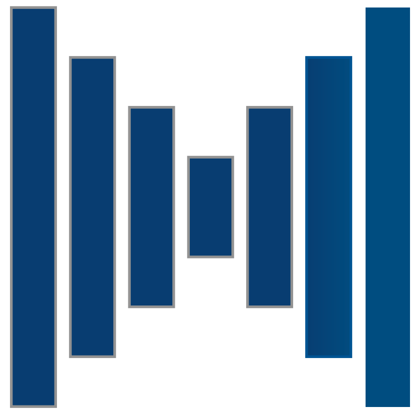
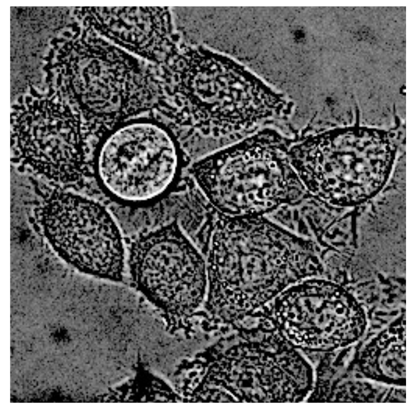


Image to analyse



raw

Prediction

Rule-based

Model-based

Machine Learning

Deep Learning

Trained Models

Foundation Models

Selection of a Trained Model

V. Uhlmann, L. Donati, D. Sage, IEEE SPM 2022

Rule-based

Model-based

Machine Learning

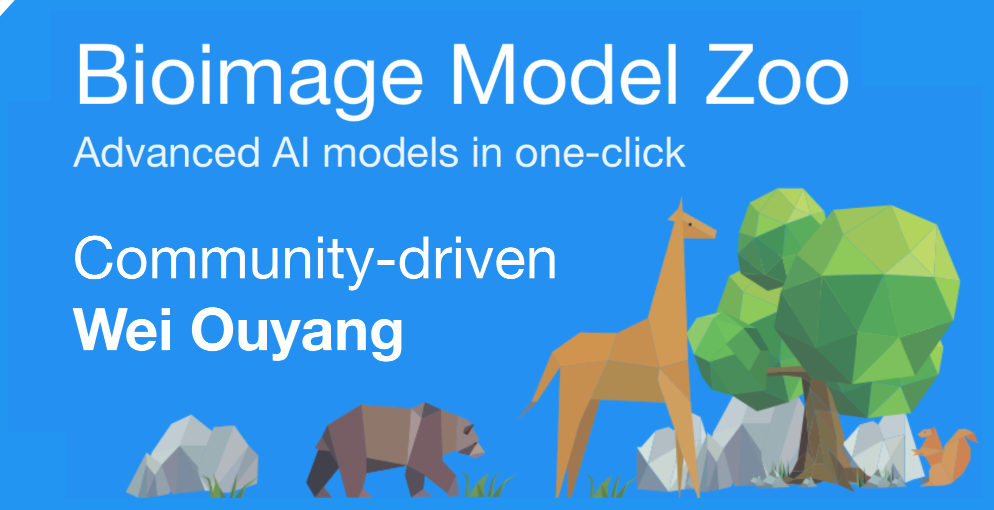
Deep Learning

Trained Models

Foundation Models

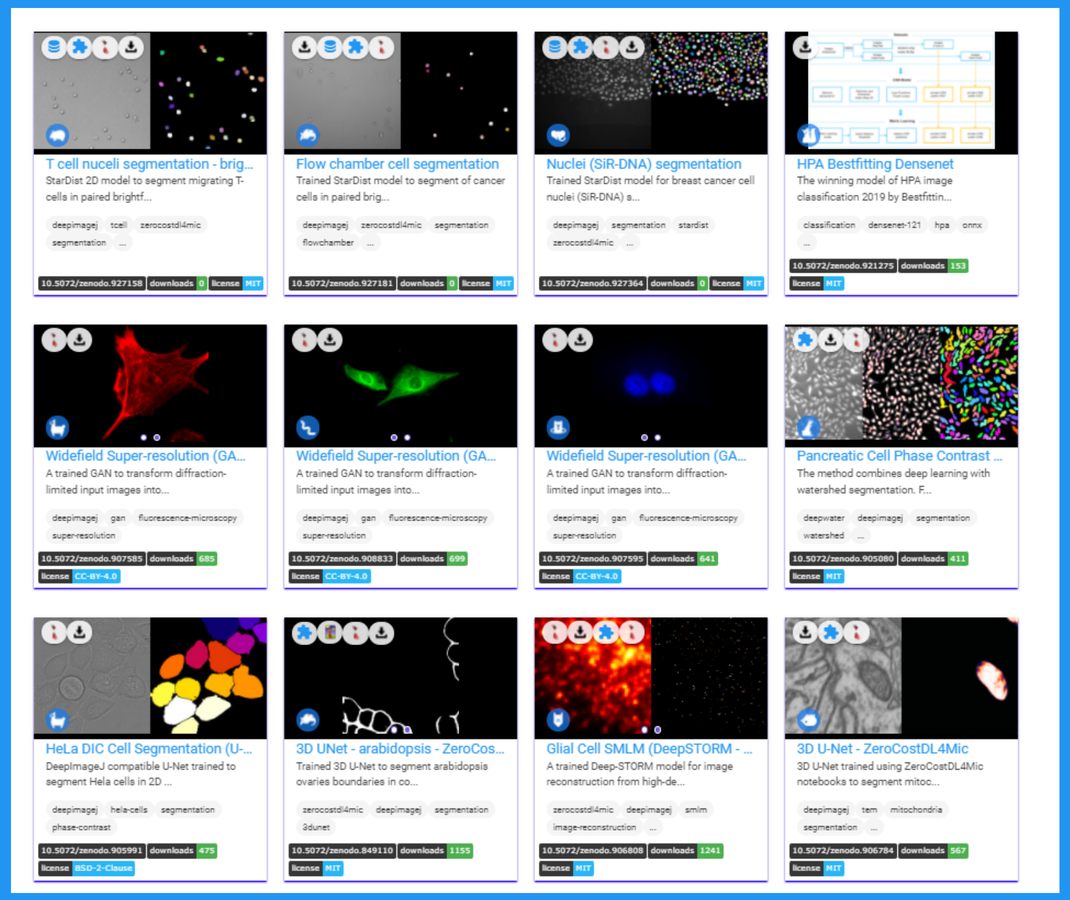
Bioimage Model Zoo

https://bioimage.io/



Community-driven
Wei Ouyang

W Ouyang et al. biorxiv, 2022



Facets

The BioImage Model Zoo and FAIR data principles are core facets of the AI4Life project.

AI4Life

Partners

#ZeroCostDL4Mic

DL4MicEverywhere

Interoperability | **Standardization** | **Reproducibility** | **FAIR Open**

Rule-based

Model-based

Machine Learning

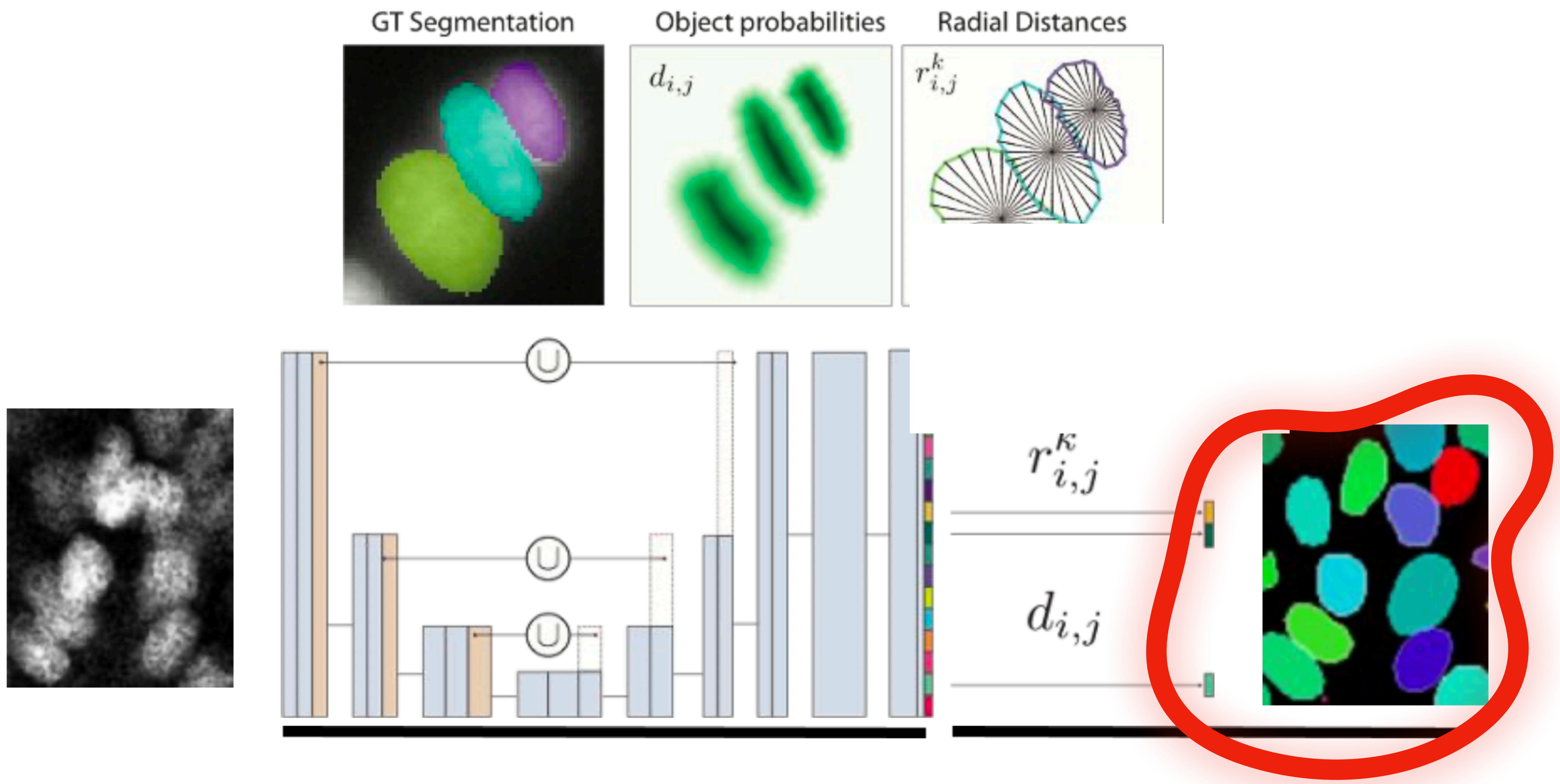
Deep Learning

Trained Models

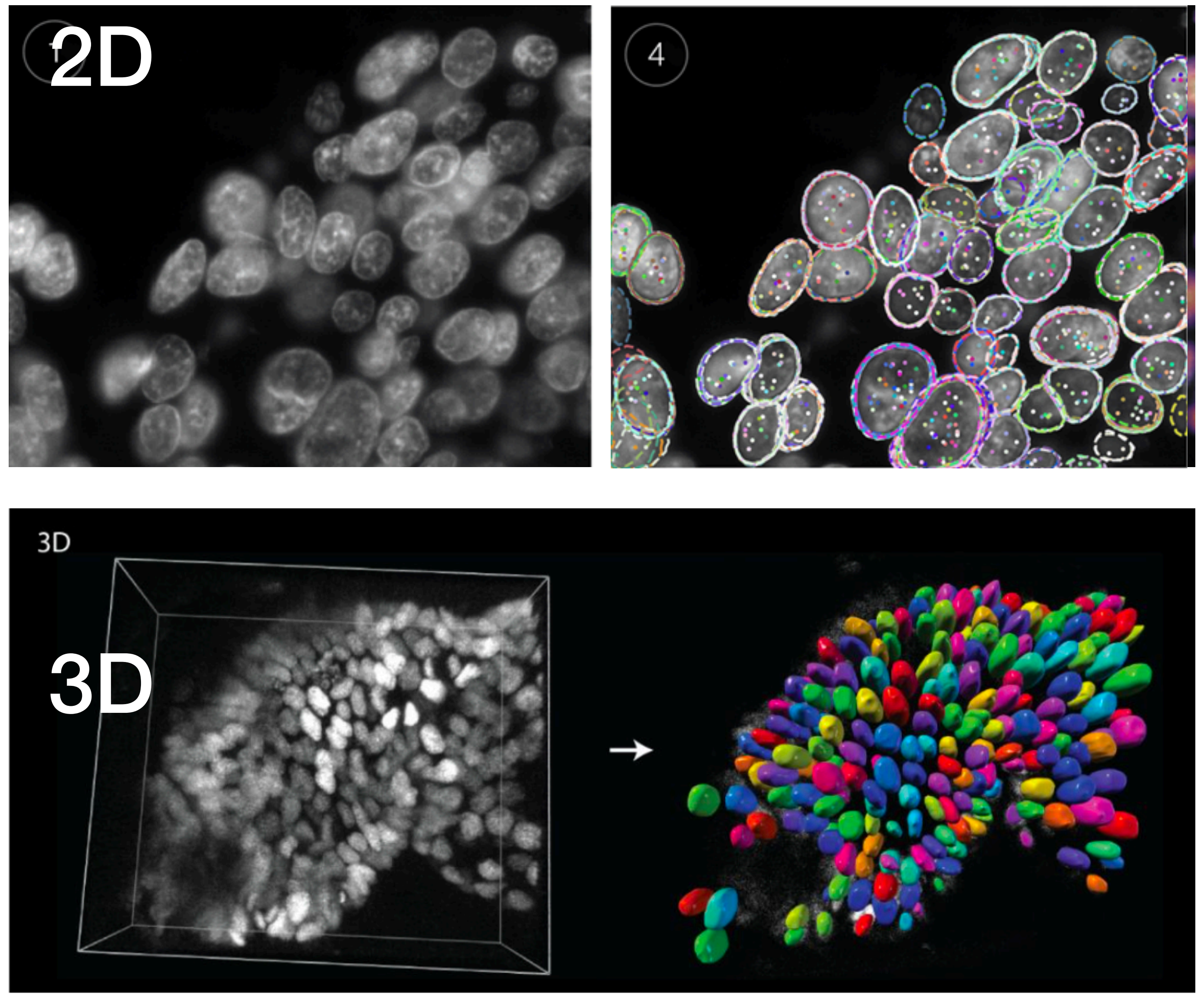
Foun-dation Models

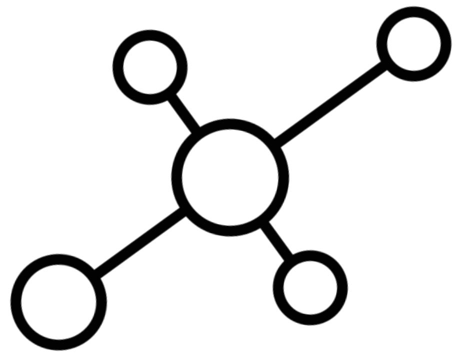
Stardist

A cell/nuclei detection method for images with star-convex shape priors (2D/3D) [Schmidt et al. MICCAI 2018]



POST-PROCESSING
NON-MAXIMUM SUPRESSION





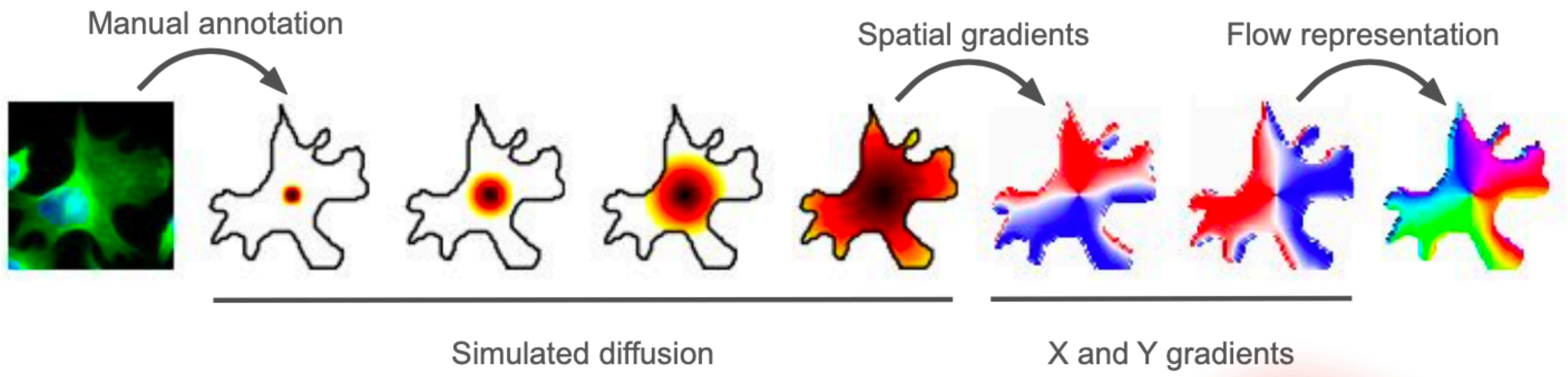
Python package, napari
QuPath, Fiji
CellProfiler
ZeroCostDL4Mic

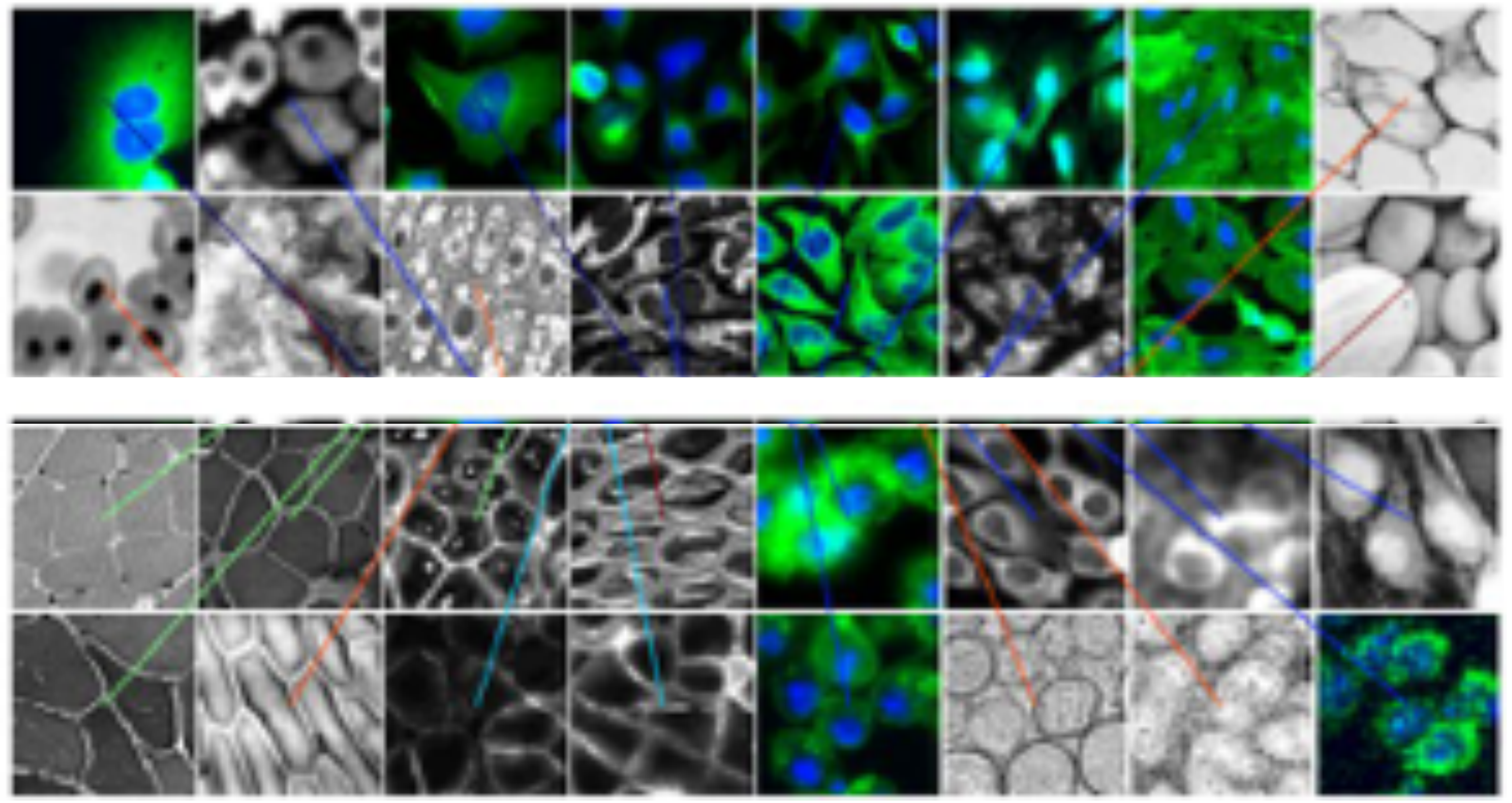
Rule-based
Model-based
Machine Learning
Deep Learning
Trained Models
Foun-dation Models

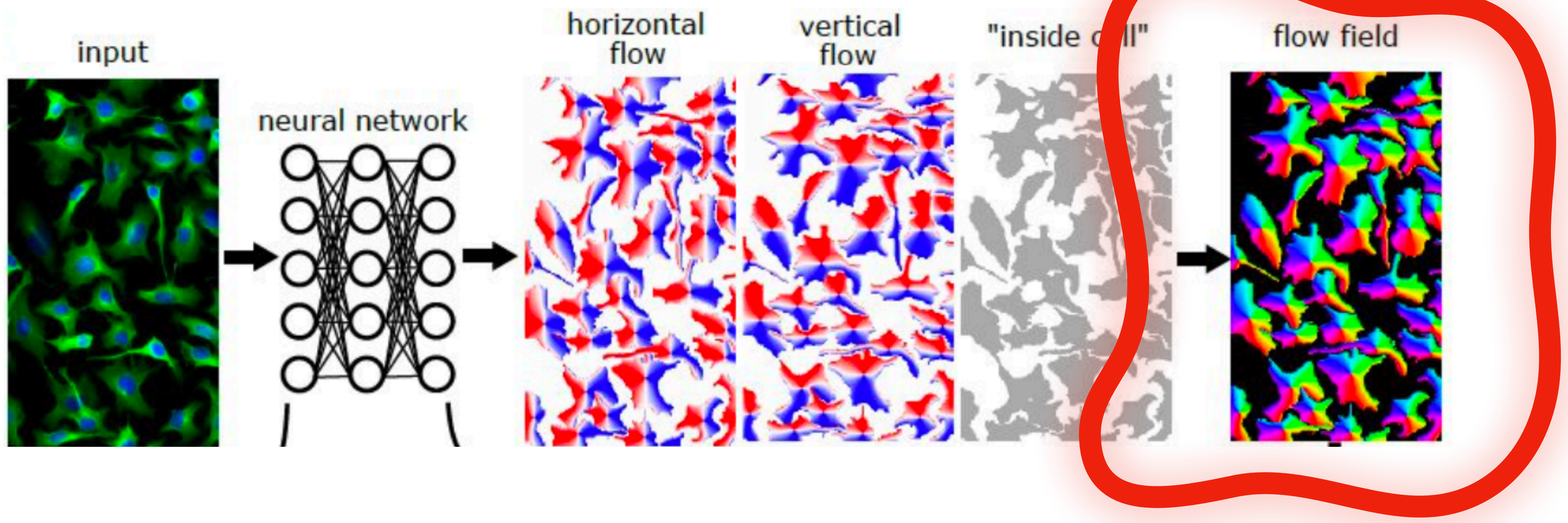
Cellpose

- Pre-trained model a wide range of images
- 70,000 manual annotated objects

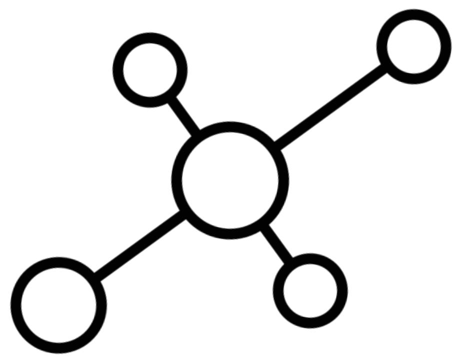
Stringer et al., "Cellpose: a generalist algorithm for cellular segmentation", Nature Methods 2021







POST-PROCESSING: GRADIENT-FLOW

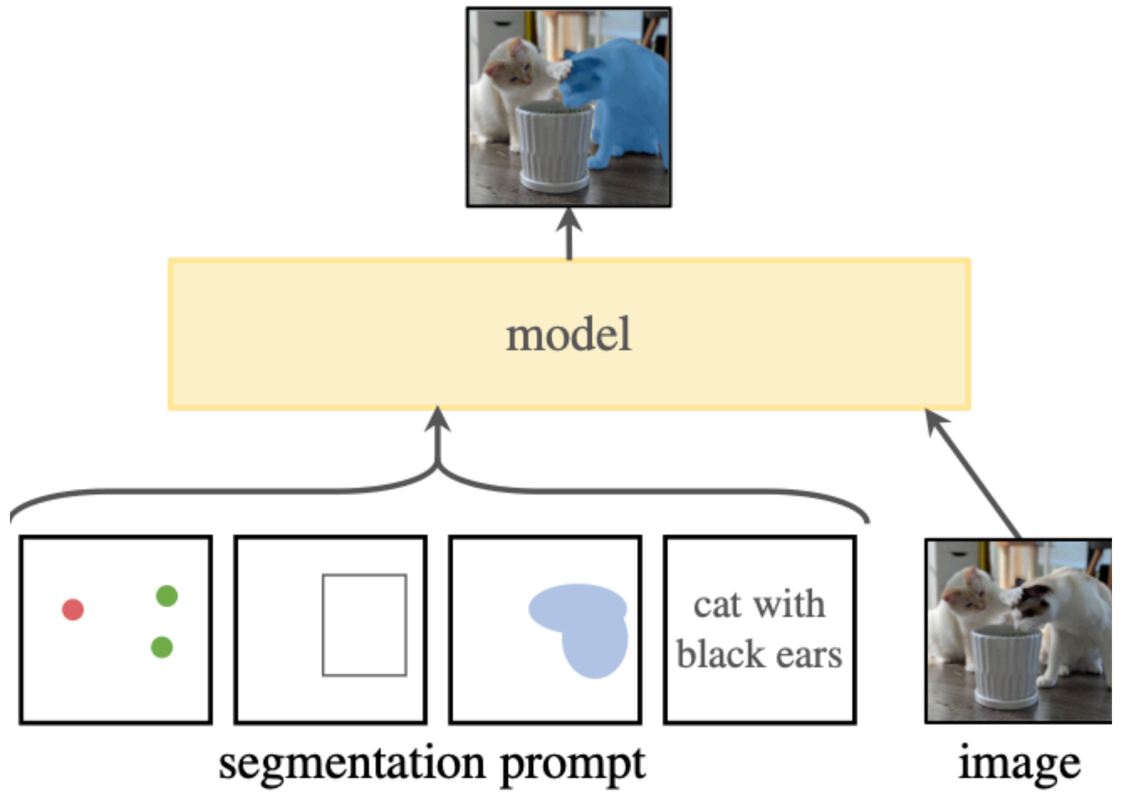


Web interface
Python package
Fiji, Napari
ZeroCostDL4Mic

Sketchpose — Omnipose

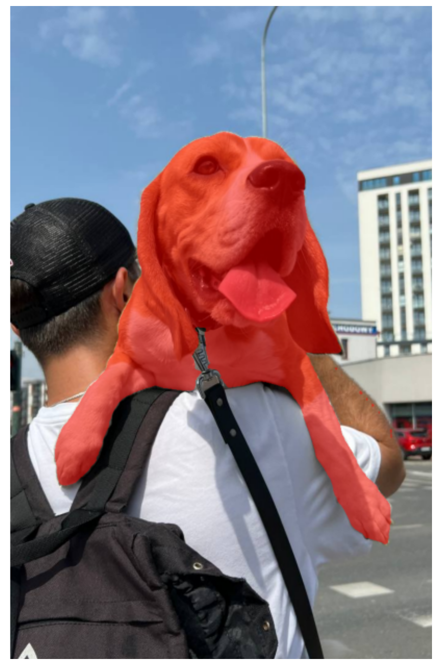
Segment-Anything Model (SAM)

- Foundation model from MetaAI
- Transform: encoding / decoding



PROMPT

The ChatGPT of the Computer Vision



BIG DATA

Model SA-1B

- Natural photographies
- Huge model (~1GB)
- 11M diverse, high-res. images
- 1.1B segmentation masks
- Open, privacy

Alexander Kirillov et al. IEEE/CVF, 2023, 2700 citations

Rule-based

Model-based

Machine Learning

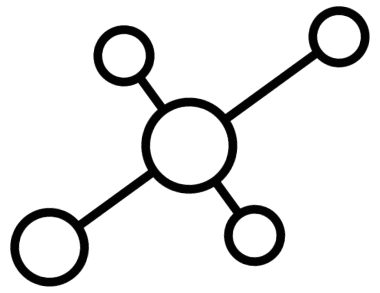
Deep Learning

Trained Models

Foun-dation Models

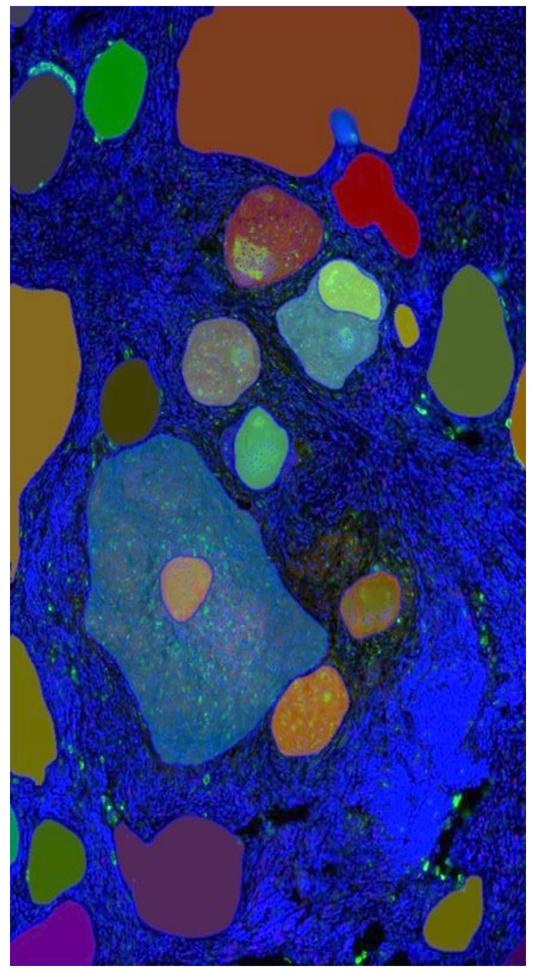
SAM for Science?

Web interface
Python package
QuPath
Napari
Fiji

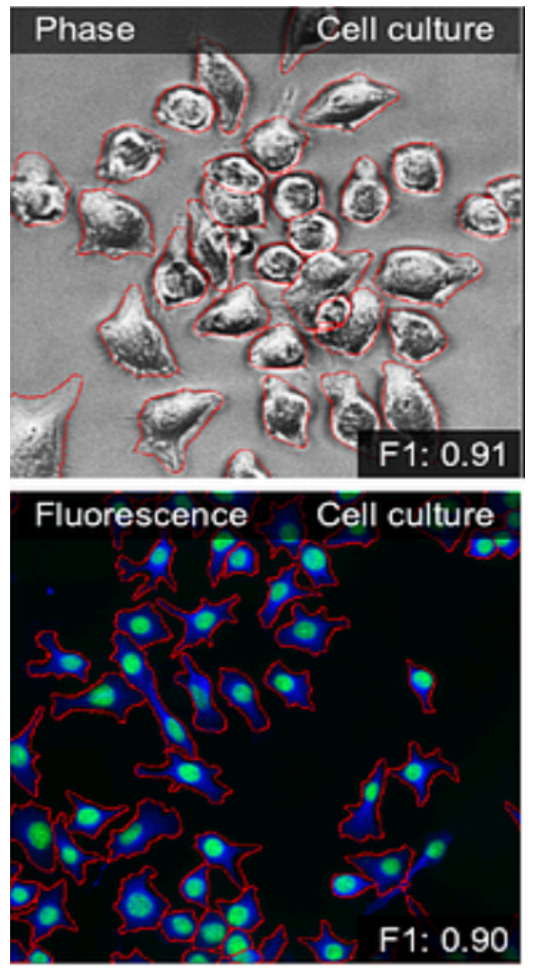


Variants of SAM Models

- MicroSAM
- MedSAM
- CellSAM
- EfficientSAM
- MobileSAM
- …

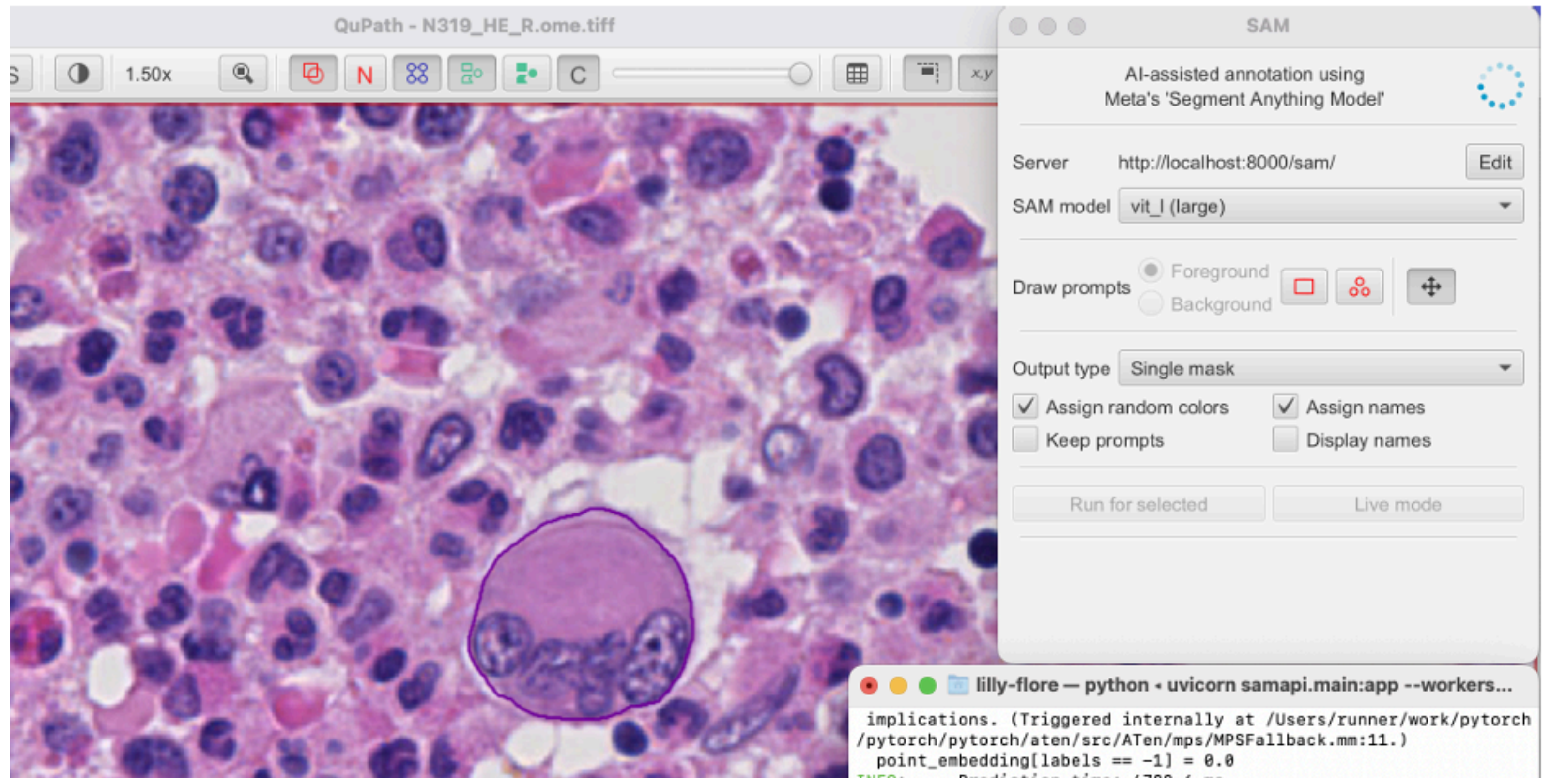


MicroSAM
C. Pape



CellSAM

ARCGIS



Acceleration of annotations

Megakaryocotes on human biopsis

SAM Large model
SAM extension of WuPath
SAM on server
R. Sarkis, CHUV, L-F. Celma, EPF
April 2024

Two Paradigms of Image Analysis

Model-based		Machine Learning
Based on a-priori model		Based on data
Physical rule biophysics law, prior knowledge of the sample	Physics vs. Data	**Adaptivity to the data** no explicit model 🙂
Insight of the structure of the problem Convergence prove, intuitive parameters	Process Control	**Blackbox** no guarantees, validation on data
Results explainable 🙂 Mathematics, error analysis	Explainability	**Not explainable** Working in progress, tools
Generalization Under known assumptions	Generalizability	**Strong claim of IA** Questionnable aspect
Very efficient ... Very inefficient from simple algorithms to iterative algorithms	Computing	**Training: slow / GPU** Prediction: efficient

Difficult to manually tweak the parameters of the model because a DNN has millions of parameters inside

Expert-Driven

Data-Driven

Introduction to Microscopy Image Analysis

Lecture for the workshop AI4Life given by Daniel Sage, 10 June 2024

CONTEXT — BIOIMAGE INFORMATICS

METHODS — MODEL-BASED VS. DATA-DRIVEN

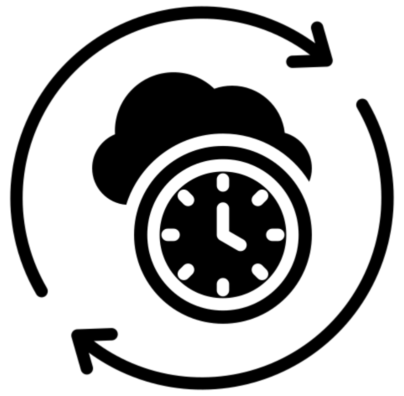
LEARNING — DATA AND TOOLS

WRAP UP — BIOIMAGE ANALYSIS

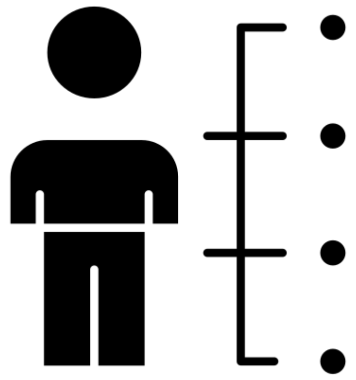
When Shall I Use Deep Learning?

No explicit model of the objects of interest

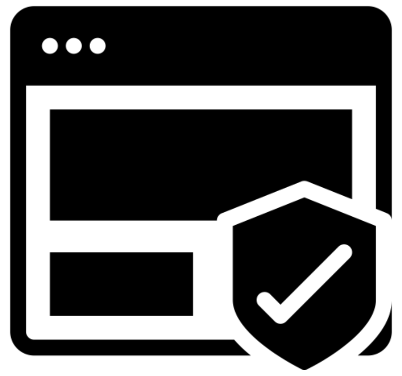
- Abstract features
- Complex relationships
- Hierarchical features
- No physical rules / engineering



How much time and resources?



What are the required skills?



How evaluate the accuracy?

PROMISES OF DEEP LEARNING

☞ **End-to-end learning**

☞ **Continual improvement**

☞ **Ability to generalization**

Importance of the Data

Building a training dataset is a scientific process including preparation, unbiased, curation, annotation, validation, integrity, open-access.

Dataset size

Few ground-truth **Overfitting**

Normalized data

Misaligned raw data **Divergence**

Model complexity

Few # parameters **Underfitting**

Representativity

Mismatch conception **Dataset shift**

Data selection

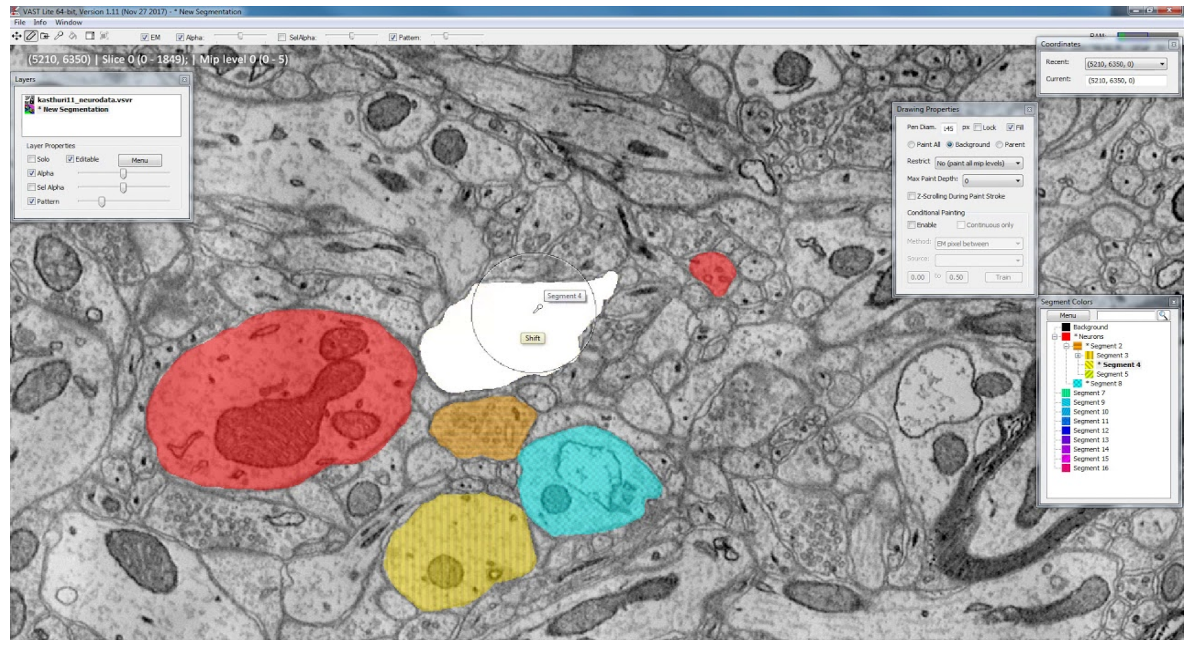
Exclude phenotypes **Bias**

Distribution

Imbalance classes **Ignore minority**

Ground-Truth Data

Annotation

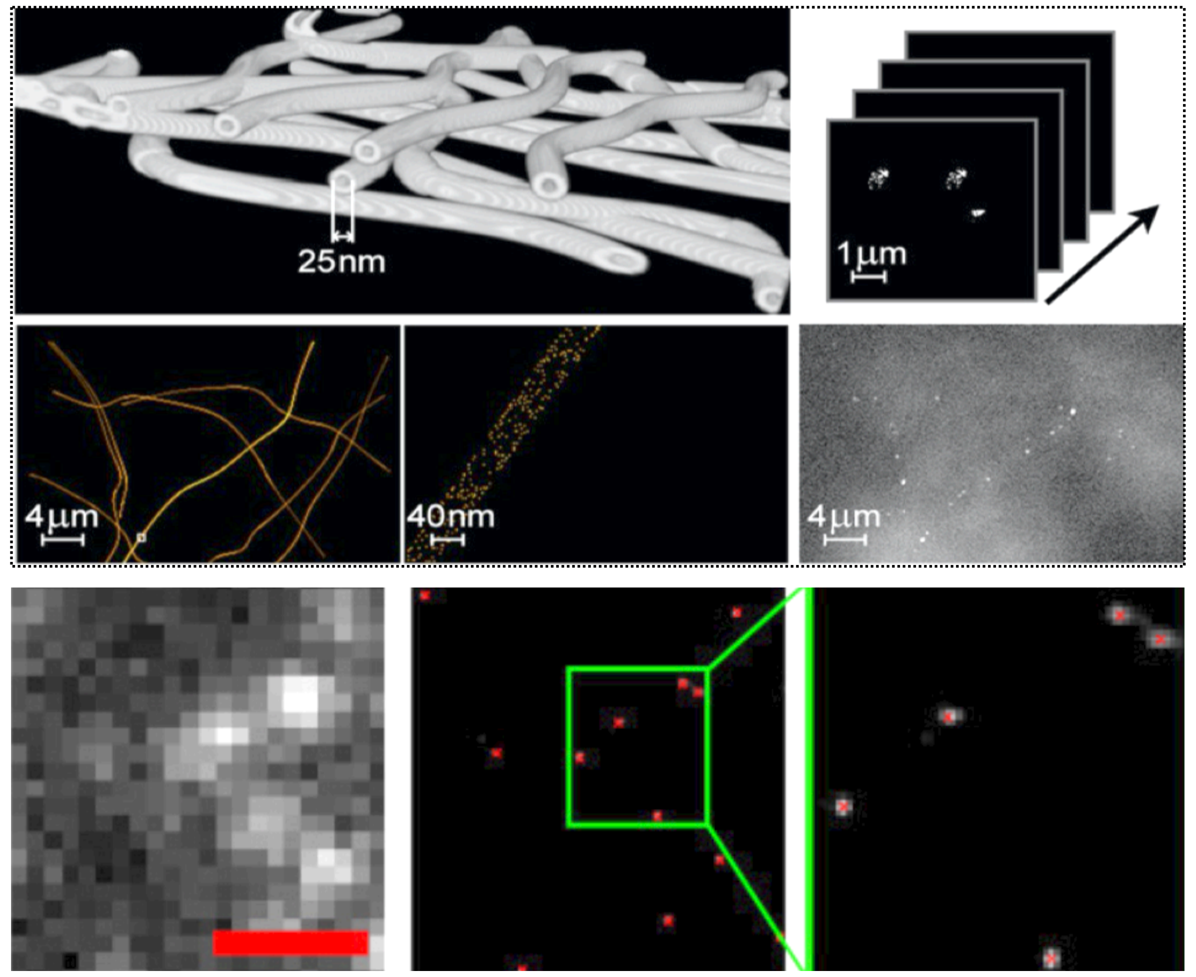


Crowd-source

Public Databases

ImageNet, Bioimage Archive, Broad

Simulation

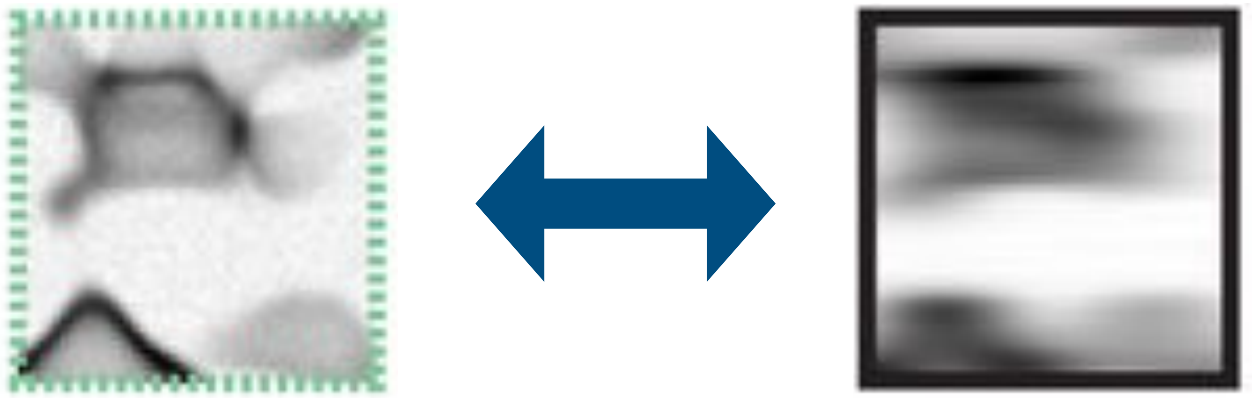


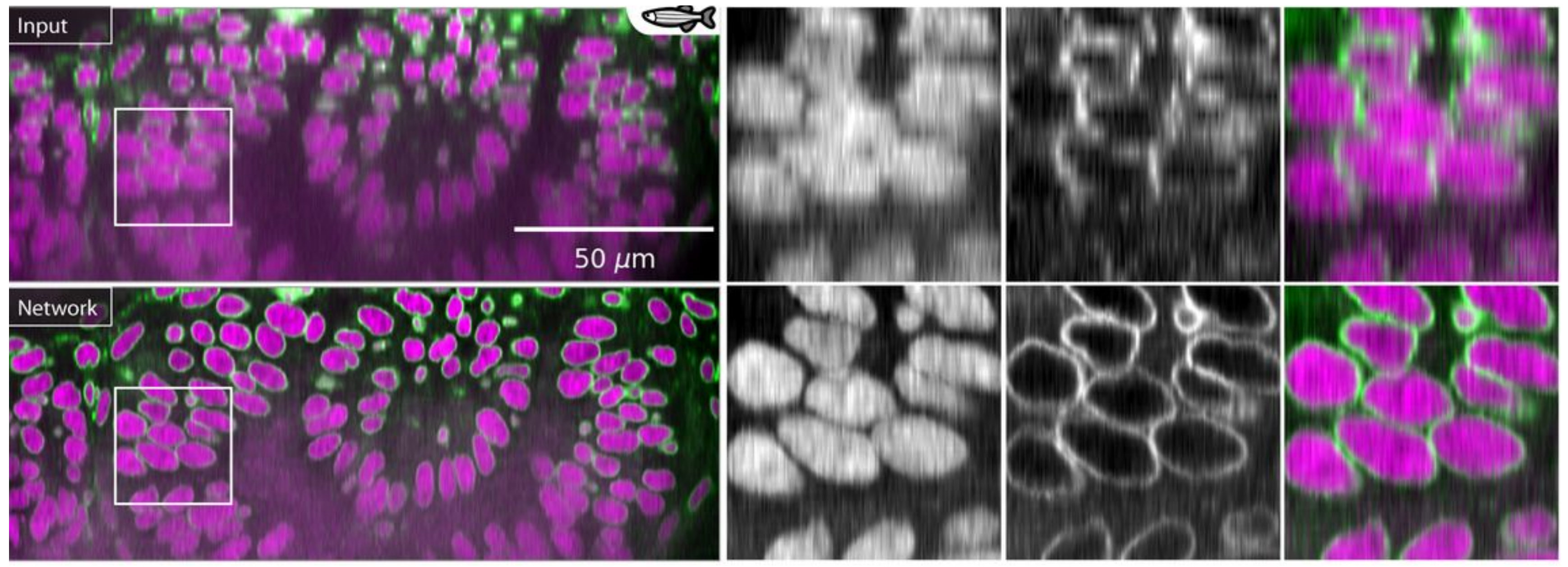
Content-Aware

Smart Acquisition

- Low SNR and high SNR
- SR and resolution standard
- Different focal plane
- Axial vs. Lateral

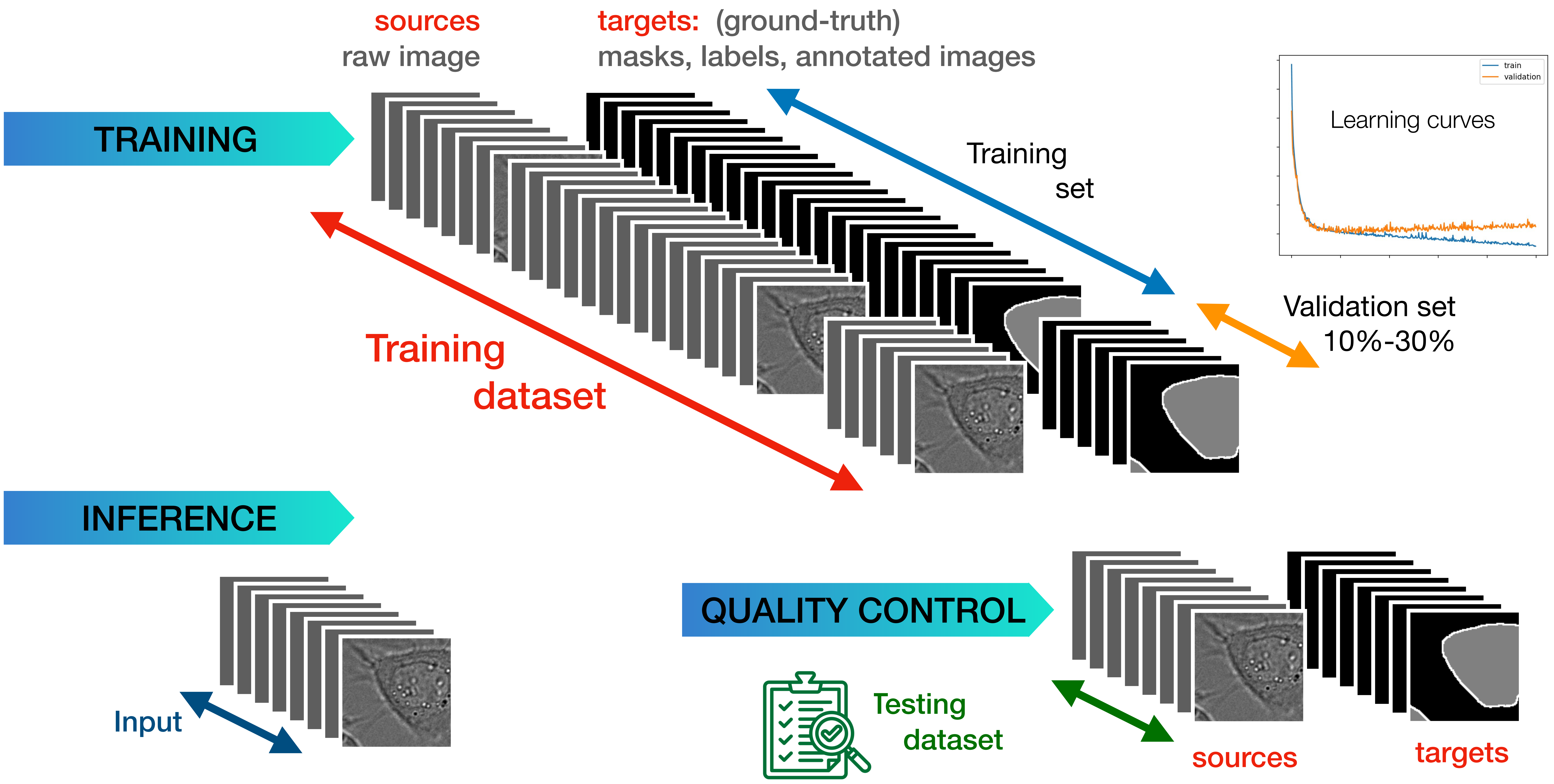
Lateral plane **Axial plane**





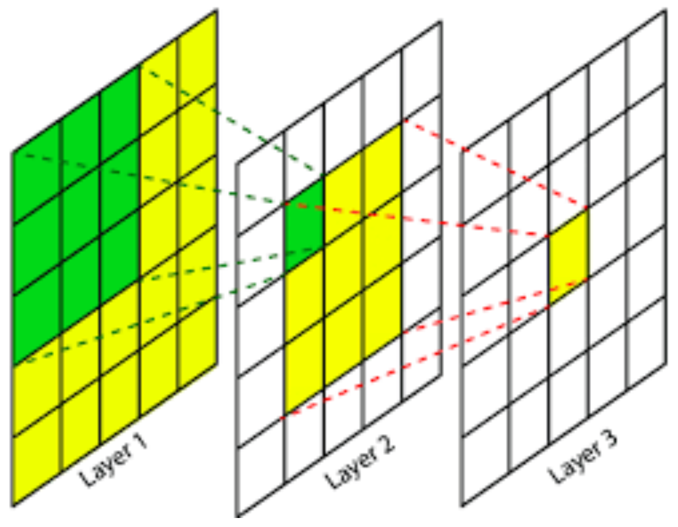
CARE, M. Weigert, 2019

Training Datasets



Deep-Learning Model?

MODEL = ARCHITECTURE + LEARNT PARAMETERS

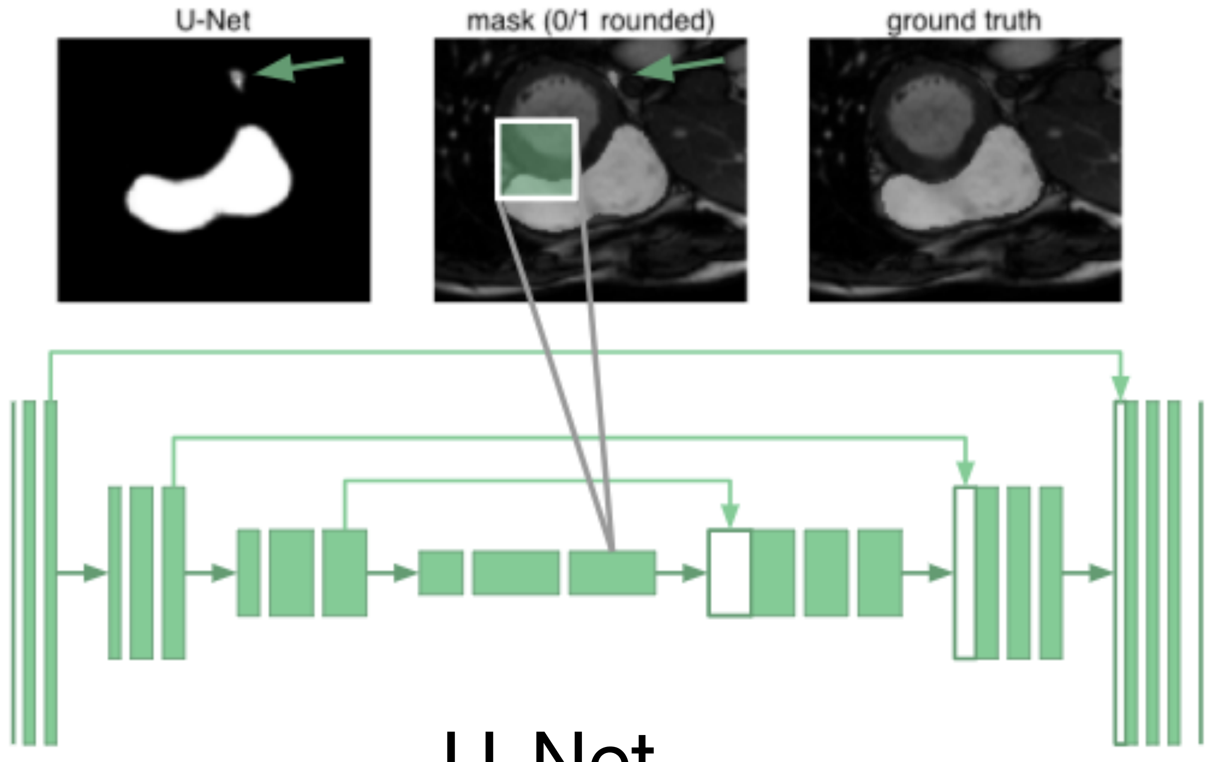


Neural Network Architecture

- **CNN Convolutional neural network**
- **Deep** multiple layers that gradually extracts higher-level features

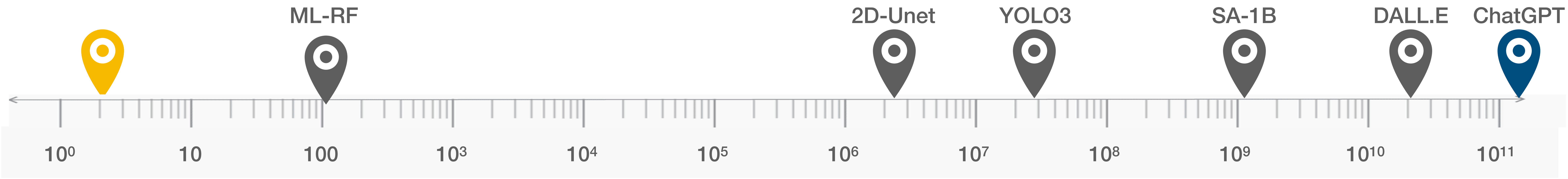
Parameters

- Weights of convolution kernel + bias of non-linearity
- **Trained** parameters using **data** and a criteria to minimize (**metric**)

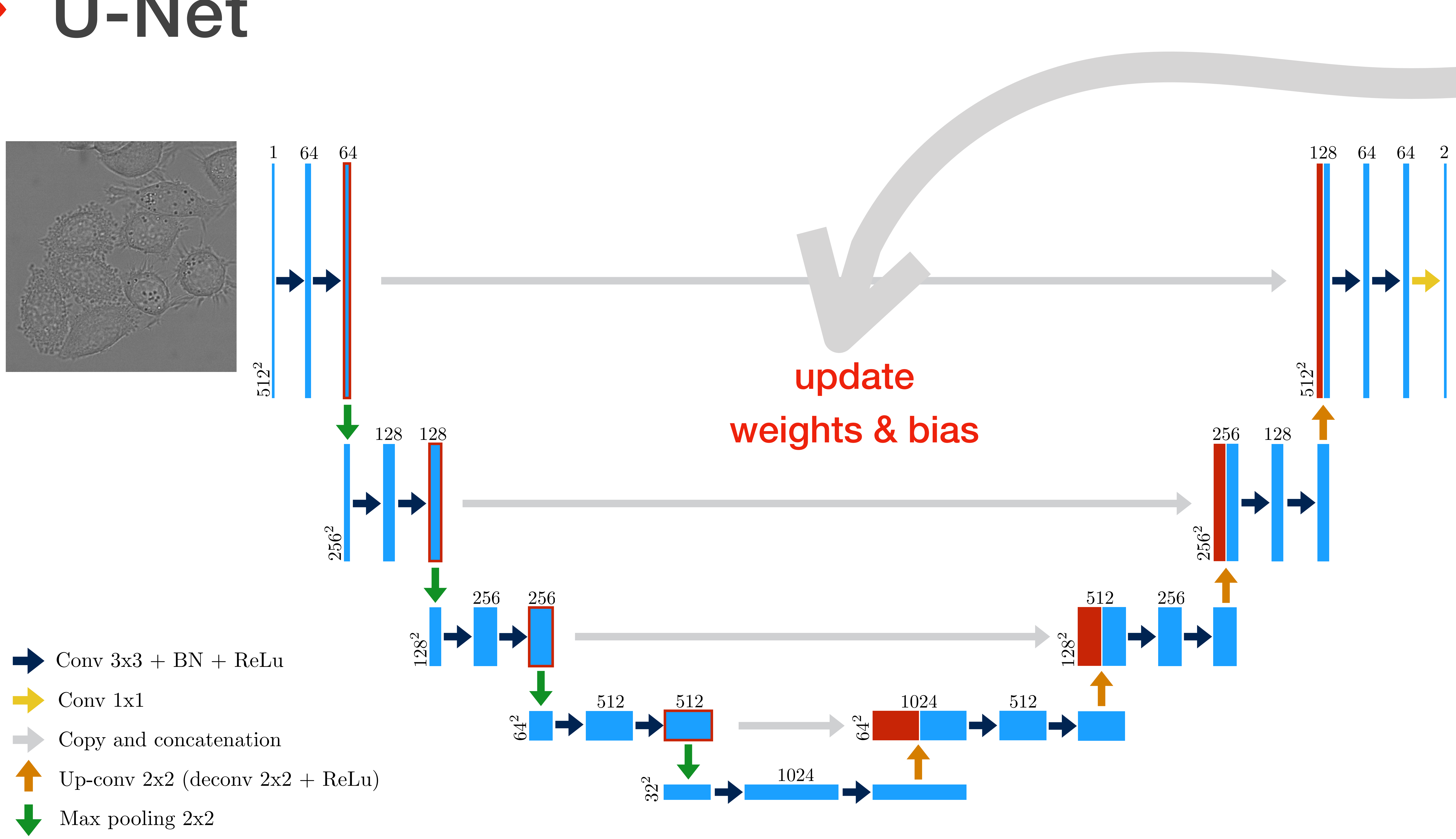


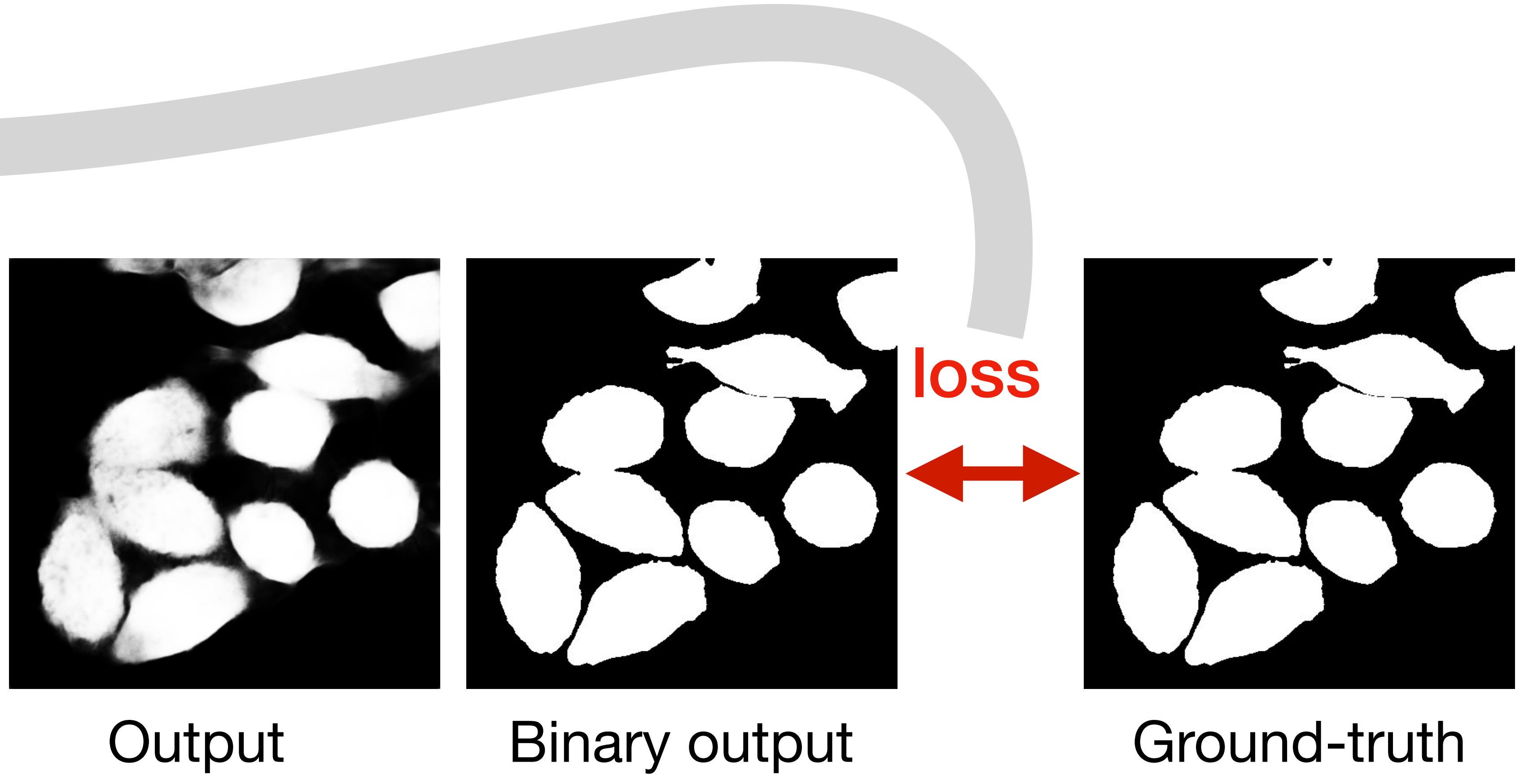
U-Net

+ METADATA **interoperability, open**

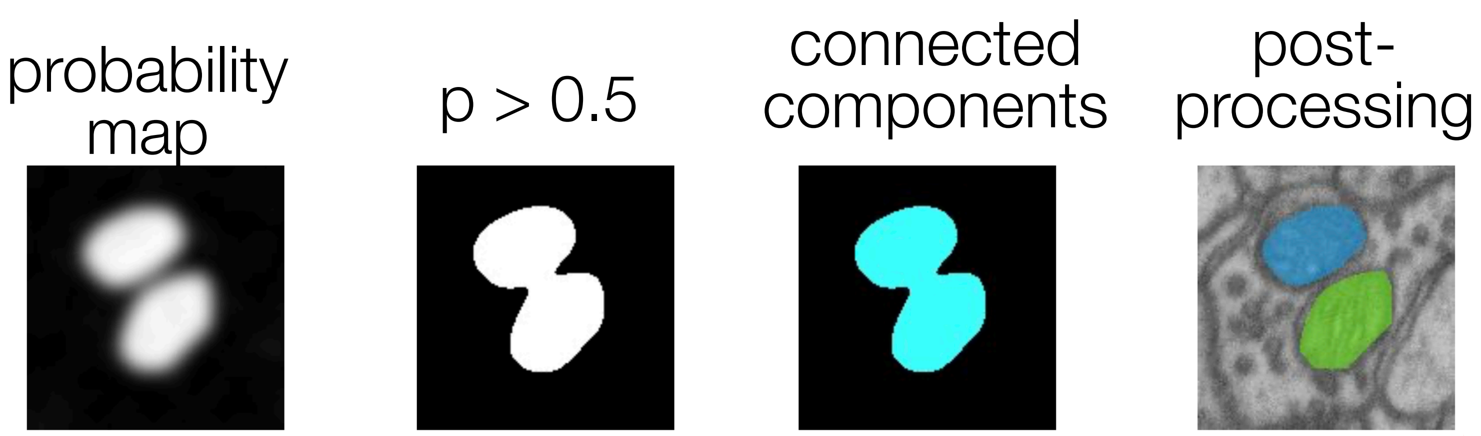


U-Net

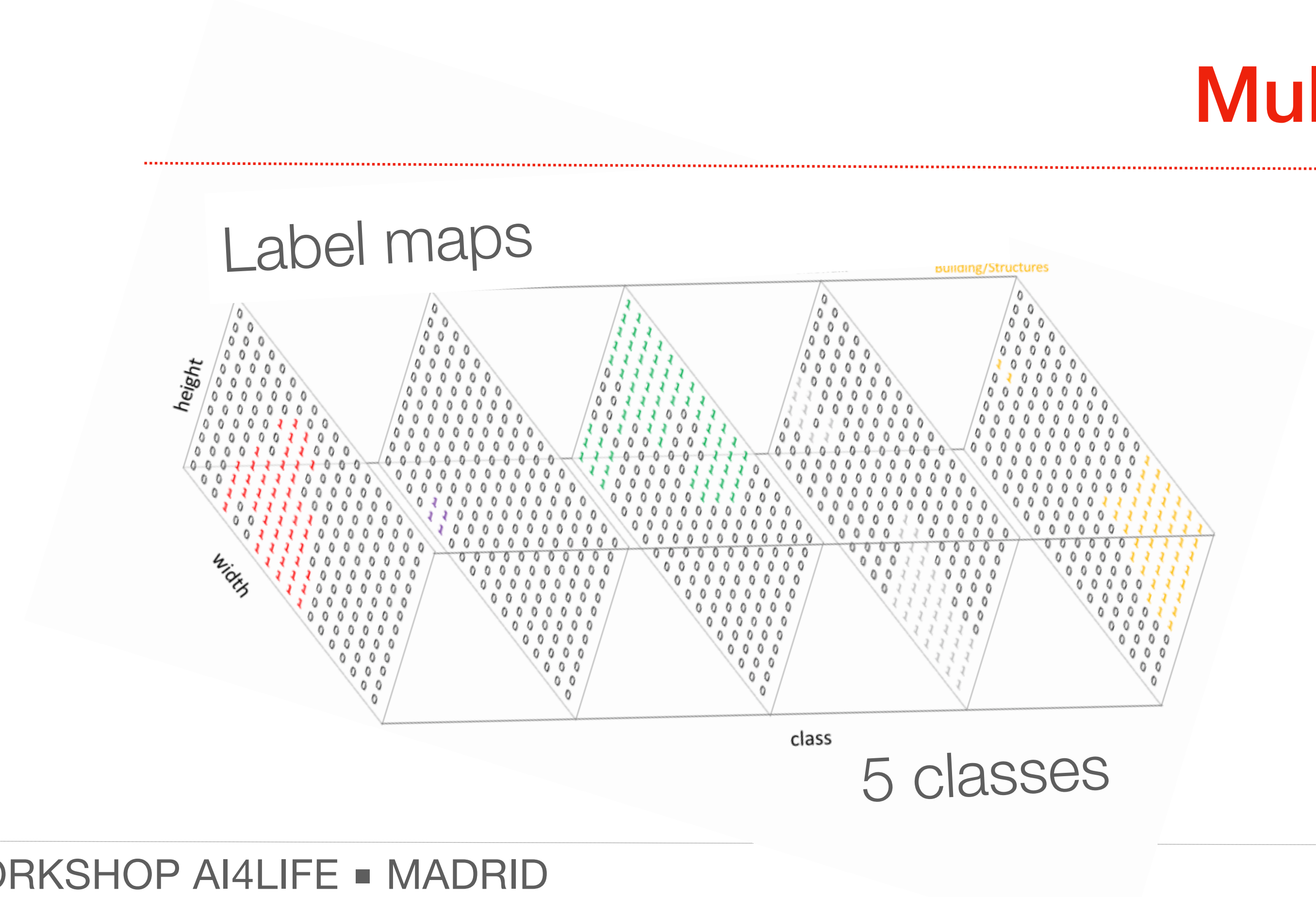


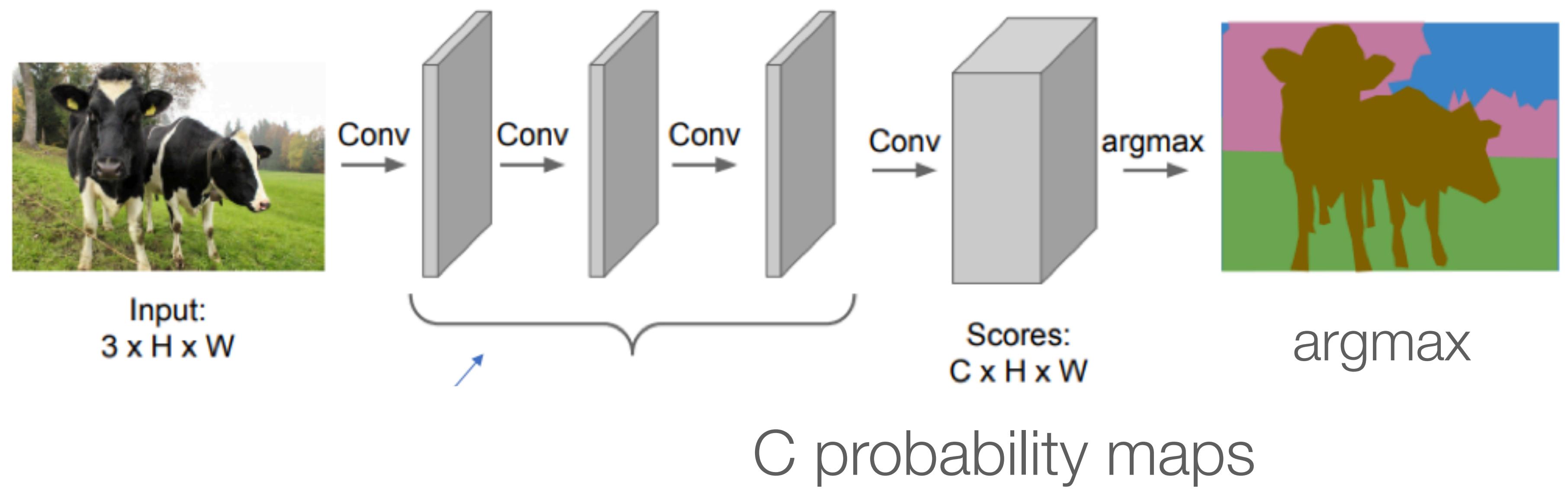


Binary classification



Multiple label classification

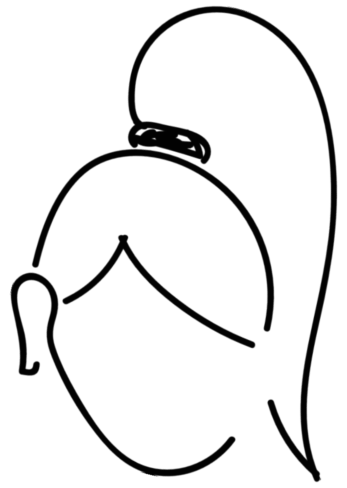




U-Net — pixel classification

Ronneberger et al., U-Net, MICCAI, 2015

Model Producers



- Data science
- IT skills
- Programming
- Domain knowledge

CPU

GPU
NVIDIA ↗Gb RAMM

HPC
Computing server

COM

CLOUD
Computing service

binder aws Microsoft Azure colab

Deep Learning Frameworks

PyTorch

Primarily developed by Facebook

- Tensor -> numpy
- Convenient for research

TensorFlow

Primarily developed by Google

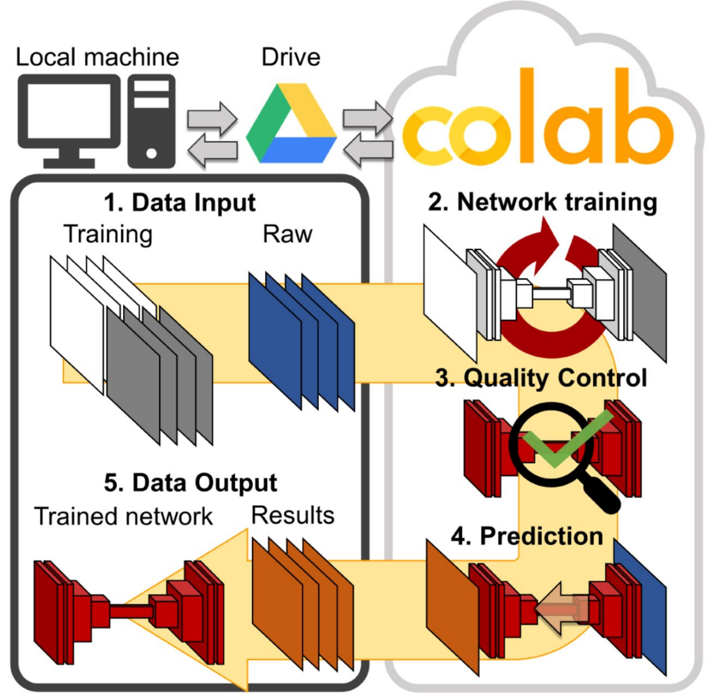
- Major changes in TF2
- Keras

ONNX

Open format to represent ML models

ZeroCostDL4Mic

- **Self-explanatory** Notebooks
- Running on **Google Colab** (free)
- Export to the bioimage zoo (beta)
- U-net 2D, 3D, Stardist, noise2void, ...



L. von Chamier, Nature Comm, 2021

#ZeroCostDL4Mic

Model Consumers

- Domain knowledge
- Final users
- Validation
- Trust

CPU

Deep Learning Consumers

Python

CellProfiler

Fiji

QuPath

Matlab

Icy

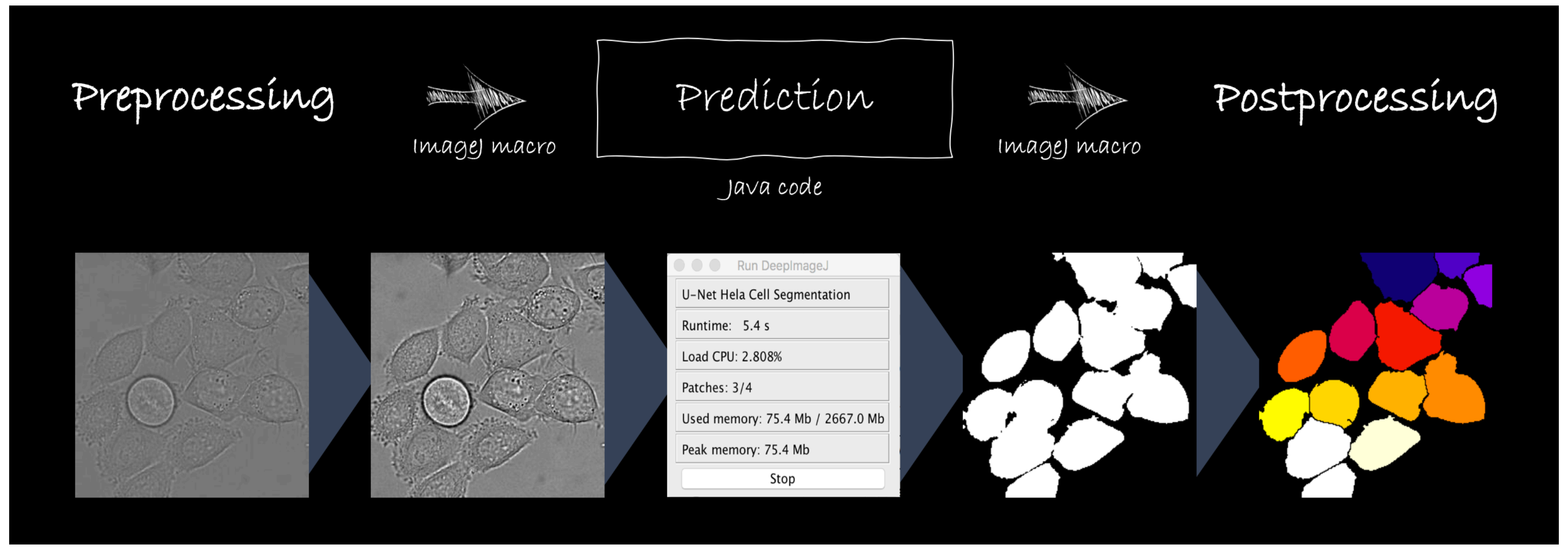
Napari

ImJoy

EPFL uc3m

deepImageJ

Run deep-learning interference in ImageJ as plugin or macro



- ✓ Macro recordable (pipeline)
- ✓ Expose models to user
- ✓ One-click installation
- ✓ Pre & post-processing
- ✓ Models from the **Bioimage Zoo**

TF 1 & 2
PT 1 & 2
ONNX

3.0

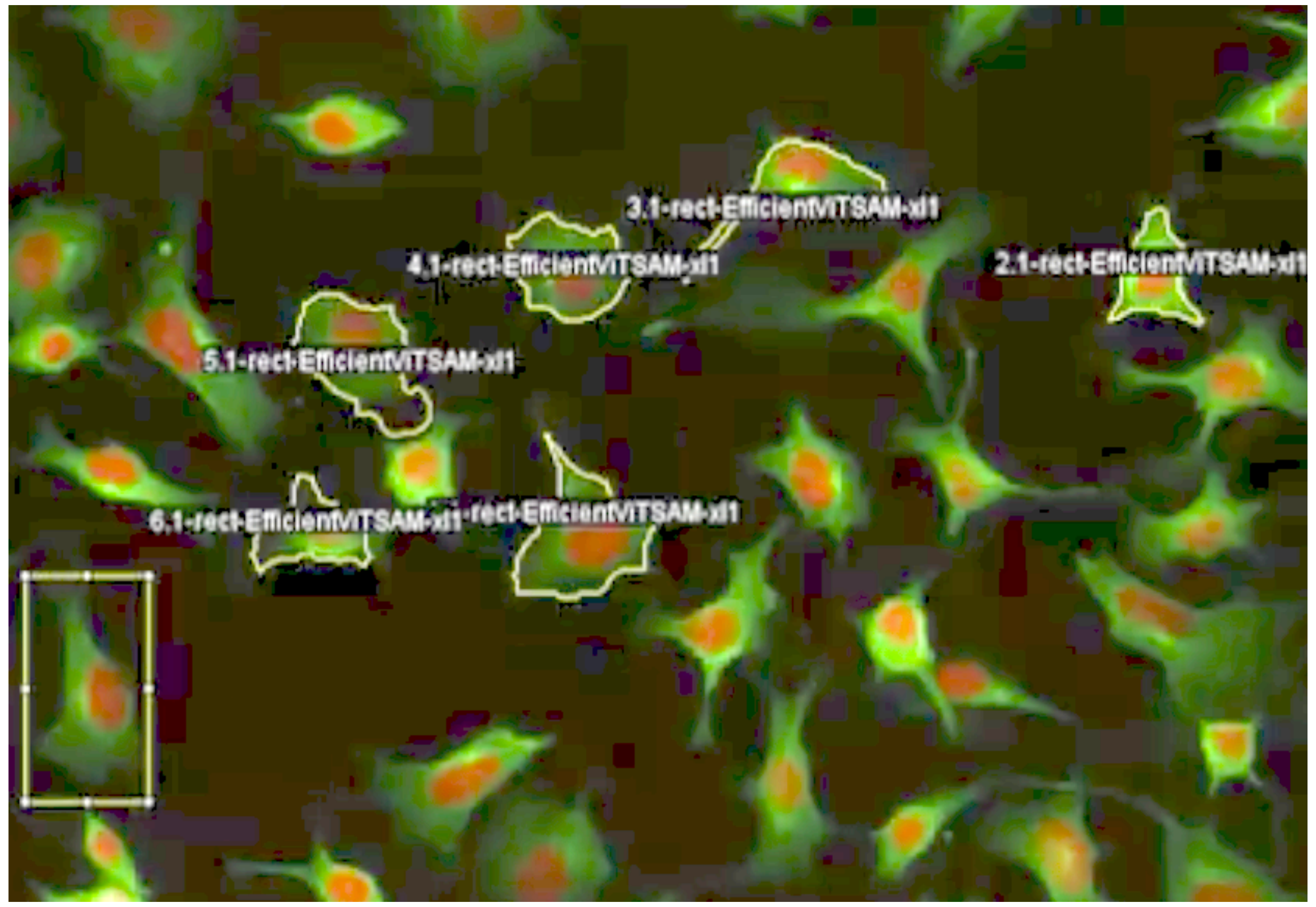
E. Gómez-de-Mariscal, Nature Methods, 2021.

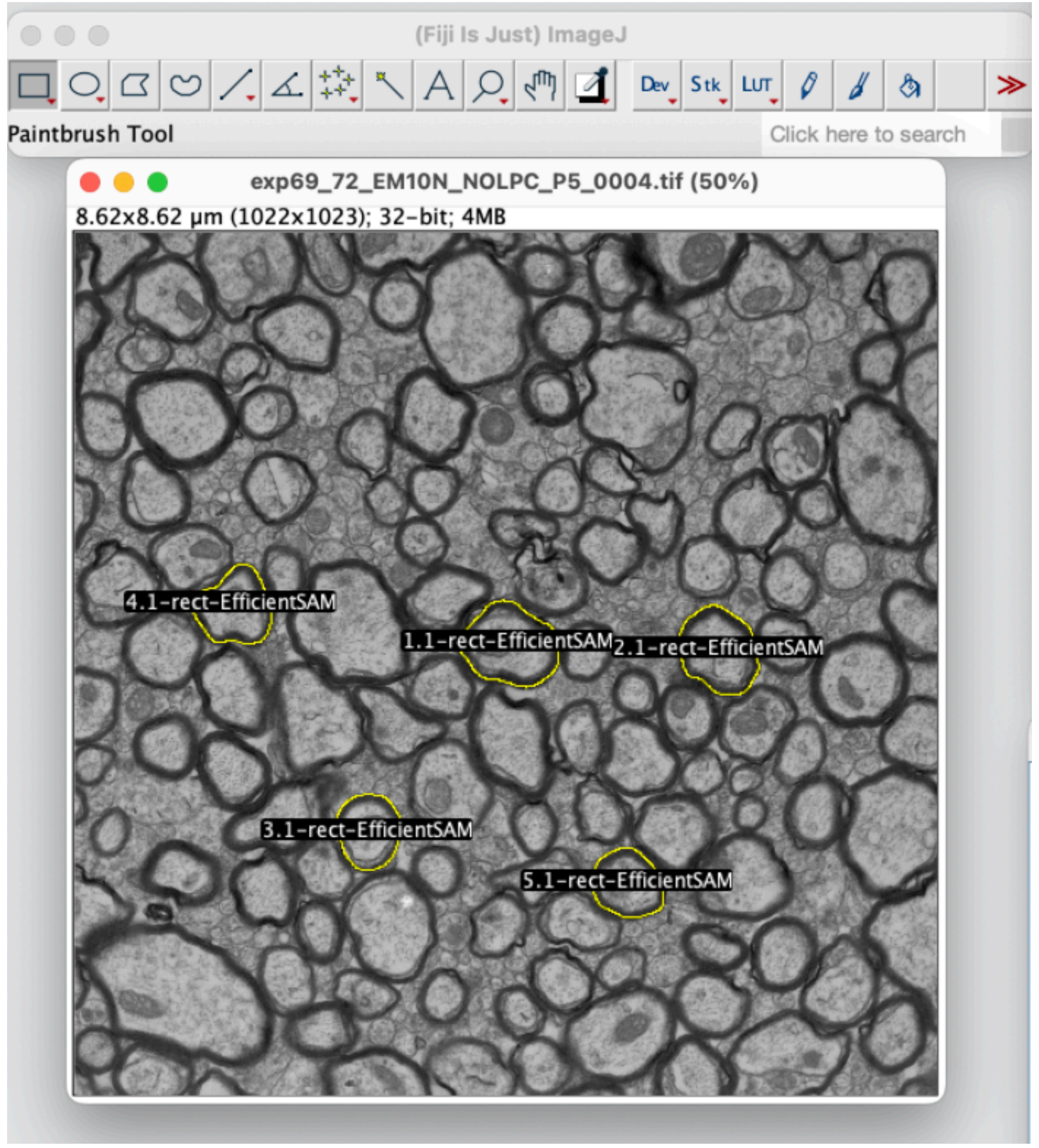
deepimagej.github.io

SAMJ Annotation with SAM on FIJI (CPU)

SAMJ

- FIJI Plugin and ICY plugin
- Model Efficient SAM (run on CPU)
- Automatic installation of the Python environment
- Smart strategy for tiling



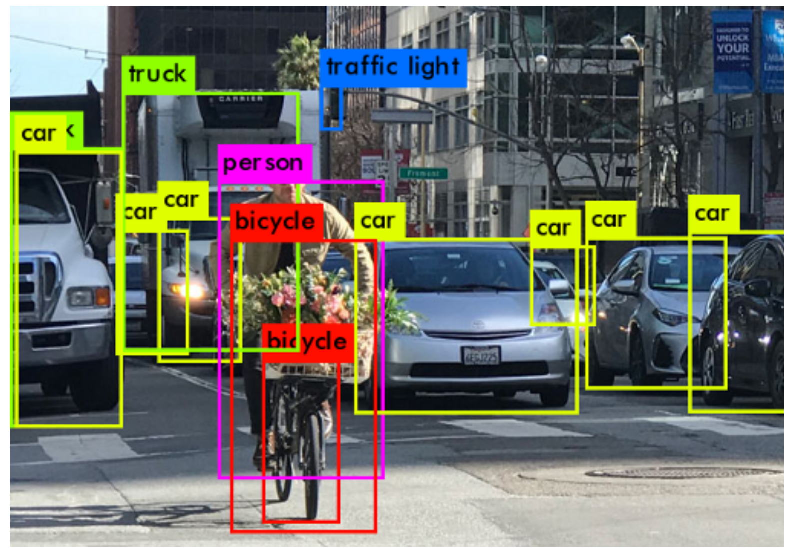


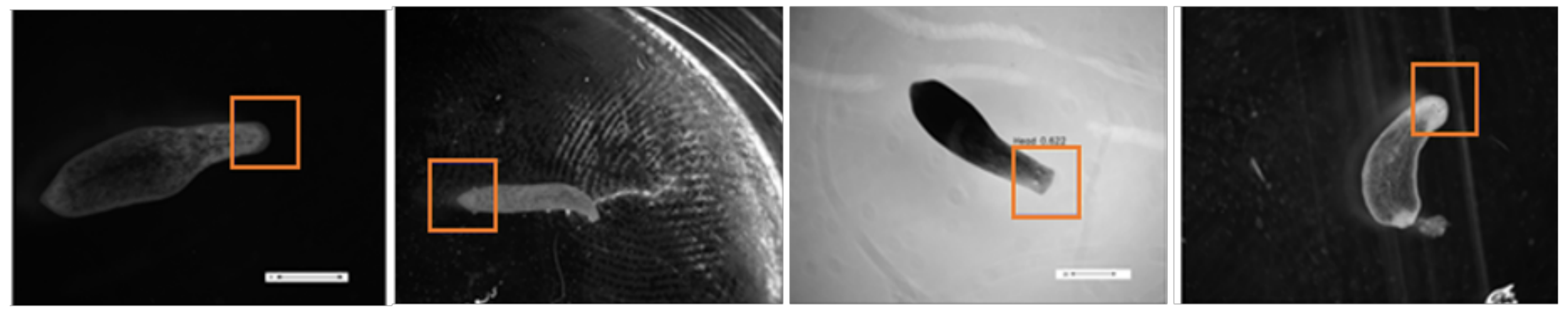
SAMJ Team: Carlos, Caterina, Arrate, Vladimir Ulman, Adrian Ines, Jonathan Heras, Curtis Rueden, Jean-Christophe, Daniel

And Many Others Tools

Object detection

YOLO

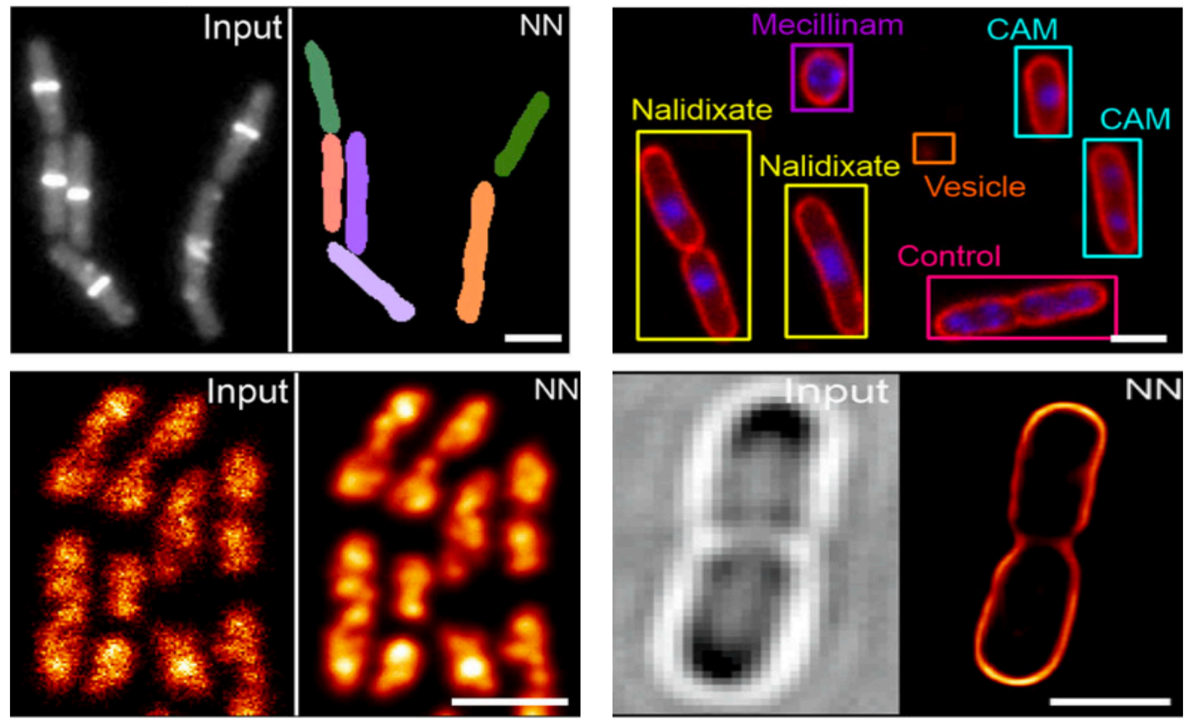




Phagocata sp. *Girardia dorotocephala* *Phagocata gracilis* *Schmidtea mediterranea*

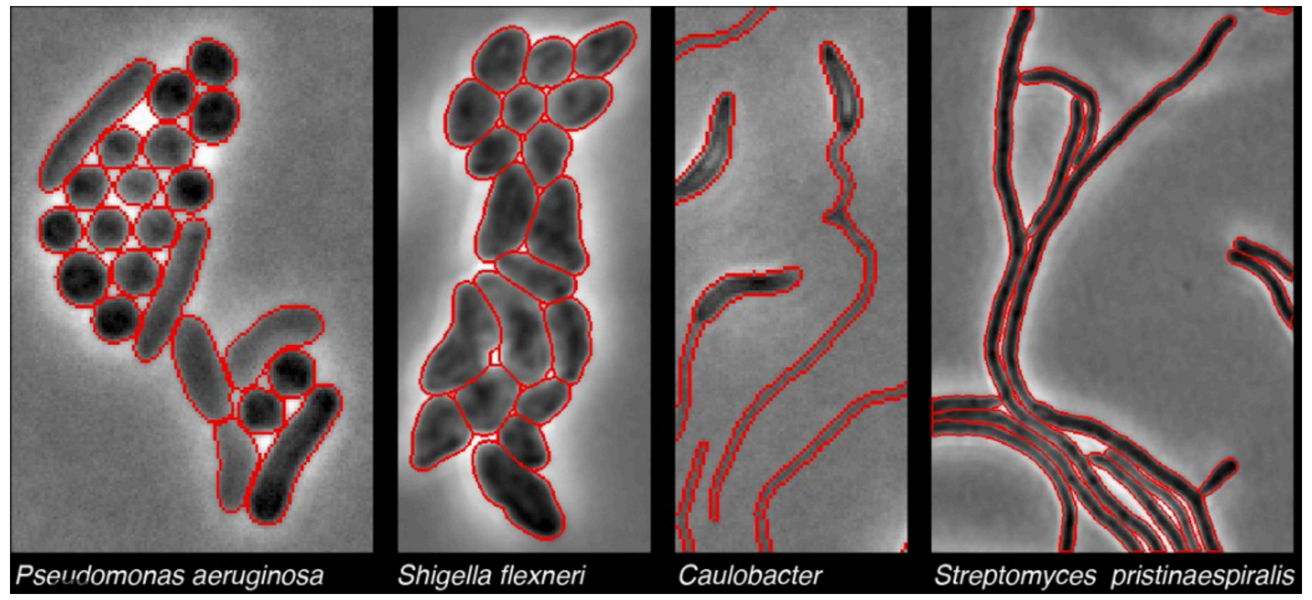
Characterization of bacteria

DEEPBACS

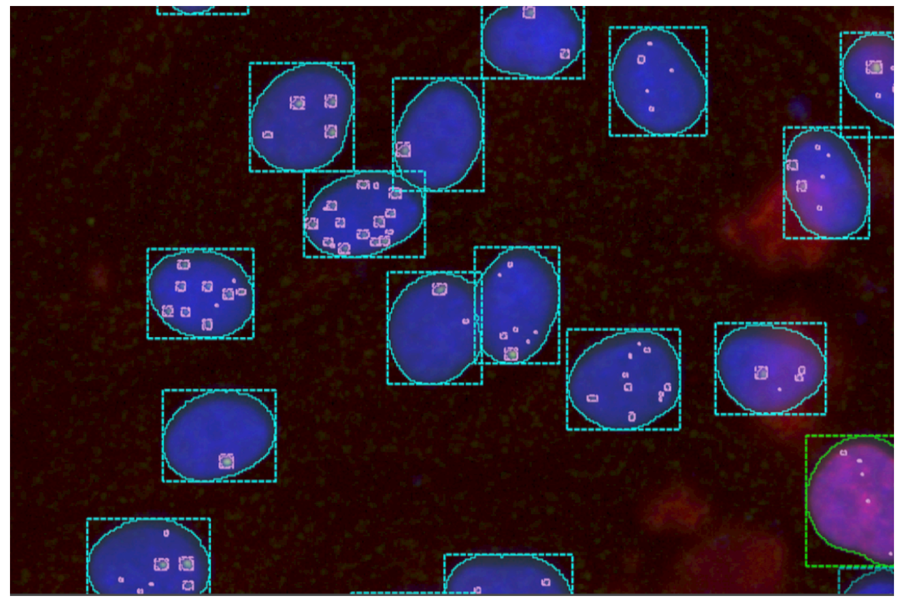


Detection of elongated cells

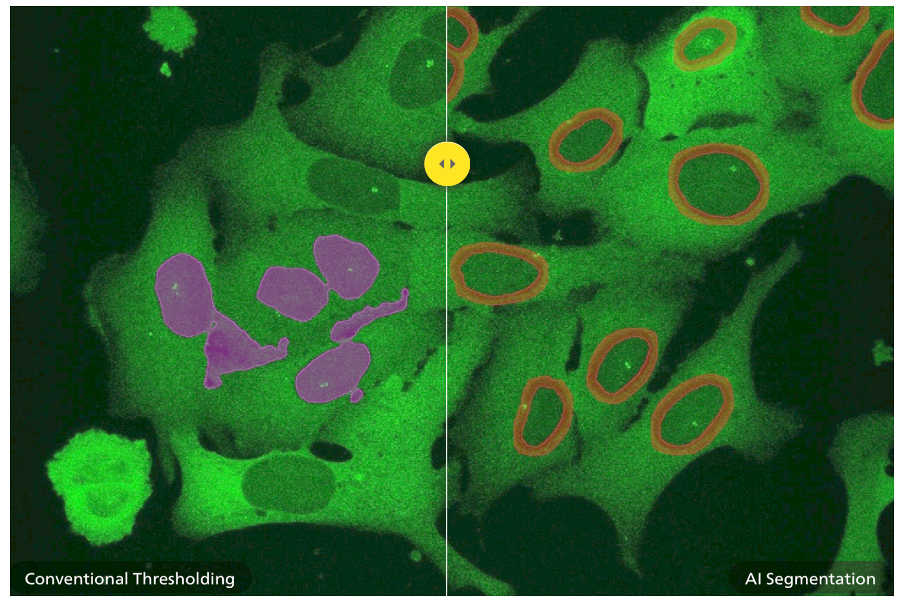
OMNIPOSE



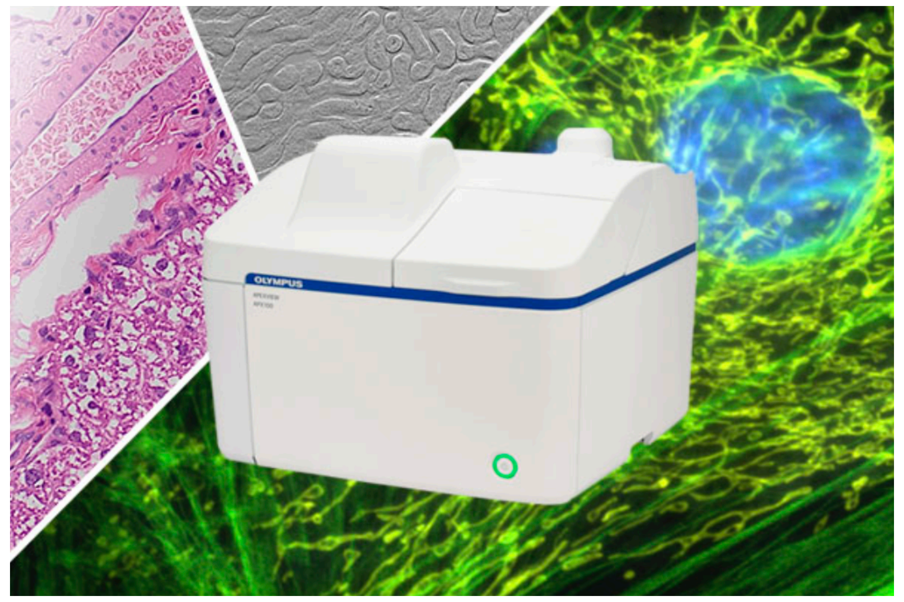
- ➡ CellProfiler
- ➡ QuPath
- ➡ Fiji (deepImageJ)
- ➡ BiaPy
- ➡ DeepMIB
- ➡ DeepTrack



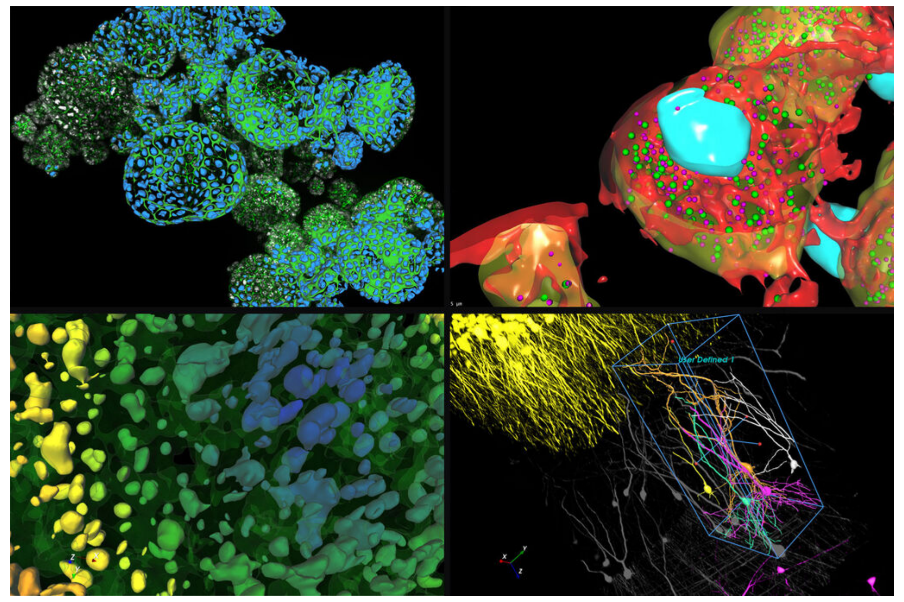
ZEISS APEER



NIKON NIS.AI



OLYMPUS APEXVIEW



LEICA AIVIA

EPFL

Introduction to Microscopy Image Analysis

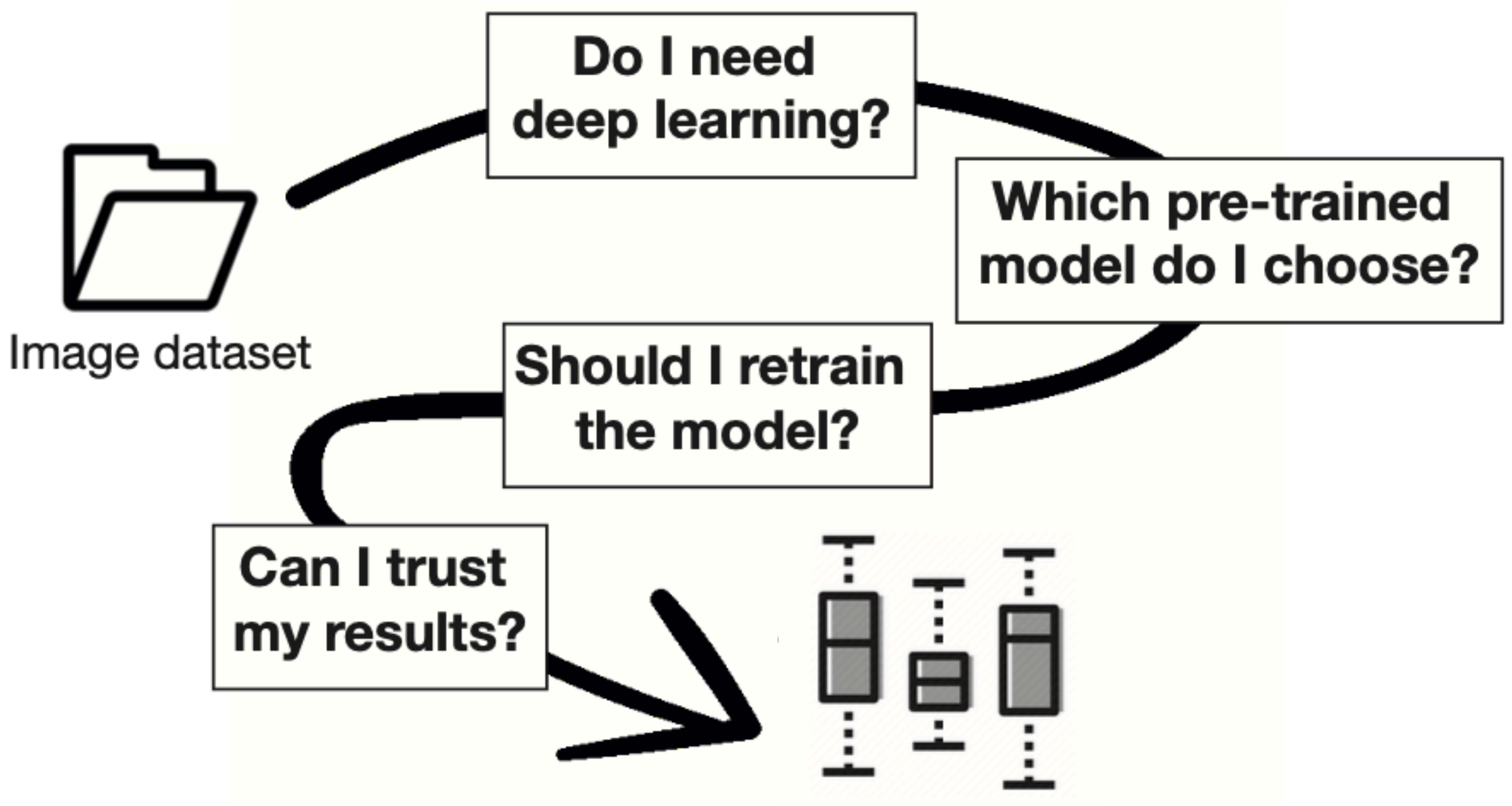
Lecture for the workshop AI4Life given by Daniel Sage, 10 June 2024

daniel.sage@epfl.ch

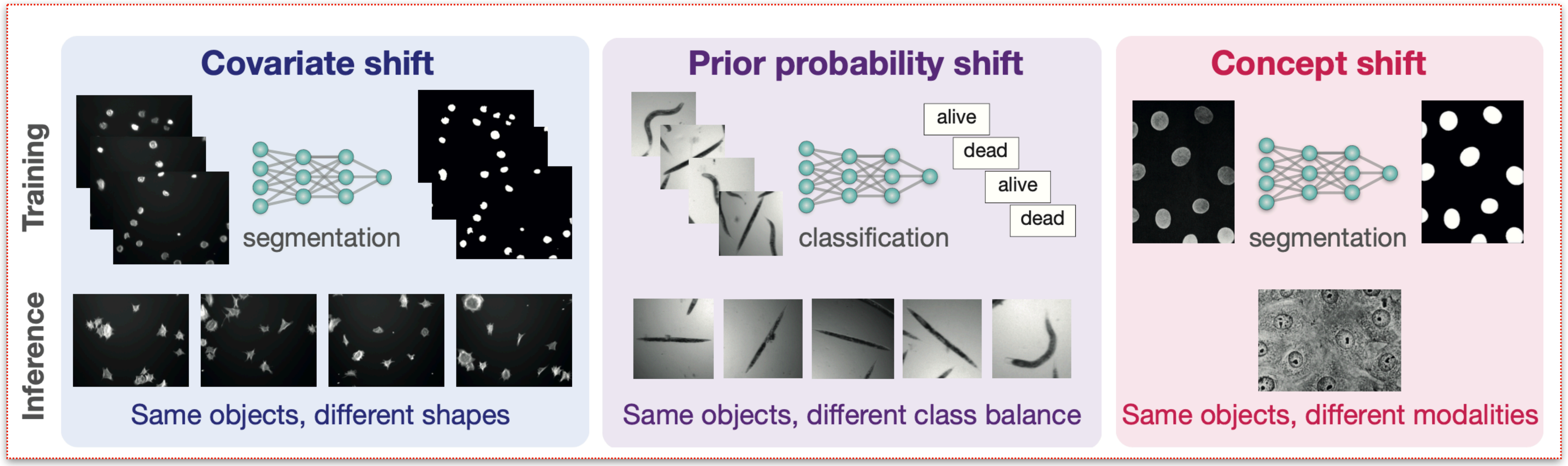
Risk and Challenges

Overuse of DL | Unnecessary complication

Education image analysis and DL



V. Uhlmann, L. Donati, D. Sage, IEEE SPM 2022



Dataset shift | Low performance, hallucination

Fine tuning

Trust in results | Overconfidence or Skepticism

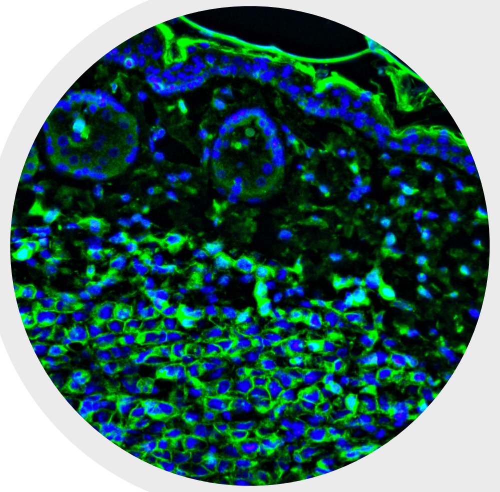
Validation / Interpretation

The data is the code.

Scientific Design of the Training Datasets

Preparation, curation, annotation, validation, unbiased, integrity, open

When you train a machine, you are actually programming the algorithm using data instead of using code.



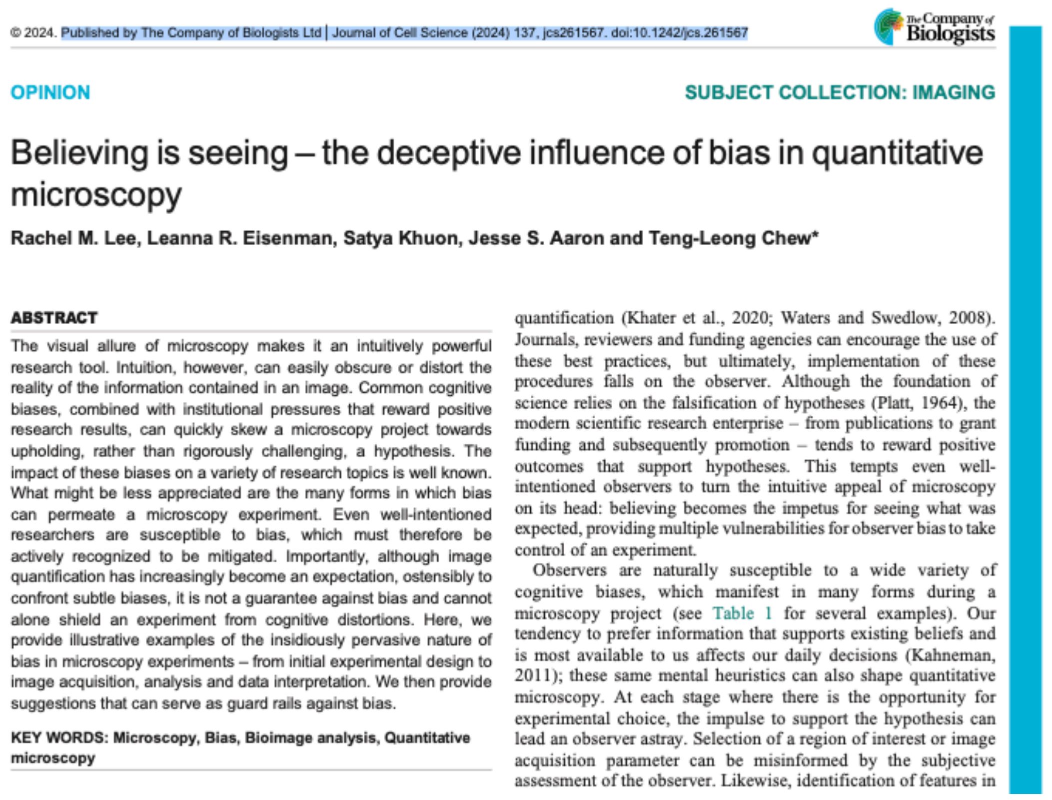
Machine learning is only as good as the training data you put into it.

Believing is Seeing

Believing is seeing – the deceptive influence of bias in quantitative microscopy

R. M. Lee et al.

Journal of Cell Science 2024


© 2024. Published by The Company of Biologists Ltd | Journal of Cell Science (2024) 137, jcs261567. doi:10.1242/jcs.261567

The Company of Biologists

OPINION

SUBJECT COLLECTION: IMAGING

Believing is seeing – the deceptive influence of bias in quantitative microscopy

Rachel M. Lee, Leanna R. Eisenman, Satya Khuon, Jesse S. Aaron and Teng-Leong Chew*

ABSTRACT

The visual allure of microscopy makes it an intuitively powerful research tool. Intuition, however, can easily obscure or distort the reality of the information contained in an image. Common cognitive biases, combined with institutional pressures that reward positive research results, can quickly skew a microscopy project towards upholding, rather than rigorously challenging, a hypothesis. The impact of these biases on a variety of research topics is well known. What might be less appreciated are the many forms in which bias can permeate a microscopy experiment. Even well-intentioned researchers are susceptible to bias, which must therefore be actively recognized to be mitigated. Importantly, although image quantification has increasingly become an expectation, ostensibly to confront subtle biases, it is not a guarantee against bias and cannot alone shield an experiment from cognitive distortions. Here, we provide illustrative examples of the insidiously pervasive nature of bias in microscopy experiments – from initial experimental design to image acquisition, analysis and data interpretation. We then provide suggestions that can serve as guard rails against bias.

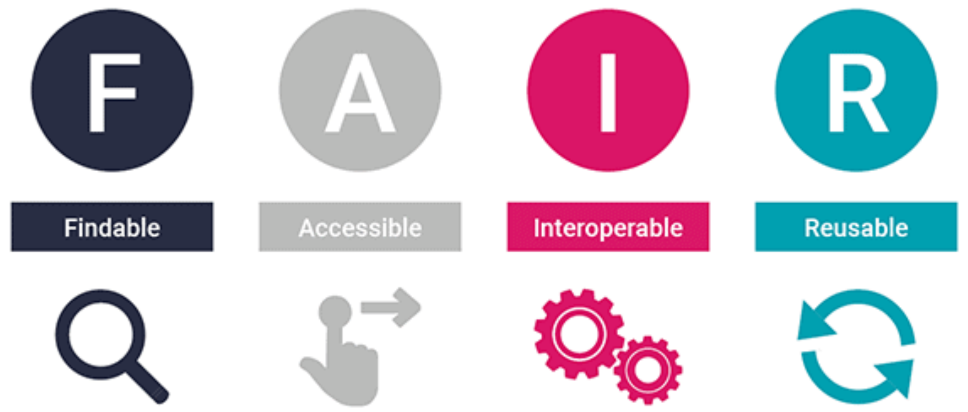
KEY WORDS: Microscopy, Bias, Bioimage analysis, Quantitative microscopy

quantification (Khater et al., 2020; Waters and Swedlow, 2008). Journals, reviewers and funding agencies can encourage the use of these best practices, but ultimately, implementation of these procedures falls on the observer. Although the foundation of science relies on the falsification of hypotheses (Platt, 1964), the modern scientific research enterprise – from publications to grant funding and subsequently promotion – tends to reward positive outcomes that support hypotheses. This tempts even well-intentioned observers to turn the intuitive appeal of microscopy on its head: believing becomes the impetus for seeing what was expected, providing multiple vulnerabilities for observer bias to take control of an experiment.

Observers are naturally susceptible to a wide variety of cognitive biases, which manifest in many forms during a microscopy project (see Table 1 for several examples). Our tendency to prefer information that supports existing beliefs and is most available to us affects our daily decisions (Kahneman, 2011); these same mental heuristics can also shape quantitative microscopy. At each stage where there is the opportunity for experimental choice, the impulse to support the hypothesis can lead an observer astray. Selection of a region of interest or image acquisition parameter can be misinformed by the subjective assessment of the observer. Likewise, identification of features in

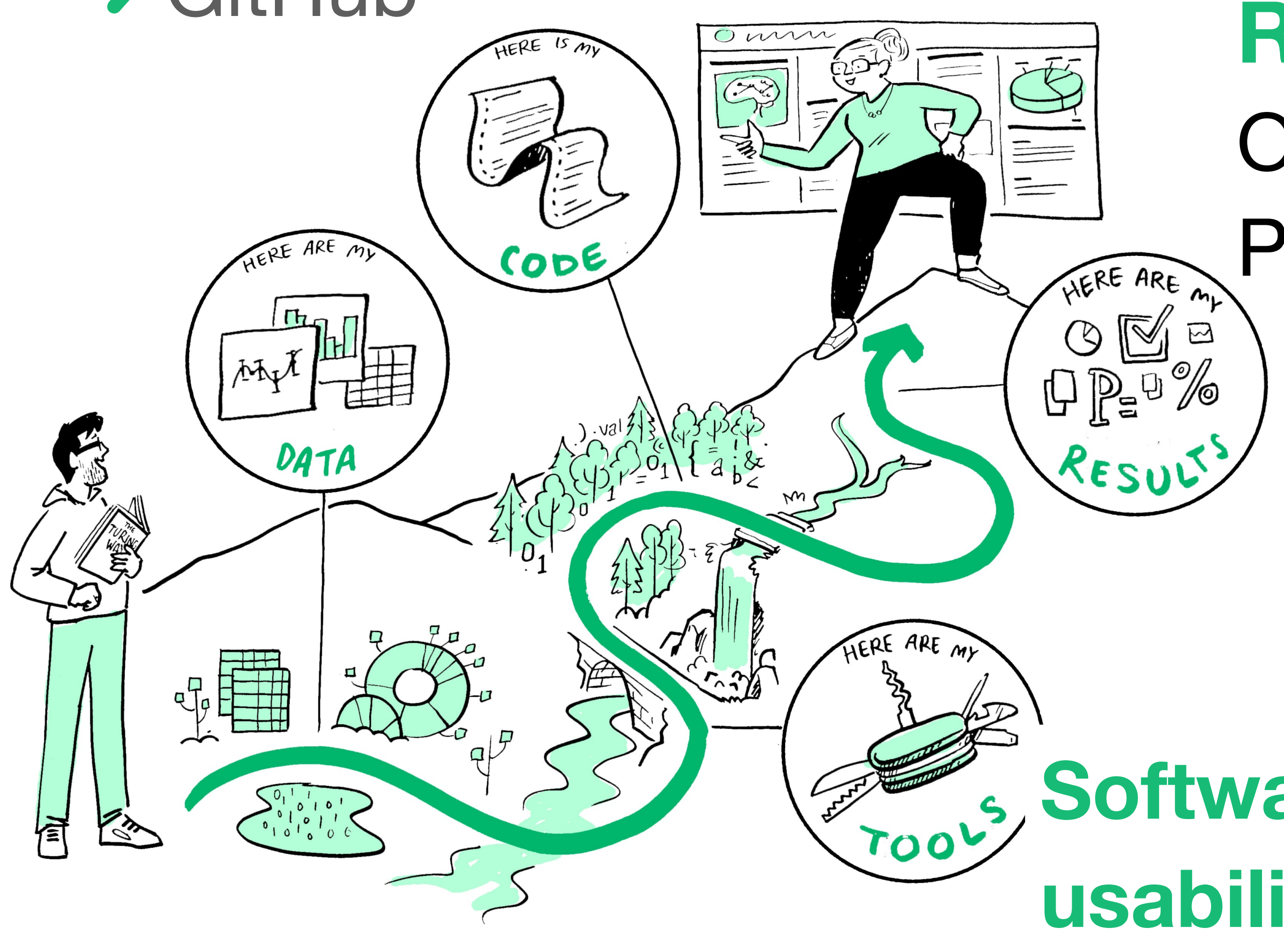
Clustering illusion	Seeing groups in time or space as significant when they are random
Color perception	Illusions due to misleading perception of colors
Confirmation bias	Favoring information that supports existing beliefs
Congruence bias	Not testing alternative hypotheses for the observed data.
Contrast effect	Over- or under-estimating a feature based on spatial-temporal surroundings
Illusory correlation	Seeing a relationship where there is no underlying correlation
Pareidolia	Seeing patterns that do not exist
Publication bias	Withholding negative results from publication
Recency bias	Giving greater weight to more recent observations
Selection bias	Focusing on a sample that is not representative of the population population
Survivorship bias	Overlooking data that does not survive a selection process

Open Imaging and Responsability



Code
Open-source, versioning, test
➜ GitHub

Data
Accessible
Meta-data
➜ Zenodo

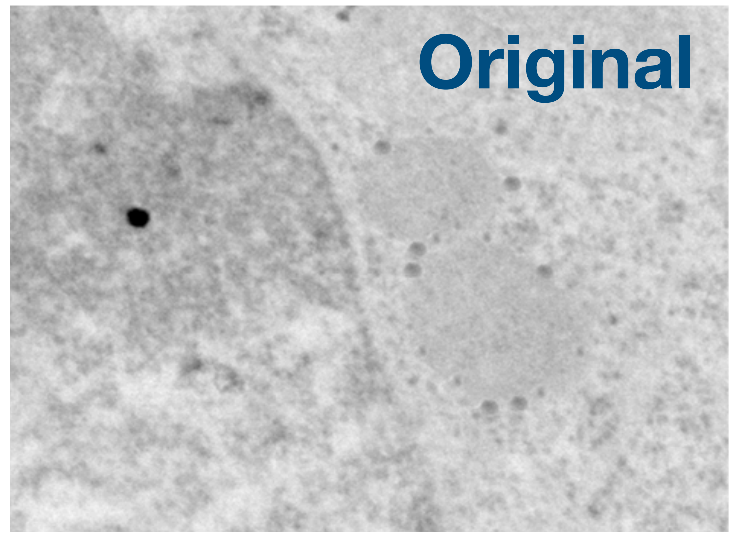


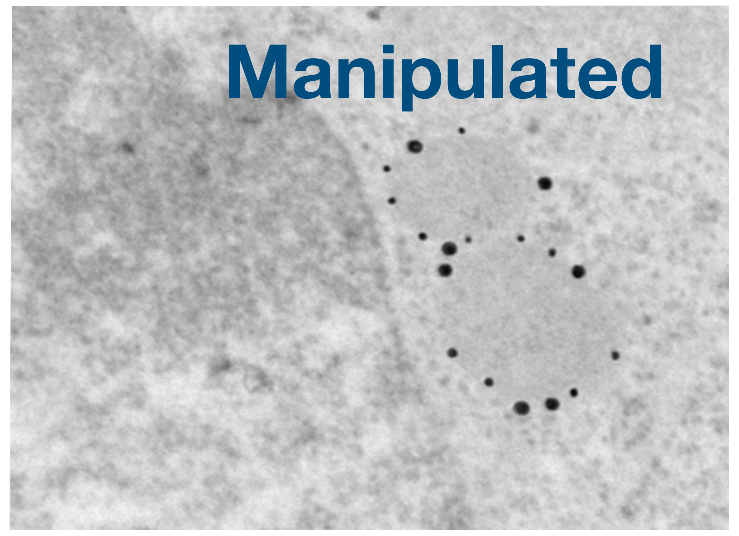
Results
Open Access
Peer-reviewed publication

Software usability

Ethical Guidelines

1. Keeping raw data, open the data
2. No image compression
3. Simple adjustment acceptable
4. Cropping acceptable
5. Digital filtering is not encouraged
6. Combining images if no hiding
7. No local alteration
8. Compare in the same conditions
9. Image should be documented
10. Reporting the analysis script

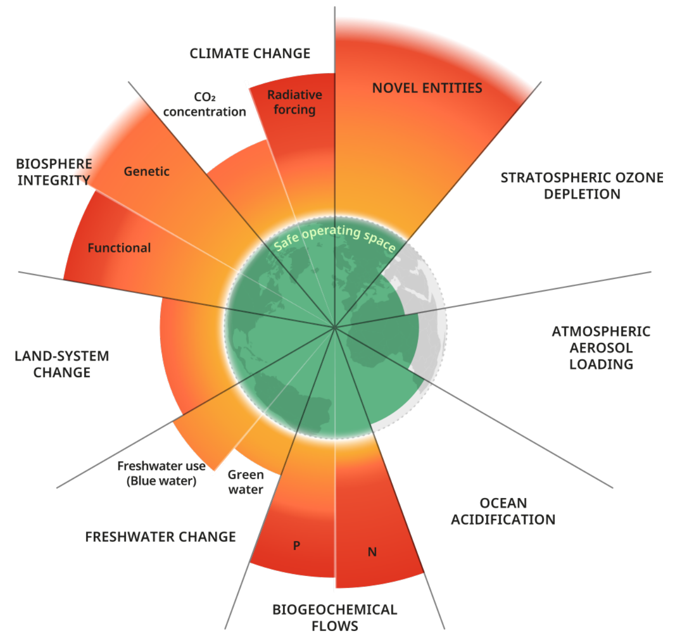




Manipulation and misconduct
C. Blatt, Plant Physiology, 2013.

Avoiding twisted pixels
D. Cromey, 2010.

Mishandling and Misconducts
K. Miura, S. Nørrelykke, 2021.



Sustainability
No overshot of the planetary boundaries
Digital: Life cycle of equipments (gray energy)
➜ green-algorithms.org

Illustration from the online book The Turing Way

AI4Life in a nutshell

Goals

To empower life science researchers to harness the full potential of AI and ML methods for bioimage analysis

Facets

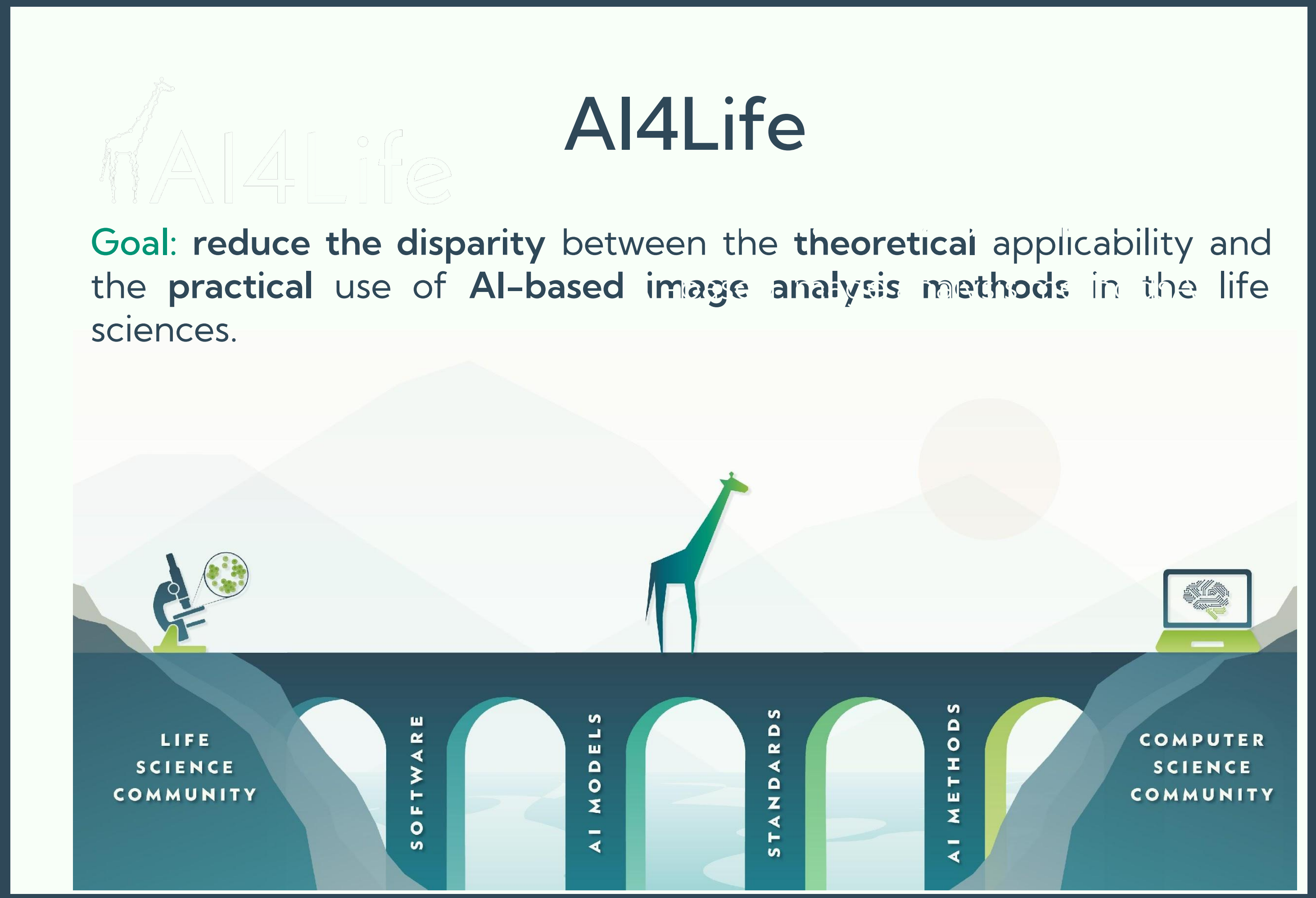
The BioImage Model Zoo and FAIR data principles are core facets of the AI4Life project.

Outcome

Services provided and solution developed to solve microscopy image analysis problems

Horizon Europe funded project

Bringing together the computational and life science communities



www.ai4life.eurobioimaging.eu

Democratized availability of AI-based image analysis methods

Simple model deployment, sharing, and dissemination through a new developer-facing service

Organize **Open Calls and Challenges** for image analysis problems

Establish standards for the submission, storage and FAIR access

Empower **common image analysis** platforms with **AI integration**

Organizing outreach and training events courses/workshops and participation in conferences