# Machine Learning
(in bioimage analysis from practical point-of-view)

Lassi Paavolainen, PhD, PI
Bioimage Profiling group, FIMM

Bioimage Analysis Course 2022
8.3.2022

**FIMM**
Institute for Molecular Medicine Finland
Nordic EMBL Partnership for Molecular Medicine
**HiLIFE UNIT**

**UNIVERSITY OF HELSINKI**

# CONTENTS

1. Bioimage Profiling research group

2. Introduction to machine / deep learning in bioimage analysis

3. Introduction to various ML / DL methods and software packages for bioimage analysis

# BIOIMAGE PROFILING RESEARCH GROUP

Our research group focuses on uncovering complex information from bioimages using machine learning

## We are interested in…

› developing novel (deep learning) solutions for bioimage analysis

› studying various learning approaches to create general models

› applying these methods and models to profile cancer cell and tissue samples imaged using microscopy

## We are…

Lassi Paavolainen, PI

Gantugs Atarsaikhan

Isabel Mogollon Figueroa

Mo Baikoghli

https://www2.helsinki.fi/en/hilife-fimm/bioimage-profiling-paavolainen

3

# WHAT IS MACHINE LEARNING?

Machine learning is an approach where a model to solve a task is learned from data as opposed to explicitly defining how to solve this task

For instance, a linear model

$$f(x) = ax + b$$

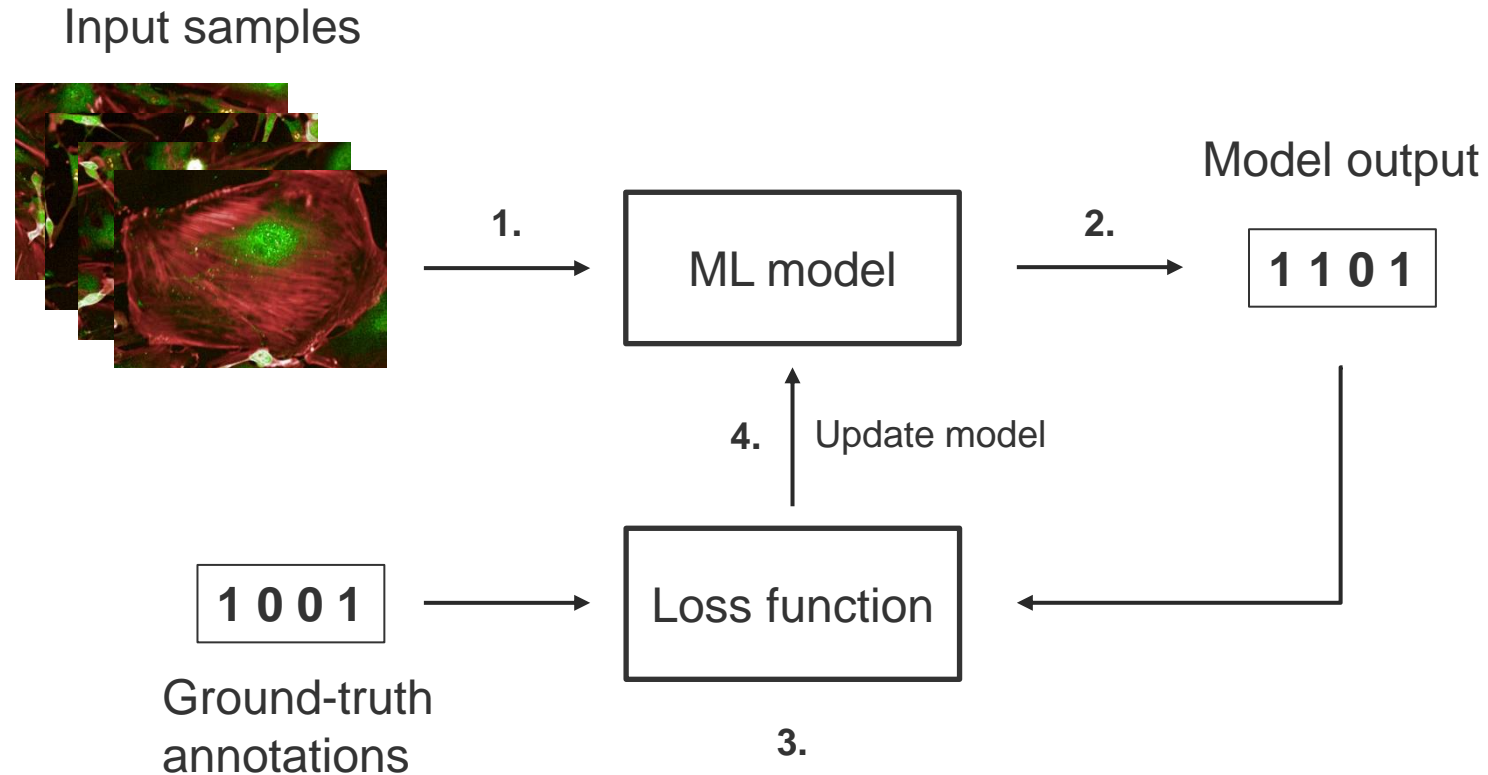can be learned from examples

$$f(0) = 3 \text{ and } f(3) = 18$$

or defined explicitly by the user (or method developer) as

$$f(x) = 5x + 3, \text{ where}$$
$$a = 5 \text{ and } b = 3$$

In practice, model (over)fitted to training data generalizes poorly to new data
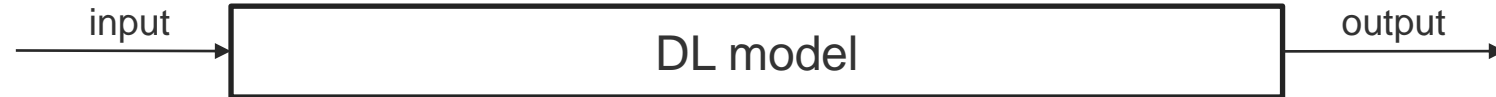
# HOW THE MODEL LEARNS FROM THE DATA



Input samples

1. → ML model

2. → Model output **1 1 0 1**

4. Update model

**1 0 0 1** → Loss function

Ground-truth annotations

3.

# WHAT IS DEEP LEARNING?

› Subset of machine learning based on deep artificial neural networks

› Used in most recent achievements in machine learning (also in image analysis)
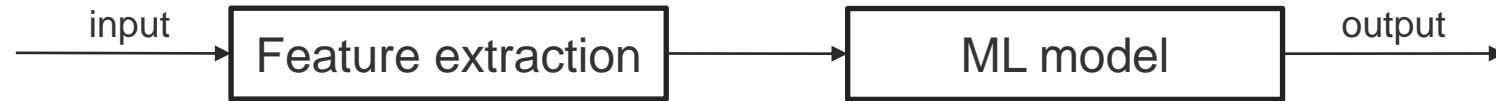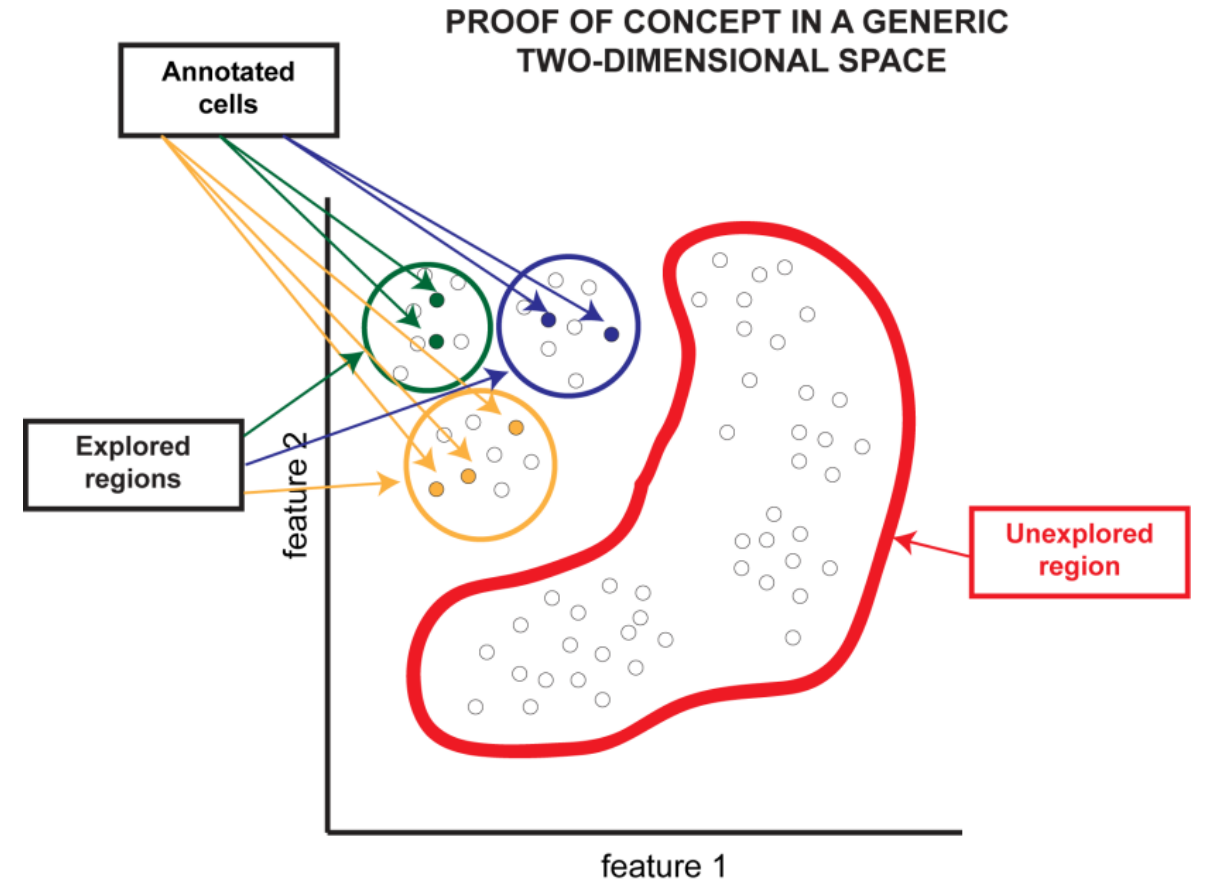
# SUPERVISED LEARNING

## Idea

1. Define a problem to solve, e.g.
   - Classes of cell phenotypes
   - Classes of images into high and low quality
   - Image foreground and background regions
2. Annotate training data to solve the problem
3. Train a model using annotated data
4. Use the model to predict rest of the data
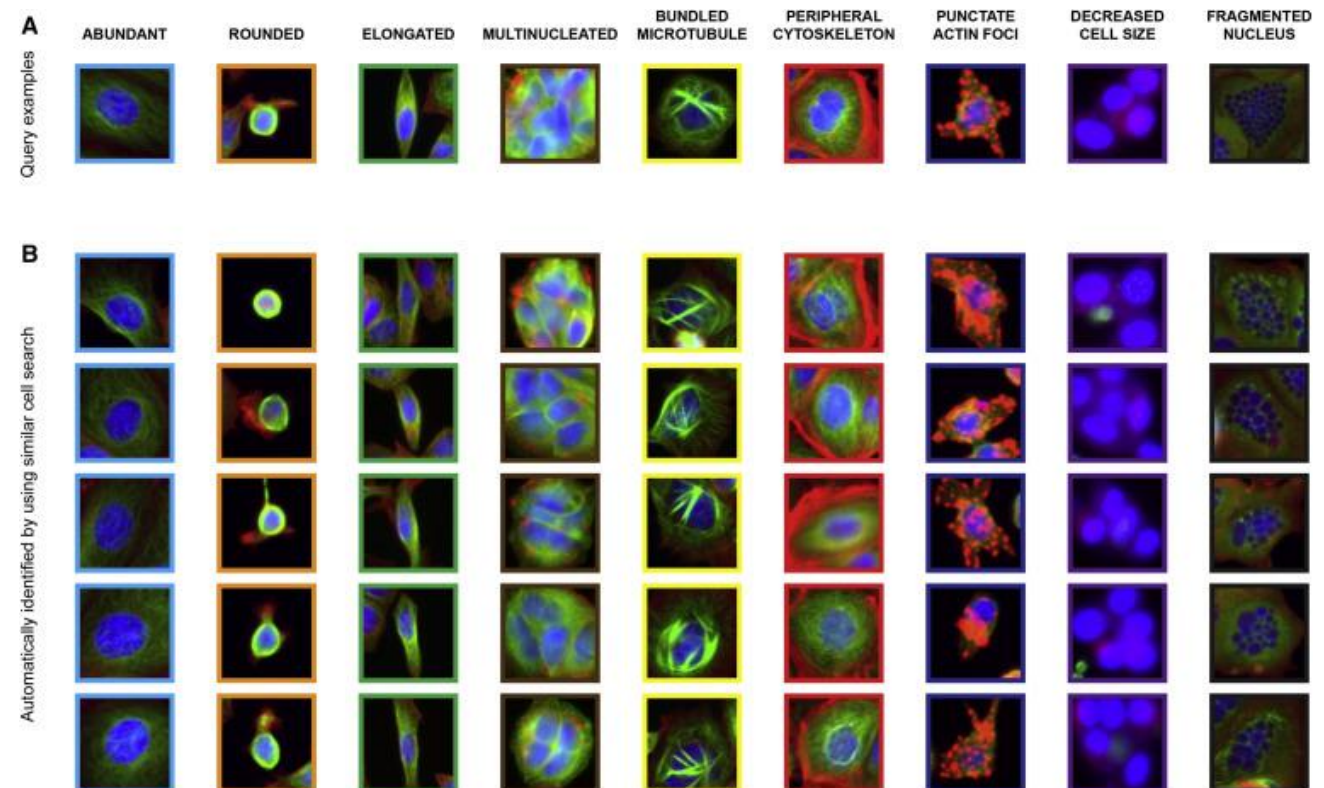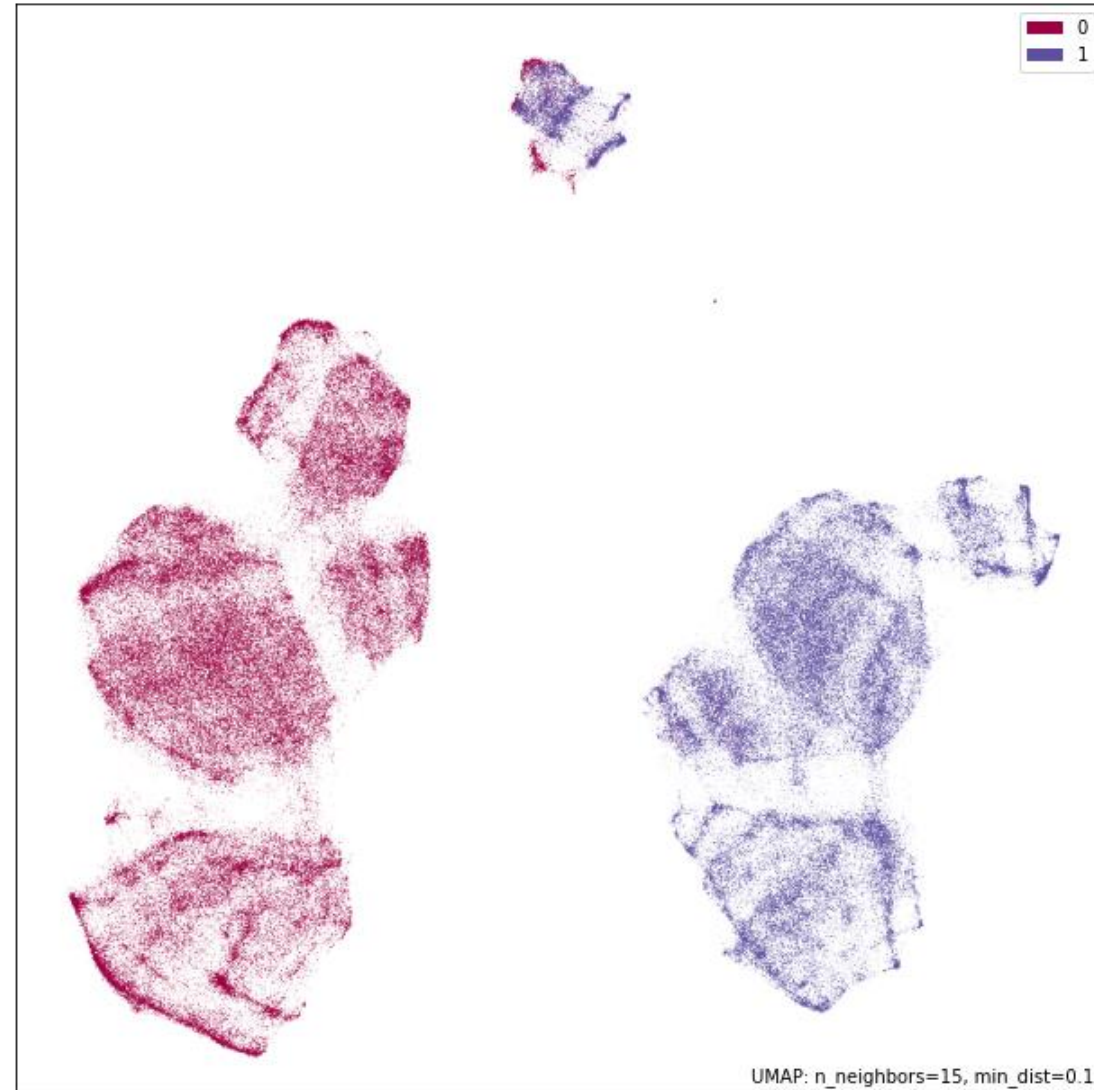
# ISSUES WITH SUPERVISED LEARNING

› Resources required to annotate samples

› Expertise needed for annotation

› Can only answer to known questions

› Annotations may include bias
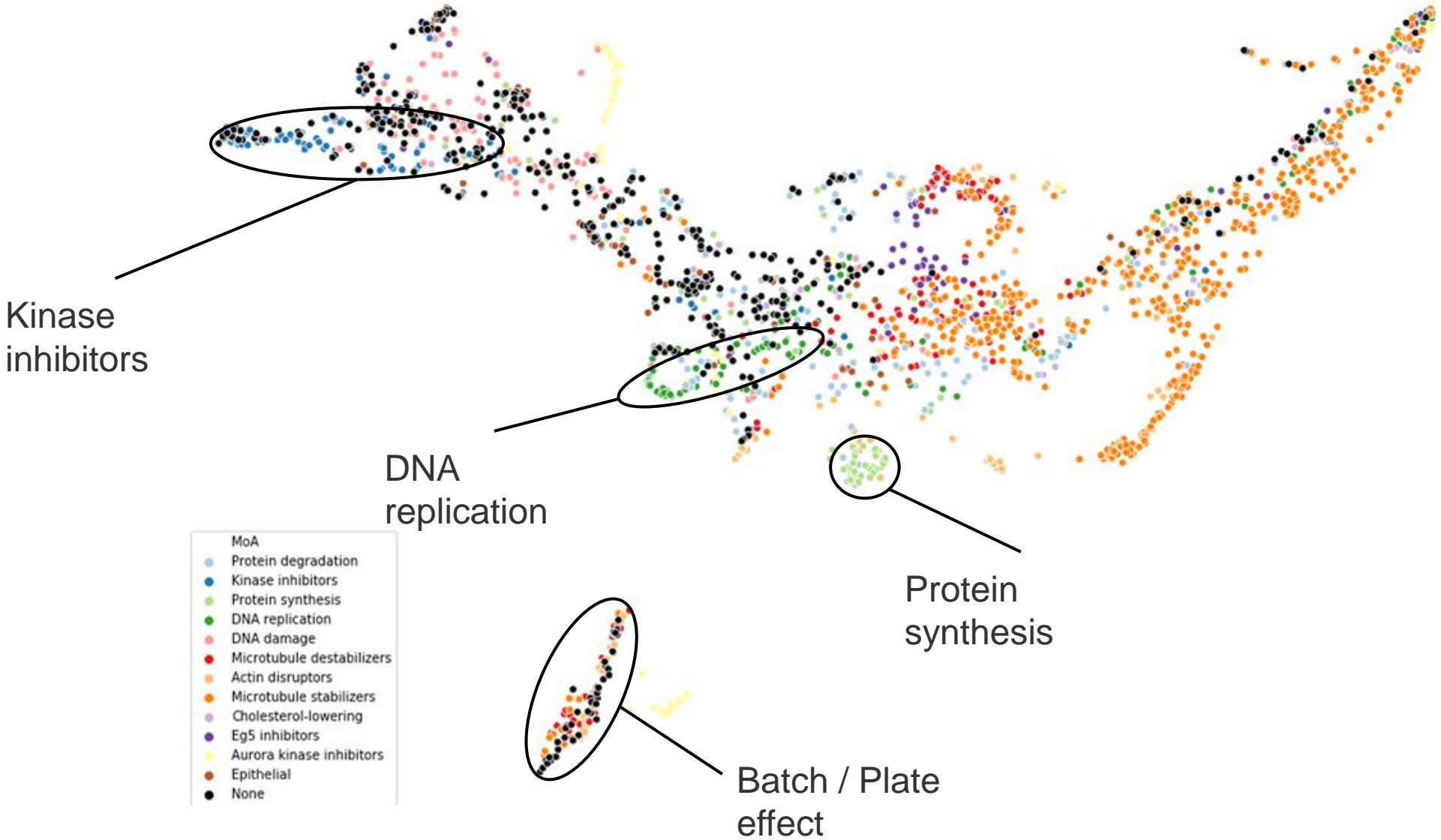
› Overfitting to training data

# UNSUPERVISED LEARNING

Find patterns in data without annotations:

- Clustering methods
- PCA
- UMAP / T-SNE

# Unsupervised learning / Explorative methods

# OTHER TYPES OF LEARNING APPROACHES

› Semi-supervised learning
  ▪ Use labeled and unlabeled data to learn patterns in data

› Transfer learning
  ▪ Use a model pre-trained with other data

› Self-supervised learning
  ▪ Learn representation of the data without labels

› Weakly-supervised learning
  ▪ Use noisy labels to learn

› Active learning
  ▪ Actively annotate uncertain samples

› Reinforcement learning
  ▪ Learn using feedback from the process
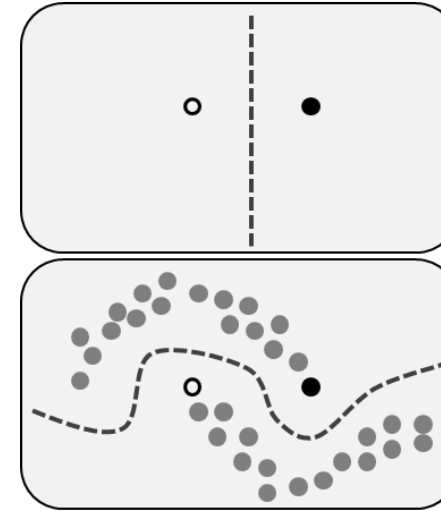


By Techerin - Own work, CC BY-SA 3.0,
https://commons.wikimedia.org/w/index.php?curid=19514958
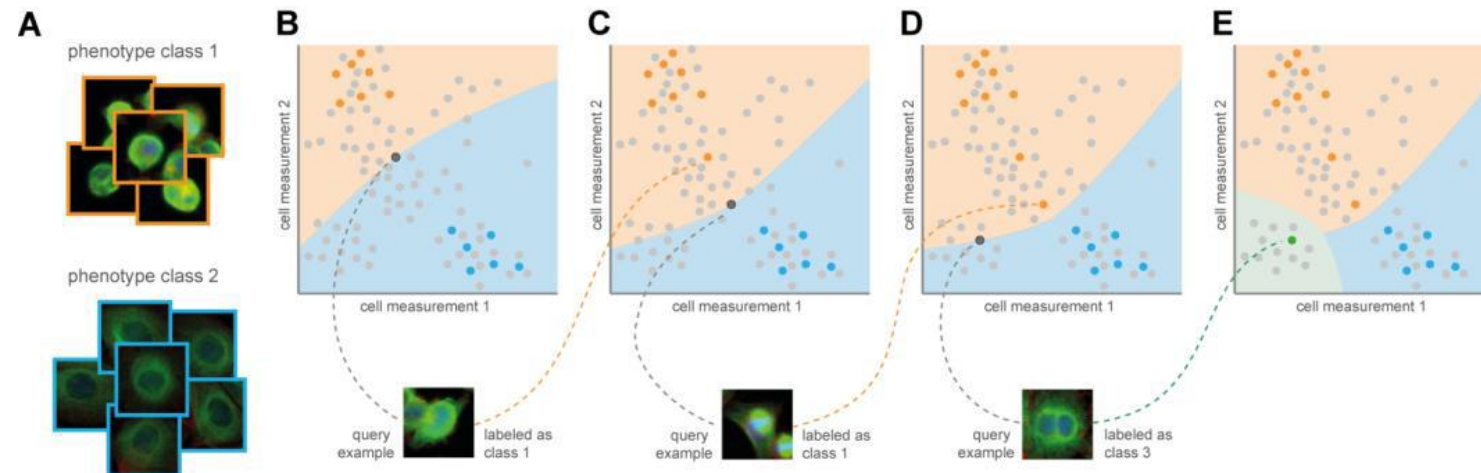
# TYPICAL ML TASKS IN BIOIMAGE ANALYSIS

› Cell classification (phenotypes, cell types)

› Image classification (QC)

› Segmentation

- Semantic (regions)
- Instance (individual nuclei/cells/spots etc.)

› Preprocessing tasks

- Denoising
- Restoration / Deconvolution

› Feature extraction

- Features learned from the data as opposed to hand-crafted features

# OBJECT CLASSIFICATION

## Classical ML approach

1. Segment objects or detect bounding boxes

2. Extract (hand-crafted) features from segmented regions

3. Annote objects to classes

4. Use extracted features and annotations to train a classifier

5. Run prediction with the rest of features

## Deep learning approach

1. Segment objects or detect bounding boxes

2. Annotate objects to classes

3. Train DL classifier using segmented regions and annotations

4. Run prediction with the rest of objects

# CELL CLASSIFICATION SOFTWARE

› Advanced Cell Classifier

  ▪ Classification and various annotation tools

  ▪ Piccinini et al. (2017). Advanced Cell Classifier: User-Friendly Machine-Learning-Based Software for Discovering Phenotypes in High-Content Imaging Data. Cell Systems, 4(6): 651-655.

  ▪ https://www.cellclassifier.org/

› CellProfiler Analyst

  ▪ Classification and feature exploration

  ▪ Stirling et al. (2021). CellProfiler Analyst 3.0: accessible data exploration and machine learning for image analysis. Bioinformatics, 37(21): 3992-3994.

  ▪ https://cellprofileranalyst.org/

› Piximi

  ▪ DL image classification on browser
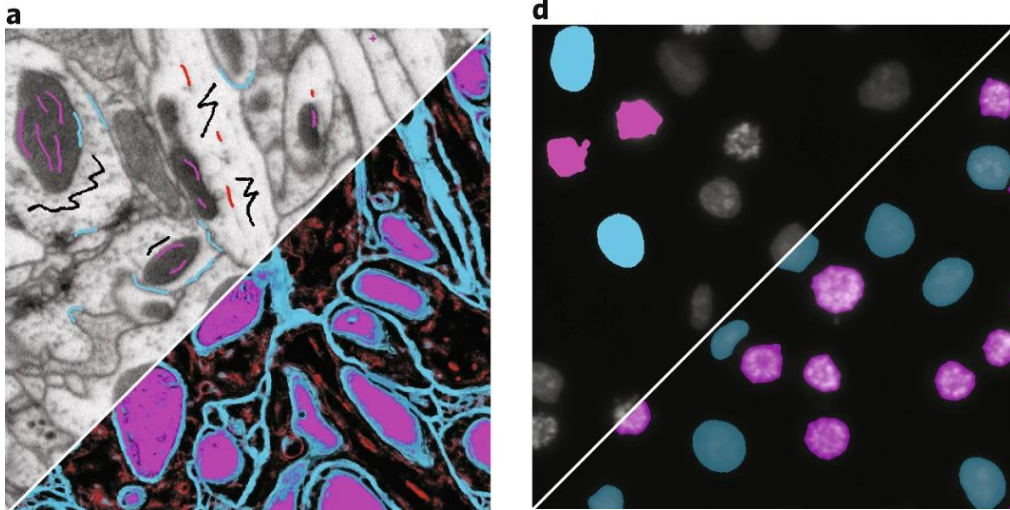
  ▪ https://github.com/piximi
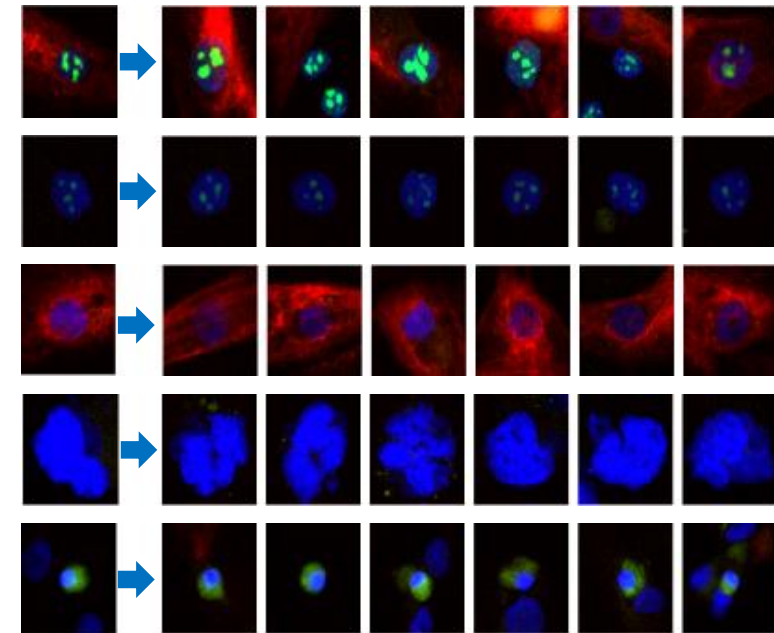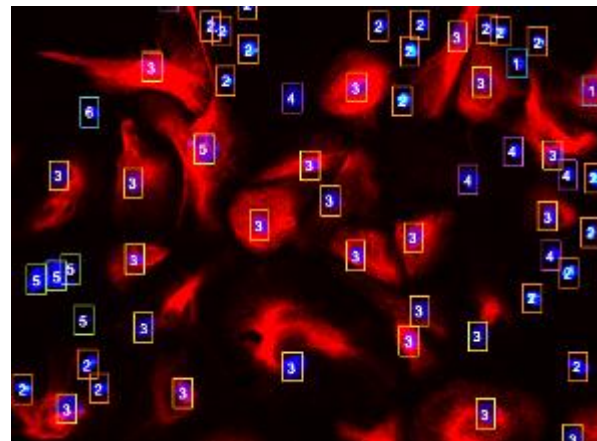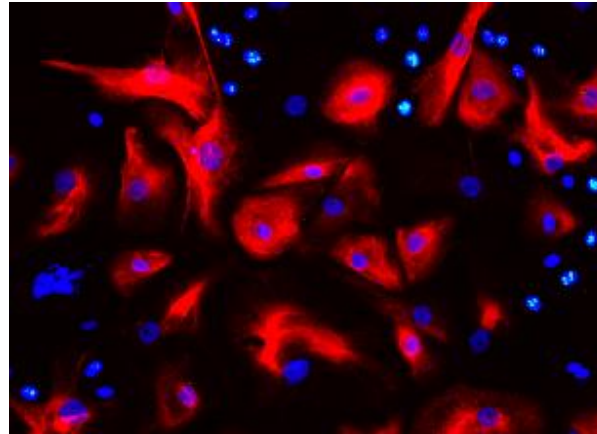
# PIXEL CLASSIFICATION FOR SEGMENTATION
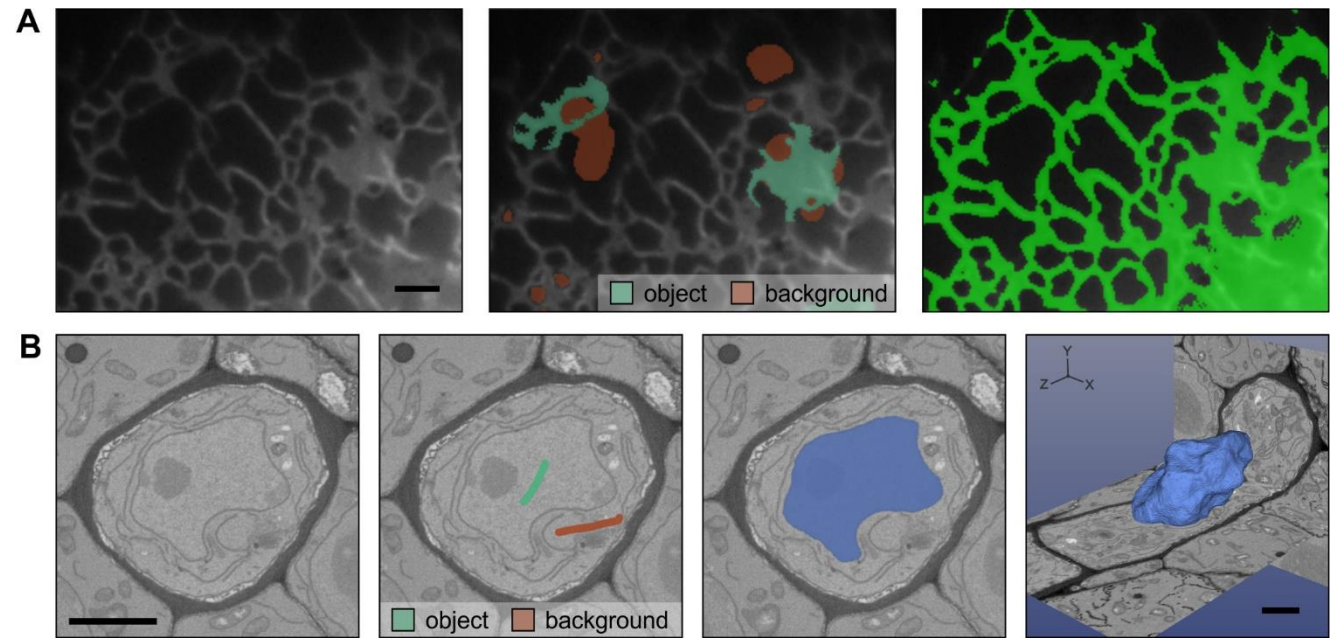
Idea:

1. Annotate pixels belonging to different classes

2. Define features from the image

3. Train a model based on features and annotated pixels

4. Predict pixels in the whole dataset using the model

› Typically requires only small amount of annotations

› Solutions are typically based on classical machine learning algorithms such as Random Forest

› Does not transfer well for other datasets

**FiMM**
**Institute for Molecular Medicine Finland**
Nordic EMBL Partnership for Molecular Medicine
**HiLIFE** UNIT

UNIVERSITY OF HELSINKI

# PIXEL CLASSIFICATION SOFTWARE

### ilastik



Berg et al. (2019). ilastik: interactive machine learning for (bio)image analysis. Nature Methods, 16: 1226-1232.

### Microscopy Image Browser



Belevich et al. (2016). Microscopy Image Browser: A Platform for Segmentation and Analysis of Multidimensional Datasets. PLoS Biology, 14(1): e1002340.
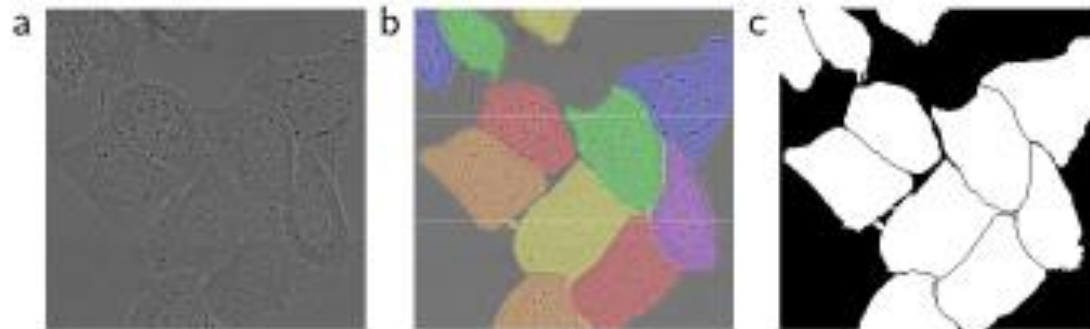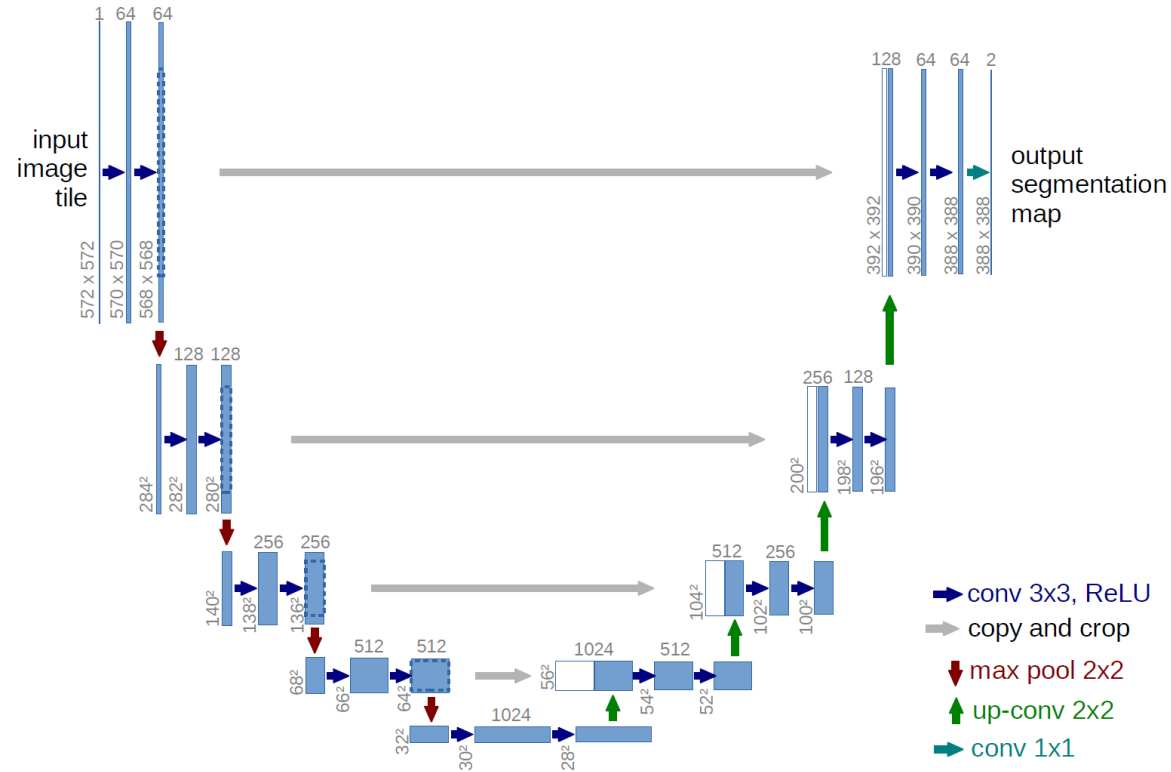
# OBJECT SEGMENTATION

› Same idea as in pixel classification for segmentation

 ▪ Most solutions require all objects annotated in the image (dense vs sparse annotation)

› Mostly based on deep learning

 ▪ Requires large amounts of training data (for robust and generalizable model)

 • Models trained using openly shared datasets, or

 • Using pre-trained models directly (the most common approach)

 ▪ Various models available for different organelles

 • Nucleus

 • Cytoplasm
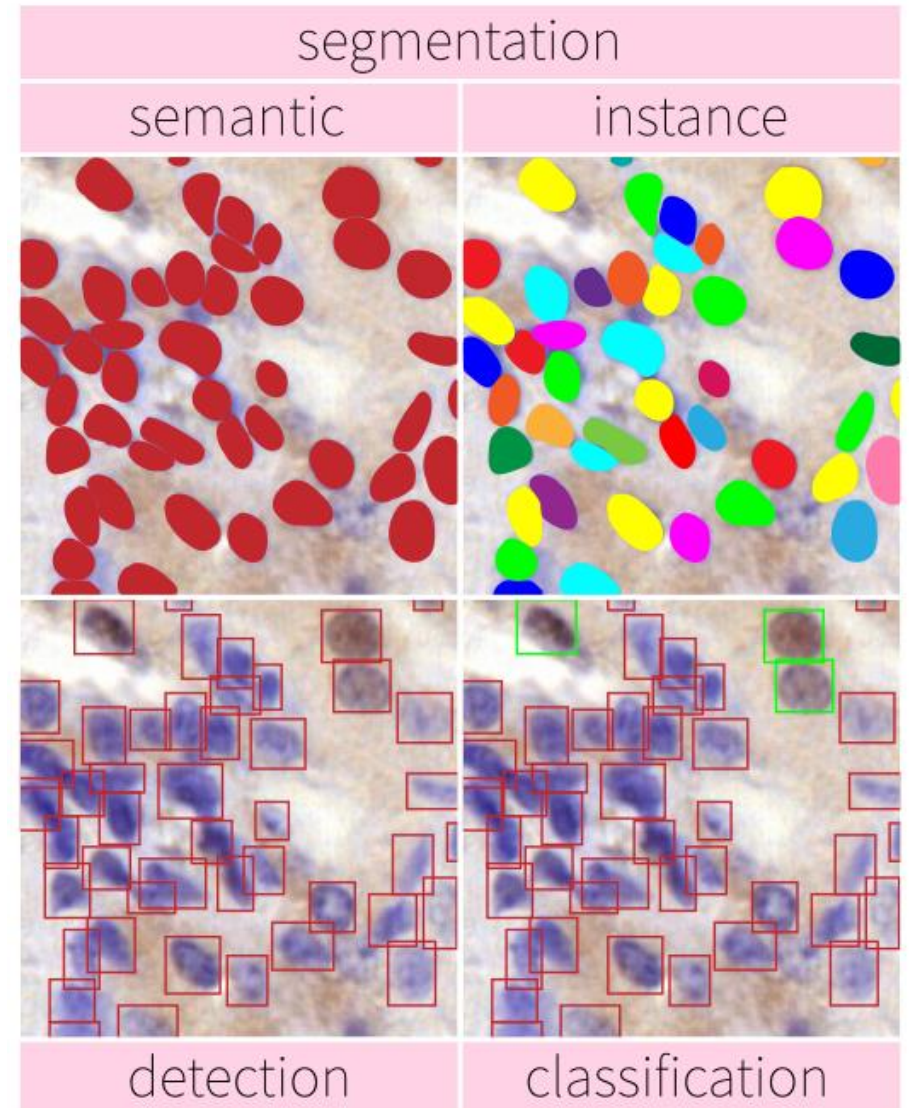


Recent review on nuclei segmentation:
Hollandi et al. (2022). Nucleus segmentation: towards automated solutions. Trends in Cell Biology.
https://doi.org/10.1016/j.tcb.2021.12.004

# U-NET

Ronneberger et al. (2015). U-Net: Convolutional Networks for Biomedical Image Segmentation. Medical Image Computing and Computer-Assisted Intervention (MICCAI), Springer, LNCS, Vol.9351: 234-241, available at arXiv:1505.04597

# nucleAIzer: A Parameter-free Deep Learning Framework for Nucleus Segmentation Using Image Style Transfer

Reka Hollandi [1], Abel Szkalisity [1], Timea Toth [1, 2], Ervin Tasnadi [1, 3], Csaba Molnar [1, 3], Botond Mathe [1], Istvan Grexa [1, 4], Jozsef Molnar [1], Arpad Balind [1], Mate Gorbe [1], Maria Kovacs [1], Ede Migh [1], Allen Goodman [6], Tamas Balassa [1, 5], Krisztian Koos [1], Wenyu Wang [7], Juan Carlos Caicedo [6], Norbert Bara [1, 8], Ferenc Kovacs [1, 8], Lassi Paavolainen [7], Tivadar Danka [1], Andras Kriston [1, 8], Anne Elizabeth Carpenter [6], Kevin Smith [9, 10], Peter Horvath [1, 7, 8, 11] ✉
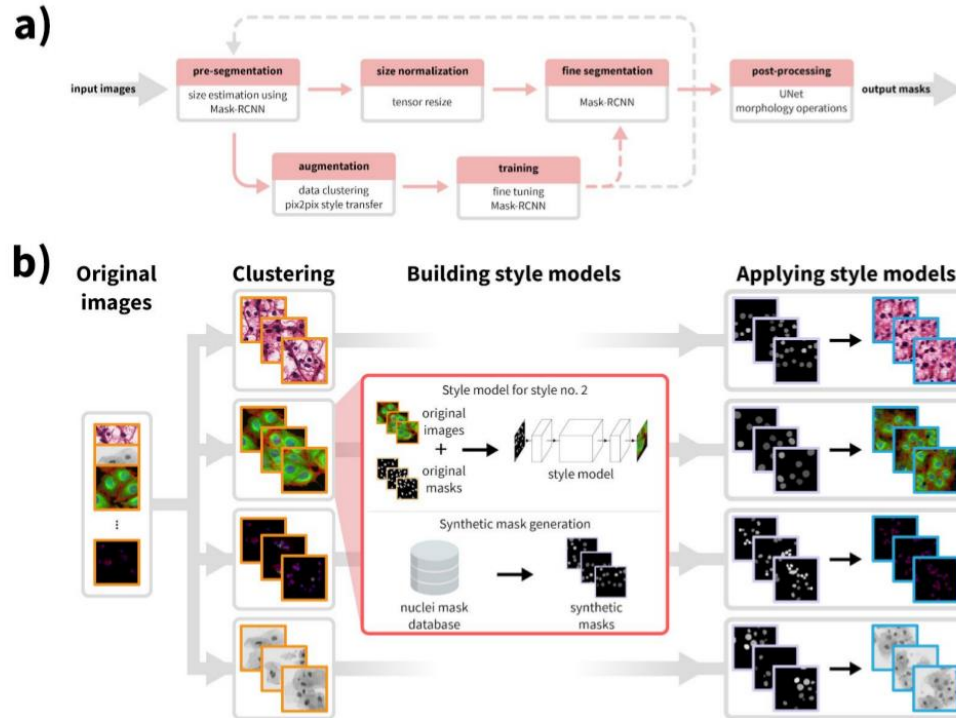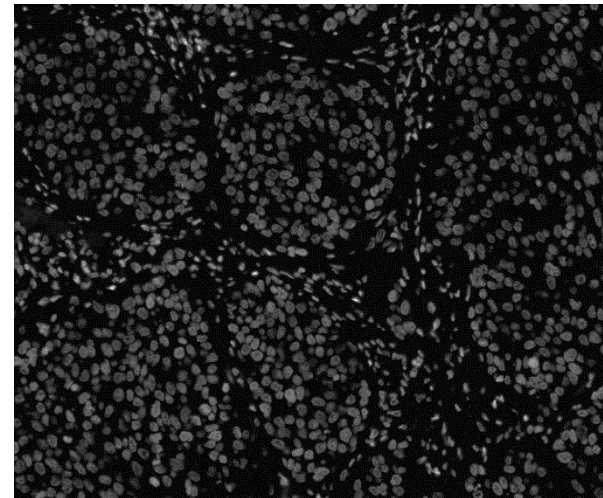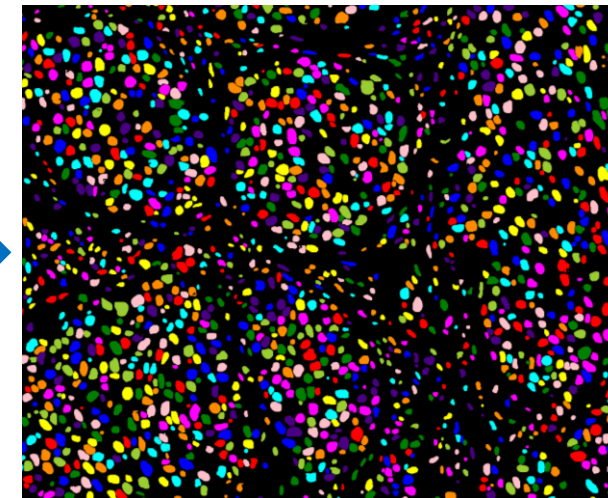
**Figure 1. Overview of our approach.** a) Upper row of boxes presents the nucleus segmentation and pre-processing; an initial Mask R-CNN network estimates typical nucleus sizes, then images are rescaled such that mean nucleus size is uniform and a Mask R-CNN network trained on images with uniform nucleus size predicts segmentations. A contour refinement step using a U-Net based network with a morphology operation is applied to obtain the final segmentation

Hollandi et al. (2020). nucleAIzer: A parameter-free deep learning framework for nucleus segmentation using image style transfer. Cell Systems, 10(5): 453–458.

## 2018 Data Science Bowl
Find the nuclei in divergent images to advance medical discovery
Featured • 2 years ago • 3634 Teams

DAPI / Hoechst channel
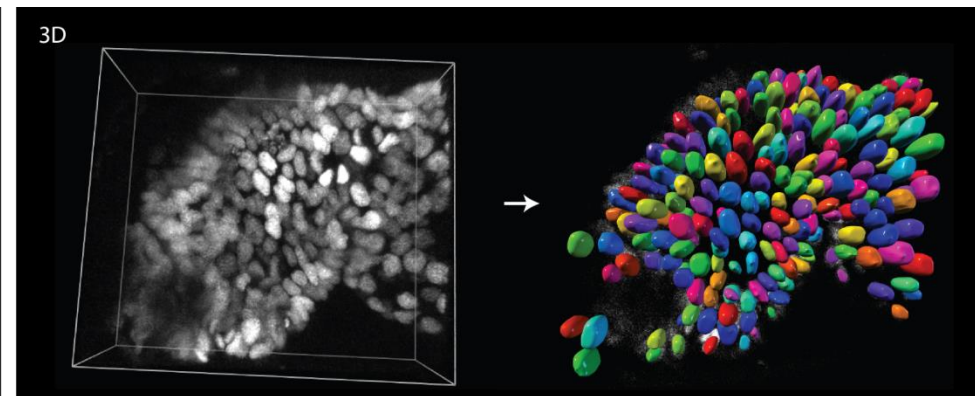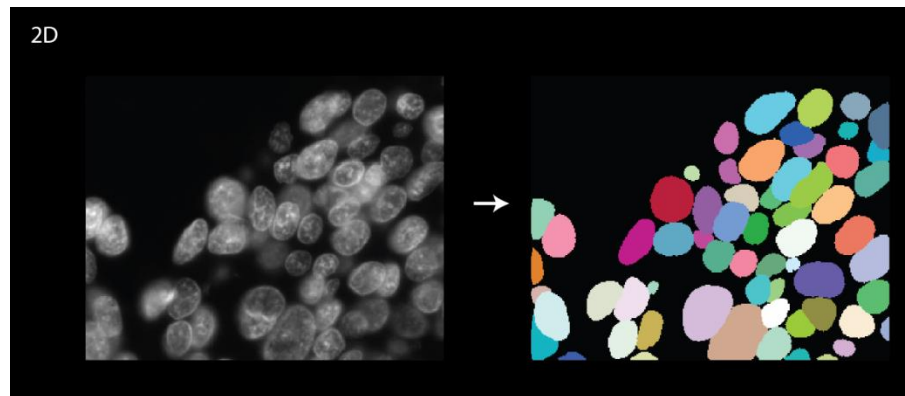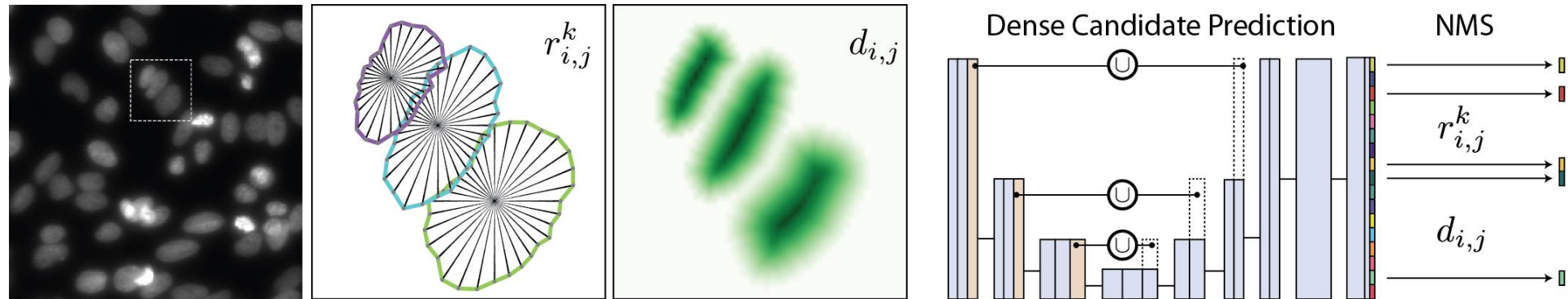
Segmentation



Available at:
http://www.nucleaizer.org/
https://github.com/lopaavol/nuclei_segmentation

# STARDIST - OBJECT DETECTION WITH STAR-CONVEX SHAPES



Schmidt et al. (2018). Cell Detection with Star-Convex Polygons. Medical Image Computing and Computer-Assisted Intervention (MICCAI), LNCS, Vol.11071: 265-275.

Available at:
https://github.com/stardist/stardist

# Cellpose: a generalist algorithm for cellular segmentation

**Carsen Stringer[†], Tim Wang, Michalis Michaelos, Marius Pachitariu[†]**

HHMI Janelia Research Campus, Ashburn, VA, USA

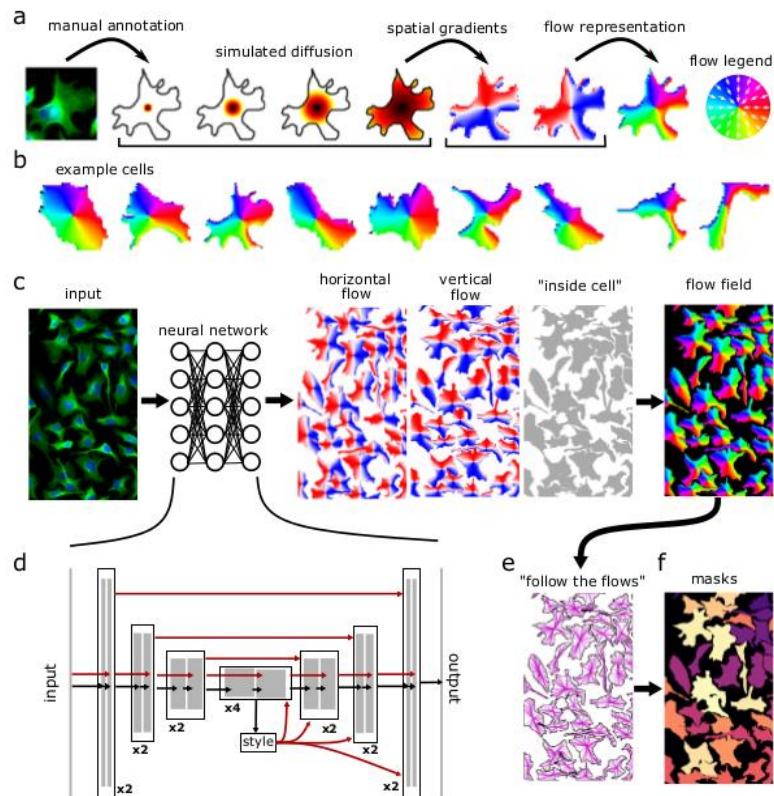[†] correspondence to (stringerc, pachitarium) @ janelia.hhmi.org

**Figure 1: Model architecture. a,** Procedure for transforming manually annotated masks into a vector flow representation that can be predicted by a neural network. A simulated diffusion process started at the center of the mask is used to derive spatial gradients that point towards the center of the cell, potentially indirectly around corners. The X and Y gradients are combined into a single normalized direction from $0°$ to $360°$. **b,** Example spatial flows for cells from the training dataset. **cd,** A neural network is trained to predict the horizontal and vertical flows, as well as whether a pixel belongs to any cell. The three predicted maps are combined into a flow field. **d** shows the details of the neural network which contains a standard backbone neural network that downsamples and then upsamples the feature maps, contains skip connections between layers of the same size, and global skip connections from the image styles, computed at the lowest resolution, to all the successive computations. **e,** At test time, the predicted flow fields are used to construct a dynamical system with fixed points whose basins of attraction represent the predicted masks. Informally, every pixel "follows the flows" along the predicted flow fields towards their eventual fixed point. **f,** All the pixels that converge to the same fixed point are assigned to the same mask.
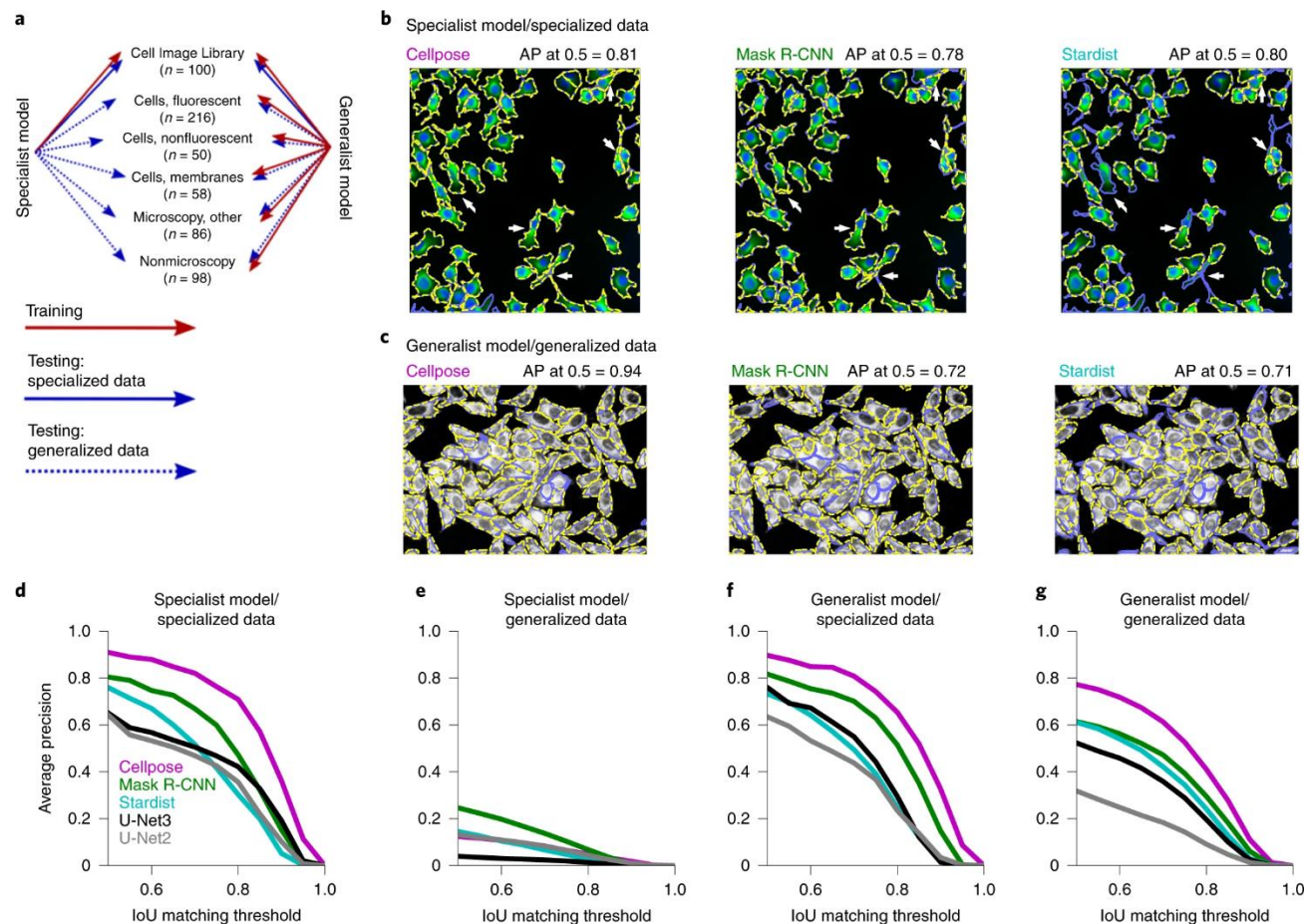
Stringer et al. (2021). Cellpose: a generalist algorithm for cellular segmentation. Nature Methods, 18: 100-106.

Available at:
https://www.cellpose.org/
https://github.com/MouseLand/cellpose

# IMAGE DENOISING, NOISE2NOISE & NOISE2VOID

Lehtinen et al. (2018). Noise2Noise: Learning Image Restoration without Clean Data. arXiv: 1803.04189
Krull et al. (2019). Noise2Void - Learning Denoising From Single Noisy Images. Proceedings of the
IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2129-2137

Available at:
https://github.com/NVlabs/noise2noise
https://github.com/juglab/n2v
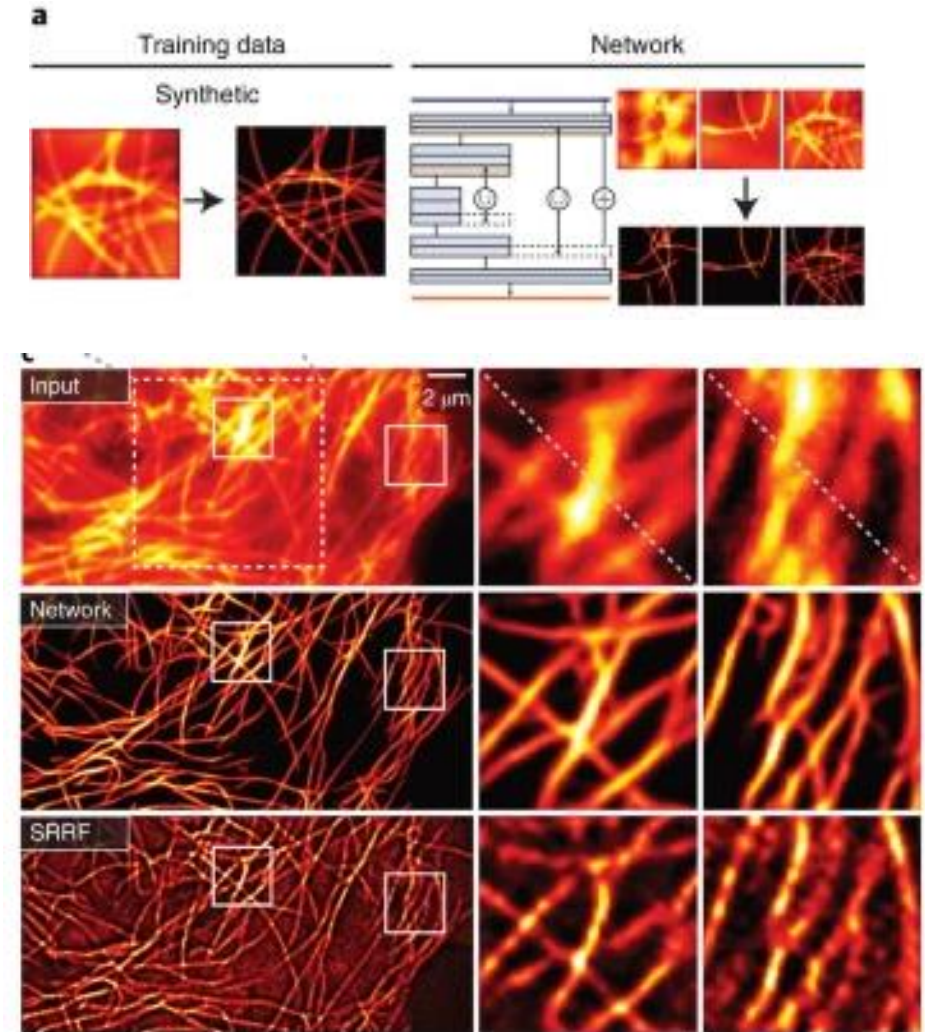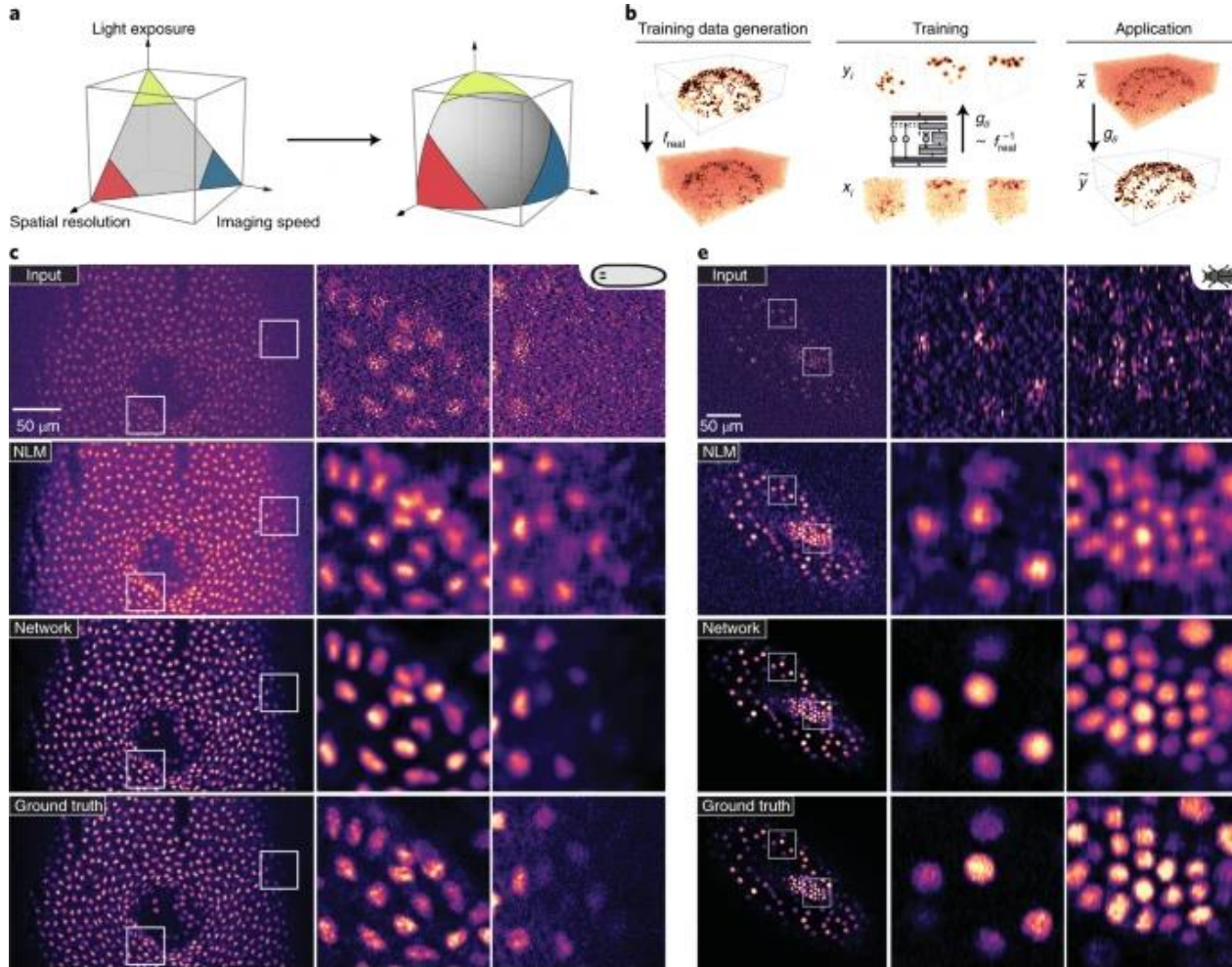
# CONTENT-AWARE IMAGE RESTORATION (CARE)



Weigert et al. (2018). Content-aware image restoration: pushing the limits of fluorescence microscopy. Nature Methods 15, 1090-1097.

Available at:
https://github.com/CSBDeep/CSBDeep

23

# METHOD / MODEL LIBRARIES

› ZeroCostDL4Mic
  - Collection of notebooks to apply various deep learning methods
  - von Chamier et al. (2021). Democratising deep learning for microscopy with ZeroCostDL4Mic. Nature Communications, 12: 2276
  - https://github.com/HenriquesLab/ZeroCostDL4Mic

› BioImage.IO (available in DeepImageJ)
  - Model library
  - https://bioimage.io/



a) Data upload    b) Network training

d) Data download    c) Inference

Notebook    Training dataset    Trained network    Dataset    Results    Data transfer
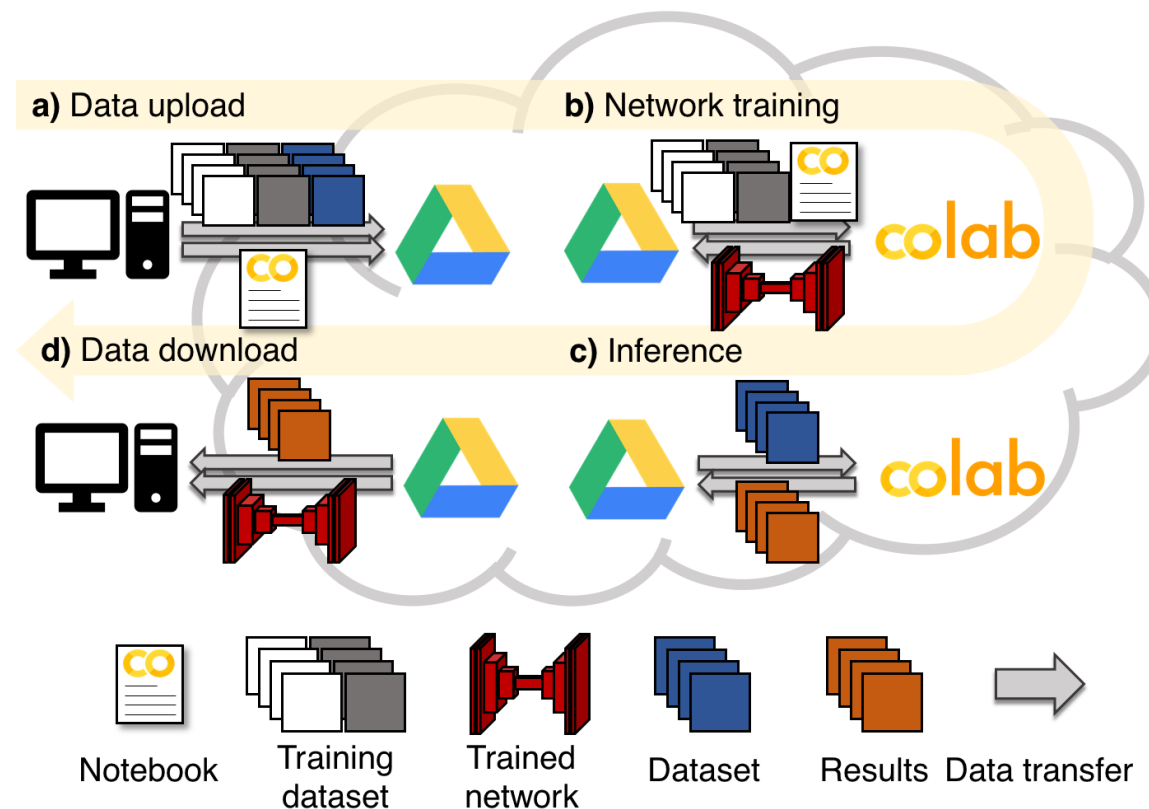
# PLATFORMS / MODEL LIBRARIES

› CDeep3M
  - Haberl et al. (2018). CDeep3M—Plug-and-Play cloud-based deep learning for image segmentation. Nature Methods, 15: 677-680.
  - https://cdeep3m.crbs.ucsd.edu/cdeep3m

› ImJoy
  - Ouyang et al. (2019). ImJoy: an open-source computational platform for the deep learning era. Nature Methods, 16: 1199-1200.
  - https://imjoy.io/

› Deep Cell Kiosk
  - Bannon et al. (2021). DeepCell Kiosk: scaling deep learning–enabled cellular image analysis with Kubernetes. Nature Methods, 18: 43-45.
  - https://deepcell.org/

› nucleAIzer
  - Hollandi et al. (2020). nucleAIzer: A Parameter-free Deep Learning Framework for Nucleus Segmentation Using Image Style Transfer. Cell Systems, 10(5): 453-458.
  - http://www.nucleaizer.org/

FiMM
Institute for Molecular Medicine Finland
Nordic EMBL Partnership for Molecular Medicine
HiLIFE UNIT

UNIVERSITY OF HELSINKI

# SUMMARY

› Machine learning can be easy to apply

› Most recent achievements are based on deep learning

› Deep learning can solve image analysis problems previously considered extremely challenging

› Various software packages exist optimized for different applications

› General models are available though often some training is still needed to optimize for user's own data

# REFERENCES

› Bannon et al. (2021). DeepCell Kiosk: scaling deep learning–enabled cellular image analysis with Kubernetes. Nature Methods, 18: 43-45.

› Belevich et al. (2016). Microscopy Image Browser: A Platform for Segmentation and Analysis of Multidimensional Datasets. PLoS Biology 14(1): e1002340.

› Berg et al. (2019). ilastik: interactive machine learning for (bio)image analysis. Nature Methods 16, 1226–1232.

› Haberl et al. (2018). CDeep3M—Plug-and-Play cloud-based deep learning for image segmentation. Nature Methods, 15: 677-680.

› Hollandi et al. (2020). nucleAIzer: A parameter-free deep learning framework for nucleus segmentation using image style transfer. Cell Systems, 10(5): 453–458.

› Hollandi et al. (2022). Nucleus segmentation: towards automated solutions. Trends in Cell Biology.

› Krull et al. (2019). Noise2Void - Learning Denoising From Single Noisy Images. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2129-2137.

› Lehtinen et al. (2018). Noise2Noise: Learning Image Restoration without Clean Data. arXiv: 1803.04189.

› Ouyang et al. (2019). ImJoy: an open-source computational platform for the deep learning era. Nature Methods, 16: 1199-1200.

› Piccinini et al. (2017). Advanced Cell Classifier: User-Friendly Machine-Learning-Based Software for Discovering Phenotypes in High-Content Imaging Data. Cell Systems, 4(6): 651-655.

› Ronneberger et al. (2015). U-Net: Convolutional Networks for Biomedical Image Segmentation. Medical Image Computing and Computer-Assisted Intervention (MICCAI), LNCS, Vol.9351: 234-241, available at arXiv:1505.04597

› Schmidt et al. (2018). Cell Detection with Star-Convex Polygons. Medical Image Computing and Computer-Assisted Intervention (MICCAI), LNCS, Vol.11071: 265-275.

› Stirling et al. (2021). CellProfiler Analyst 3.0: accessible data exploration and machine learning for image analysis. Bioinformatics, 37(21): 3992-3994.

› Stringer et al. (2021). Cellpose: a generalist algorithm for cellular segmentation. Nature Methods, 18: 100-106.

› von Chamier et al. (2021). Democratising deep learning for microscopy with ZeroCostDL4Mic. Nature Communications, 12: 2276

› Weigert et al. (2018). Content-aware image restoration: pushing the limits of fluorescence microscopy. Nature Methods 15, 1090-1097.

FiMM

**Institute for Molecular Medicine Finland**
Nordic EMBL Partnership for Molecular Medicine
HiLIFE UNIT

UNIVERSITY OF HELSINKI