Consiglio Nazionale delle Ricerche

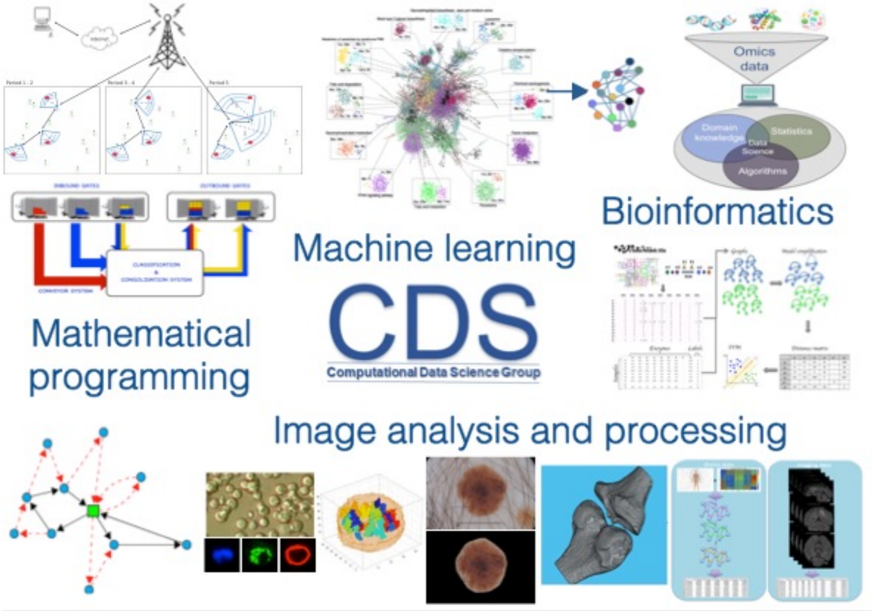
Artificial Intelligence for Cell Biology Imaging

Lucia Maddalena

National Research Council (CNR)

Inst. for High-Performance Computing and Networking (ICAR)

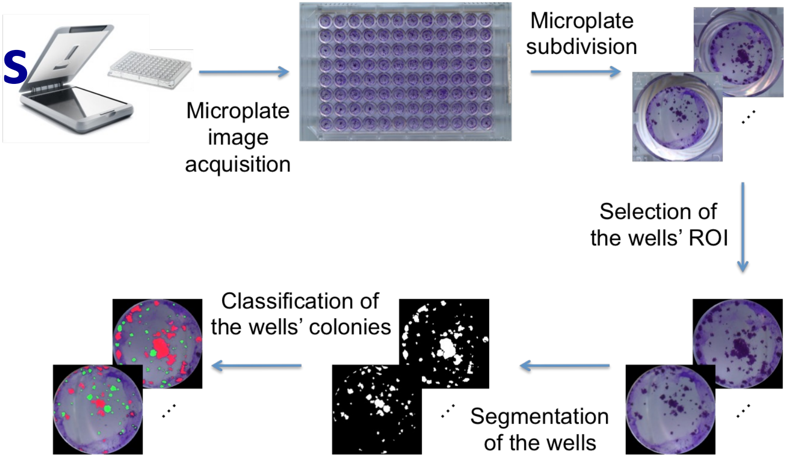
CDS-group



Outline

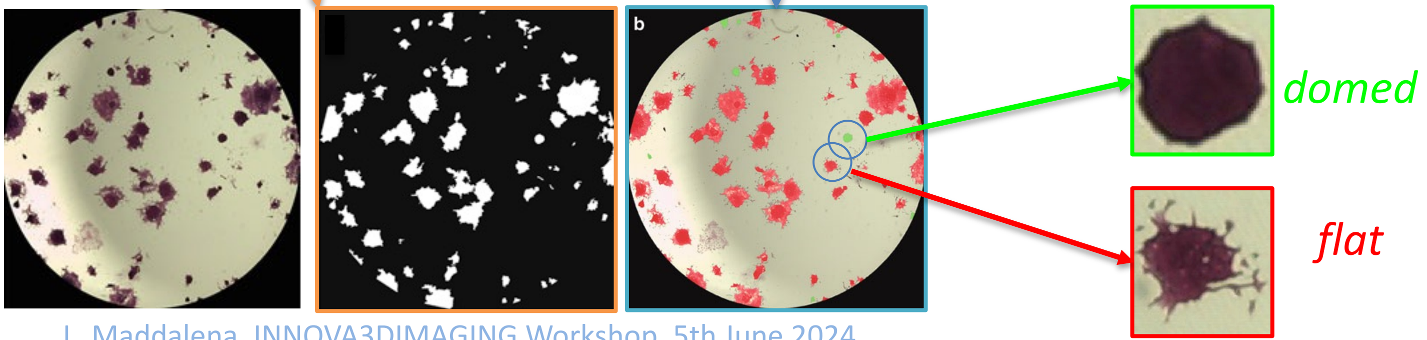
- An example-based introduction to approaches
  - Image processing-based
  - Machine learning-based
  - Deep learning-based
- Some *issues* and *tricks*
- Basic problems
  - Segmentation
  - Detection
  - Tracking
  - Enhancement
- Software
- Data

High-throughput screening of mESCs



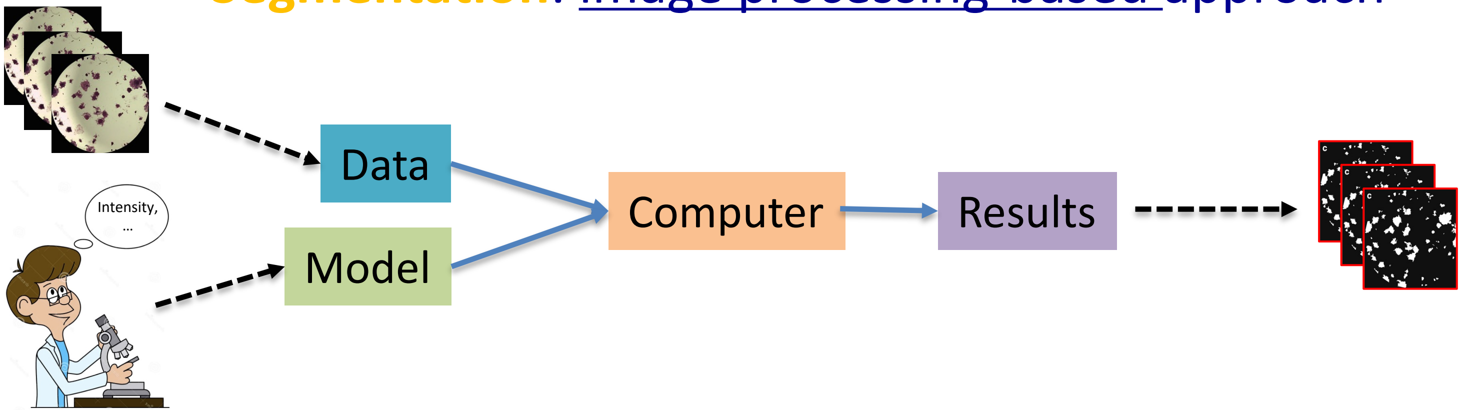
Segmentation based on intensity thresholding

Classification based on hand-crafted features (shape and texture) and various classifiers



https://doi.org/10.1007/978-3-319-23497-7_2 (2015)

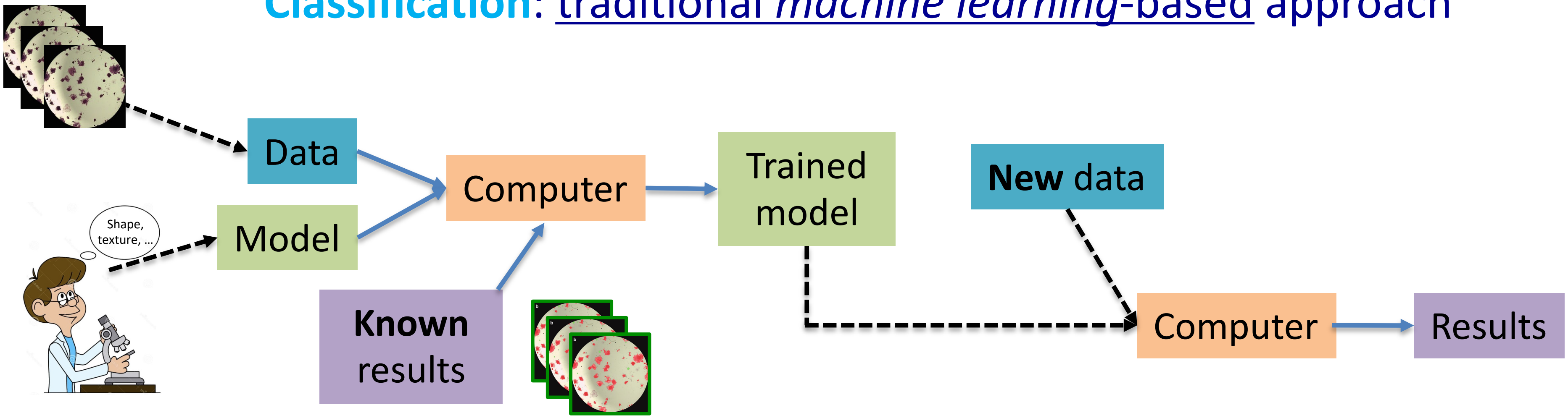
Segmentation: image processing-based approach



- Exploit image processing expertise to predefine the desired operations (e.g, thresholding, edge detection, filtering, etc.) as well as their parameters (threshold value, filter radius, etc.) to perform the task

https://doi.org/10.1007/978-3-319-23497-7_2 (2015)

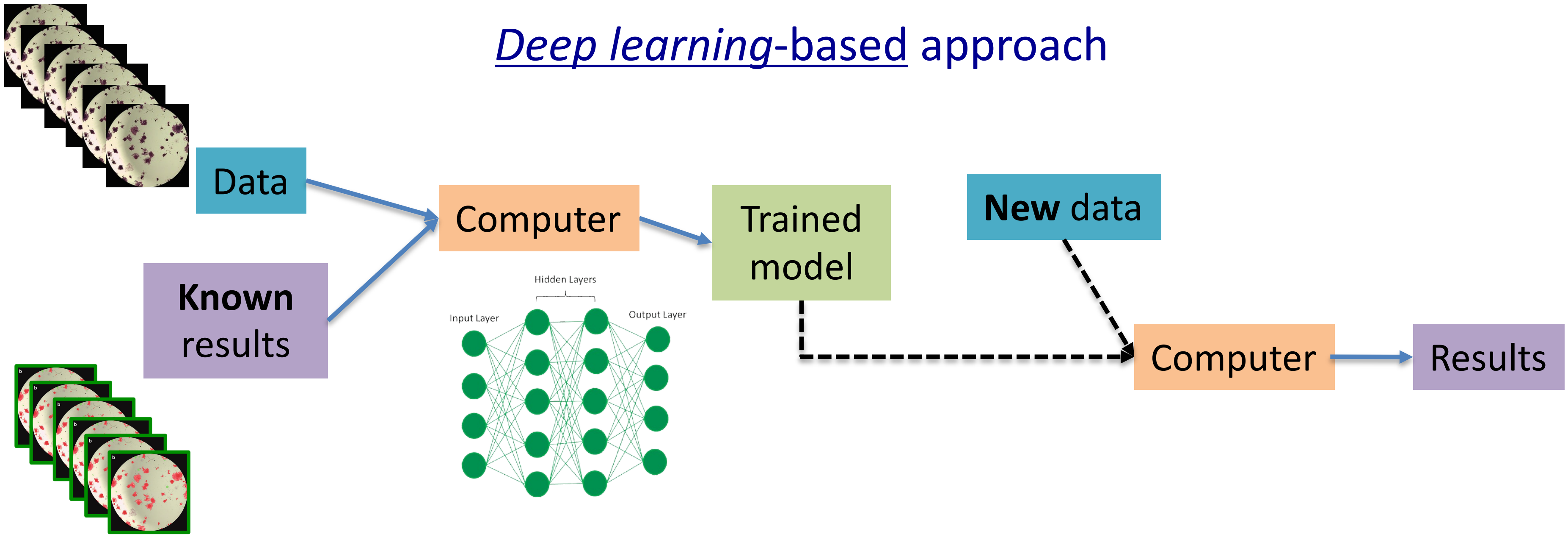
Classification: traditional *machine learning*-based approach



- Automate the model configuration, optimizing method parameters
- Need for *annotated data* (known results) for model training

https://doi.org/10.1007/978-3-319-23497-7_2 (2015)

Deep learning-based approach



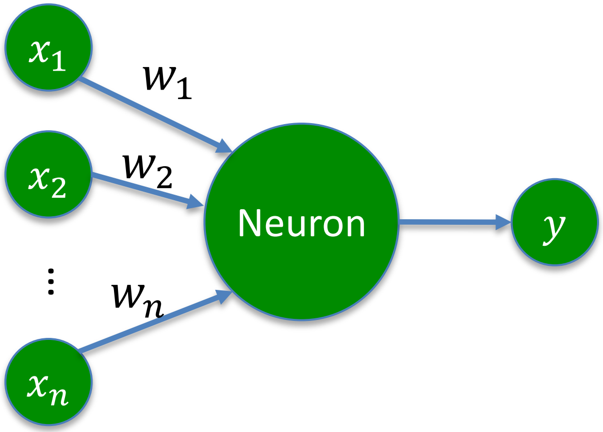
- Automatic feature learning from data and optimized method parameters (weights)
- Need for
  - High volumes of *annotated data* for model training
  - Hyperparameters’ tuning (architectural choices)

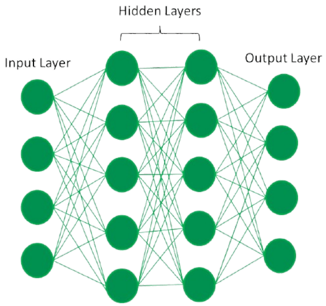
Artificial Neural Networks

$$y = \sigma\left(\sum_{i=1}^{n} x_i w_i + b\right)$$

Artificial neuron

x_i = input variables (features)
w_i = learnable weights
b = learnable bias term
σ = non-linear activation function

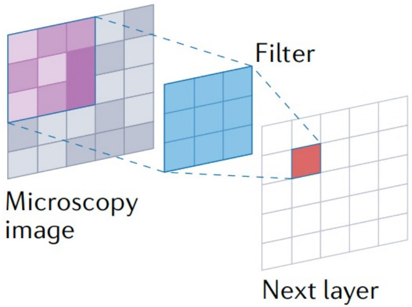




- Arranged in layers (the output of one layer is the input to the next)
- Different arrangements lead to different *architectures*
- Weights are the main adjustable parameters; their optimization (by backpropagation) leads to the *NN training*
- Basic layout: layers arranged in a fully connected fashion

https://doi.org/10.1038/s41580-021-00407-0 (2022)

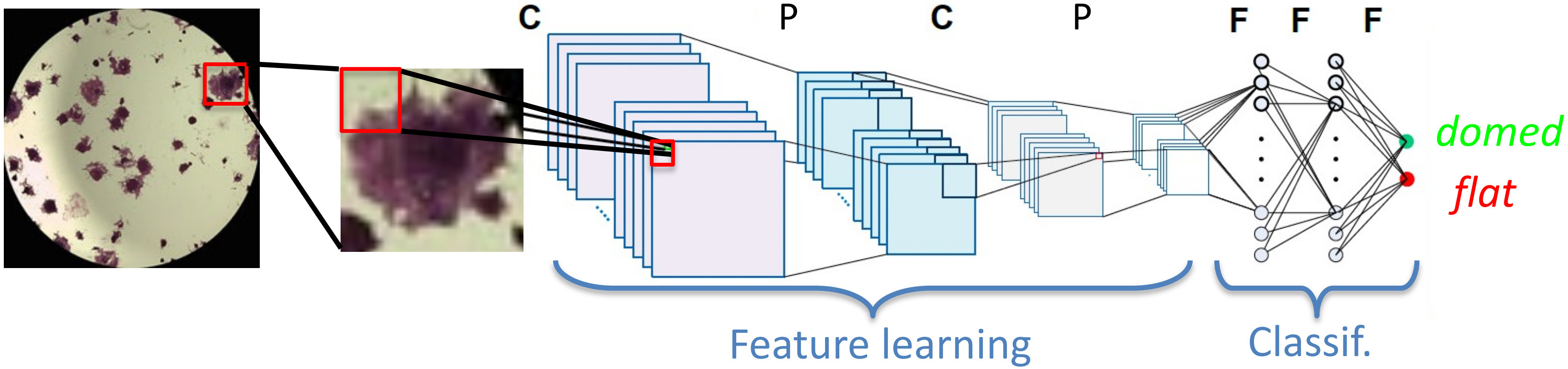
CNNs



- *Convolutional layers*: the output is the result of a small fully connected NN (filter) applied to local groups of input features
- Learn the local structure of the input data

https://doi.org/10.1038/s41580-021-00407-0 (2022)

- *Pooling layers* (P): summarize responses of neighboring regions in a feature map
- *Fully connected layers* (F): learn higher level feature representations, specific to object classes

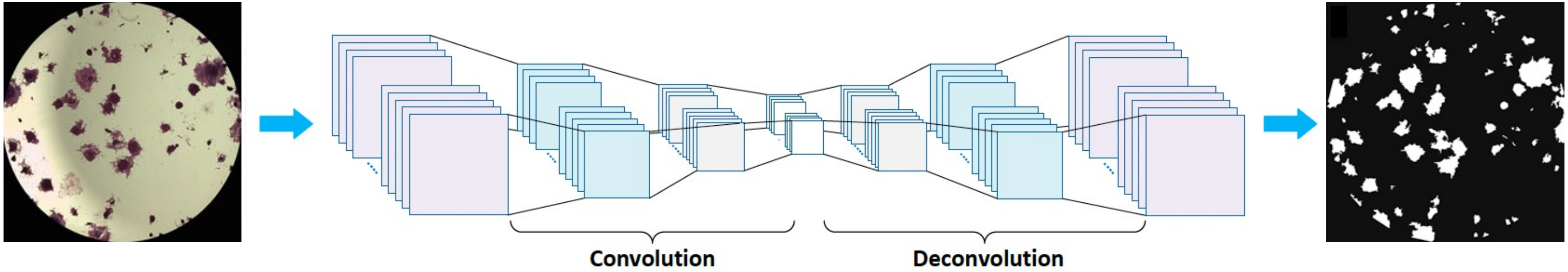


Adapted from https://doi.org/10.1109/TNNLS.2017.2766168 (2018)

FCNs

Variant of CNNs, consisting of

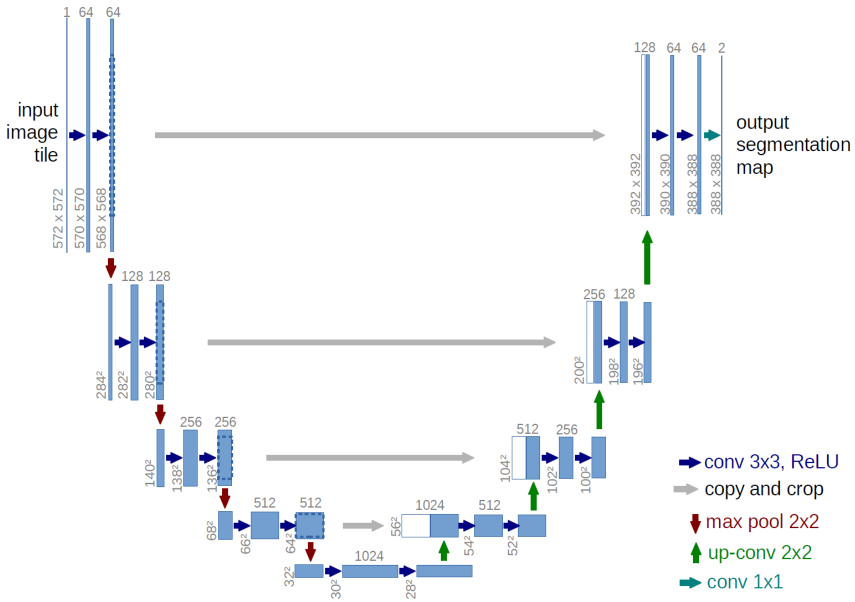
- a convolution path to learn high-level abstract feature representations and
- a deconvolution path to reconstruct fine details for segmentation



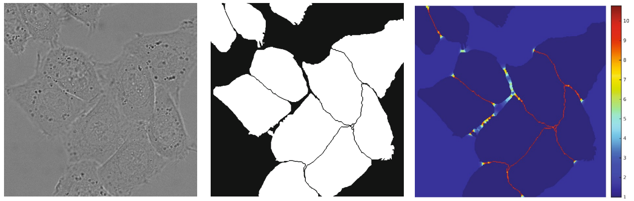
Adapted from https://doi.org/10.1109/TNNLS.2017.2766168 (2018)

U-net

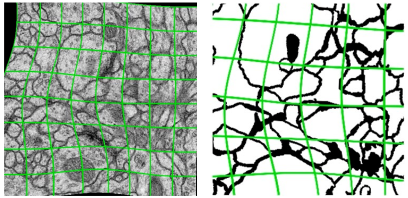
Another *encoder-decoder-style* variant of CNNs, frequently adopted for bioimaging



- pixel-wise loss weight to force the learning of border pixels



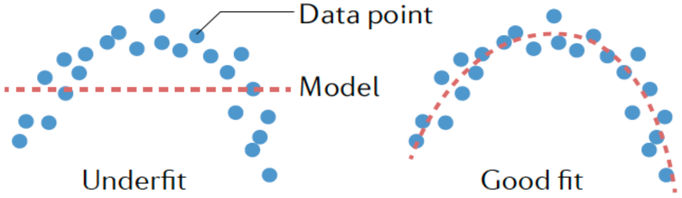
- augmentation with elastic deformations



https://doi.org/10.1007/978-3-319-24574-4_28 (U-Net, 2015)

Some ML and DL *issues*

- **Overfitting**: the method performs well on training data, less on validation data→ must have learned *shortcuts* to correct answers in training data

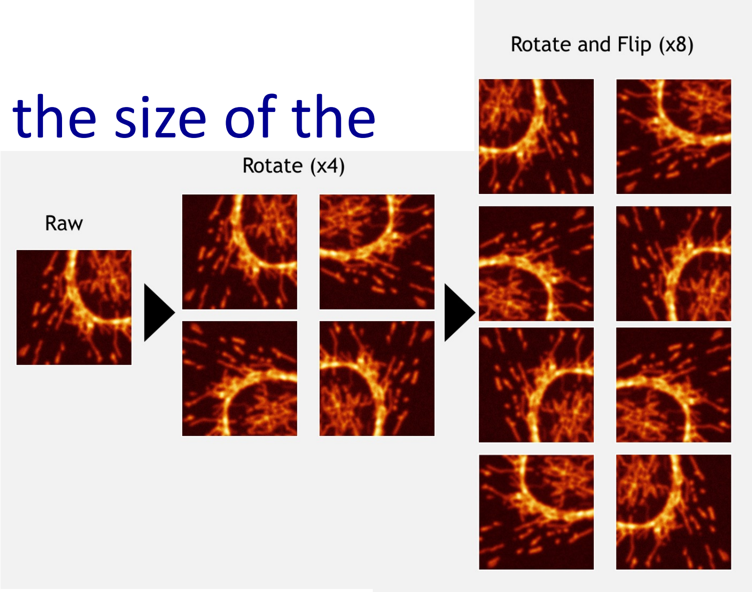


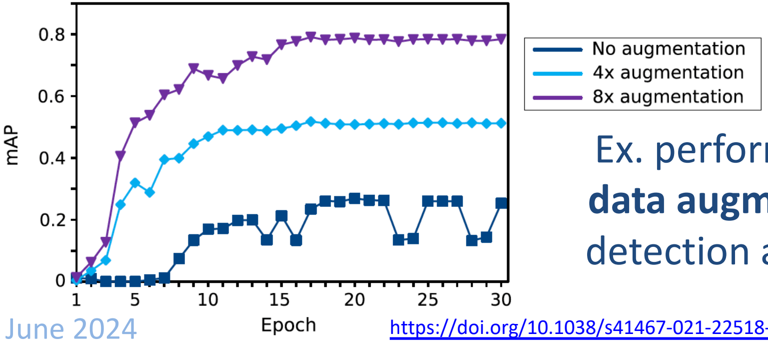
- **Learning rate**: controls the speed for updating model parameters during training. *Too slow* is time-consuming; *too high* can lead to quick convergence to a sub-optimal solution

https://doi.org/10.1038/s41580-021-00407-0 (2022)

Some ML and DL *tricks*

- **Data augmentation**: artificially increase the size of the training dataset
  - Can improve training progress by amplifying differences in the dataset
  - Useful if the available dataset is small (avoid overfitting)





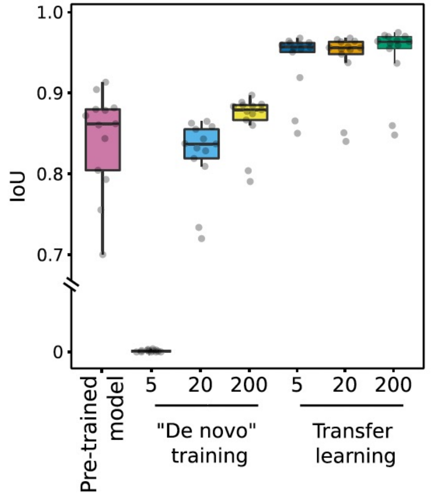
Ex. performance gain with **data augmentation** for cell detection and classification

https://doi.org/10.1038/s41467-021-22518-0 (ZeroCostDL4Mic, 2021)

Some ML and DL *tricks*

- **Transfer learning**: Exploiting a pre-trained model as a starting model (rather than initializing training with a blank model)
  - Re-use previously learned features
  - Shorten training times
  - Reduce the amount of required training data
  - Benefit from *model zoos* (e.g., *BioImage Model Zoo* https://bioimage.io)

BioImage.IO
Advanced AI models in one click



Ex. performance gain with **transfer learning** for cell segmentation

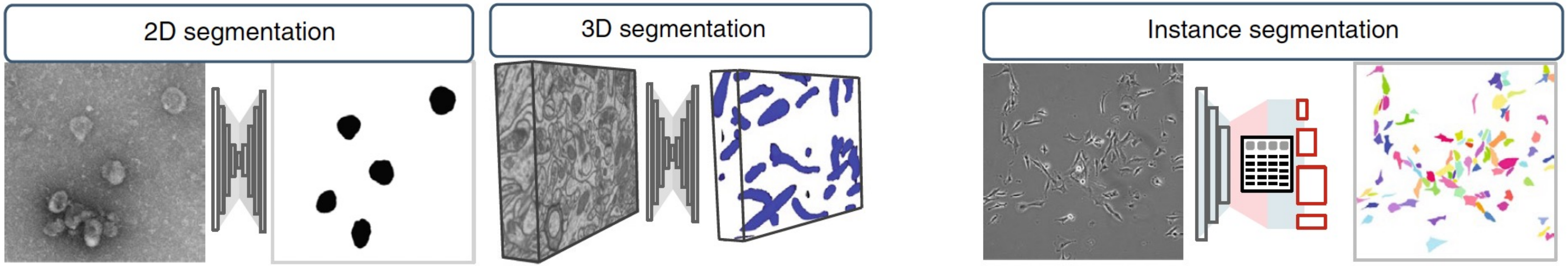
https://doi.org/10.1038/s41467-021-22518-0 (ZeroCostDL4Mic, 2021)

Image segmentation

Partitioning of images into meaningful segments

Issues: inhomogeneous background noise, low contrast, complex and varying instance structures, touching or overlapping cells, ...

- U-Net, SegNet, ...

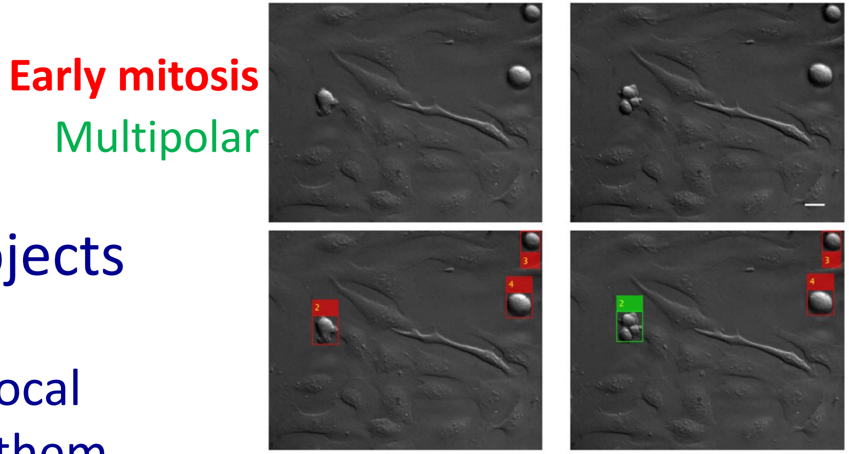


https://doi.org/10.1038/s41592-021-01262-9 (DeepImageJ, 2021)

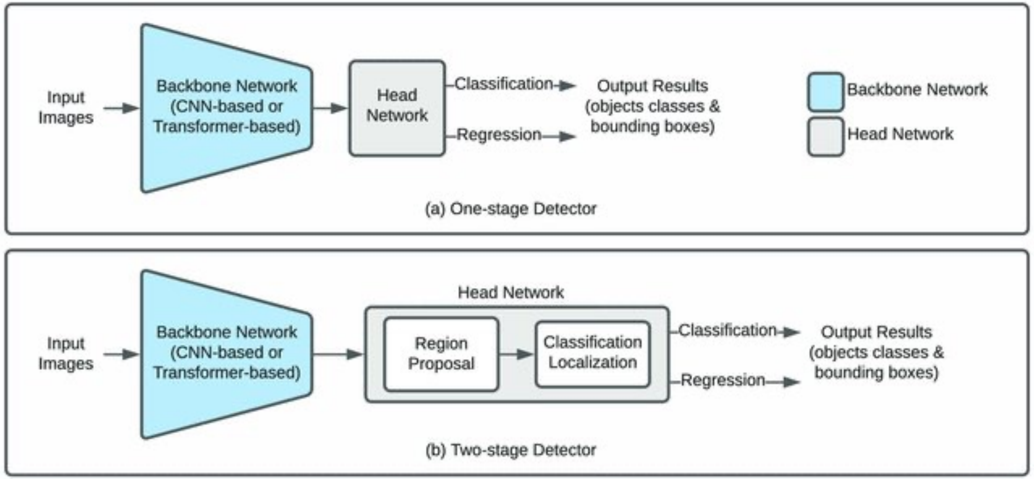
Object detection

Determine *locations* and *classes* of objects

- **Traditional approach**: extract features from local image patches and perform classification on them



https://doi.org/10.1038/s41597-023-02540-1 (ALFI, 2023)

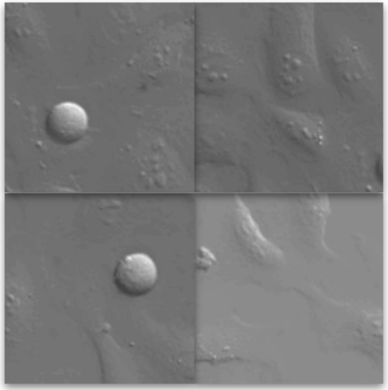


https://doi.org/10.18280/ria.370217 (2023)

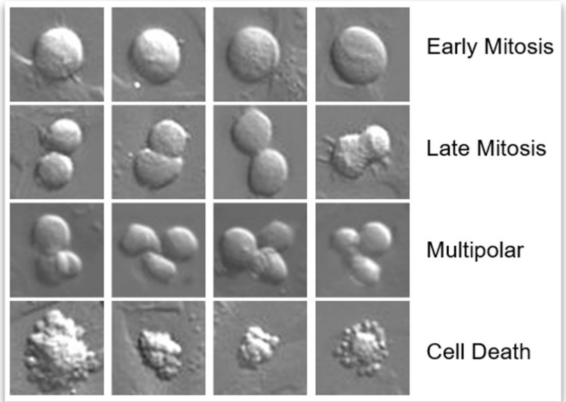
- **DL approach**
  - *one-stage*: simultaneously perform localization and classification in the head network (e.g., YOLO)
  - *two-stage*: first obtain region proposals, then perform localization and classification (e.g., R-CNN)

Example DL-based object detection

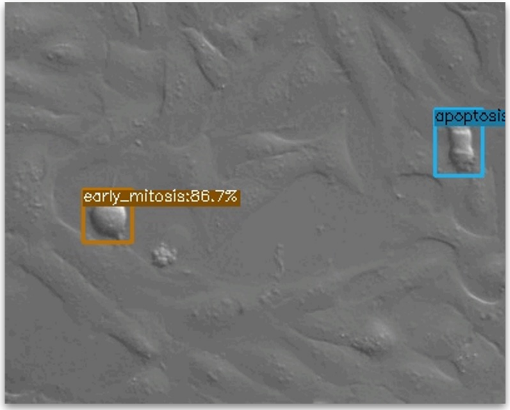
Event analysis for time-lapse microscopy



Examples of
events occuring



Identified phenotypes
(ALFI dataset)

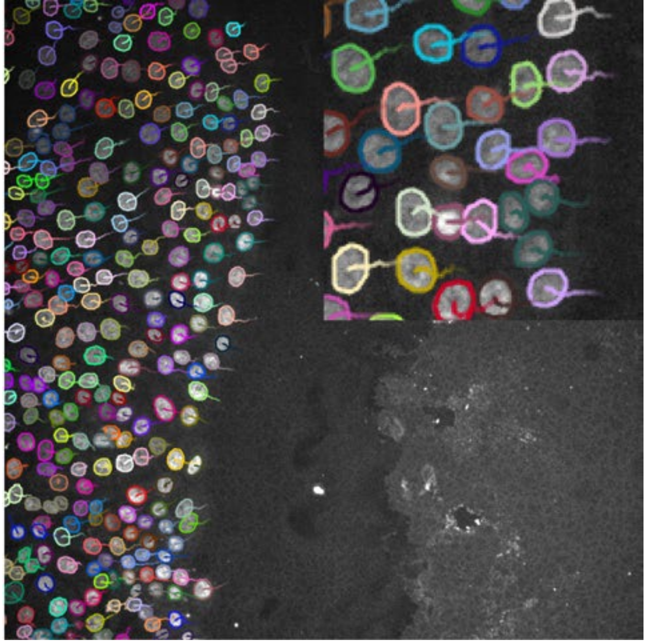


Object detection
predictions on test data

(A. Hada, Medical Imaging and Applications (MAIA) Master Thesis, 2022)

Object tracking

Follow objects through a series of time-lapse images



https://doi.org/10.1038/s41592-022-01507-1 (TrackMate7, 2022)

Issues: strongly dependent on cells detection and mitoses detection; tightly packed cells are treated as a single entity; cells change appearance in time, can appear and disappear, have erratic movement patterns; time-lapse sequences have very low temporal resolution, ...

Two major intertwined tasks:
1. Mitosis detection (tracking free or tracking-based)
2. Track establishment

Cell associations used to detect cell divisions, as well as mitotic events used as anchors for cell tracking

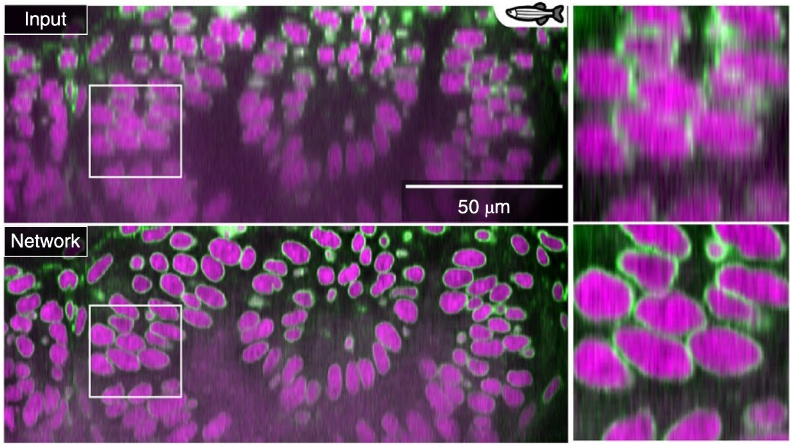
- Temporal aspects of data association and linking typically solved by traditional CV methods
- LSTM can capture longer-term dependencies among different time instances in sequential data

https://celltrackingchallenge.net (7th edition on May 27, 2024)

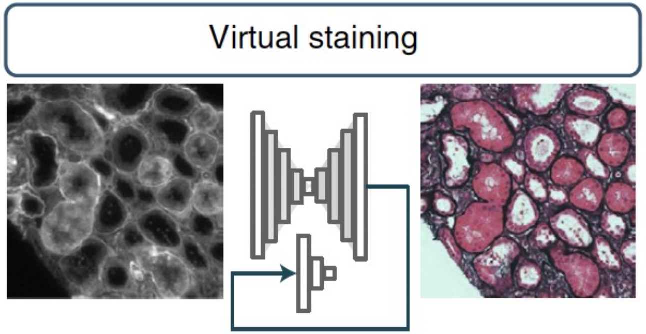
Image Enhancement

Removing artifacts and restoring essential information

- High SNR and low spatial resolution
- Denoising



https://doi.org/10.1038/s41592-018-0216-7 (Restoration, 2018)

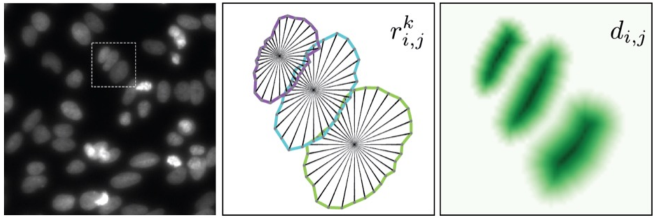


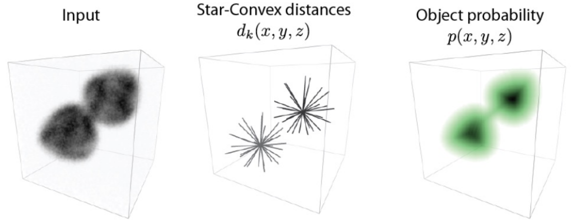
https://doi.org/10.1038/s41592-021-01262-9 (DeepImageJ, 2021)

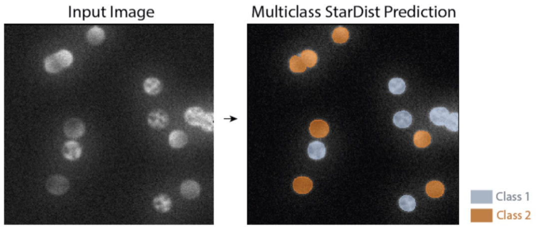
- *Cross-modality inference* or *image-to-image transformation*: transform from one type of image into another (e.g., predict fluorescent labels from transmitted-light microscopy images of unlabeled biological samples)

Publicly available software I

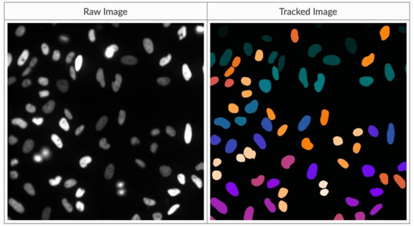
- **StarDist** https://github.com/stardist/stardist (CNN) for 2D and 3D segmentation and classification





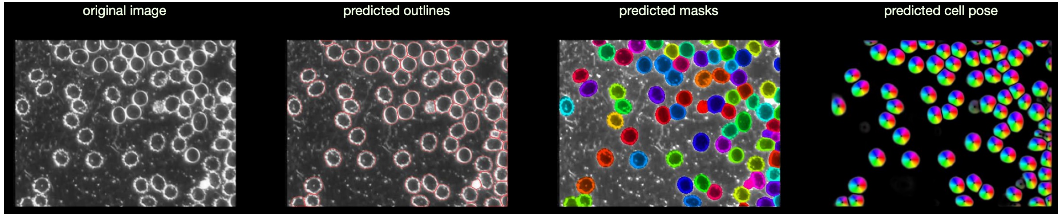


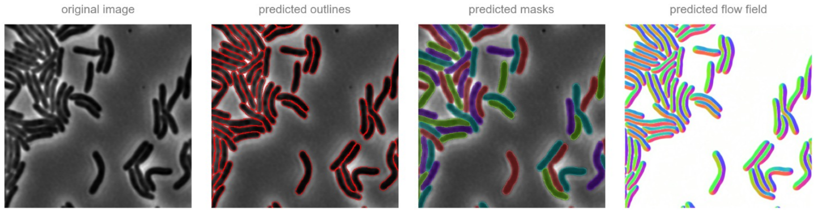
- **DeepCell** https://deepcell.org (DCNN) for segmentation, tracking, lineage, and data annotation



Publicly available software II

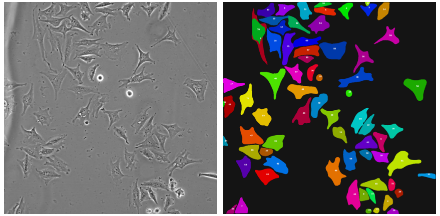
- **Cellpose** https://www.cellpose.org (U-net) for segmentation and restoration (denoising, deblurring, upsampling)





- **Omnipose** https://omnipose.readthedocs.io (U-net) for segmentation (bacteria)

- **Usiigaci** https://github.com/ElsevierSoftwareX/SOFTX_2018_158 (Mask R-CNN) for segmentation and tracking



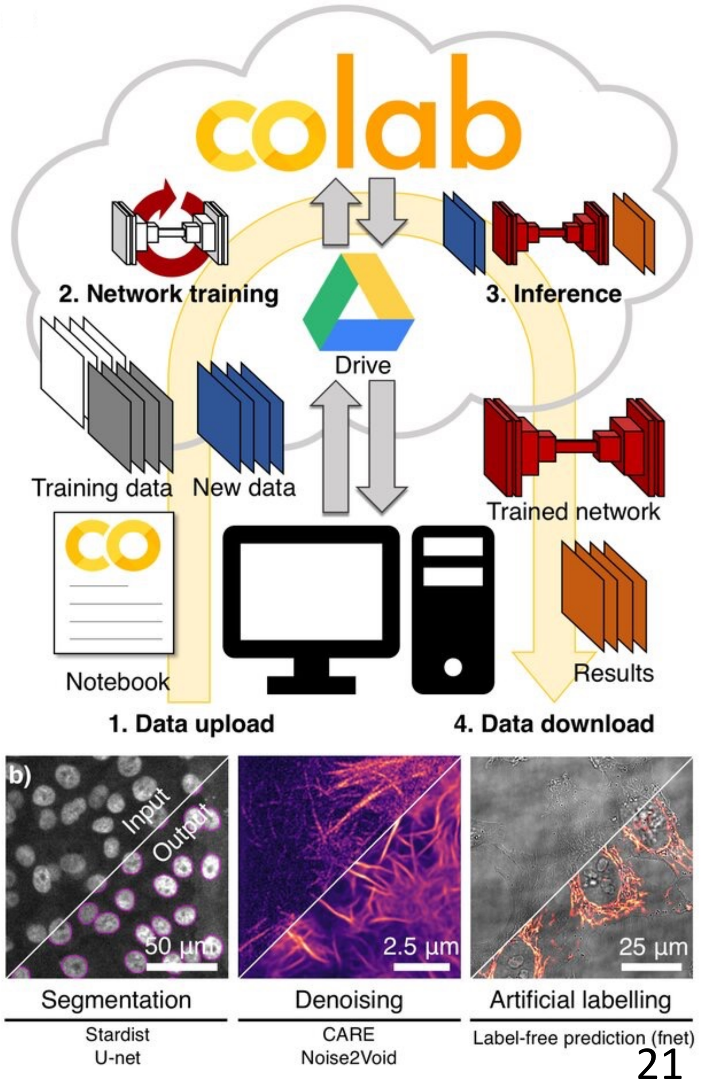
Publicly available software III

- **ZeroCostDL4Mic**
  https://github.com/HenriquesLab/ZeroCostDL4Mic
  cloud-based platform to simplify the use of DL architectures for various microscopy imaging tasks

  - 2D&3D segmentation (U-net, StarDist)
  - Object detection (YOLOv2)
  - Restoration & denoising (CARE, Noise2Void)
  - *Image-to-image translation* (e.g., fluo from BF or other fluo)

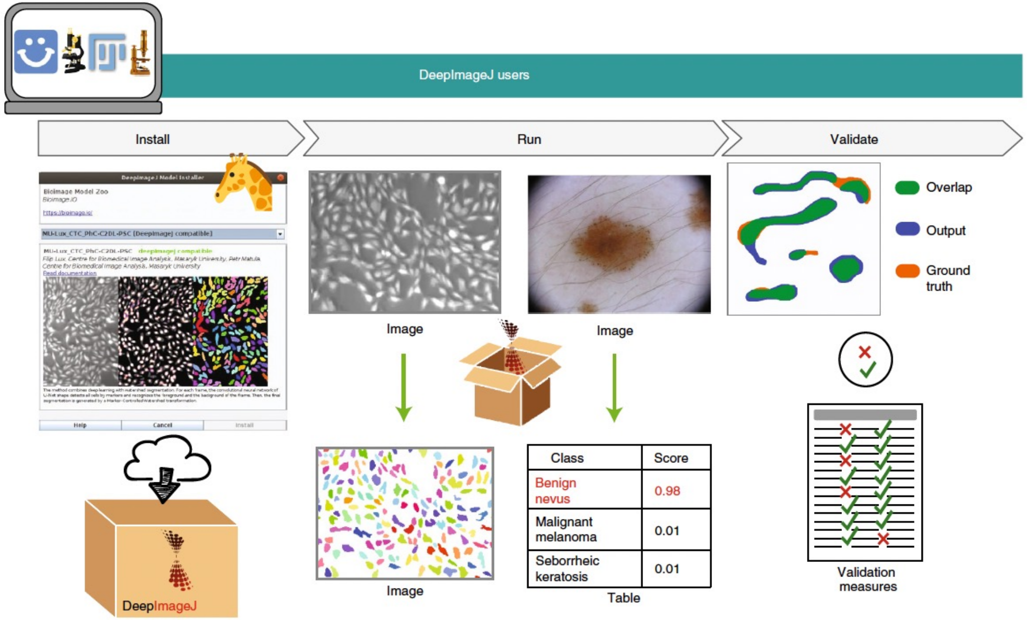
+

  - Quality control
  - Data augmentation (Augmentor)
  - Transfer learning

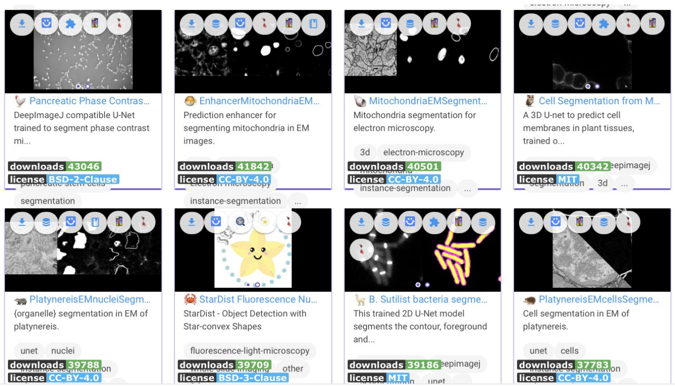


Publicly available software IV

- **DeepImageJ** https://deepimagej.github.io, environment to run DL models in ImageJ for segmentation classification, denoising, virtual staining, super-resolution



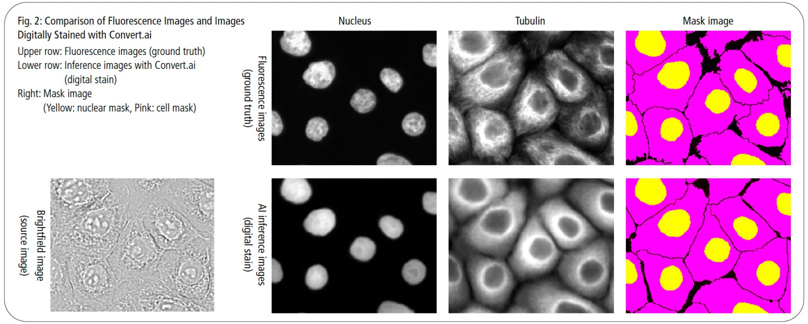
Access to pre-trained DL models from the *BioImage Model Zoo* (https://bioimage.io) repository



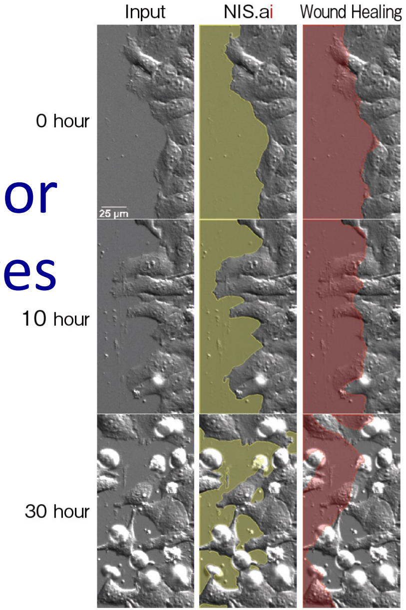
Commercial software (example)

NIS.ai **AI** **modules** for Nikon microscopes

- **Segment.ai** to extract target cells from DIC or Phase Contrast images



See later talk!

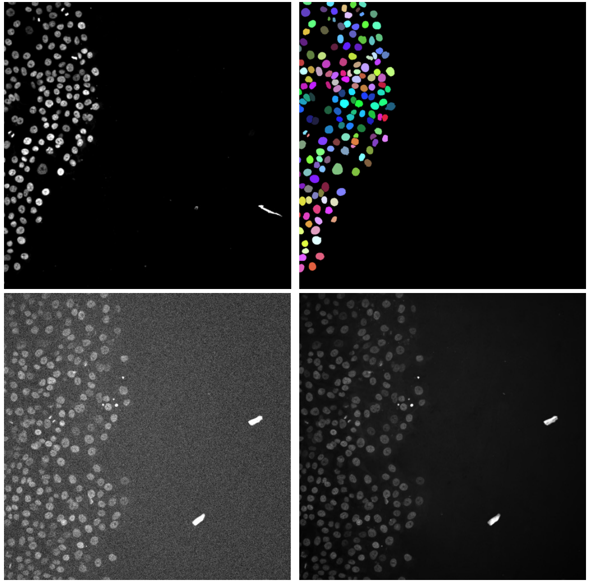


- **Convert.ai** to distinguish nuclear regions without dyeing (digital stain)

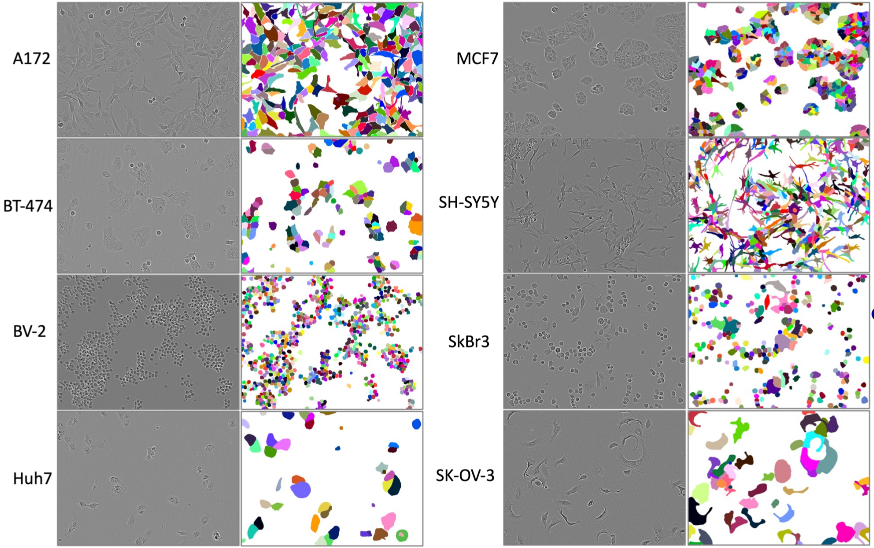
https://www.microscope.healthcare.nikon.com/it_EU/solutions/life-sciences/deep-learning-in-microscopy

Publicly available datasets I

- **StarDist**
  https://zenodo.org/records/3715492#.XnMhuXUzY5l,
  paired **fluorescence** microscopy images (SiR-DNA) and corresponding segmentation masks



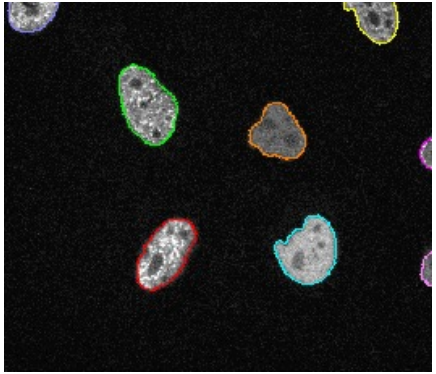
Also adapted for denoising in ZeroCostDL4Mic

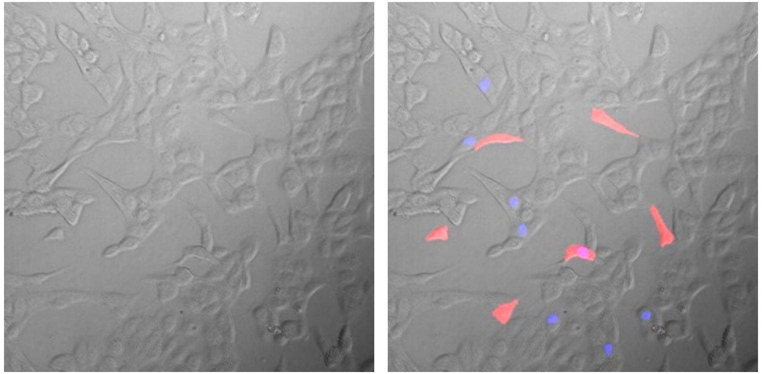


- **LIVECell**
  https://zenodo.org/records/10277106, a large-scale dataset for **label-free** live cell segmentation

Publicly available datasets II

- **Cell Tracking Challenge** https://celltrackingchallenge.net, 2D and 3D time-lapse cell segmentation and tracking benchmark, with *gold* and *silver* annotations



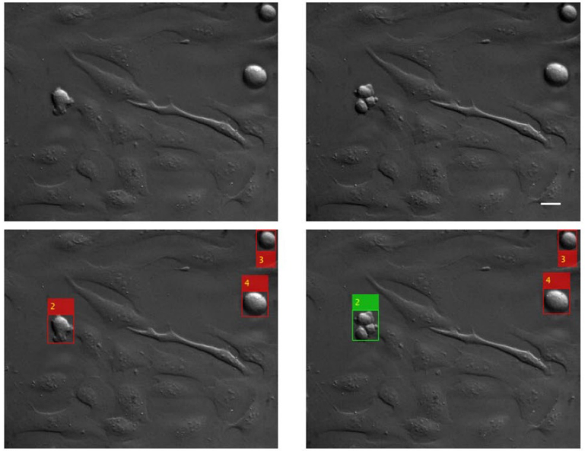


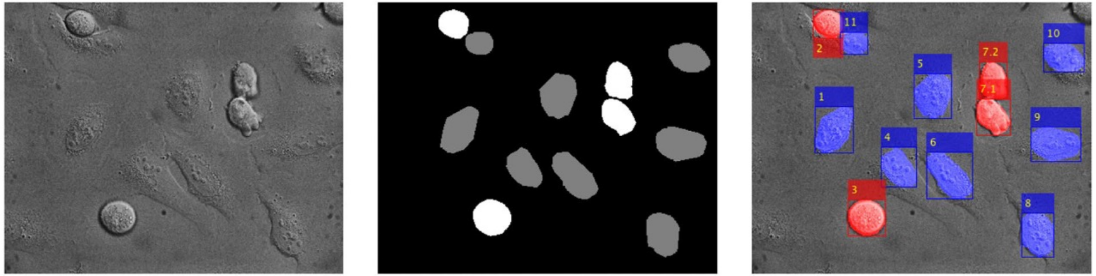
- **EVICAN (Expert VIsual Cell ANnotation) dataset** https://doi.org/10.17617/3.AJBV1S, partially annotated grayscale images of 30 different cell lines for cell and nucleus segmentation

Publicly available datasets III

- **ALFI (Annotations for Label-Free Images)**
  https://doi.org/10.6084/m9.figshare.c.6436958.v1, for segmentation, classification, tracking, and lineage

See next talk!





Acknowledgements

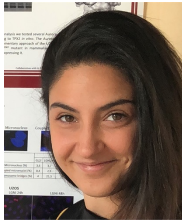
Laura Antonelli

Laura Casalino

Mario R. Guarracino

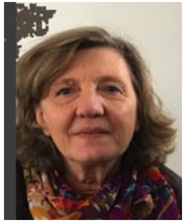
Alexandra Albu

Aroz Hada



Federica Polverino

Italia A. Asteriti



Francesca Degrassi

Giulia Guarguaglini

Istituto di Calcolo e Reti ad Alte Prestazioni

Inst. of Genetics and Byophysics, CNR

Univ. of Cassino and Souther Lazio

Inst. of Molecular Biology and Pathology, CNR

Further collaborations

CDS-group @ICAR – CNR
GNCS – INdAM
ICAR – CNR INdAM R. U.
IEOS Cancer Institute – CNR
LATNA@HSE

SLSAS-group @ICAR – CNR
The University of Sheffield
University of Campania L. Vanvitelli
University of Florida
University of Naples “Federico II” & “L’Orientale”

Some suggested reading I

- Greener et al., *A Guide to **Machine Learning** for Biologists*, Nat Rev Mol Cell Biol 23, 40–55, 2022. https://doi.org/10.1038/s41580-021-00407-0
- G. Jacquemet, ***Deep Learning** to Analyse Microscopy Images*, Biochem 43(5): 60–64, 2021. https://doi.org/10.1042/bio_2021_167
- Liu et al., *A Survey on Applications of **Deep Learning** in Microscopy Image Analysis*, Comput. Biol. Med. 134, 2021. https://doi.org/10.1016/j.compbiomed.2021.104523
- E. Meijering, *A Bird's-Eye View of **Deep Learning** in Bioimage Analysis*, CSBJ 18, 2020. https://doi.org/10.1016/j.csbj.2020.08.003
- Maddalena et al., *Artificial Intelligence for Cell Segmentation, Event Detection, and Tracking for **Label-Free Microscopy Imaging***, Algorithms 15, 2022. https://doi.org/10.3390/a15090313

Some suggested reading II

- Lucas et al., *Open-Source **Deep-Learning Software** for Bioimage **Segmentation***, Molecular Biology of the Cell 32(9), 2021. https://doi.org/10.1091/mbc.E20-10-0660
- Liu et al., ***Software Tools** for 2D Cell **Segmentation***, Cells 13(4), 2024. https://doi.org/10.3390/cells13040352

- Ma et al., *A State-of-the-Art Survey of **Object Detection** Techniques in Microorganism Image Analysis: from Classical Methods to Deep Learning Approaches*, Artif. Intell. Rev. 56, 2023. https://doi.org/10.1007/s10462-022-10209-1
- Shifat-E-Rabbi et al., *Cell Image **Classification**: A Comparative Overview*, Cytometry 97, 2020. https://doi.org/10.1002/cyto.a.23984

- Yazdi et al., *A Survey on Automated Cell **Tracking**: Challenges and Solutions*. Multimed Tools Appl, 2024. https://doi.org/10.1007/s11042-024-18697-9
- Maška et al., *The **Cell Tracking Challenge**: 10 Years of Objective Benchmarking*. Nat Methods 20, 2023. https://doi.org/10.1038/s41592-023-01879-y