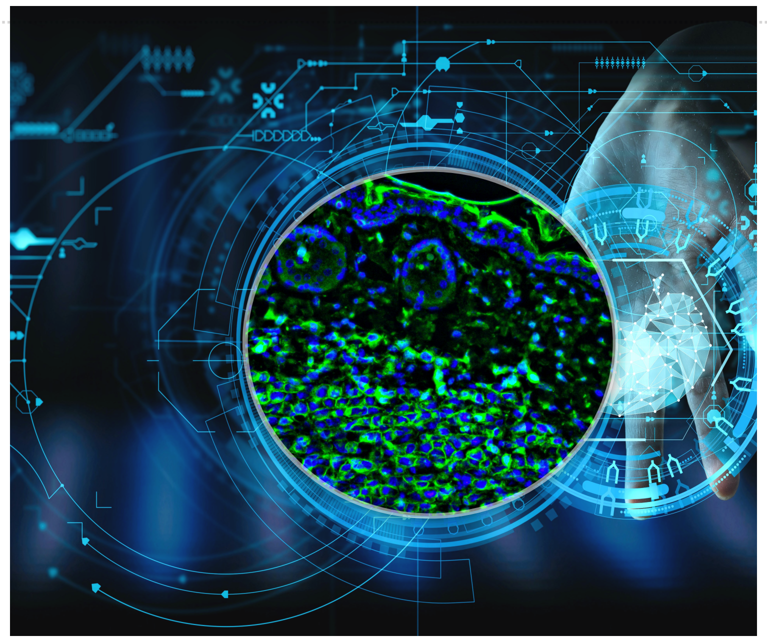
Deep Learning for Bioimage Analysis

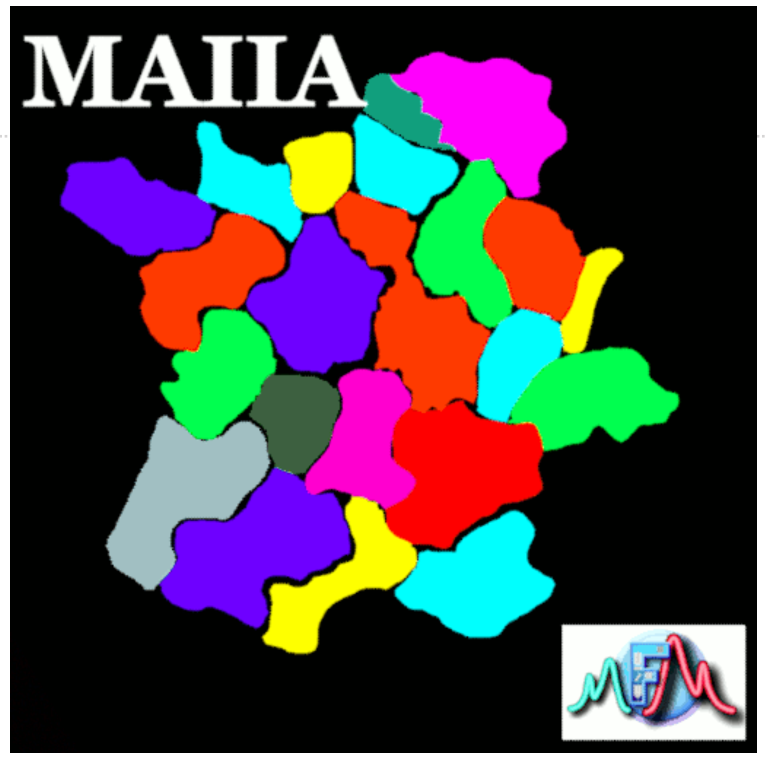
Concepts and Hands-on Neural Networks Training with a Critical Approach



Anaïs Badoual
SERPICO
INRIA Atlantique
Rennes France

Daniel Sage
Biomedical Imaging Group
EPFL Center for Imaging
Lausanne Switzerland

Héloïse Monnet
Orion CIRB
Collège de France
Paris France



Deep Learning for Bioimage Analysis

Concepts and Hands-on Neural Networks Training with a Critical Approach

A INTRODUCTION

Context
DL Tasks
Supervised learning
Fine-tuning

B WORKSHOP

Prepare datasets
Training on Colab
Analysis of results
Prediction on deepImageJ

C CONCLUSION

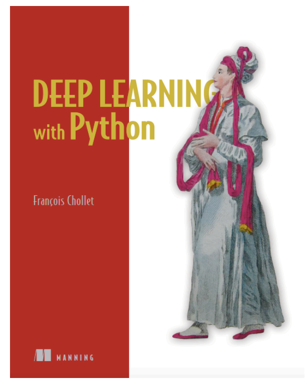
Design of datasets
Risks and challenges
Concluding remarks

Deep Learning in Bioimaging

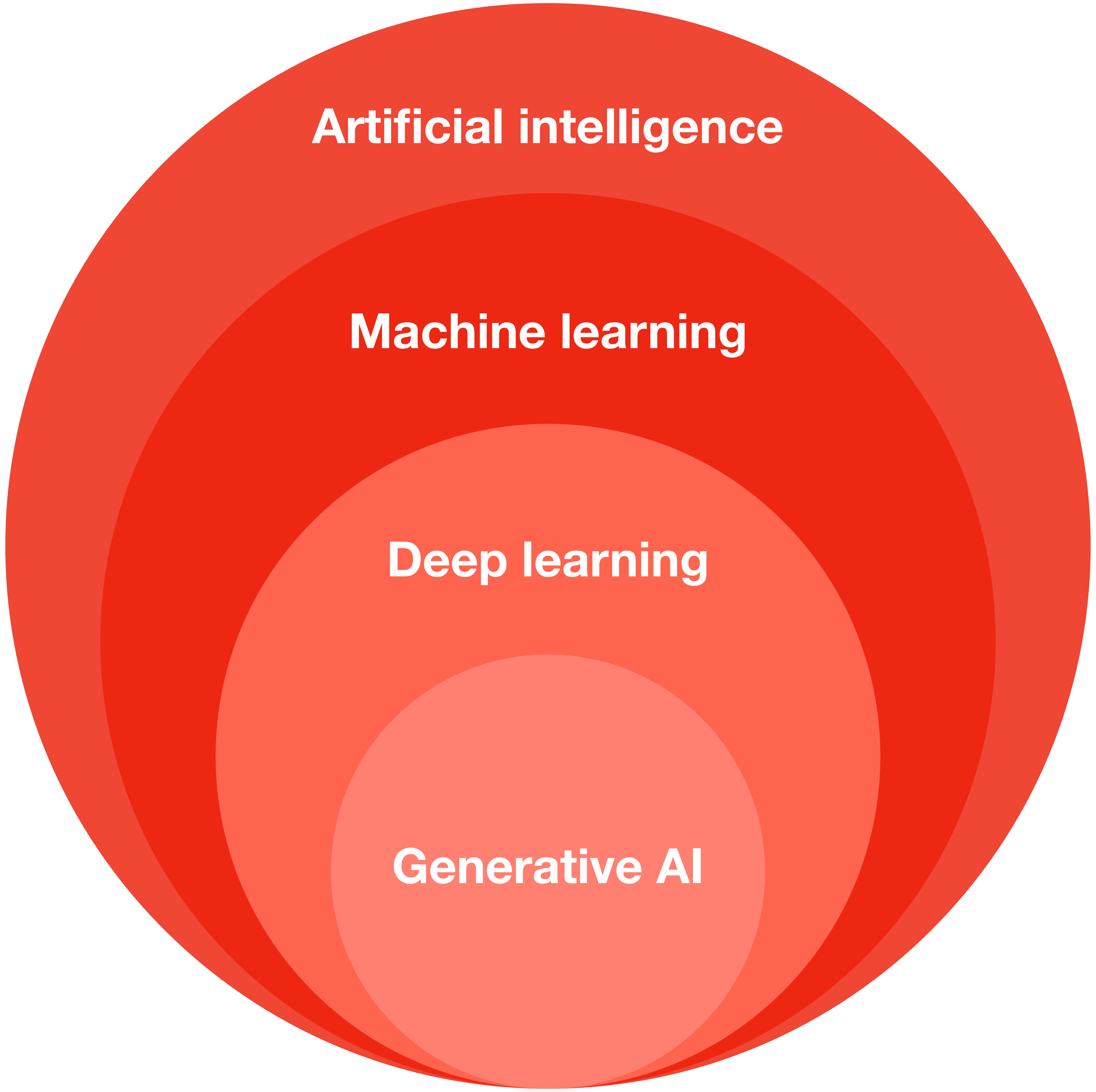
IMAGE RECONSTRUCTION	BIOIMAGE ANALYSIS
Deconvolution	Image classification
Registration	Object classification
Tomography	Object detection
Super-resolution	Image segmentation

Belthangady and Royer, Nat. Meth., 2019
McCann et al., IEEE 2017

Meijering, A bird's-eye view of Deep Learning, 2020
Hallou et al., Development, 2021



François Chollet
Deep learning with python. 2021
free PDF online



Deep Learning Tasks

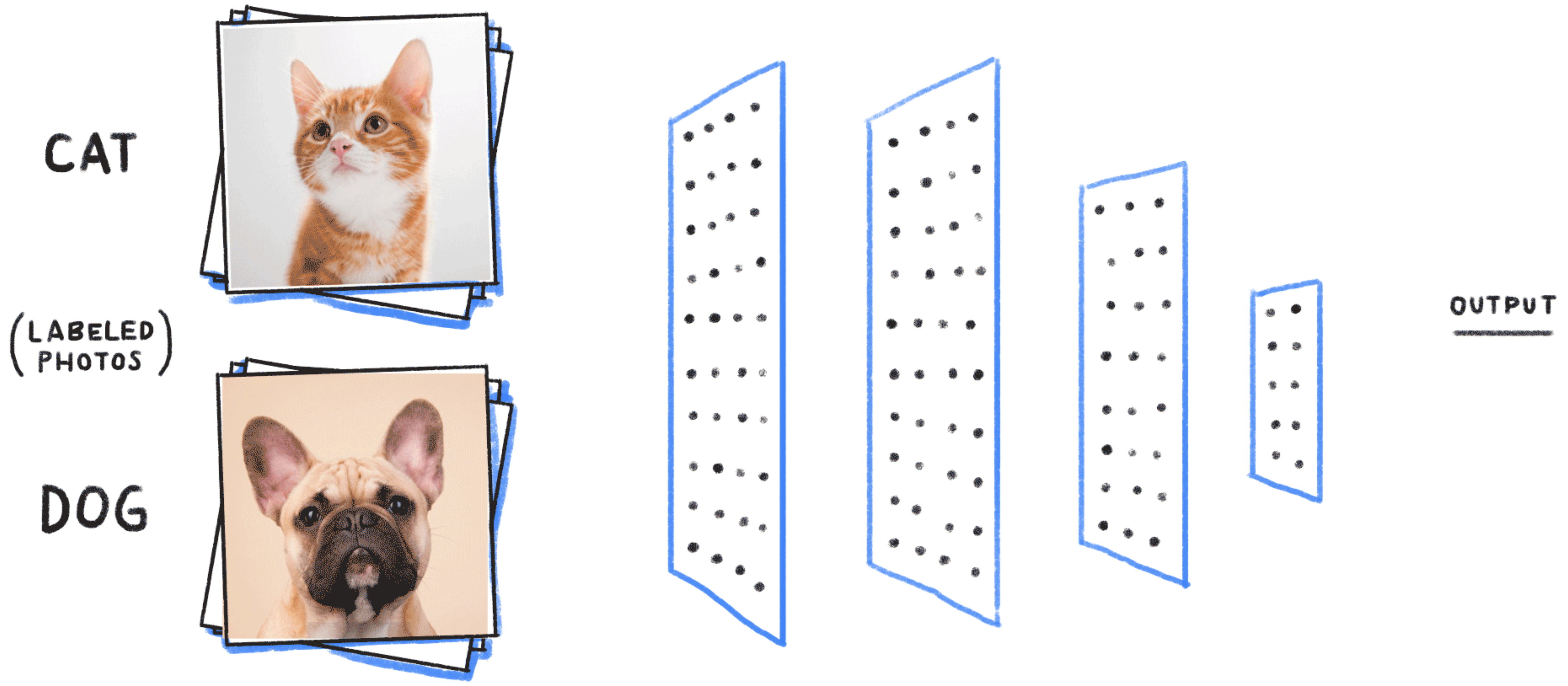
Image Classification

Style Transfer

Image Restoration

In-painting

Segmentation Detection

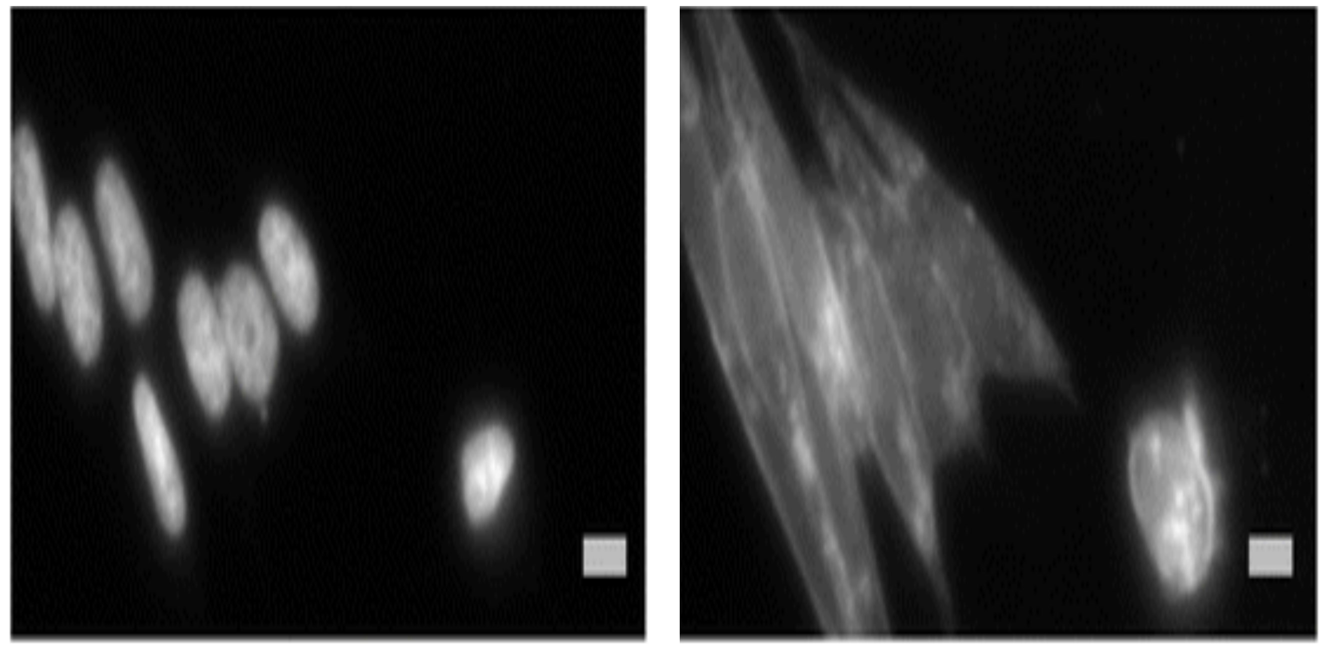


Source: https://becominghuman.ai/ Venkatesh Tata

Classic application of Computer Vision

Benign Malignant

Esteva et al., Nature, 2017



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

Deep Learning Tasks

Image Classification

Style Transfer

Image Restoration

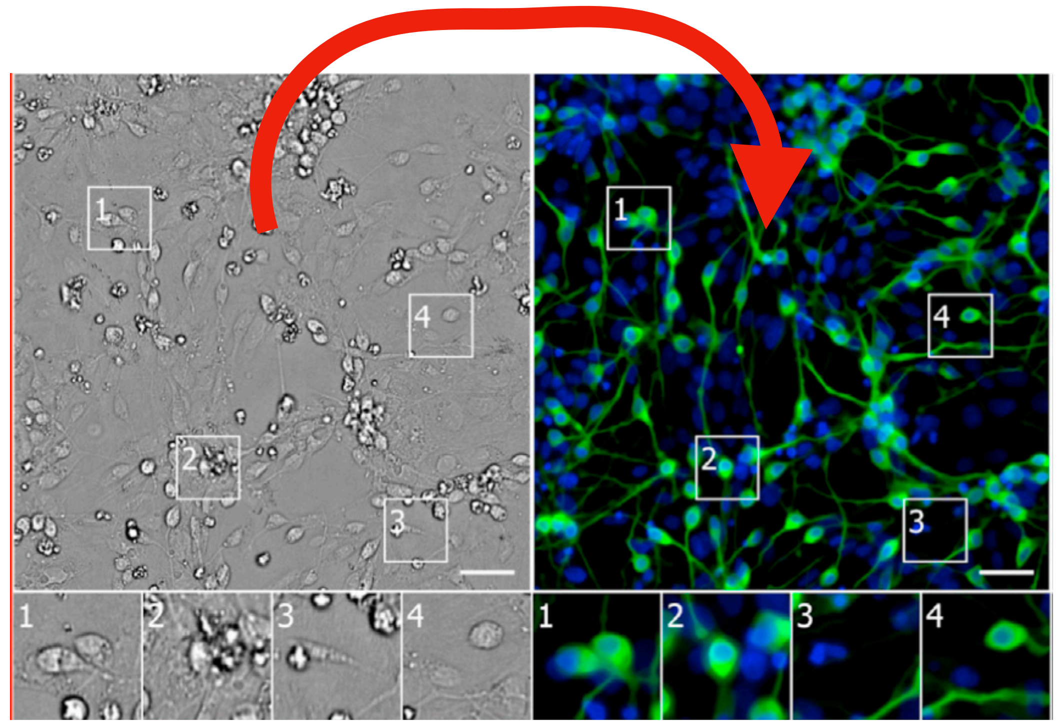
In-painting

Segmentation Detection

From Brightfield to Fluorescence

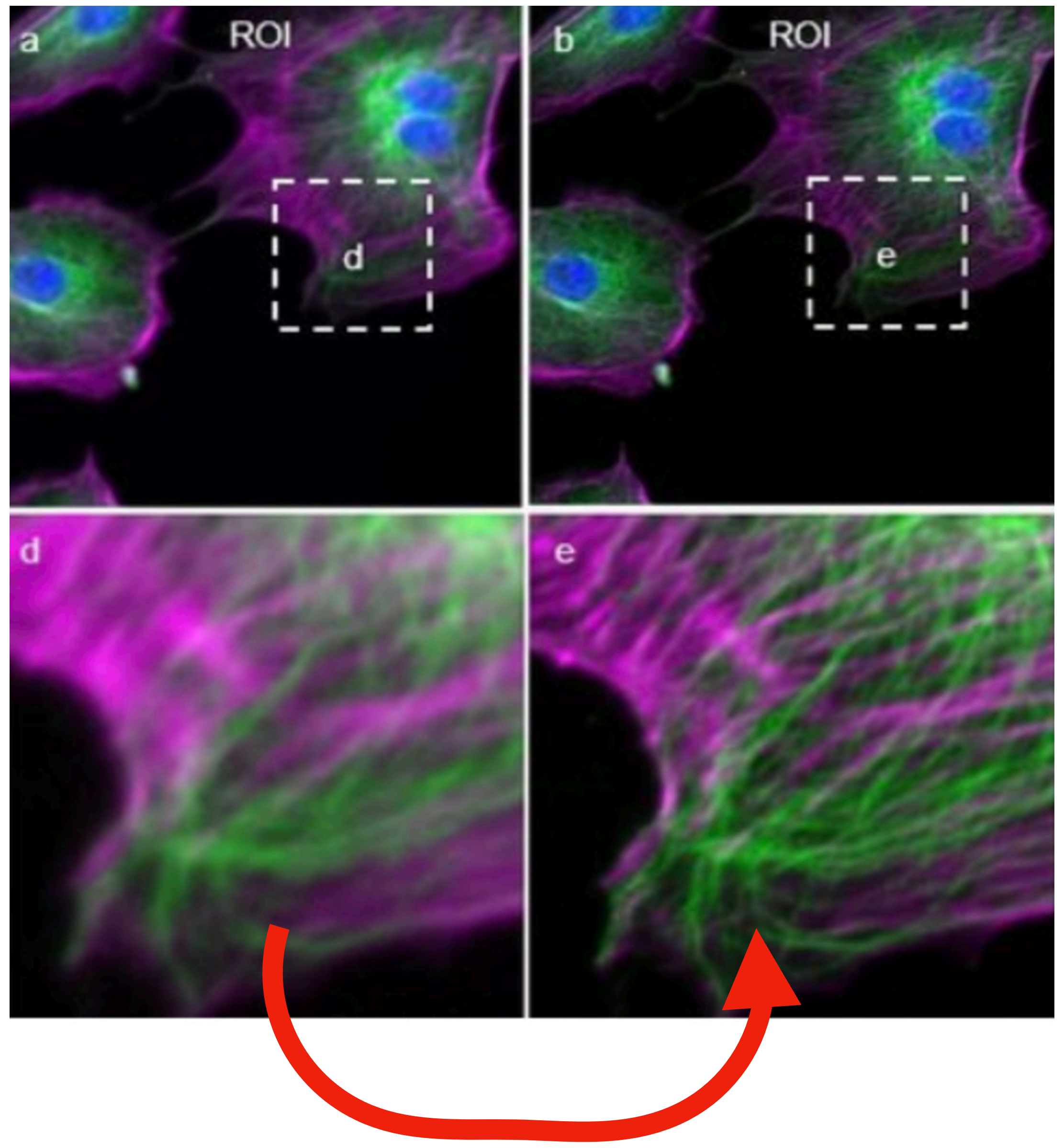
Virtual Acquisition

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



Super-resolution

H. Wang, Nature Methods, 2019,



From LR to SR

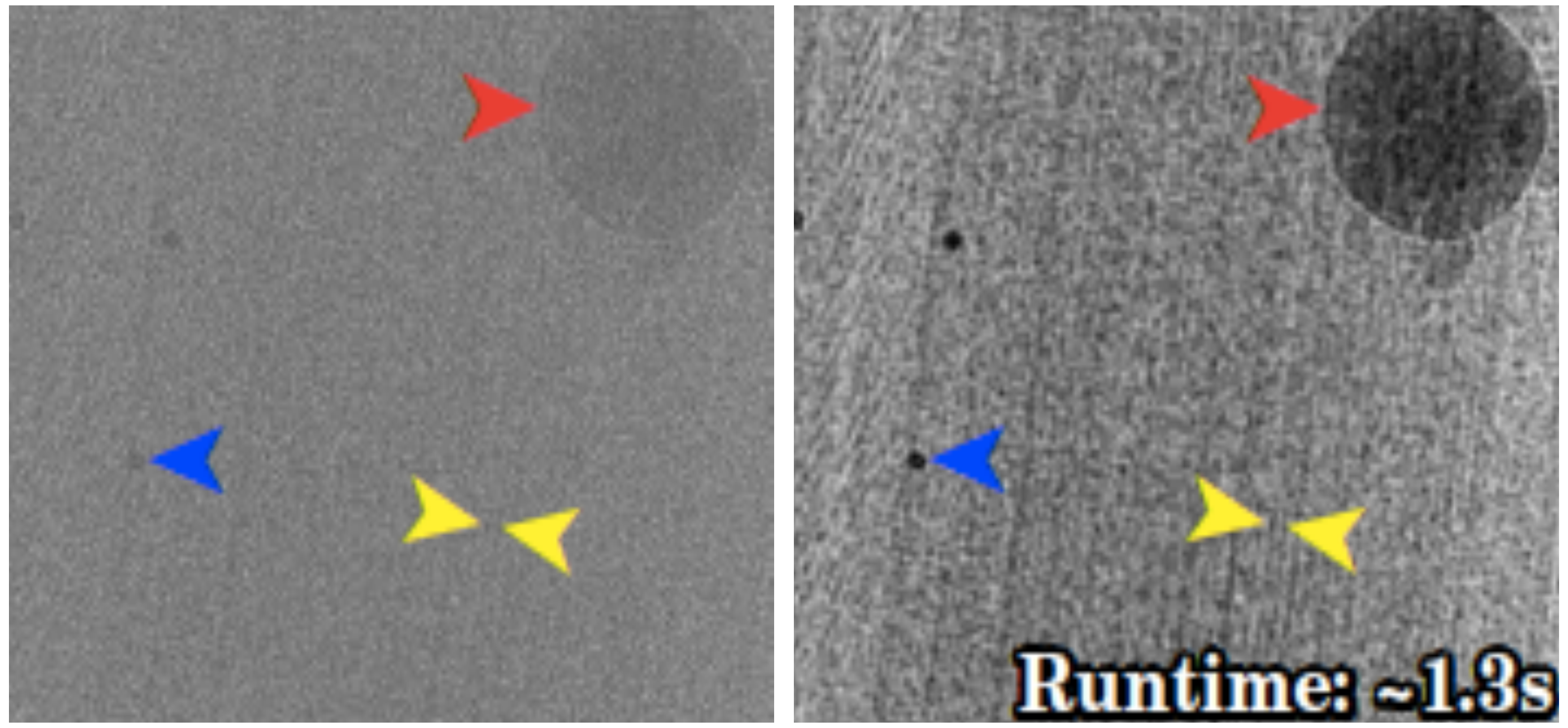
Deep Learning Tasks

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

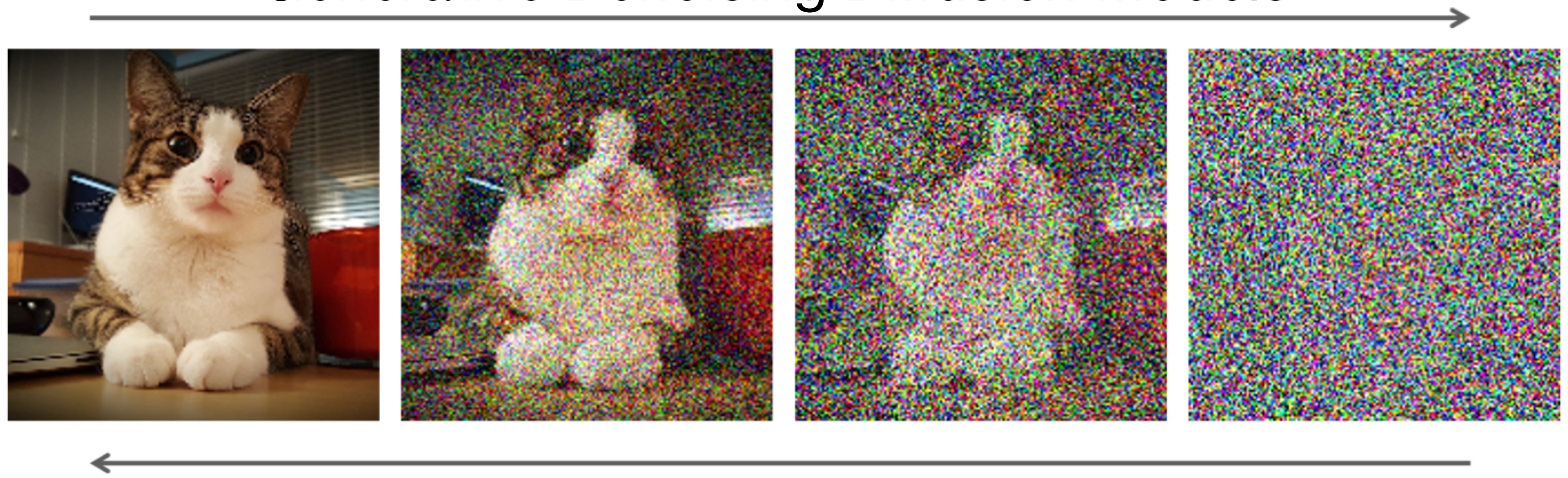
Denoising from NVIDIA

Noise2Void
Self-supervised training

Krull et al., Noise2Void -CVPR, 2019.



Generative Denoising Diffusion Models



Deep Learning Tasks

Image Classification

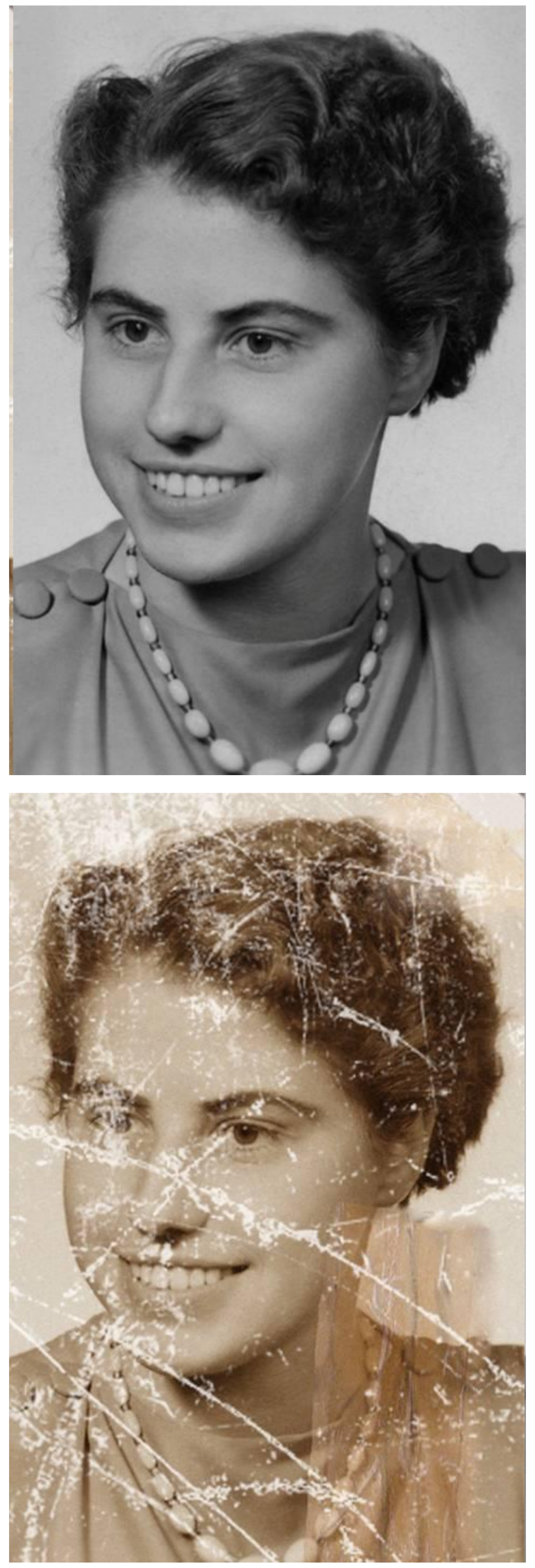
Style Transfer

Image Restoration

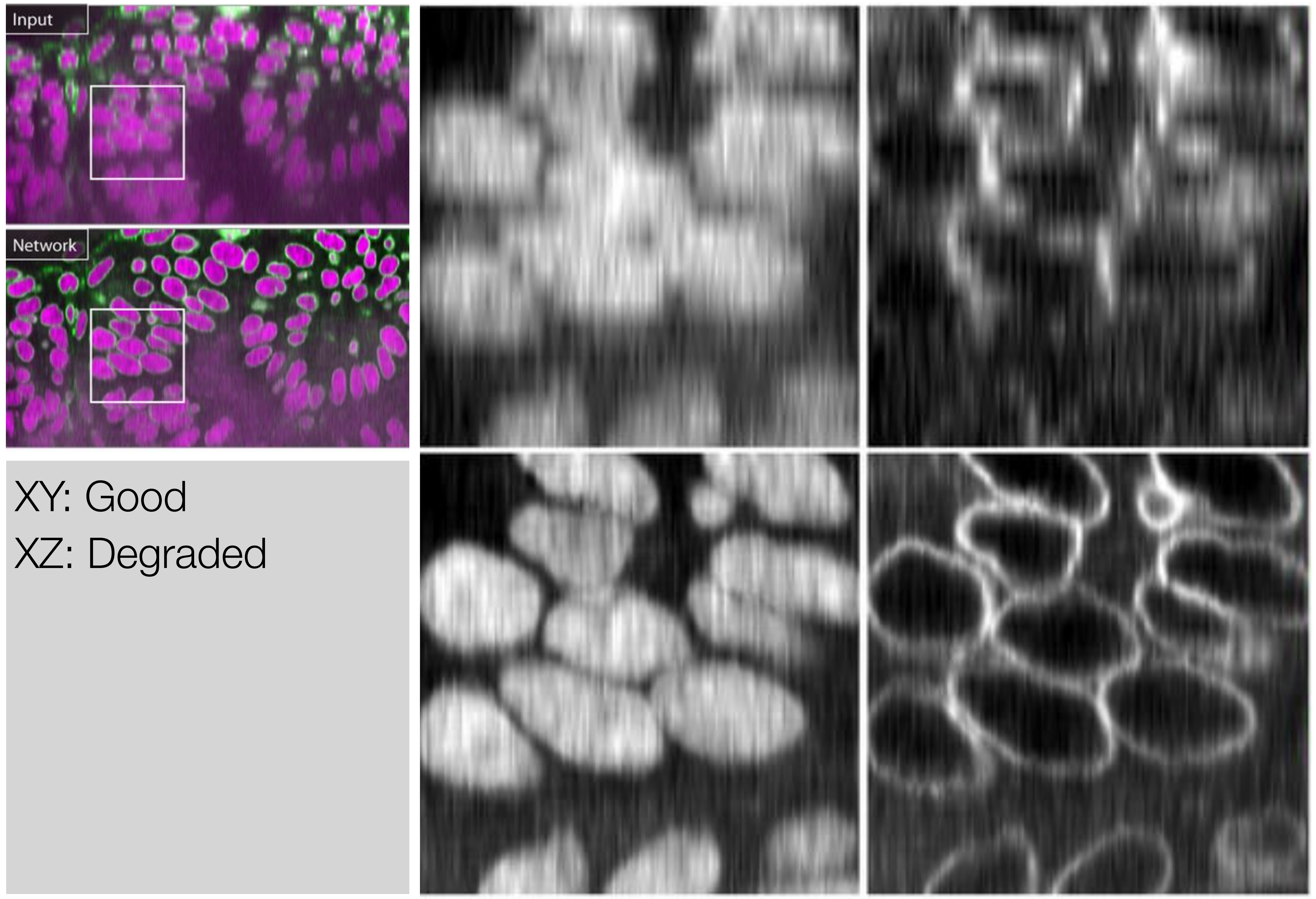
In-painting

Segmentation Detection

Fill in missing parts



Axial restoration



CARE, M. Weigert, 2019

Deep Learning Tasks

Image Classification

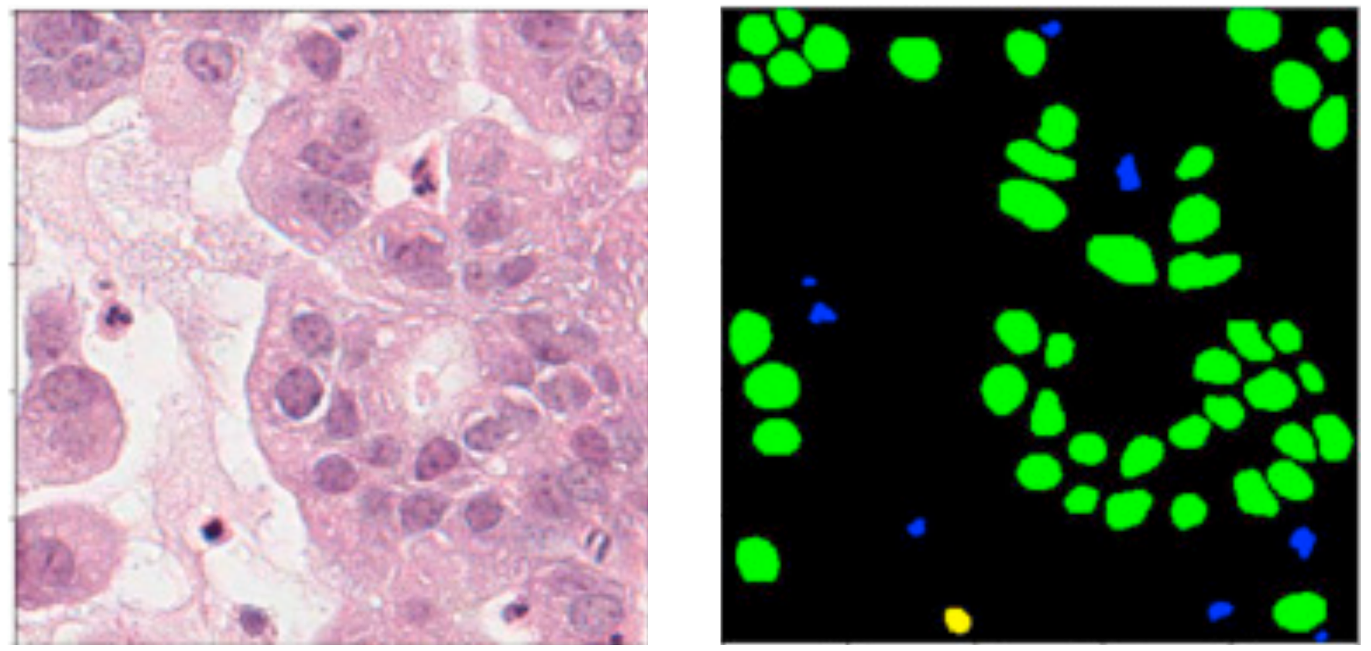
Style Transfer

Image Restoration

In-painting

Segmentation Detection

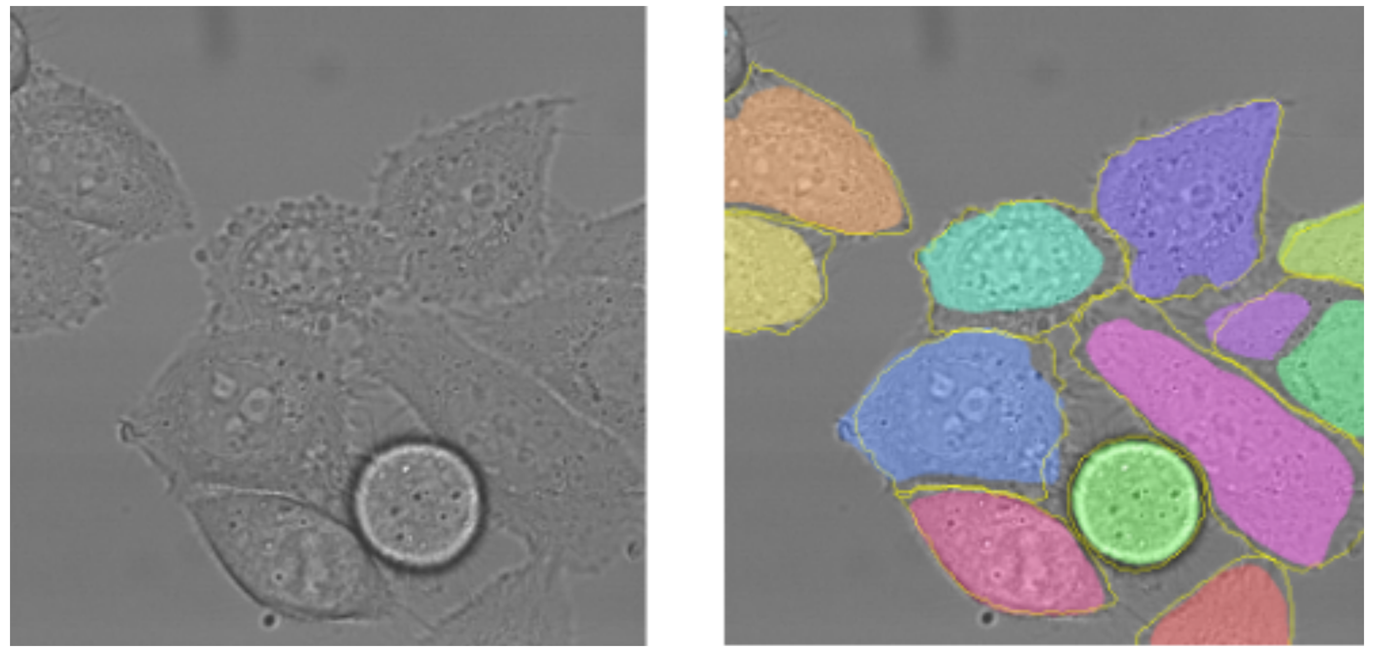
Semantic Segmentation



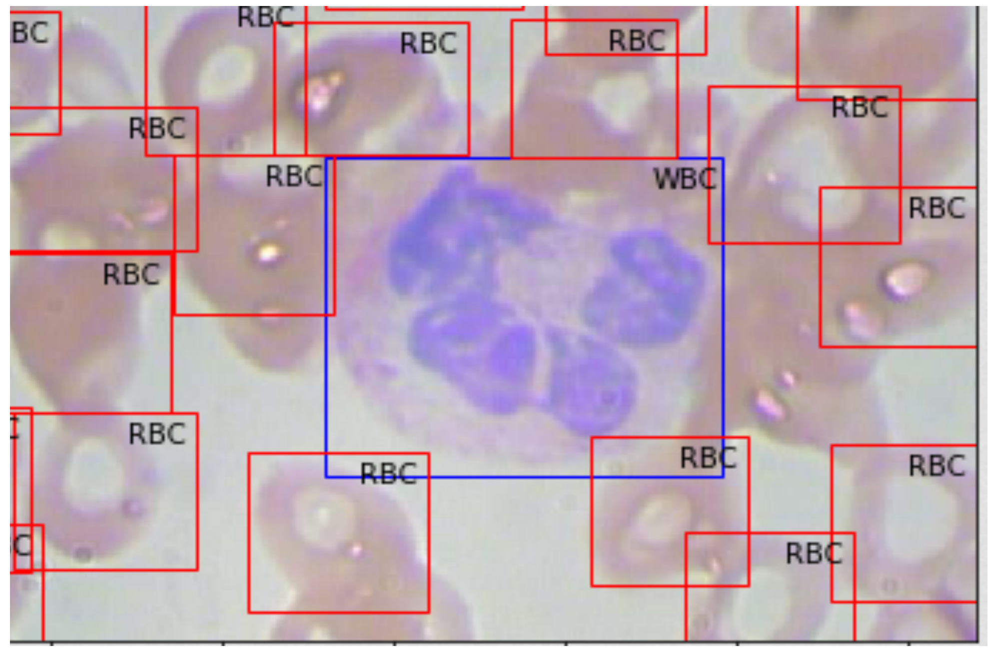
Pixel classification

- Binary
- Multiple classes

Instance Segmentation

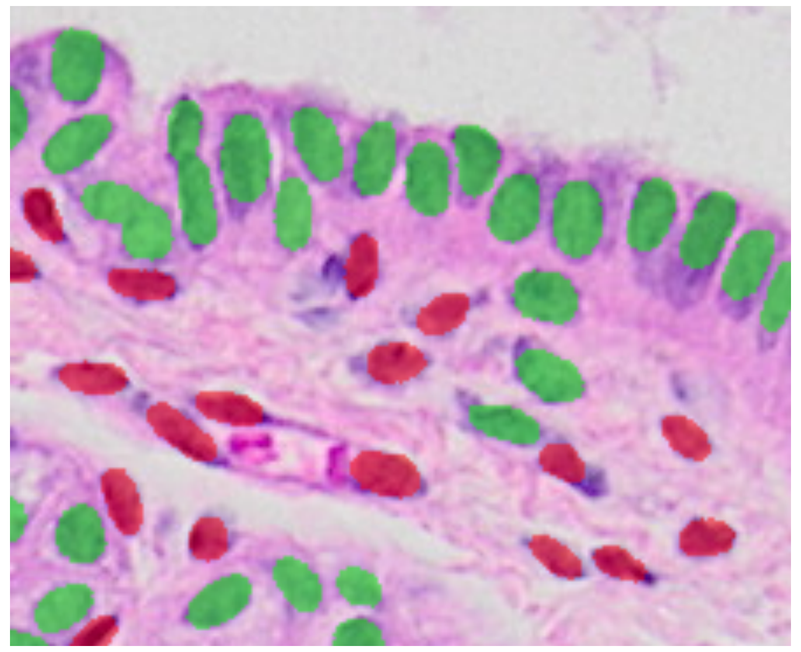


Bounding box



Blood cell detection Bounding Box
YOLOv3

Stardist + spatial class
Mouse/human cells



Juppet et al., 2021.

When Shall I Use Deep Learning?

No explicit model of the objects of interest

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

PROMISES OF DEEP LEARNING

☞ **End-to-end learning**

☞ **Continual improvement**

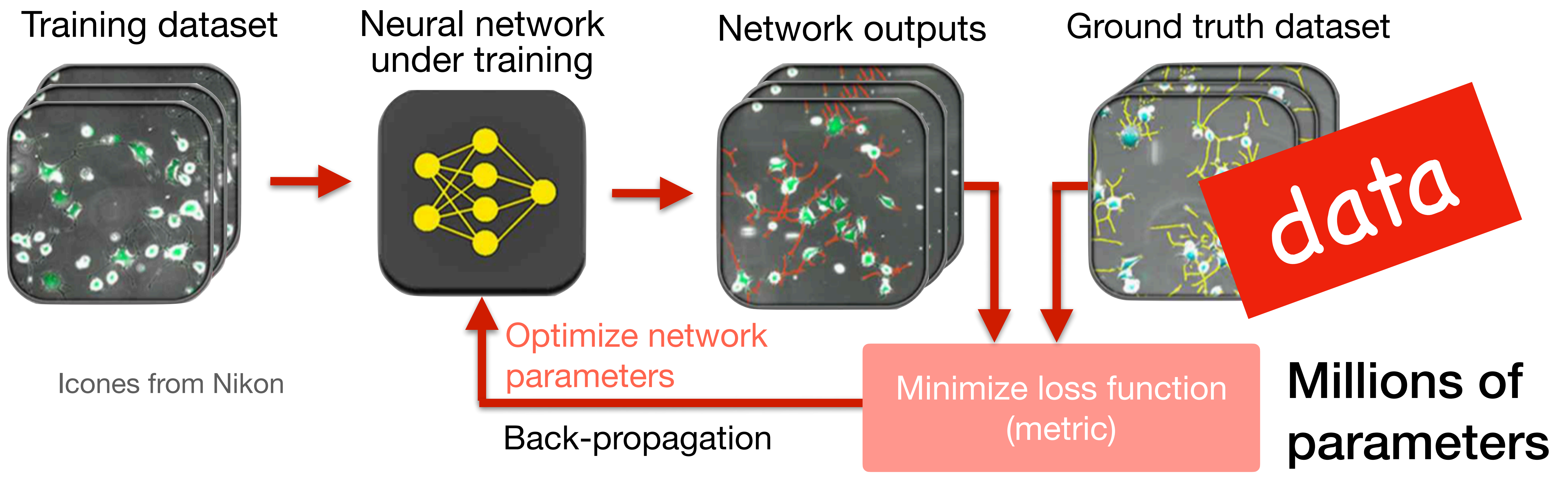
☞ **Ability to generalization**

How much time and resources?

What are the required skills?

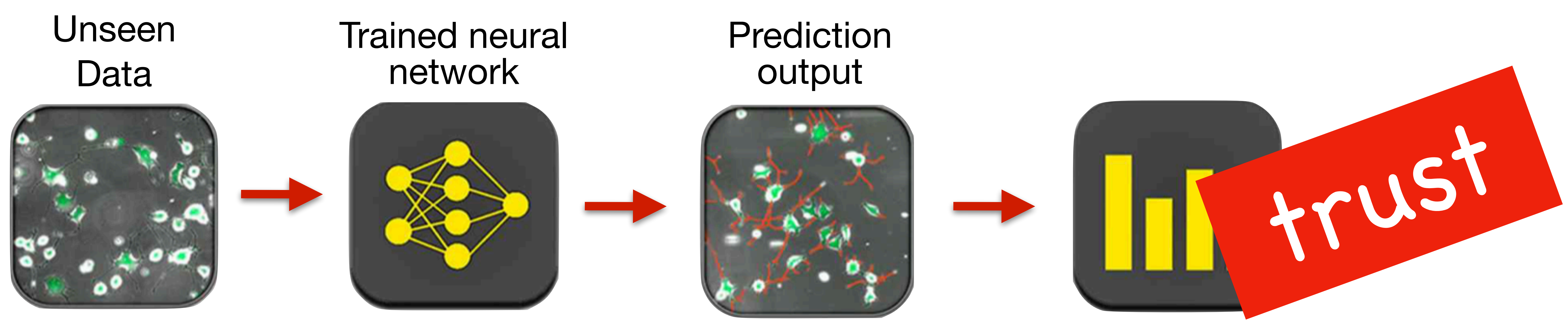
How evaluate the accuracy?

TRAINING — run 1 x time — iterative — slow



- Data science
- IT skills
- Programming
- Domain knowledge

INFERENCE — run n x time — one-shot — fast



- Domain knowledge
- Final users
- Validation
- Trust

TRAINING | **Model Producer** | **GPU**

CPU

CLOUD
Computing service

GPU
NVIDIA ↗Gb RAM

HPC
Computing server

binder aws Microsoft Azure colab

TensorFlow
PyTorch
ONNX

ZeroCostDL4Mic

#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

INFERENCE | **Model Consumer** | **CPU**

CPU

Python **CellProfiler**
QuPath **Matlab** **Fiji**
ImJoy **Napari**

deepImageJ

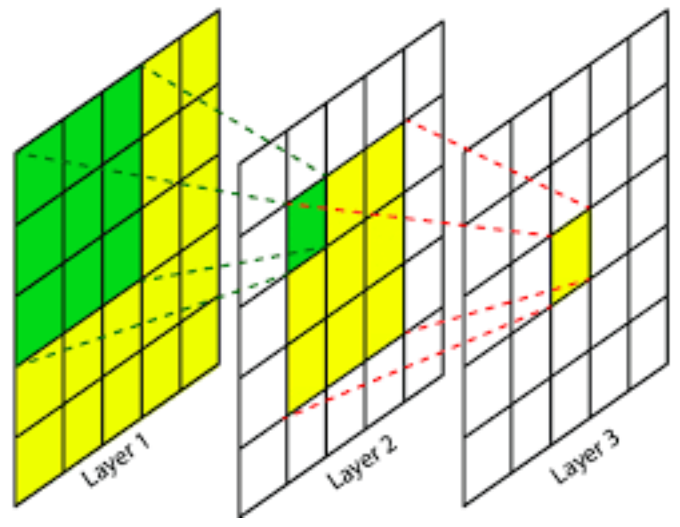
deepimagej.github.io

- Macro recordable (pipeline)
- One-click installation
- Pre & post-processing
- Models from the **Bioimage Zoo**

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

What is a Deep-Learning Model?

MODEL = ARCHITECTURE + LEARNT PARAMETERS

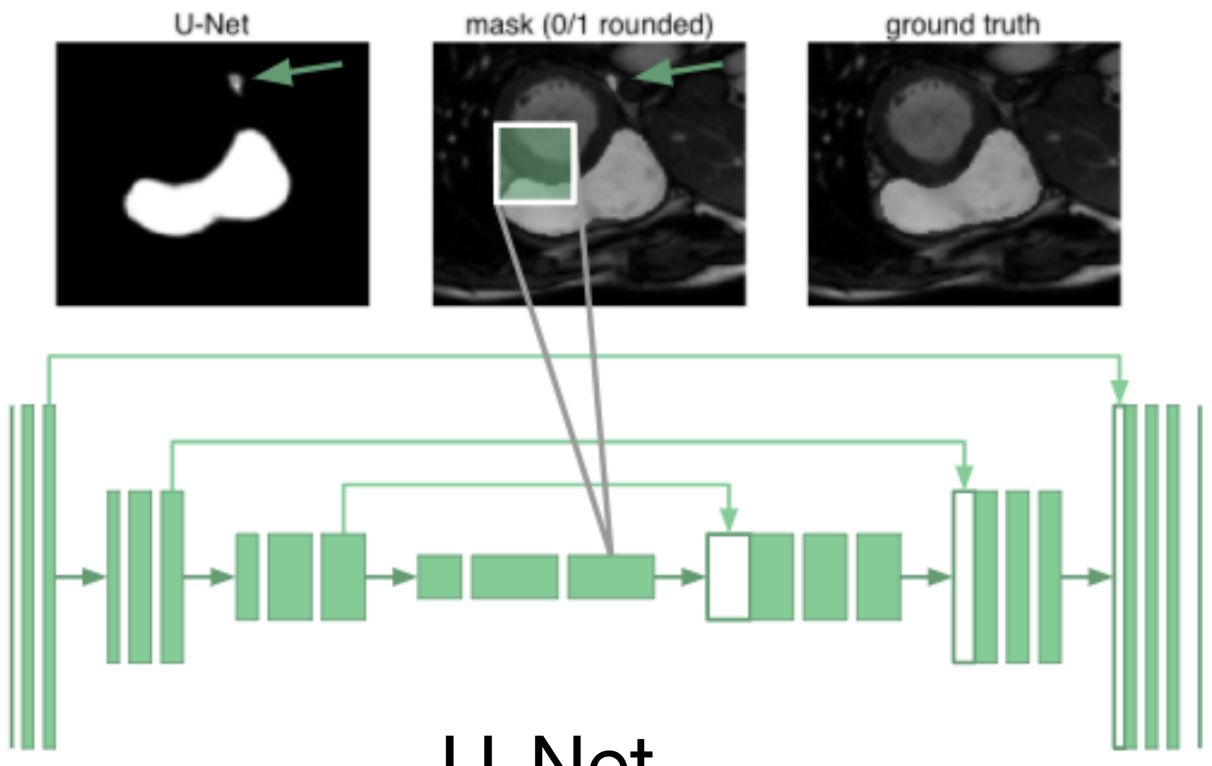


Neural Network Architecture

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

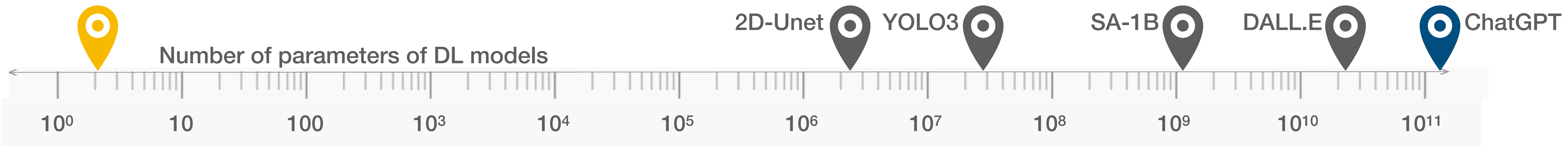
Parameters

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



U-Net

+ METADATA **interoperability, open**



How to get a Model?

Repo. of trained model (zoo)

Out-of-box trained models

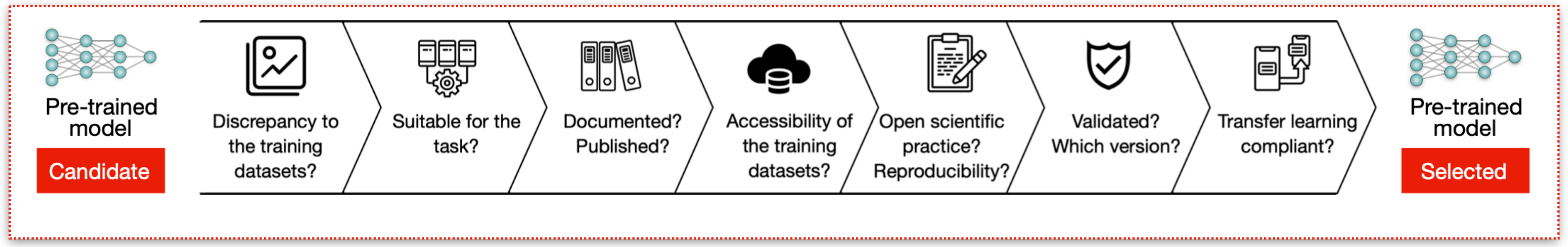
1 USE A TRAINED MODEL

+ FINE TUNING

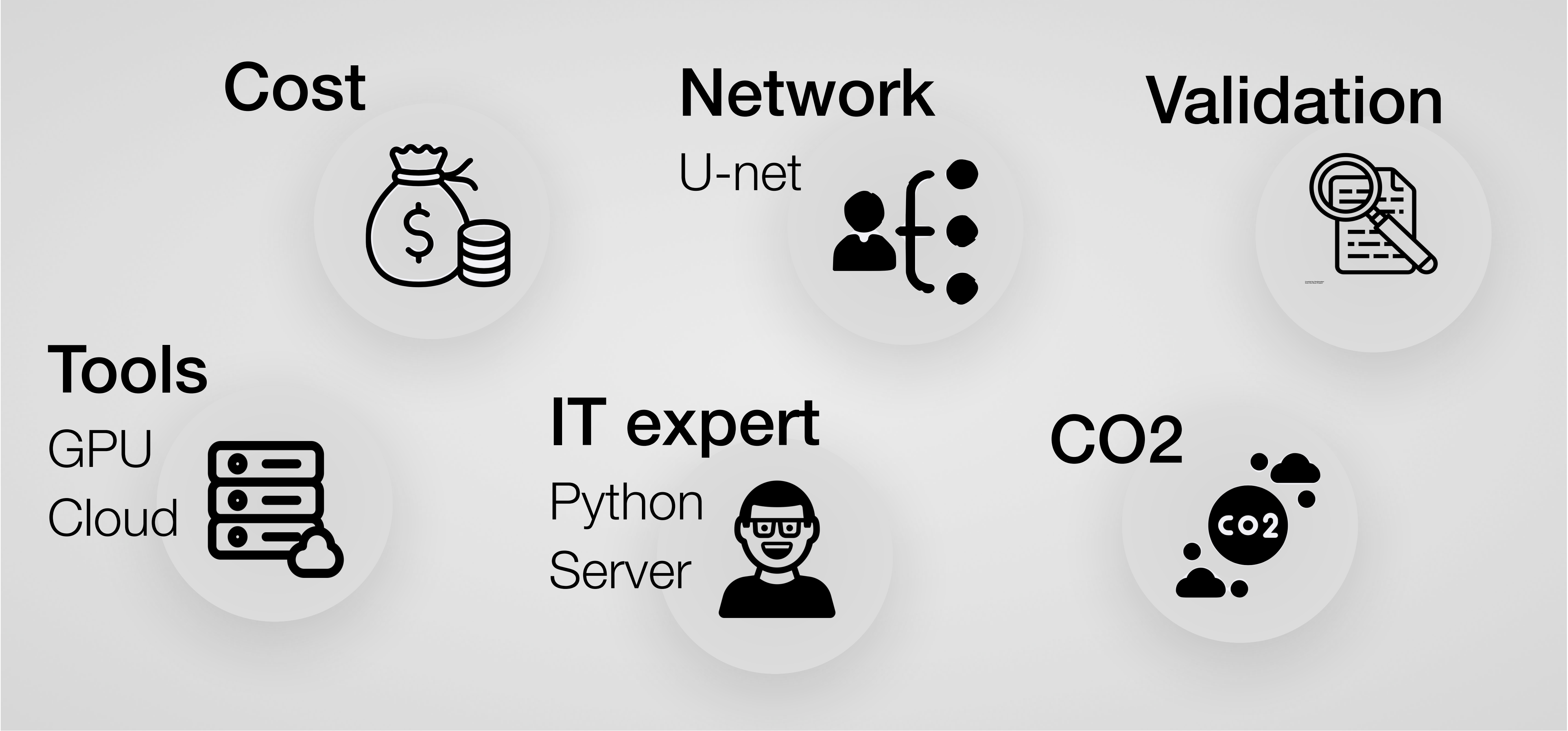
2 TRAIN YOUR OWN MODEL

Find a model

Check a model



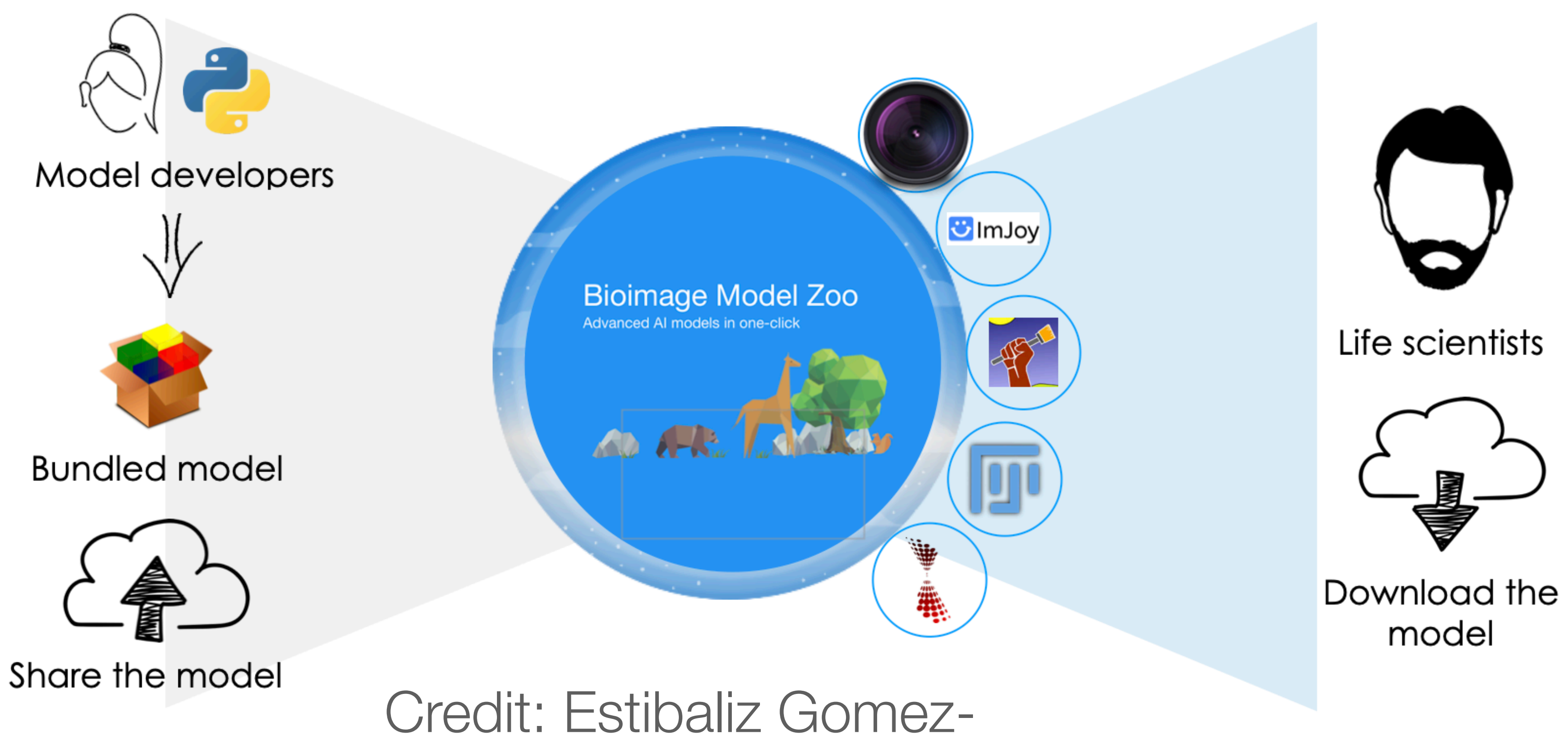
V. Uhlmann, L. Donati, D. Sage, IEEE SPM 2022



Iconic Pre-trained Models

Bioimage Model Zoo

https://bioimage.io

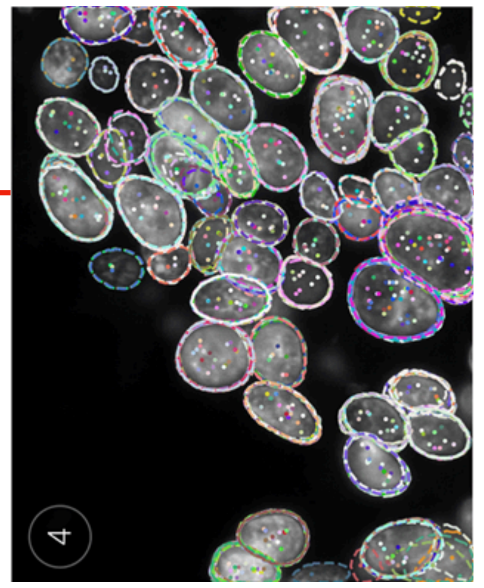


Credit: Estibaliz Gomez-de-Mariscal

✓ **Interoperability**
✓ **Open - FAIR**
✓ **Control**
✓ **Standardization**

W. Ouyang et al.
biorxiv, 2022

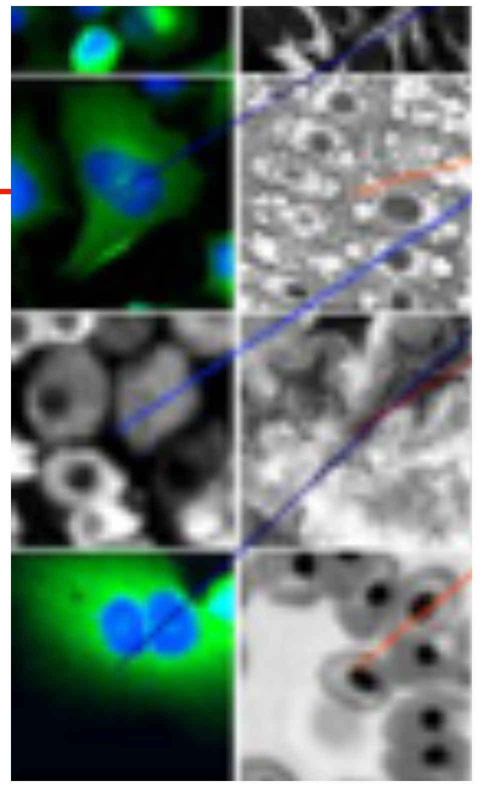
Detection of nucleus



STARDIST

star-convex shape (2D/3D)

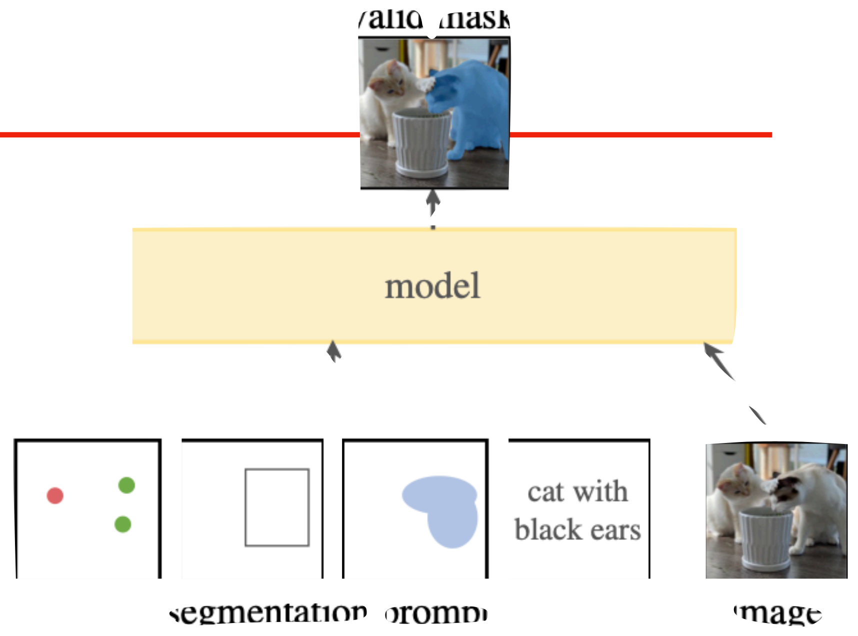
Detection of cells and more



CELLPOSE

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

Detection of anything



SAM of Meta AI

Prompt
1+ billon masks
11 million images

Deep Learning for Bioimage Analysis

Concepts and Hands-on Neural Networks Training with a Critical Approach

WORKSHOP

Prepare datasets

Training on Colab

Analysis of results

Prediction on deepImageJ

SYSTEM		WORKSHOP
Define an analytical task	Task	2D semantic segmentation, pixel classification with 2 or K classes
Generate the training and testing datasets	Data	Several datasets from the Cell Tracking Challenge (glio, HeLa), from deepBacs, and some simulations
Select the neural network and hyper parameters of training	Train	U-net, 3 scales, Cross-entropy loss, Adam, Data augmentation, validation = 10%
Evaluate the model performances and testing	Test	Compute IoU after threshold probability maps
Diagnosis performance with critical sense	Trust	Adversarial attacks, discussion

Training Systems

Define the task: what is the input? what is the output?

Architecture

U-Net

RestNet

MobileNet

GAN

DenseNets

…

Loss function

Mean Squared Error

Cross-entropy

Dice Loss

Focal Loss

Mean Absolute Error

…

Optimisation

Gradient descent

Stochastic GD

ADAM

…

Hyper-parameters

Data curation

Good ground-truth

Masks / Labels

Keypoints / Boxes

Normalisation

Protocol of acq.

Each choice affectes the behavior

Each configuration strongly affects the performances

A. Mathis Neuro 2020

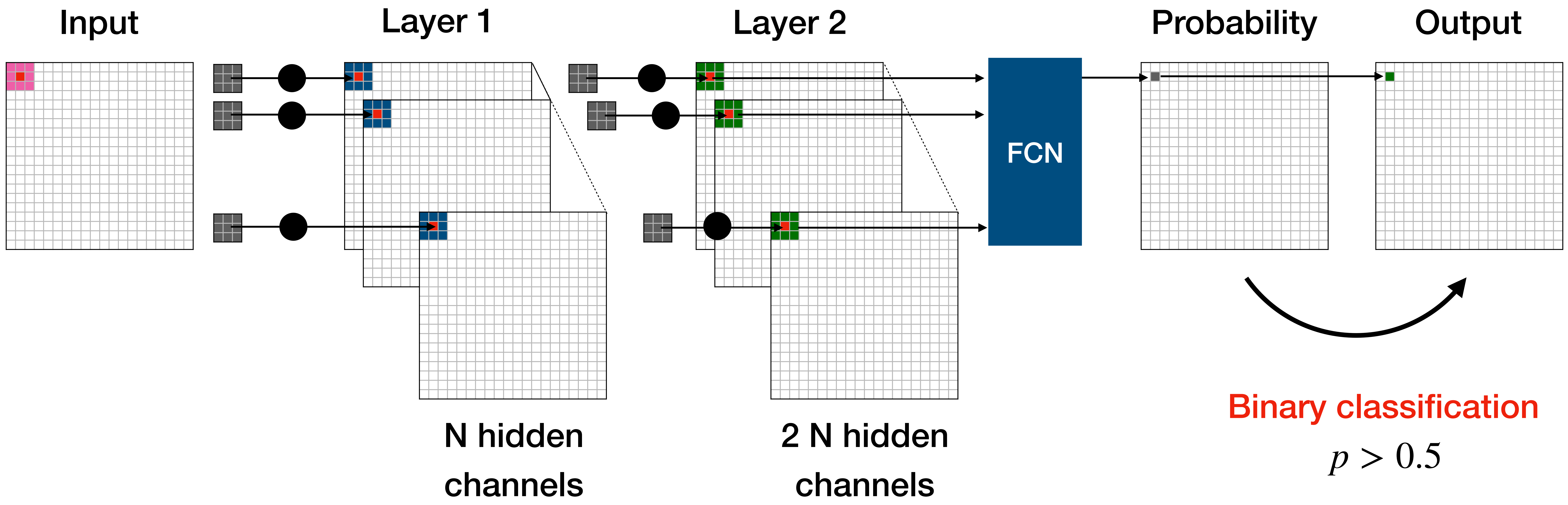
#ZeroCostDL4Mic

Multlable 2D Unet Segmentation

Colab Notebook

L. von Chamier, Nature Comm, 2021

Task: Semantic Segmentation by Pixel Classification



Binary classification

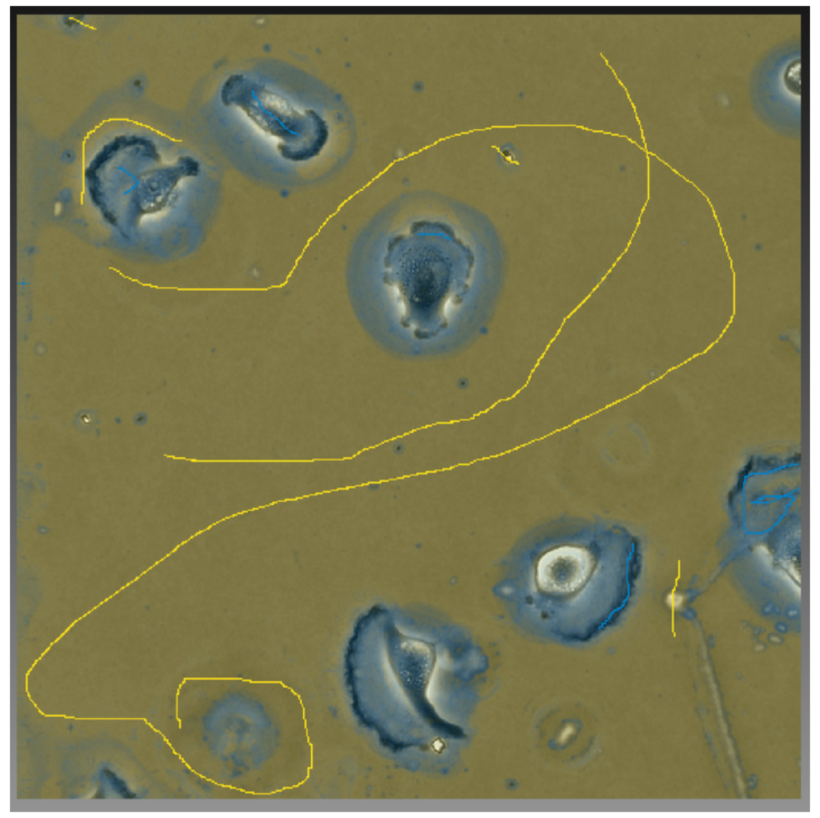
$p > 0.5$

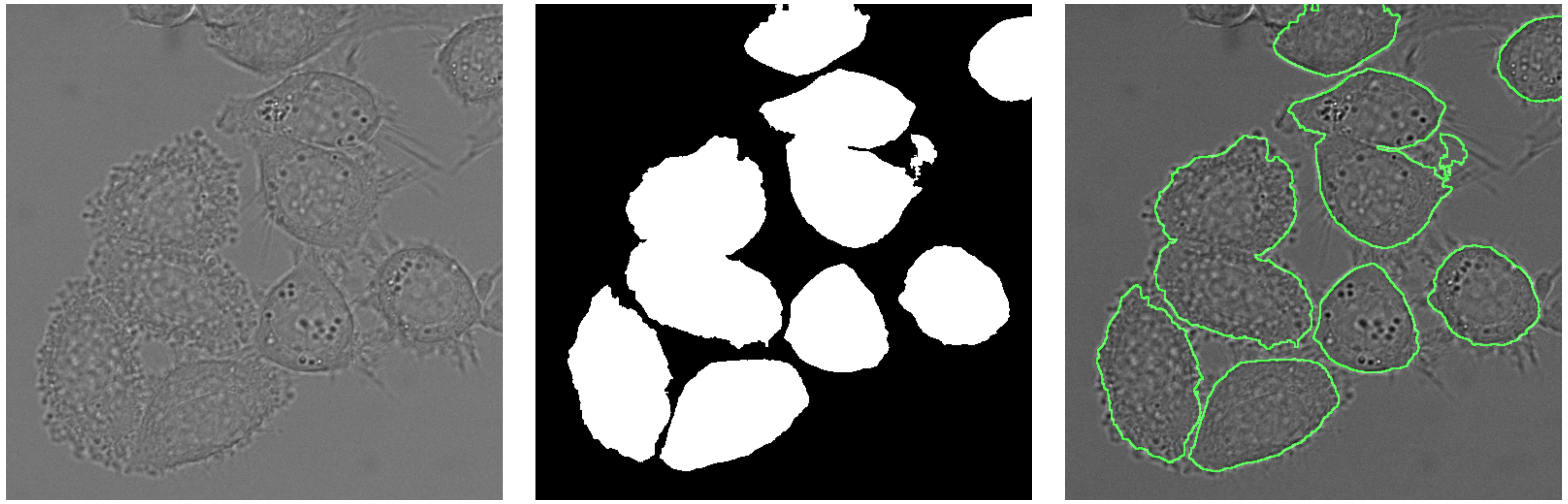
Multilabel classification

argmax(p_k)

Gliobastoma

- Cell tracking chall.
- 512 x 512
- 66 + 23 images
- Mask 2K (0/1)



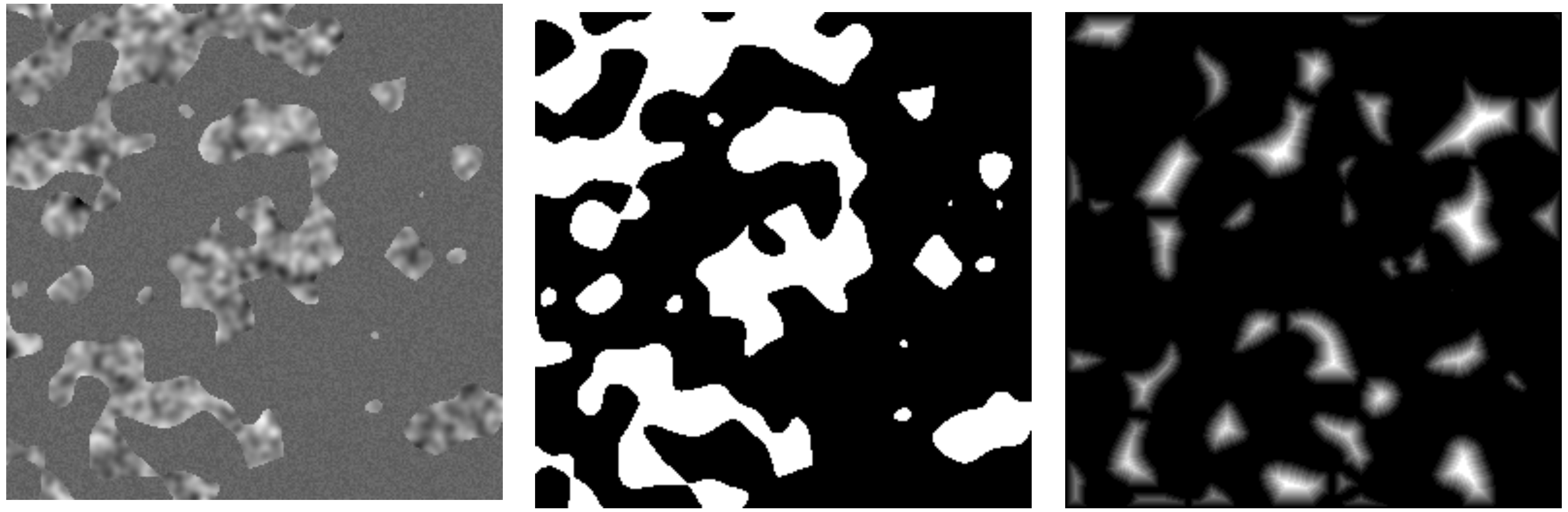


Hela

- Cell tracking chall.
- 72 + 27 images
- Mask 2K (0/1)

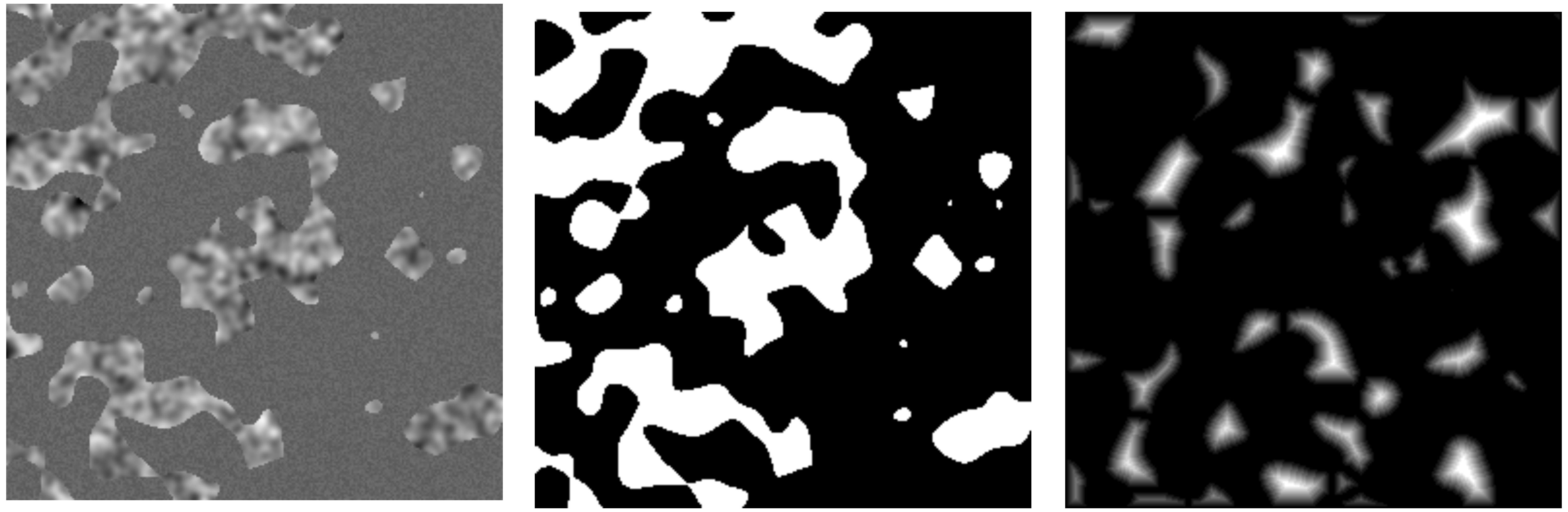
Deepbacs

- Zenodo
- 1024x1024
- 240 + 30 images
- Mask 2 K (0/1)



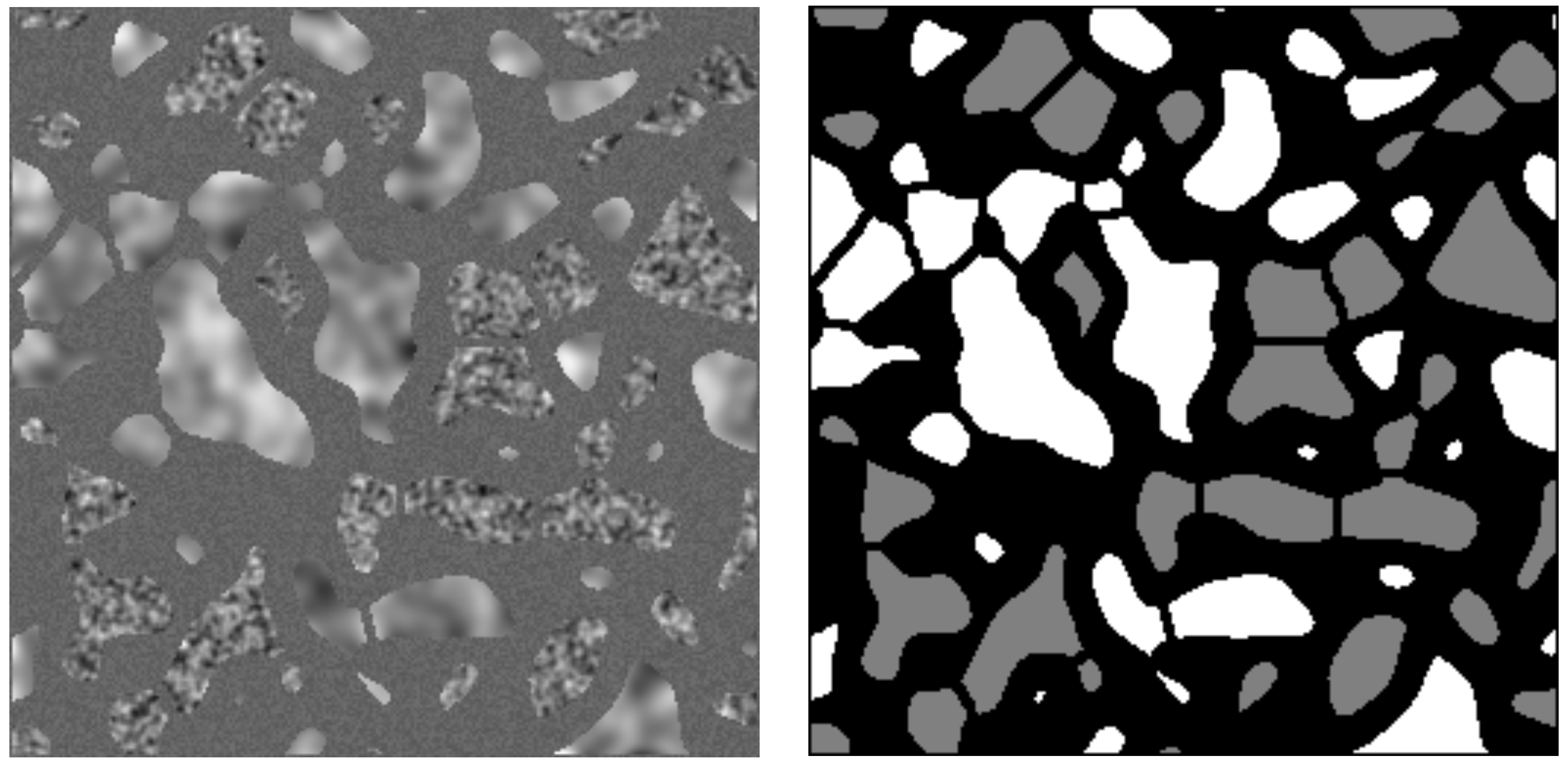
Simulation 2K 10 im

- 256 x 256
- 10 + 10 images
- Mask 2K (0/1)
- Dmaps 12K



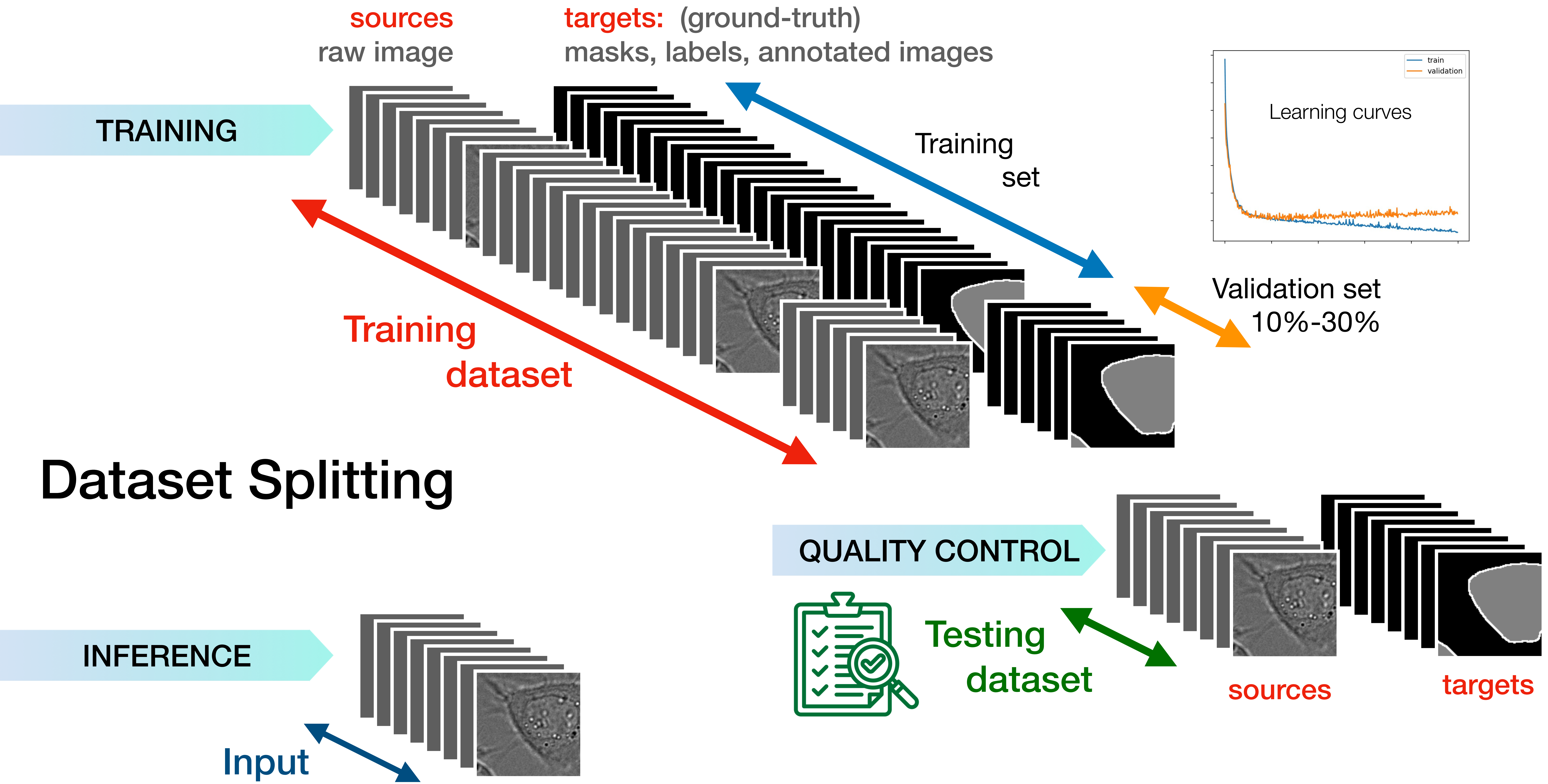
Simulation 2K 100im

- 256 x 256
- 100 + 40 images
- Mask 2K (0/1)
- Dmaps 12K



Simulation 3K 100im

- 256 x 256
- Mask 3K (0/1/2)



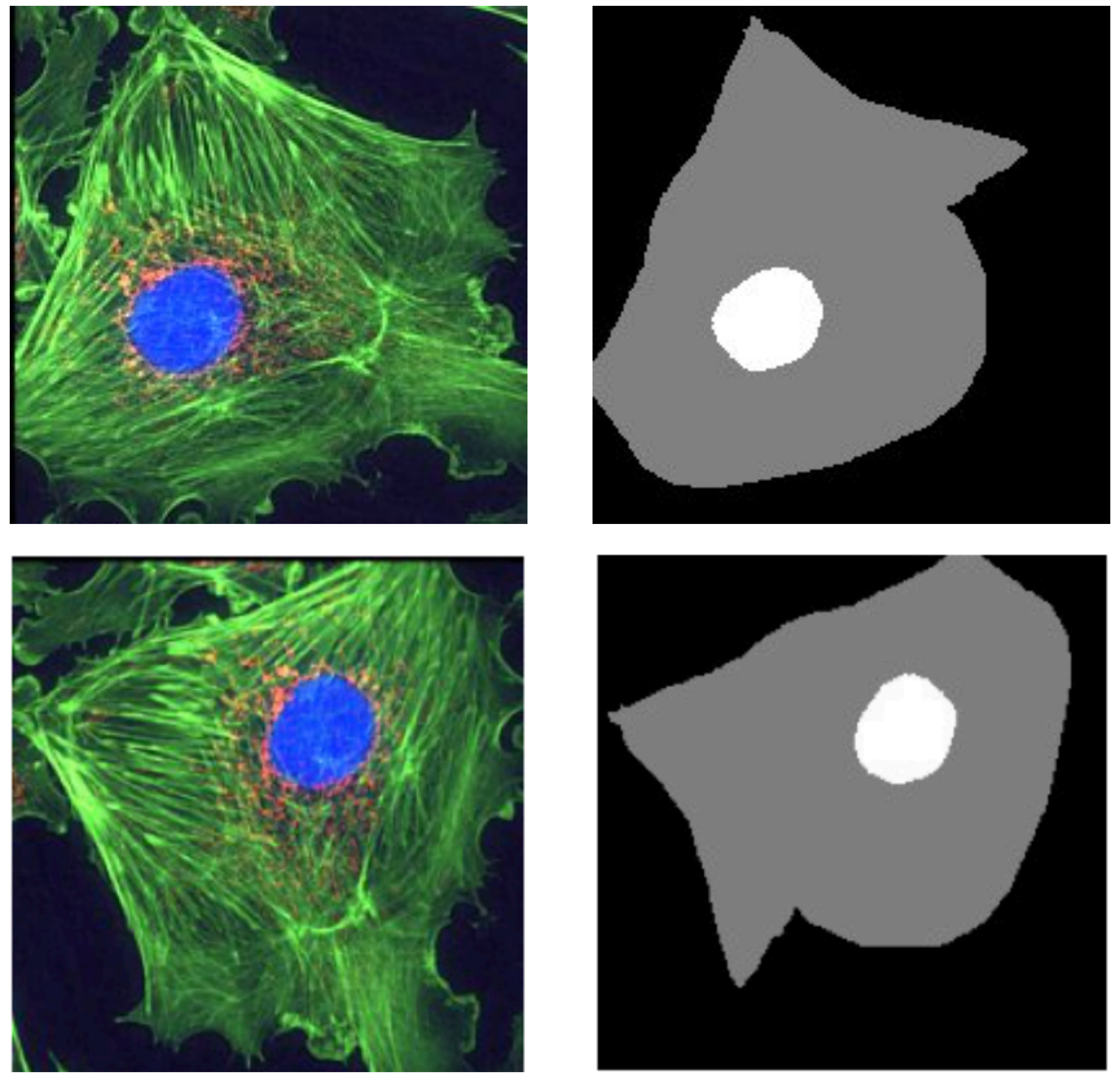
sources
raw image
targets: (ground-truth)
masks, labels, annotated images
TRAINING
Training
set
Learning curves
train
validation
Training
dataset
Validation set
10%-30%
Dataset Splitting
QUALITY CONTROL
Testing
dataset
sources
targets
INFERENCE
Input

Data Augmentation

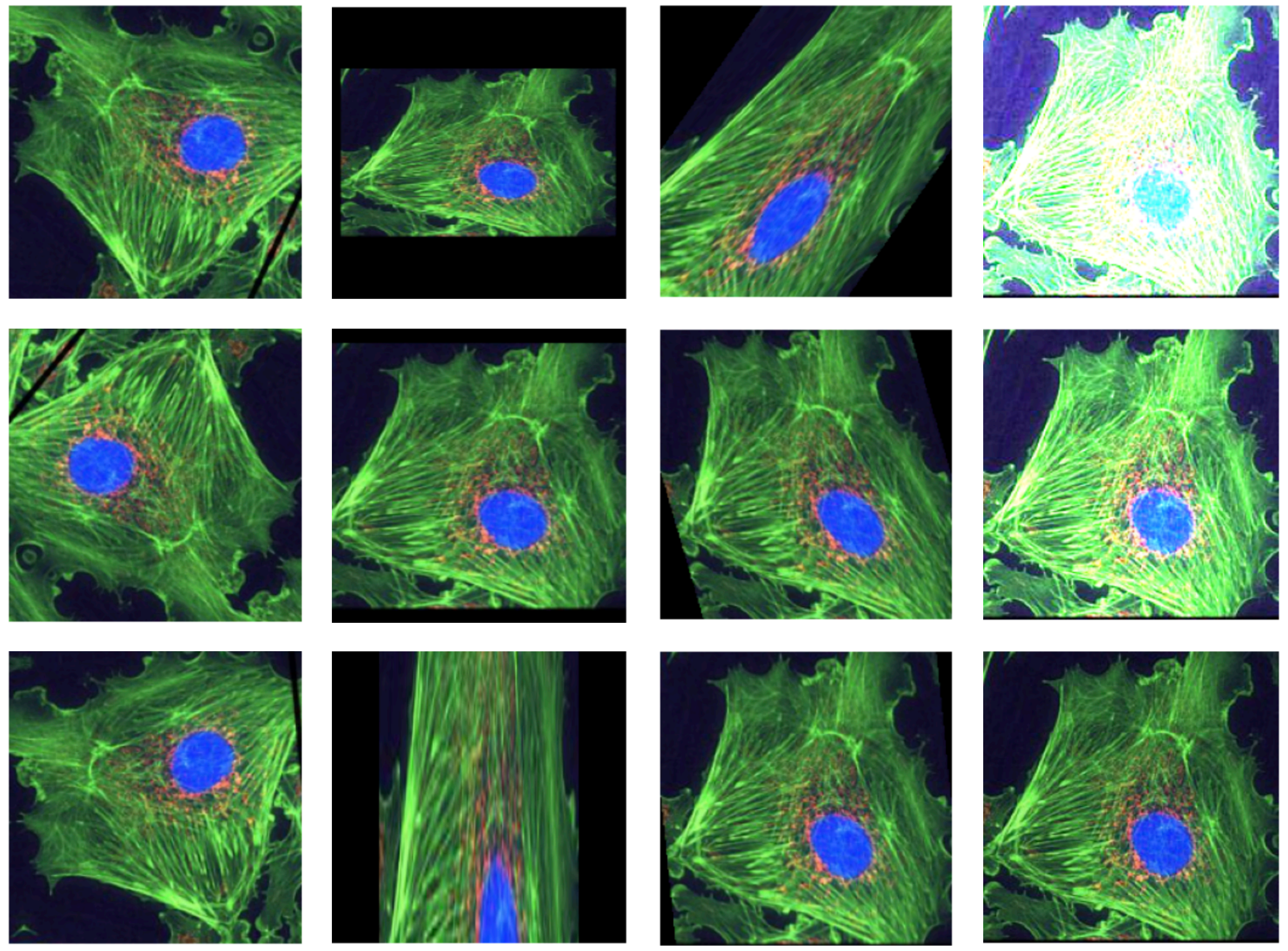
Missing data? → **Increase the diversity of your training set by applying random transformations**

- Geometric transformation
- Brightness / color / contrast
- Noise injection
- Random erasing
- Kernel filters: blur
- …
- Create your own one

➜ **Adversarial attacks**



Rigid transformation

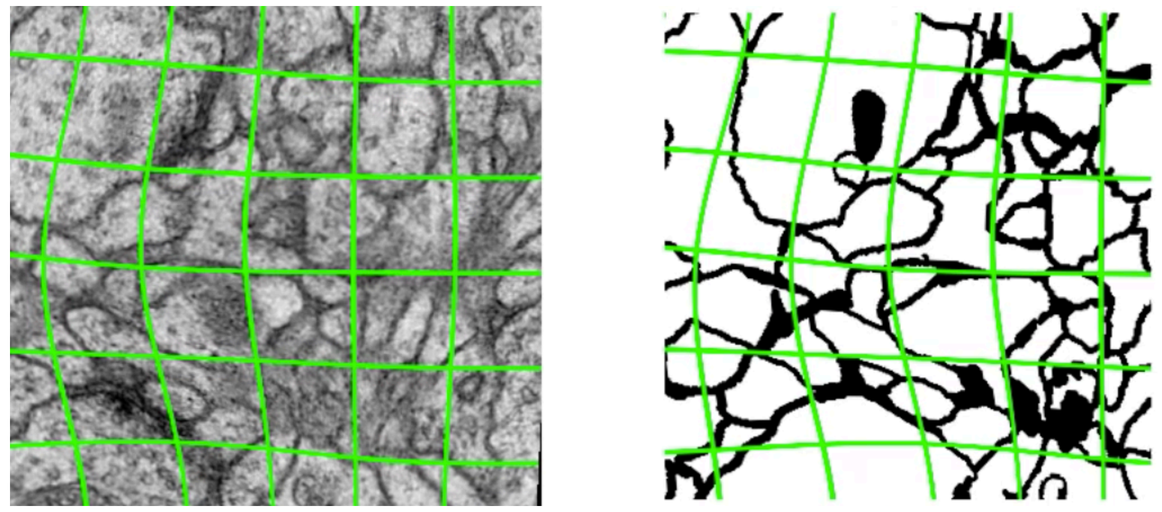


Rotation Zoom Shear Brightness

DA increases the diversity

DA squeezes better performance

DA revents overfitting



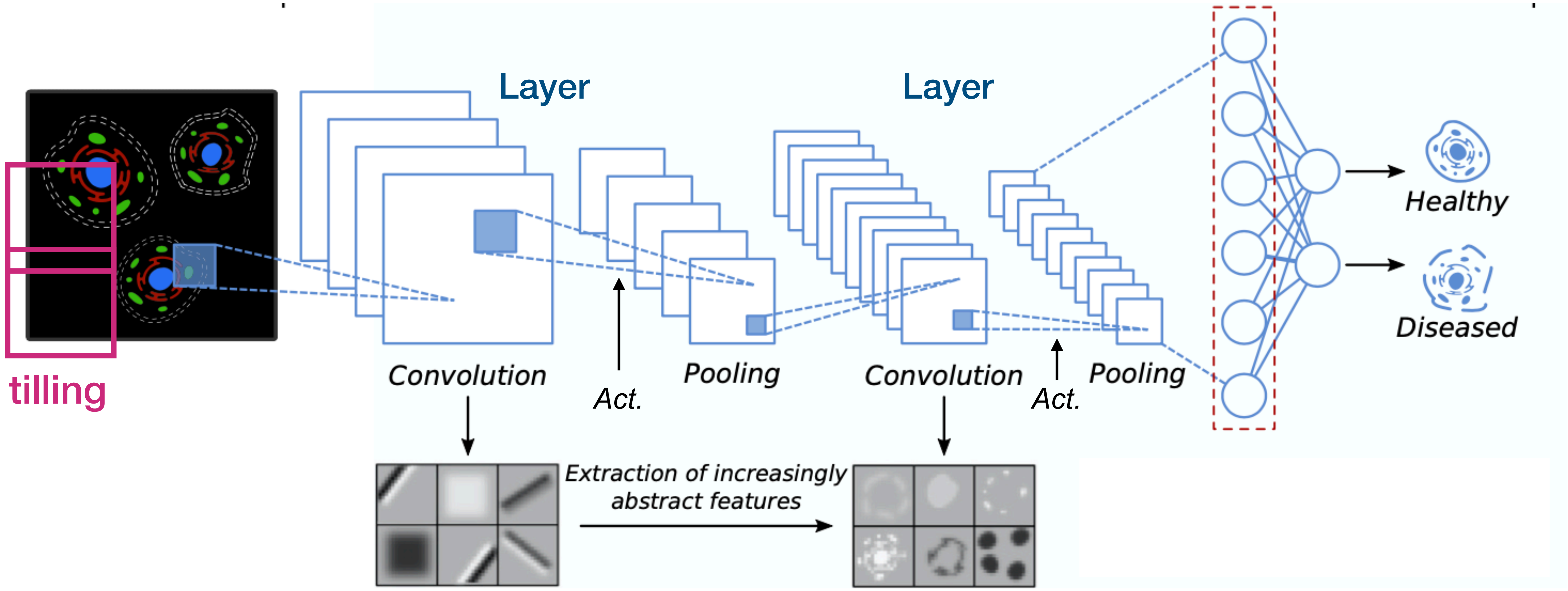
Elastic transformation

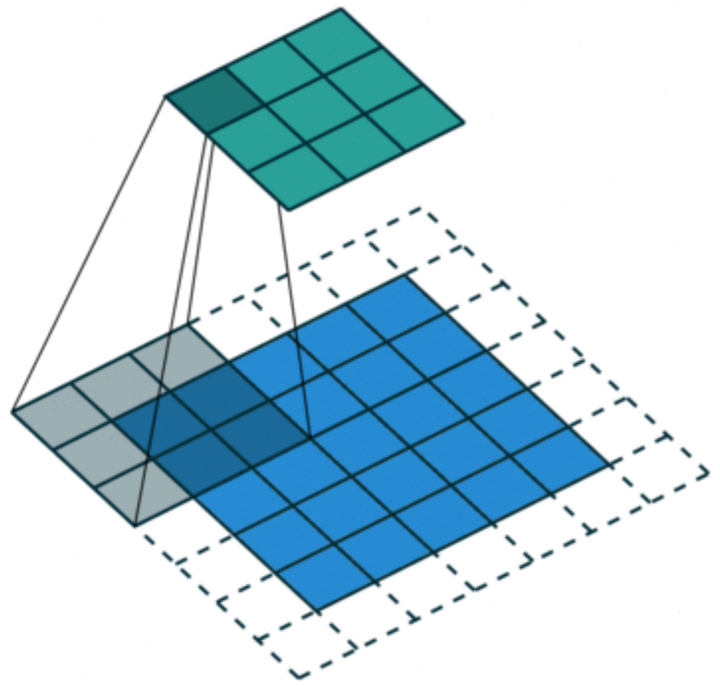
How many augmentations?

data diversity $\Longleftrightarrow$ data quality

Convolution Neural Networks (CNN)

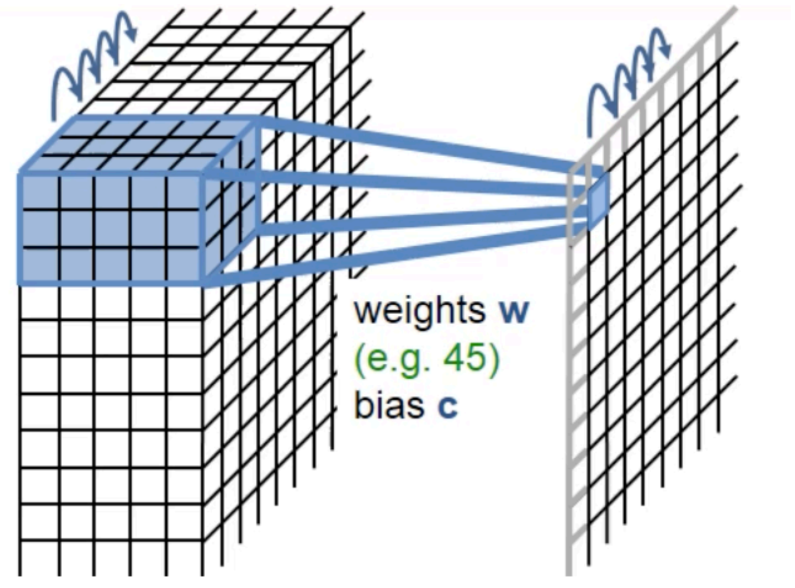
Architecture of the Network



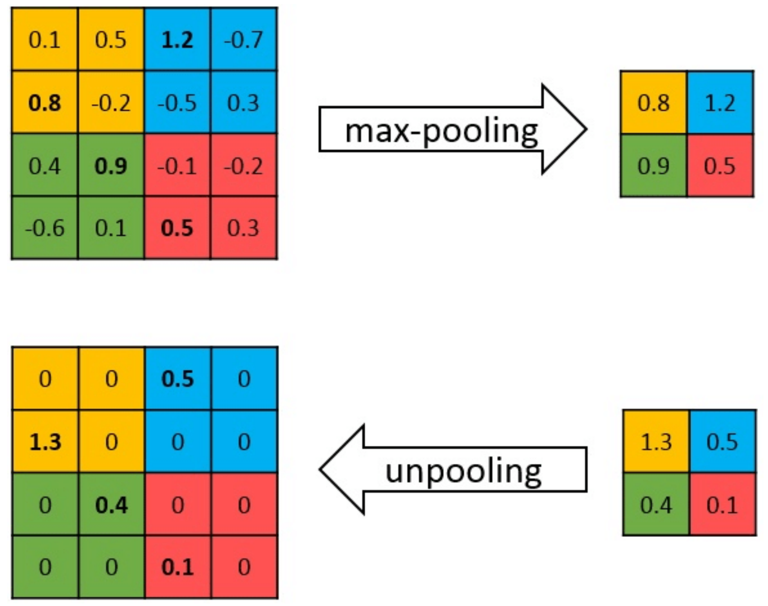


Convolution

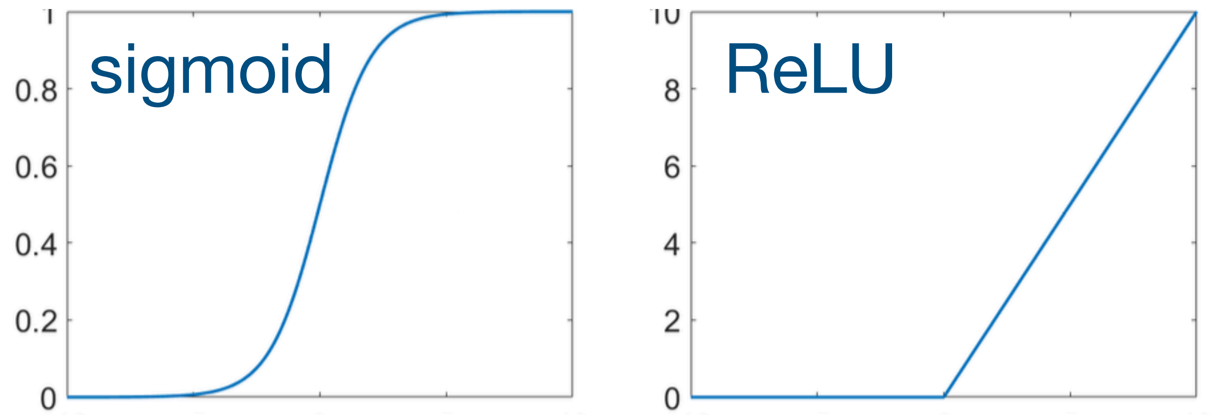
Millions of parameters to learn (weights & bias)



Channels [Deep]

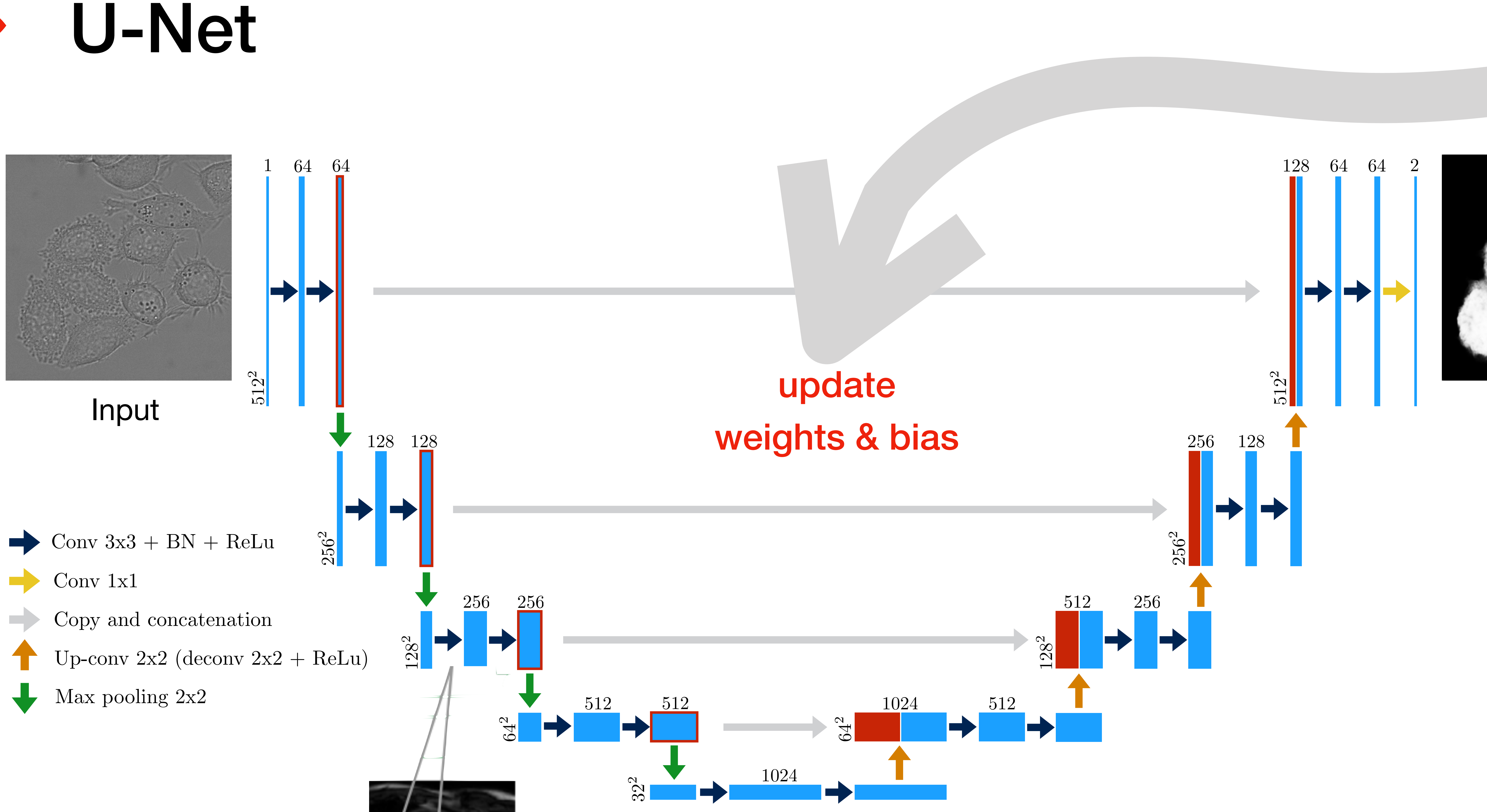


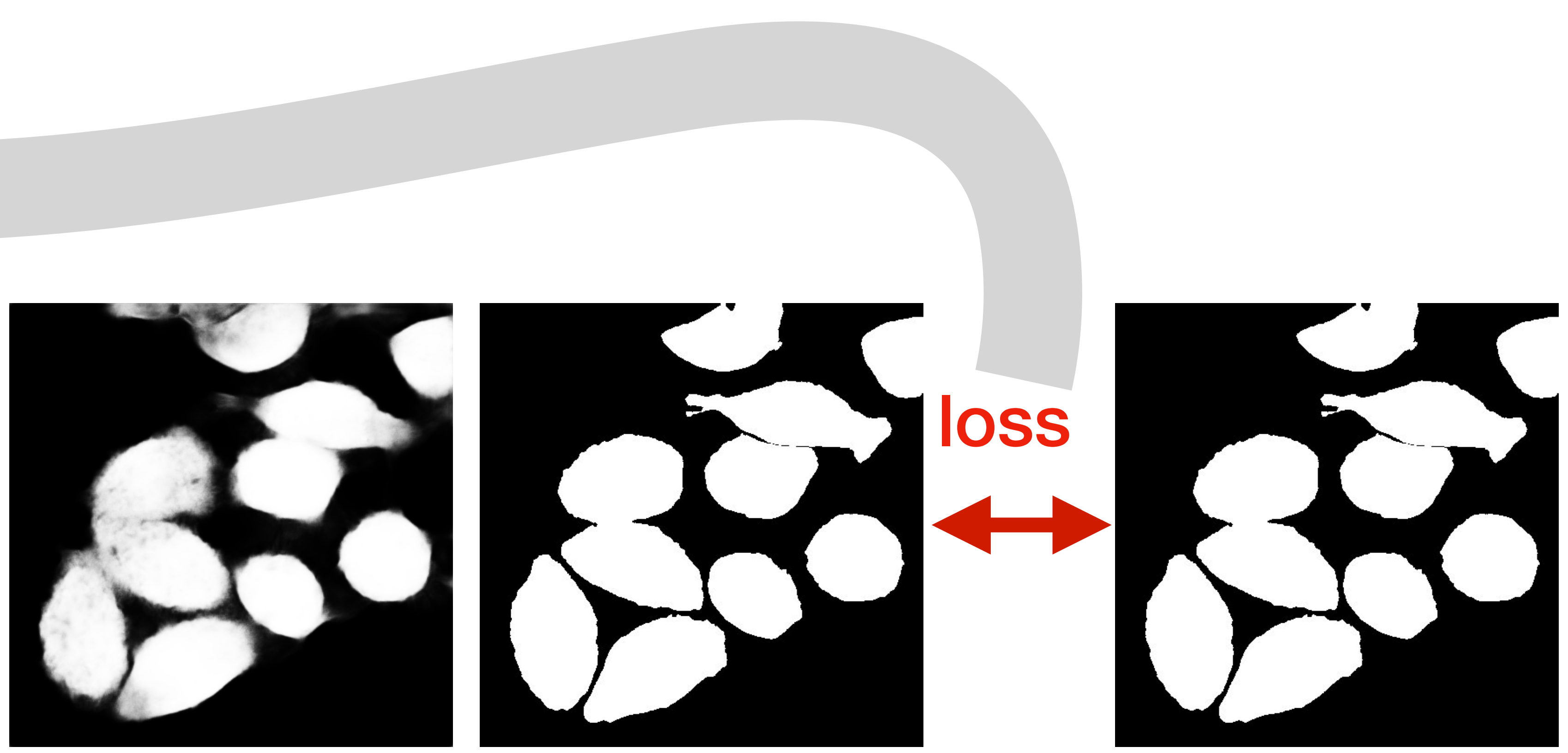
Pooling



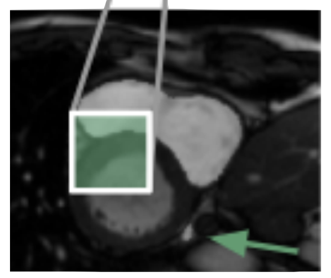
Activation

U-Net

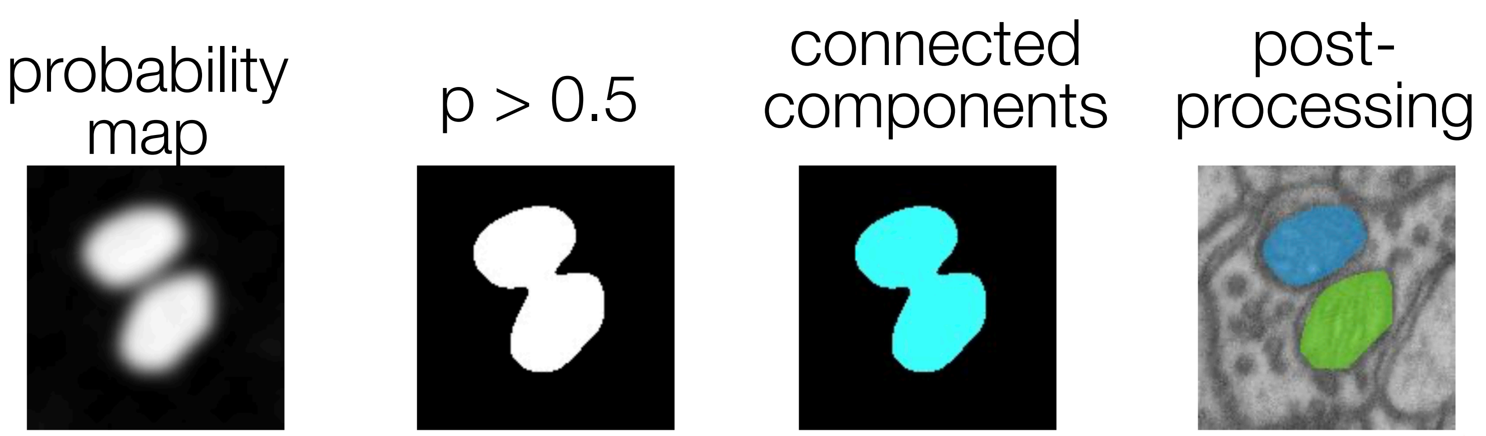




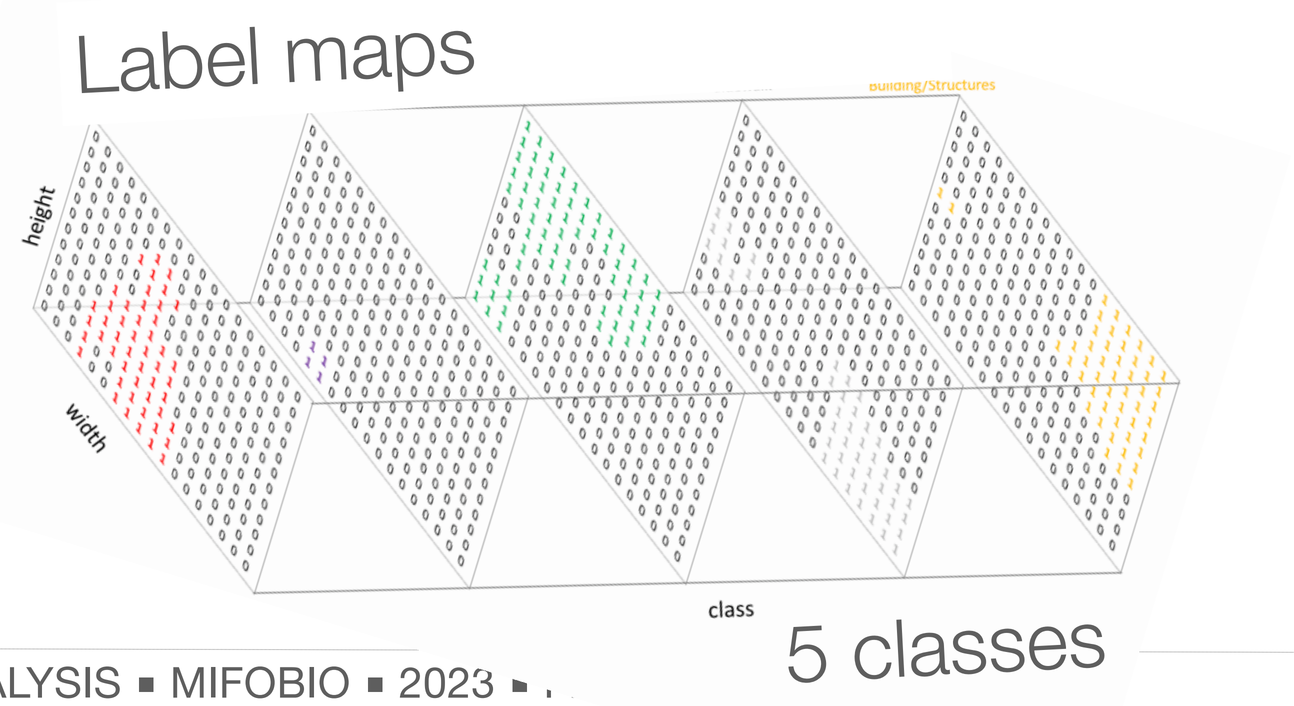
Receptive field

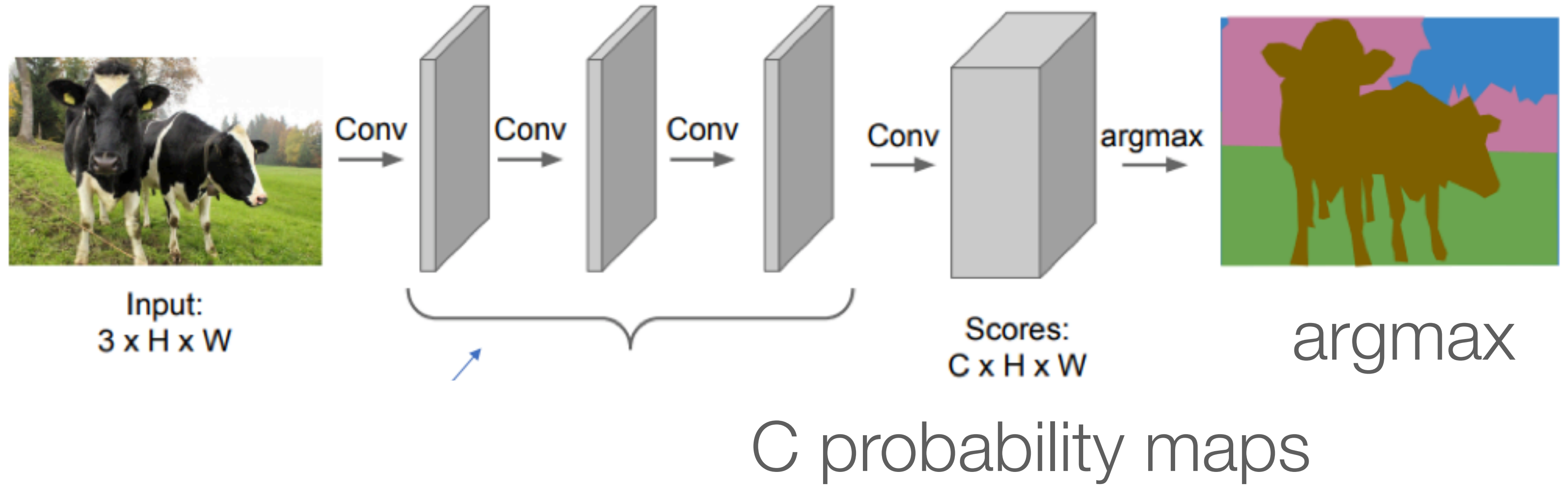


Binary classification



Multiple label classification





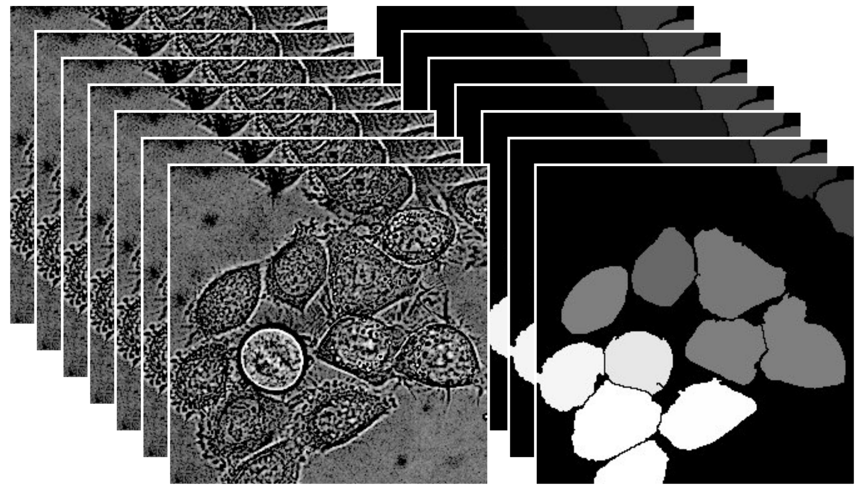
U-Net — pixel classification

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

Supervised Learning

Training

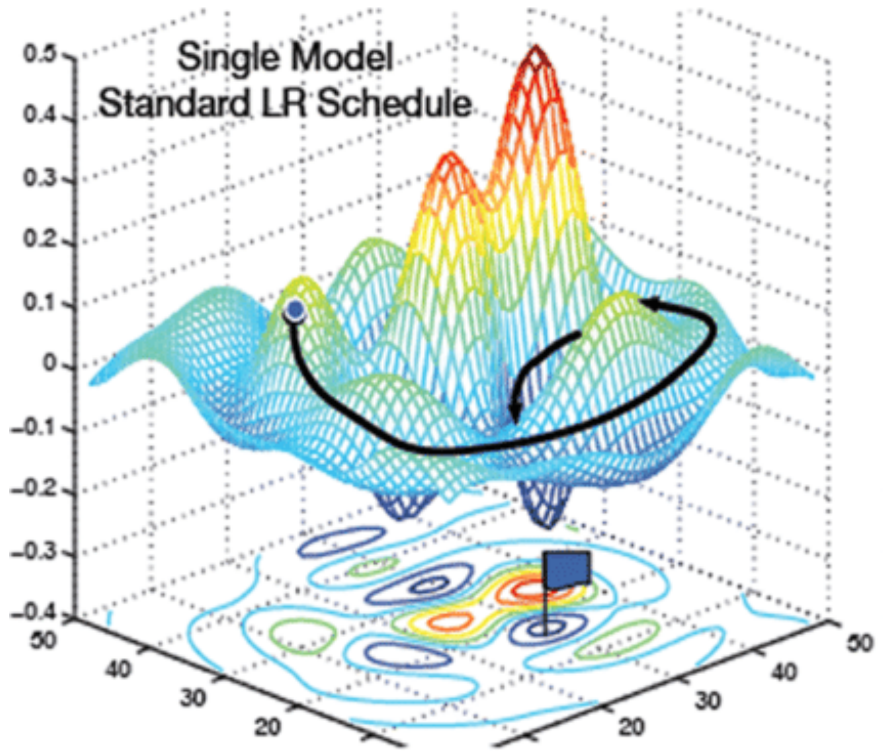
Pairs of images
Training dataset



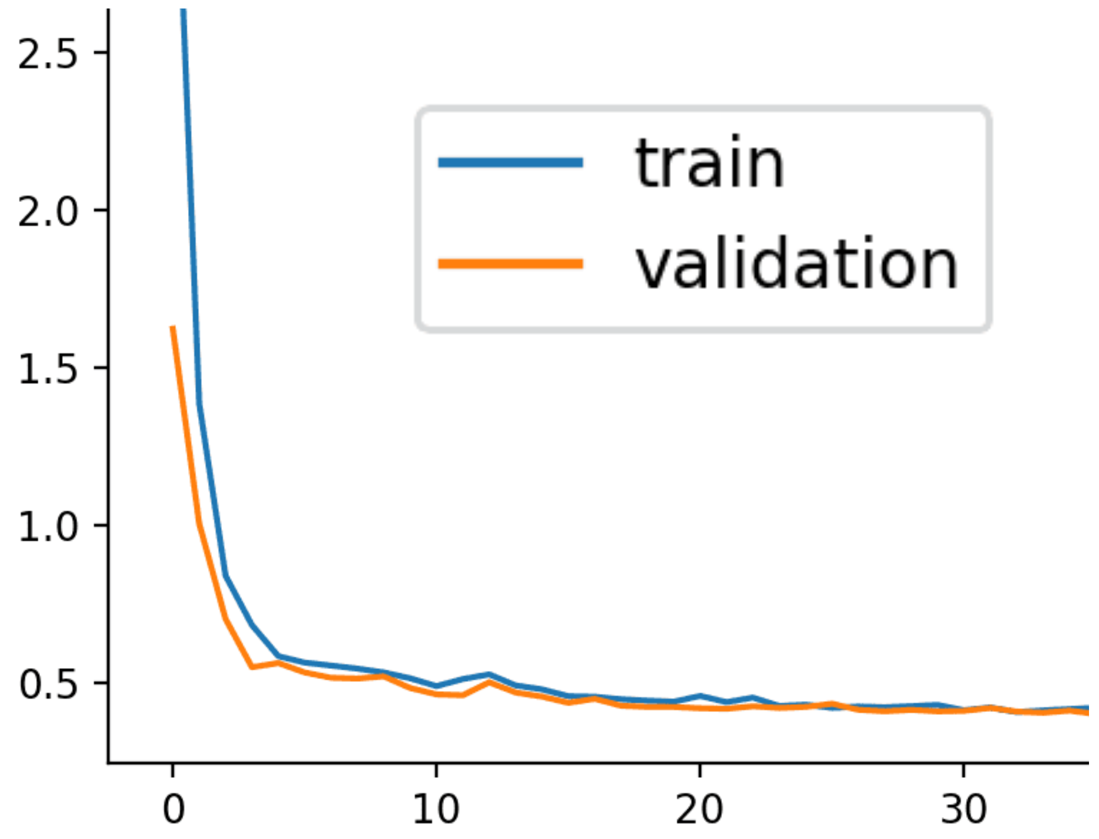
raw annotated

Trained model

Millions of parameters (weights)
of neural networks (NN)



Optimizer (learning_rate)
Adam or SGD

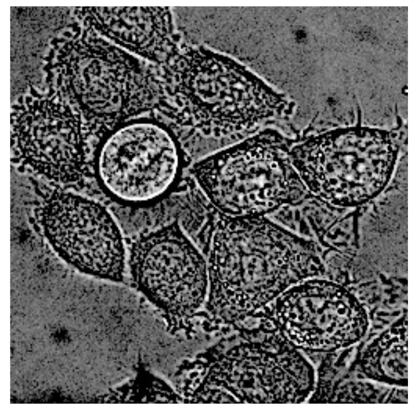


Learning curves
Story of the training: good fit

Inference

Trained model

Image to analyse



raw

Prediction

Loss Function

Regression: measure of the error

$$\textbf{MSE} = \frac{1}{N}\sum_{i}^{N}(y_i - \hat{y}_i)^2 \qquad \textbf{RMSE} = \sqrt{\textbf{MSE}}$$

$$\textbf{MAE} = \frac{1}{N}\sum_{i}^{N}\left|y_i - \hat{y}_i\right| \qquad \textbf{SNR} = 10\log_{10}\frac{\sum_i^N y_i^2}{\sum_i^N (y_i - \hat{y}_i)^2}$$

Perceptual similarity

SSIM Similar structure index for measuring

Classification

$$\textbf{CE} = -\sum_{i}^{N} y_i \log(\hat{y}_i) + (1 - y_i)\log(1 - \hat{y}_i)$$

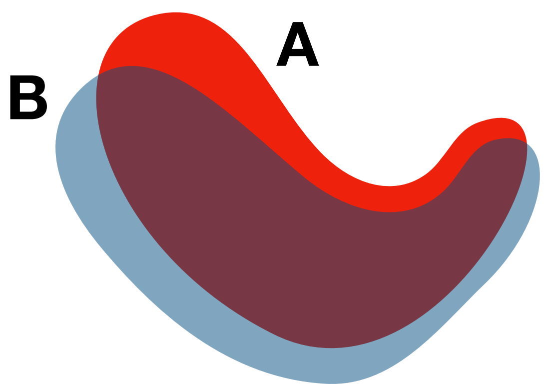
Cross-Entropy (log loss)

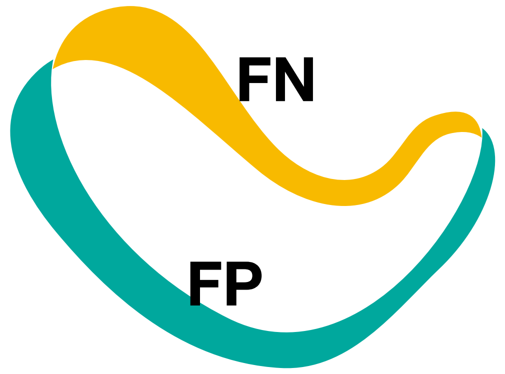
$$\textbf{CE}_{0/1} = -\sum_{i}^{N}\log(\hat{y}_i)$$

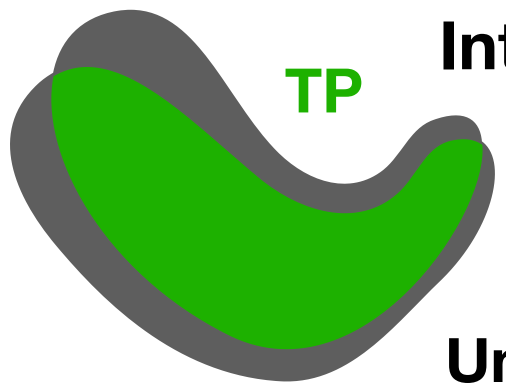
$$\textbf{WCE} = -\alpha\sum_{i}^{N} y_i \log(\hat{y}_i) + (1 - y_i)\log(1 - \hat{y}_i)$$

Weighted Cross-Entropy

Shape comparison







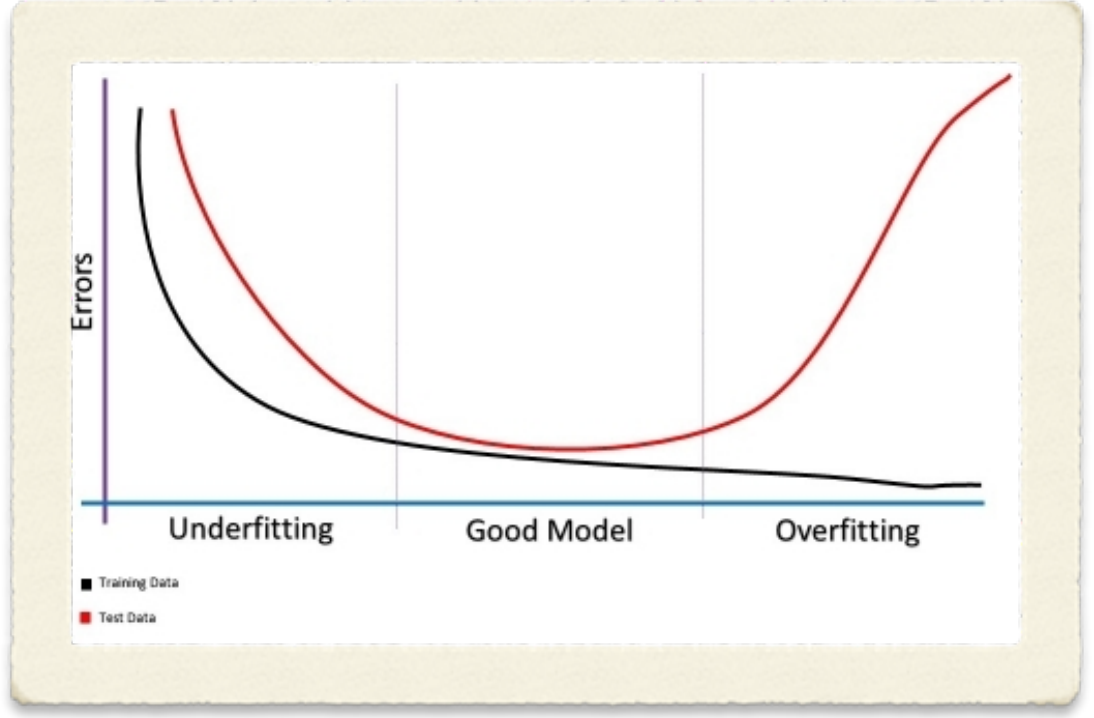
Intersection $= A \cap B =$ **TP**

$$\textbf{IoU} = \frac{A \cap B}{A \cup B}$$

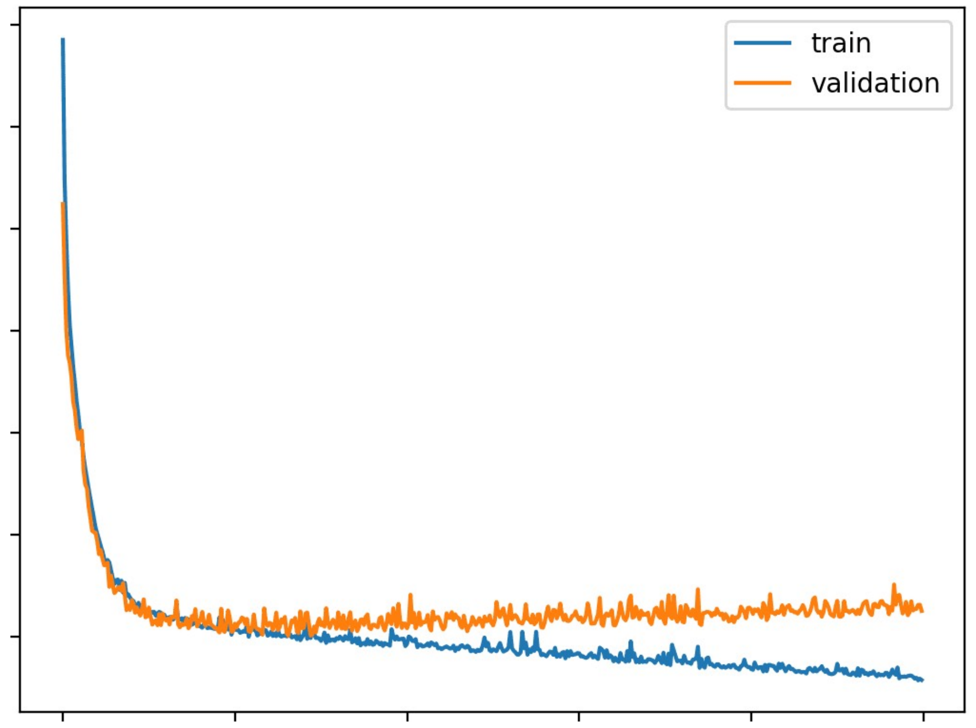
$$\textbf{IoU} = \frac{\textbf{TP}}{\textbf{TP} + \textbf{FP} + \textbf{FN}} = \textbf{Jaccard}$$

$$\textbf{Dice} = \frac{2\,\textbf{TP}}{|\textbf{A}| + |\textbf{B}|}$$

Diagnosis of the Training



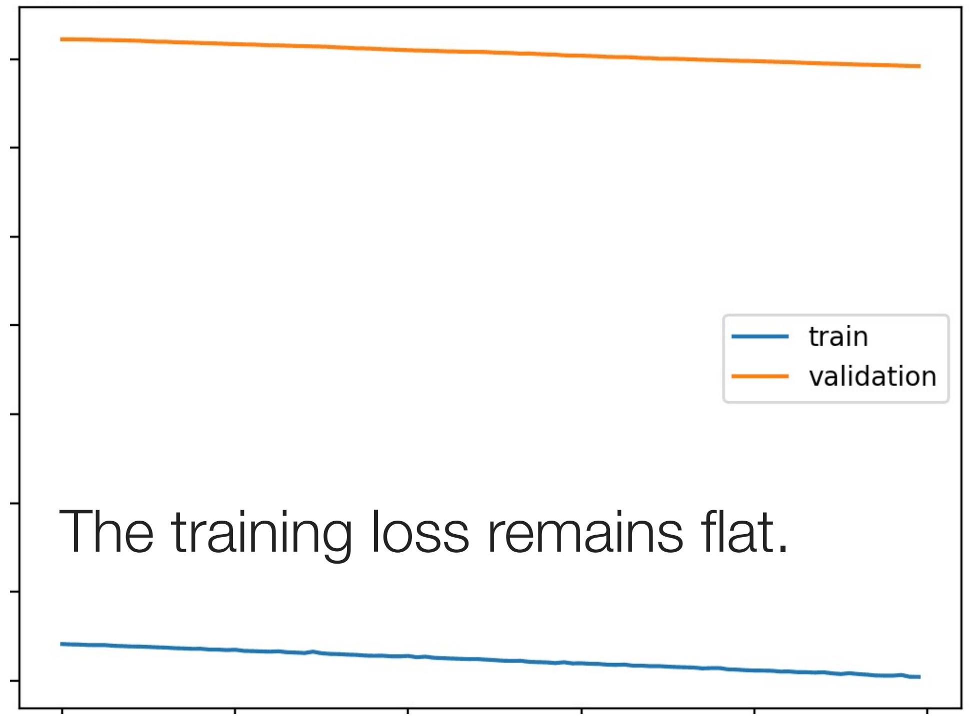
Ove[redacted]has lea[redacted]ell.

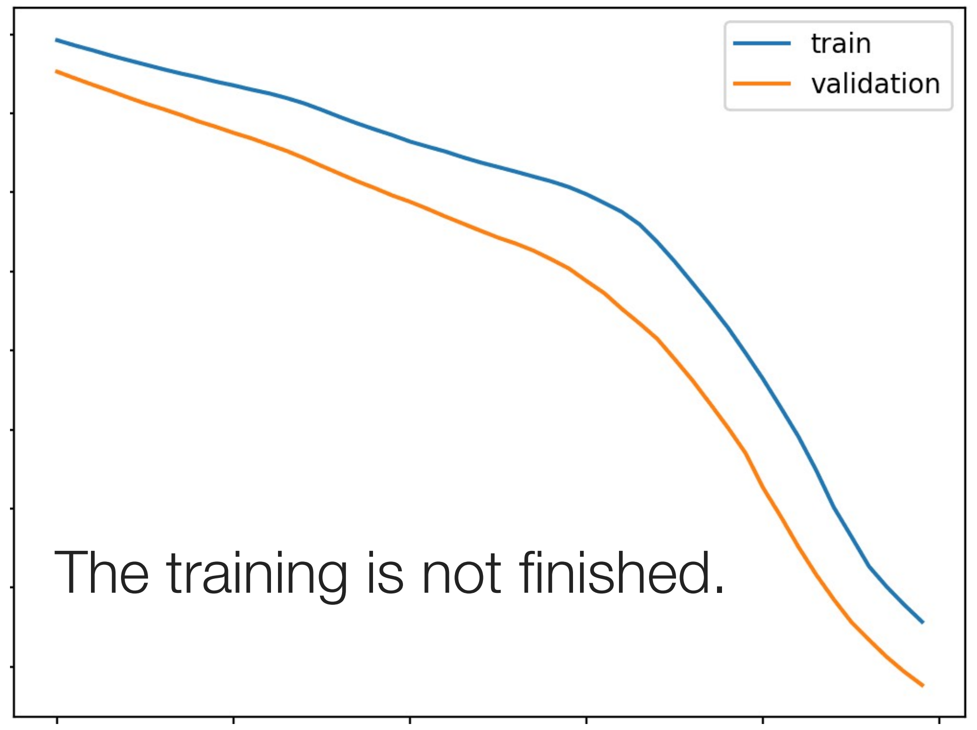


The validation loss decreases to a point and increases later.

The training loss continues to decrease with experience.

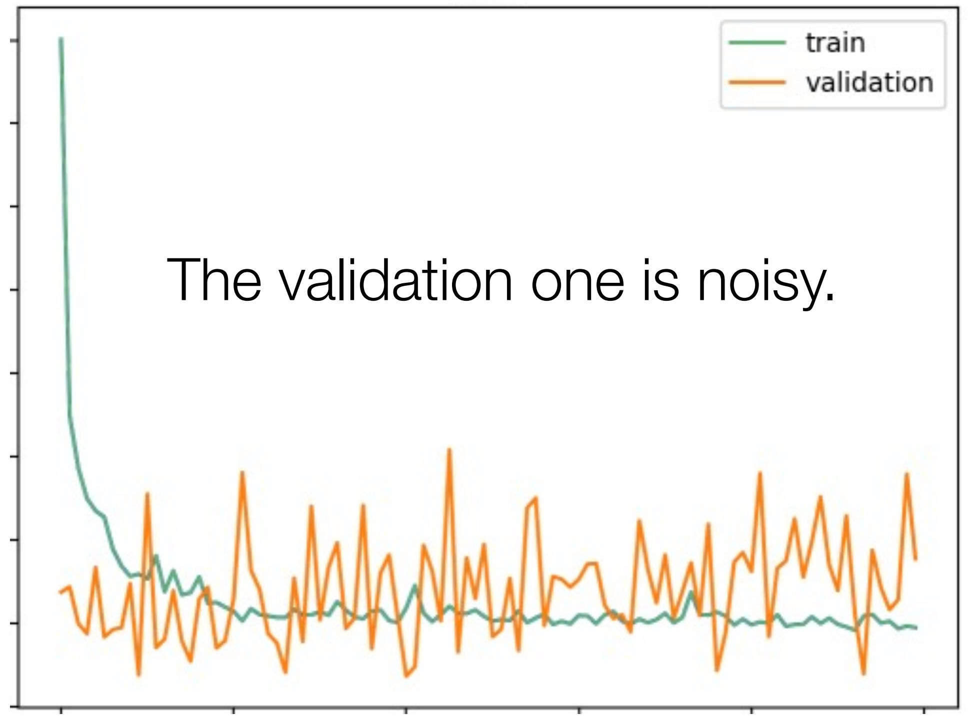
The[redacted]ble ca[redacted]e dataset.

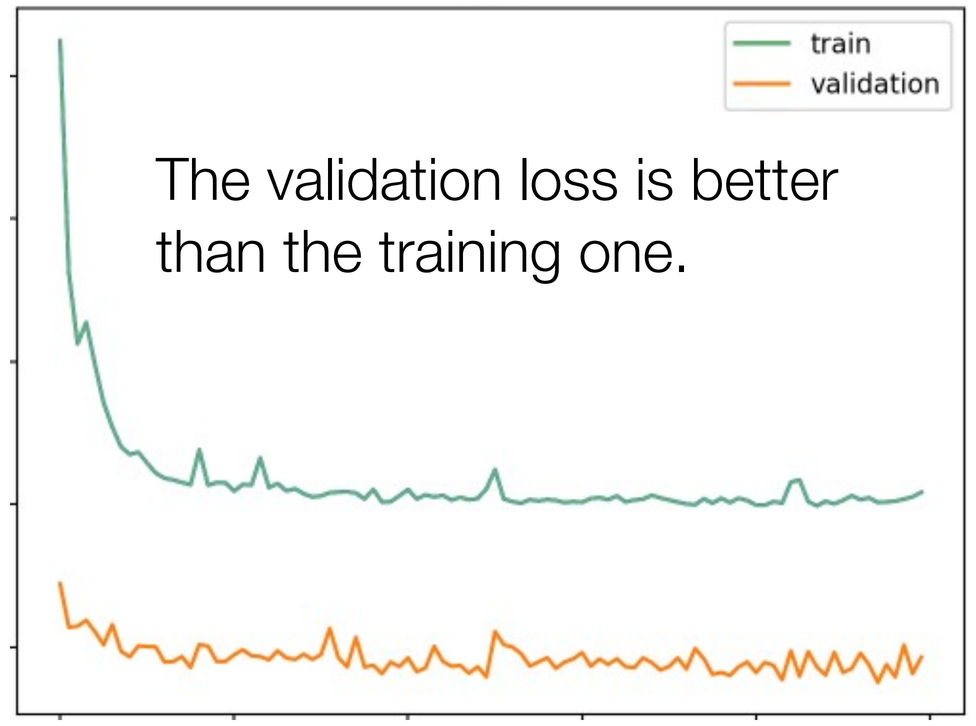




Observation of the training curve.

Th[redacted] provide enough information to be able to generalize.

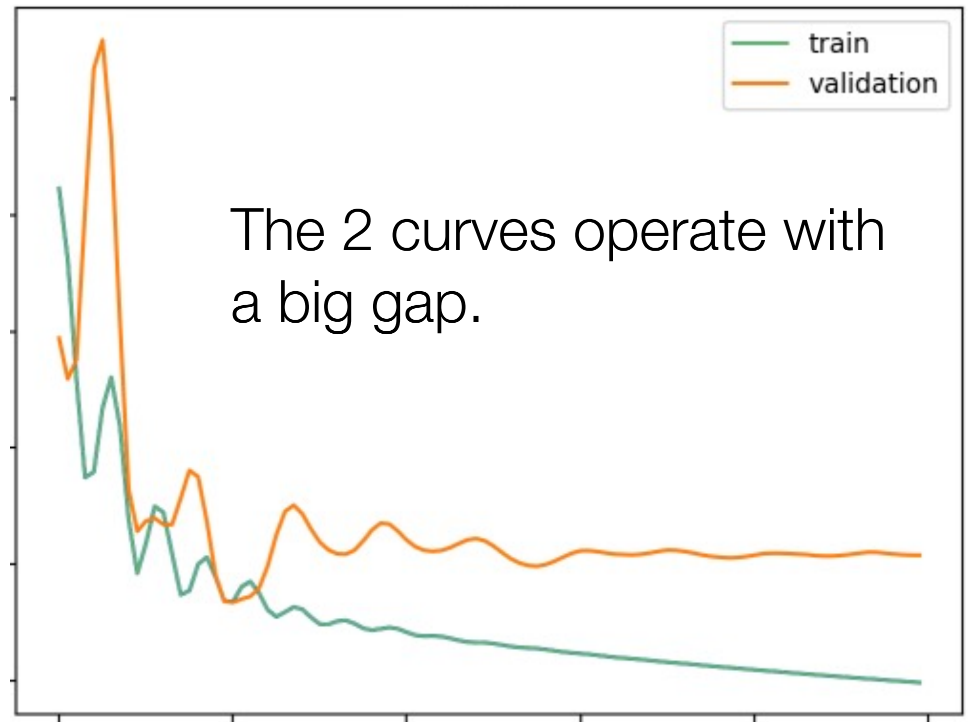




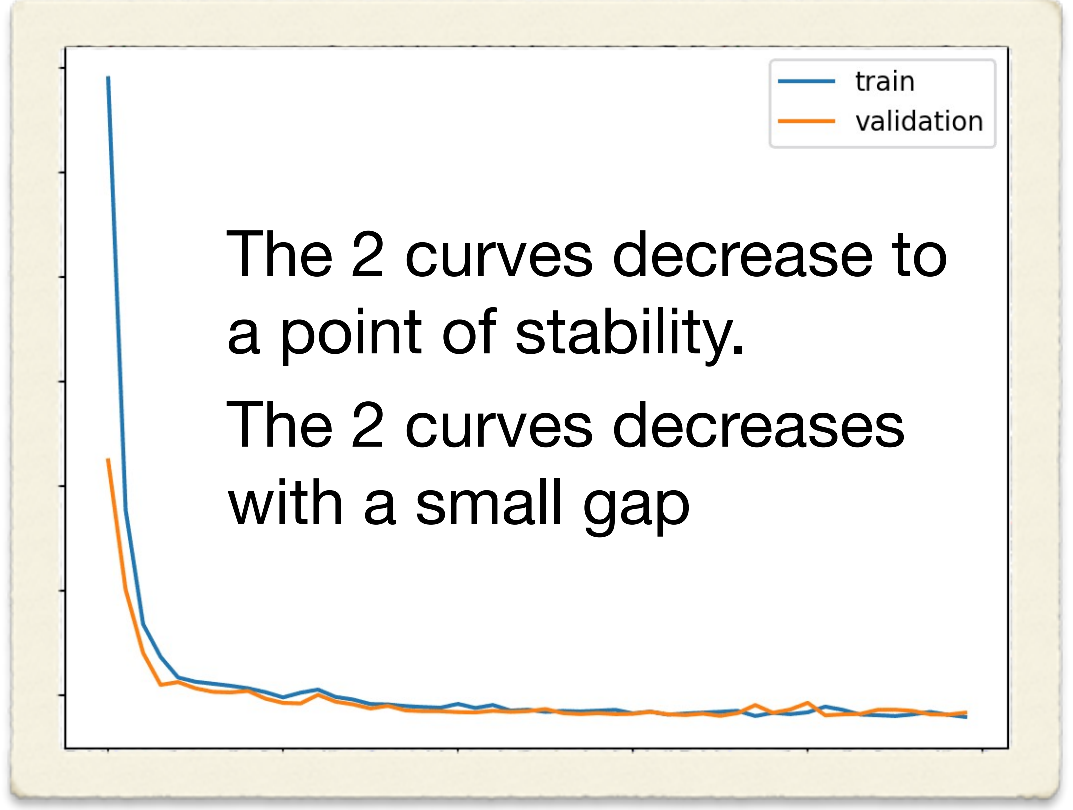
Different behaviour between training and validation.

the validation dataset is easier than the training.

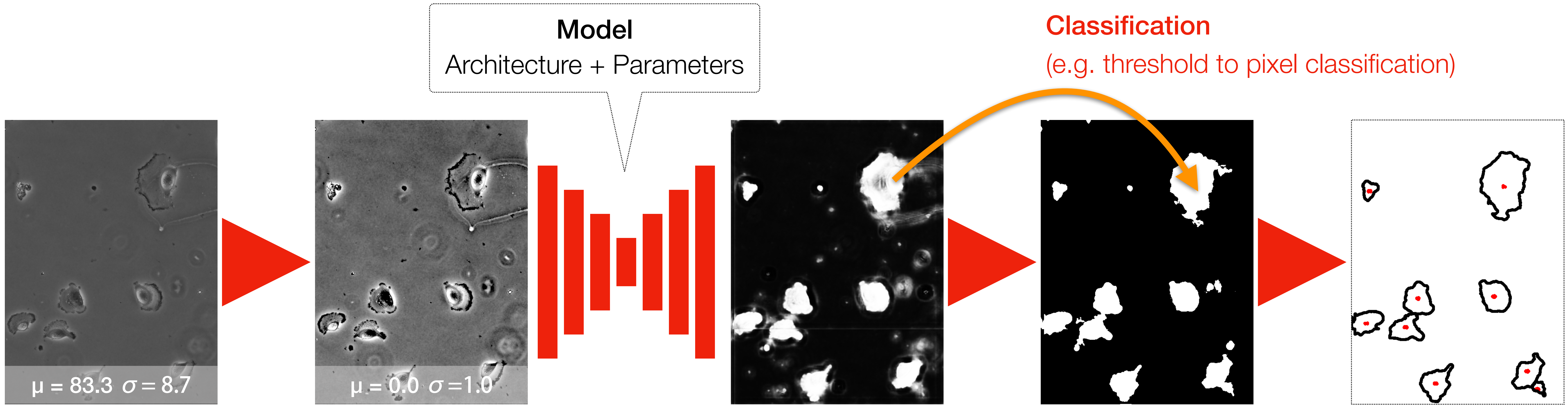
Th[redacted]to capture the model, relative to the validation dataset.



Good Fit



Prediction



Preprocessing
(e.g. Normalization)

Prediction
(e.g. map of probability)

Postprocessing
(e.g. Analyze Particles)

- Dynamic range
- Scale
- Noise level

- Object detection
- Watershed
- Measurement

Deep Learning for Bioimage Analysis

Concepts and Hands-on Neural Networks Training with a Critical Approach

CONCLUSION

Design of datasets

Risks and challenges

Concluding remarks

The data is the code.

Scientific Design of the Training Datasets

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

Dataset size
Few ground-truth **Overfitting**

Normalized data
Misaligned raw data **Divergence**

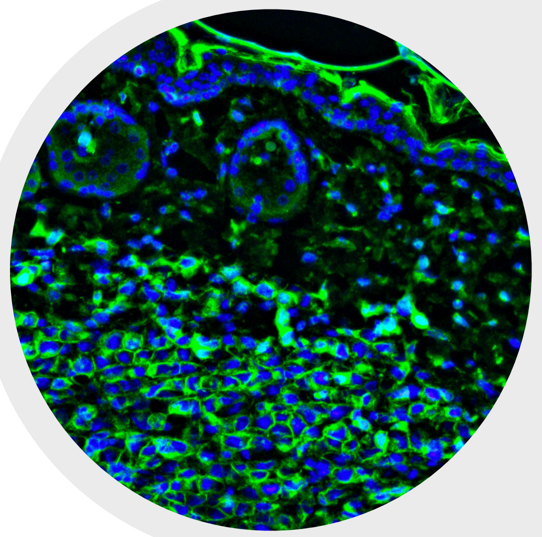
Model complexity
Few # parameters **Underfitting**

Representativity
Mismatch conception **Dataset shift**

Data selection
Favor/exclude phenotypes **Bias**

Distribution
Imbalance classes **Ignore minority**

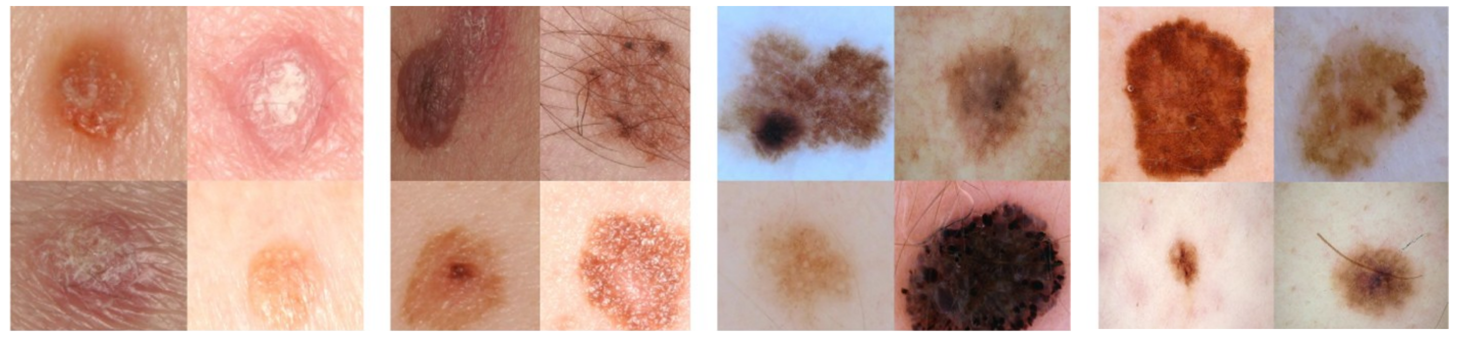
When you train a machine, you are actually programming the algorithm using data instead of using code.



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

Where to Find Ground-Truth Data?

HUMAN EXPERTS



130'000 skin **lesion** images
2000 diseases
21 dermatologists

Esteva et al., Nature, 2017.

Manual tools

AnnotatorJ [Hollandi, 2020]
Segmentor [Borland 2021]
webKnossos [Boergen 2018]

Semi-automatic tools

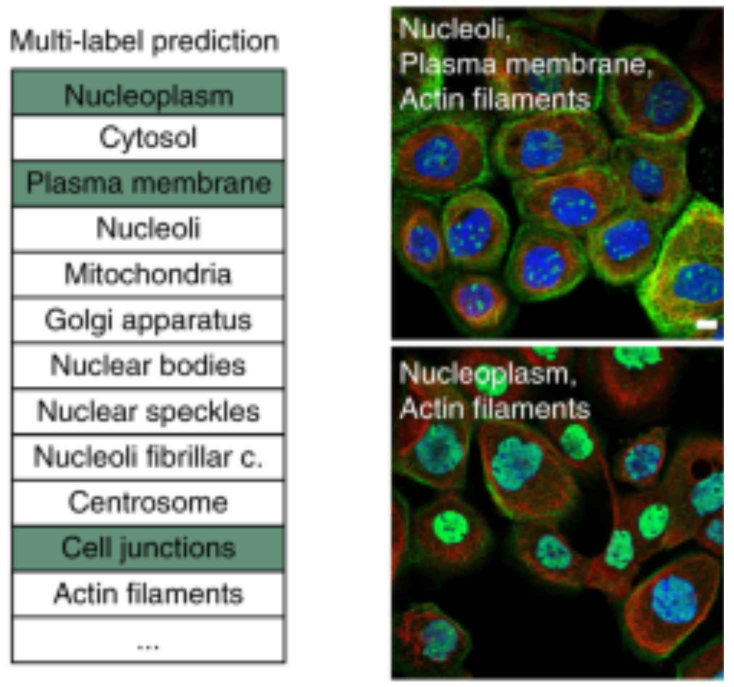
Ilastik [Berg, 2019]
Weka [Arganda, 2017]
LabKit [Fiji Team]

3D?

CROWD-SOURCE

CITIZEN SCIENCE

- Cell Atlas of the Human Protein Atlas
- Image class. task of subcellular pattern
- As a video game



PUBLIC DATABASE

ImageNet
14'000'000 images 22'0000 cat.

Kaggle Data Science Bowl
Find the nuclei in divergent images
Caicedo et al. Nature Methods 2019

Broad Bioimage Benchmark
Ljosa et al., Nature Methods, 2012

CytoImageNet
900'000 images / 900 classes
Hua, arxiv, 2021

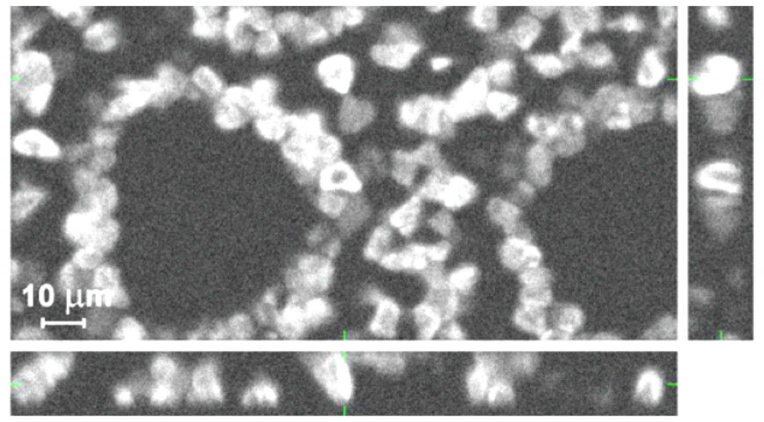
Building bioimage databases

- Consensus
- Demanding
- Make it available!

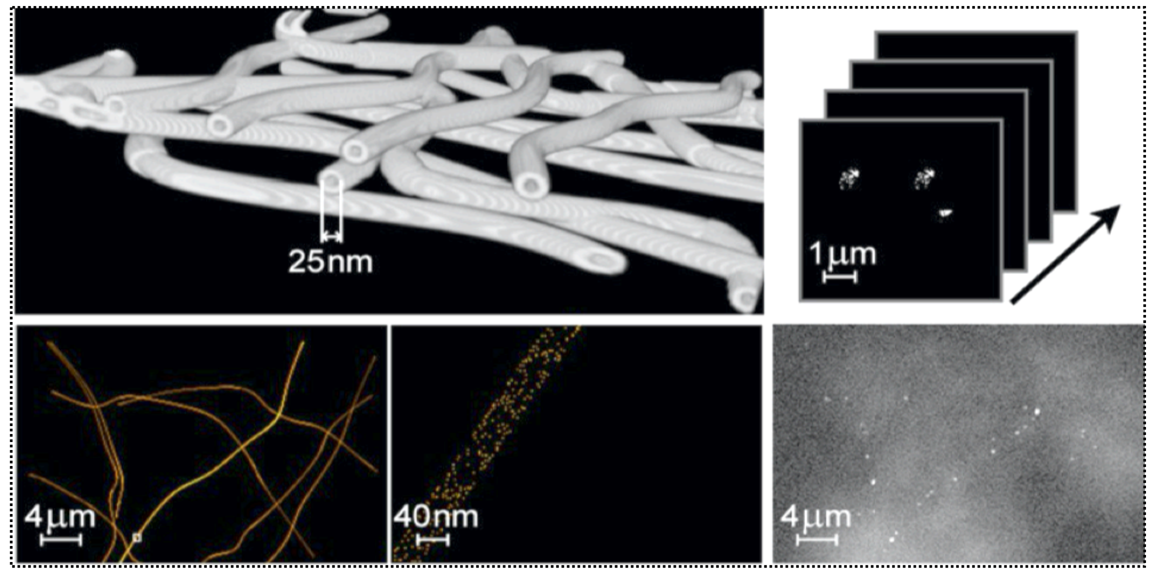
Where to Find Ground-Truth Data in Microscopy?

SIMULATION

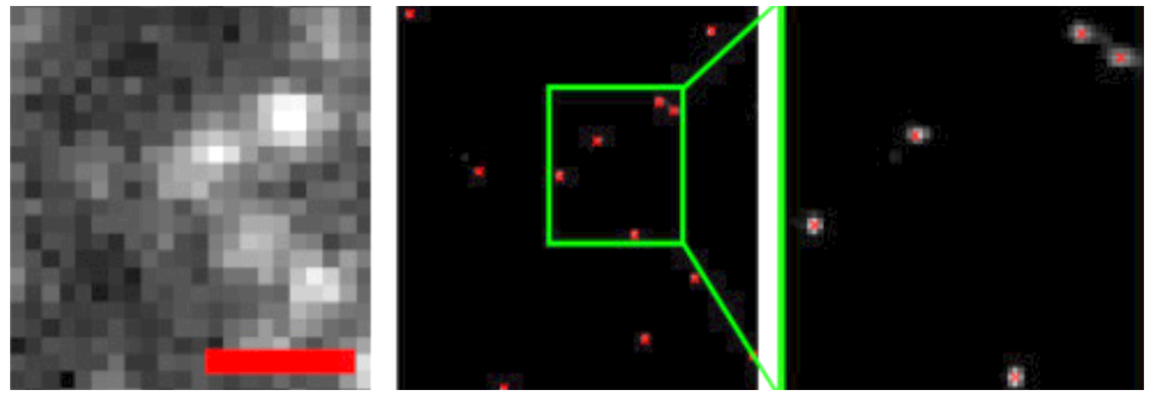
CytoPacq
D. Wiesner et al.
Bioinformatics, 2019



Light propagation



SMLM Simulator
Sage et al, Nat. Met. 2015



Deep-STORM
Nehme et al, Optica 2018

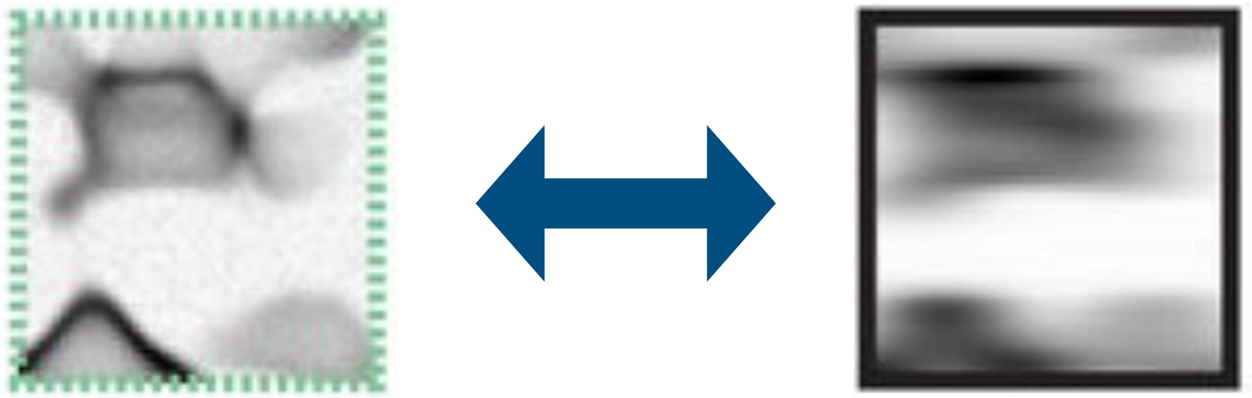
Require the knowledge of the forward model (physic)

CONTENT AWARE

Smart Acquisition
- Low SNR and high SNR
- Super-resolution and resolution standard
- Different focal plane
- Non-bleached images

Simulation the degradation
- Add artificial noise (e.g. noise2noise)
- Blur with a artificial PSF
- Artificially degraded

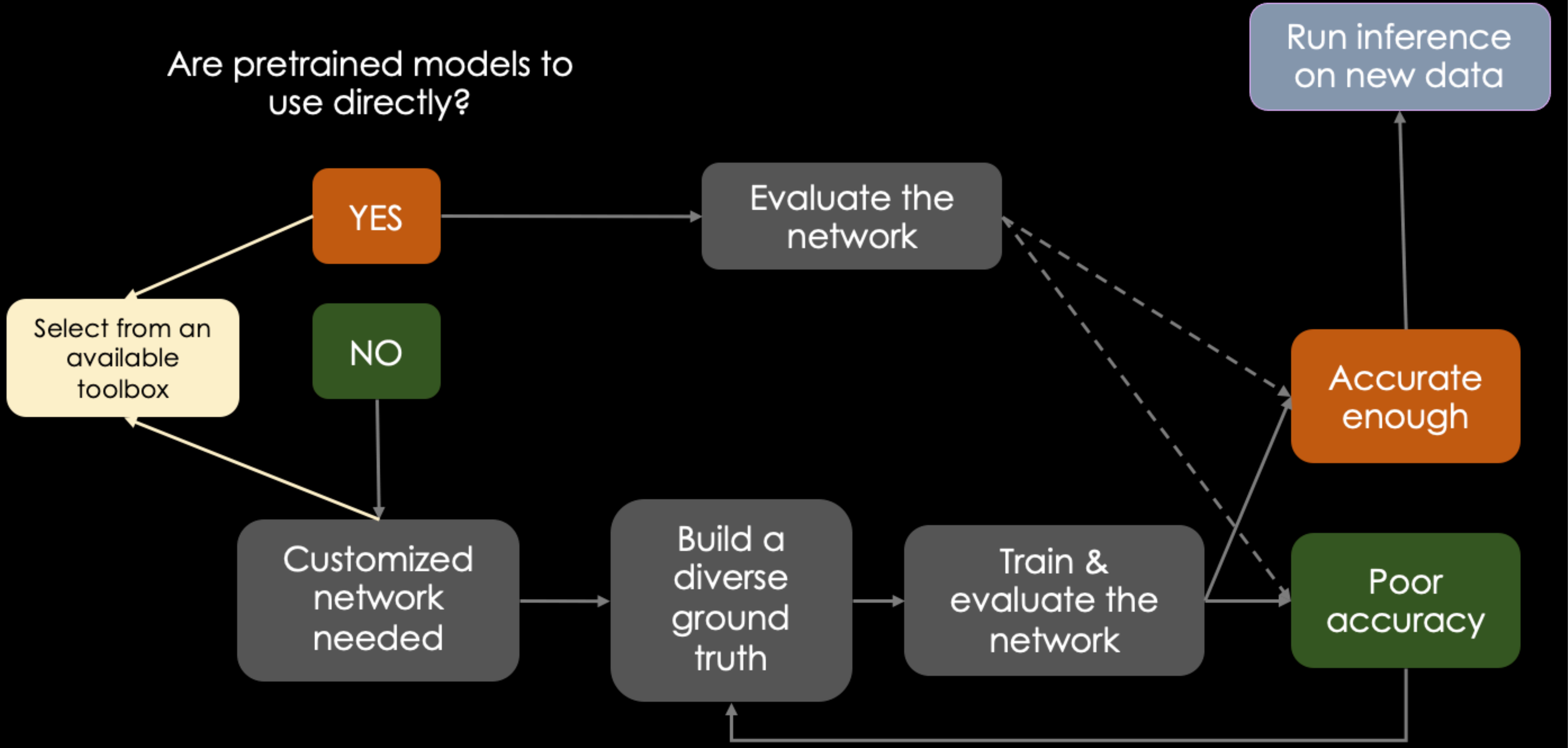
Lateral plane **Axial plane**



Deep learning systems

Identify the sample type(s) and features you need to analyse

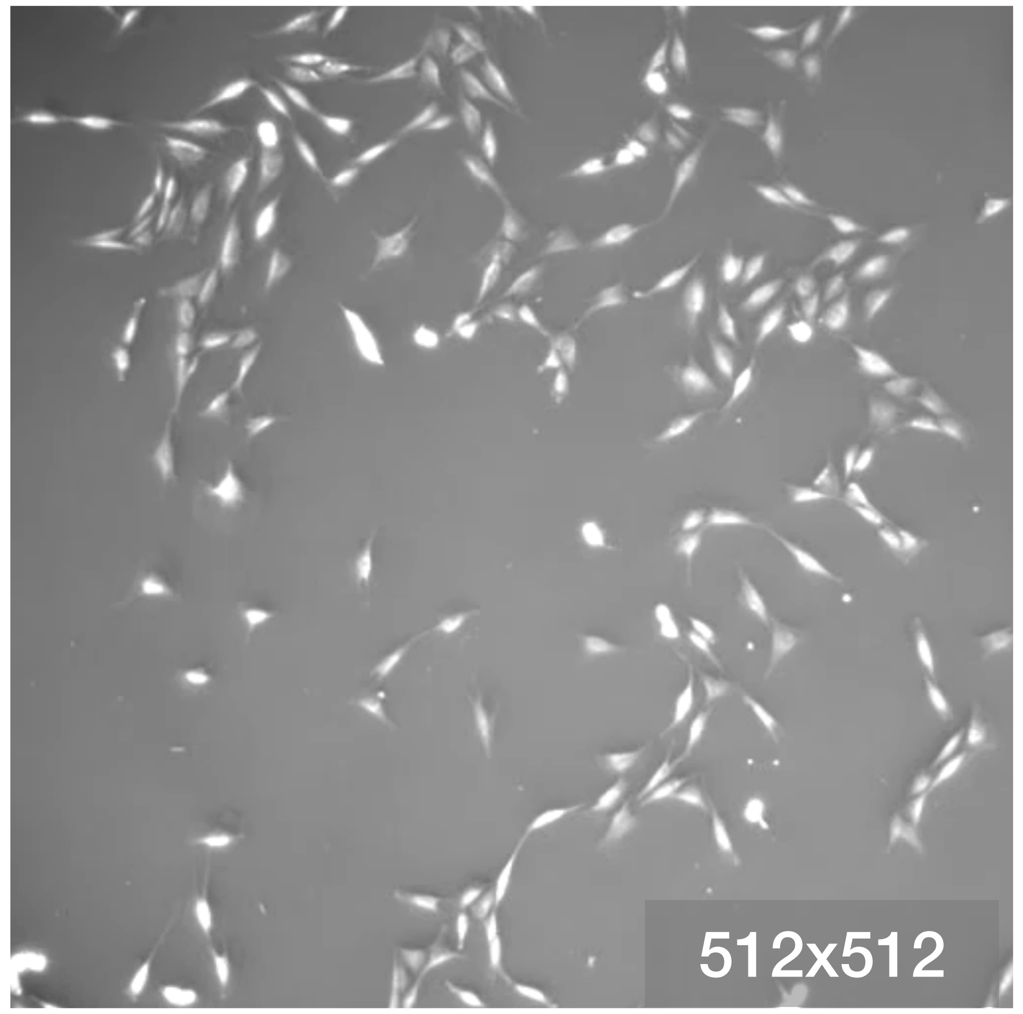
Questions to consider:
What features give me the info I need?
Live or fixed?
Highly accurate?
Multichannel?



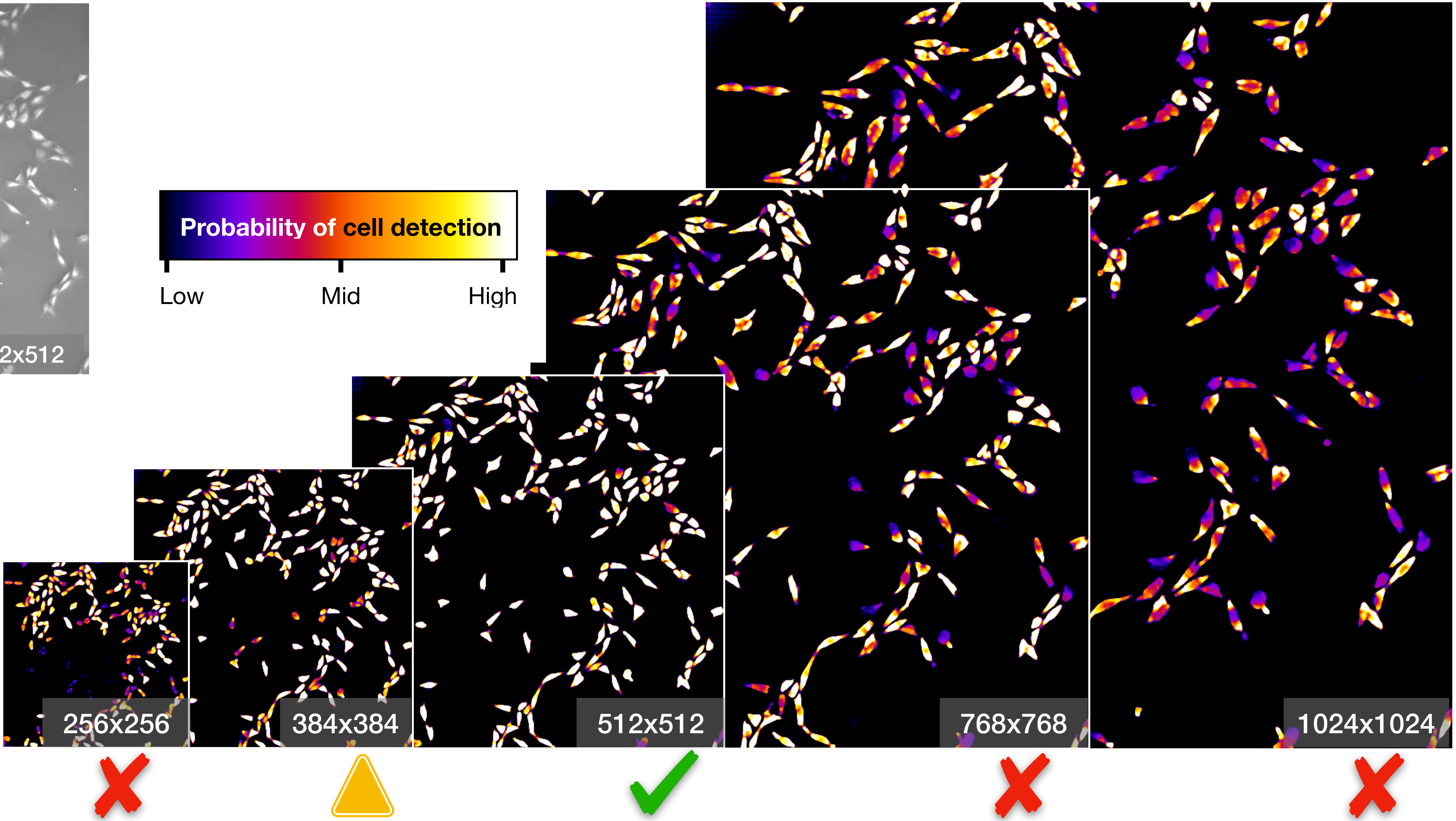
A. Mathis, et al., Neuron 2020

Robustness

deepImageJ



Input image



Hallucination

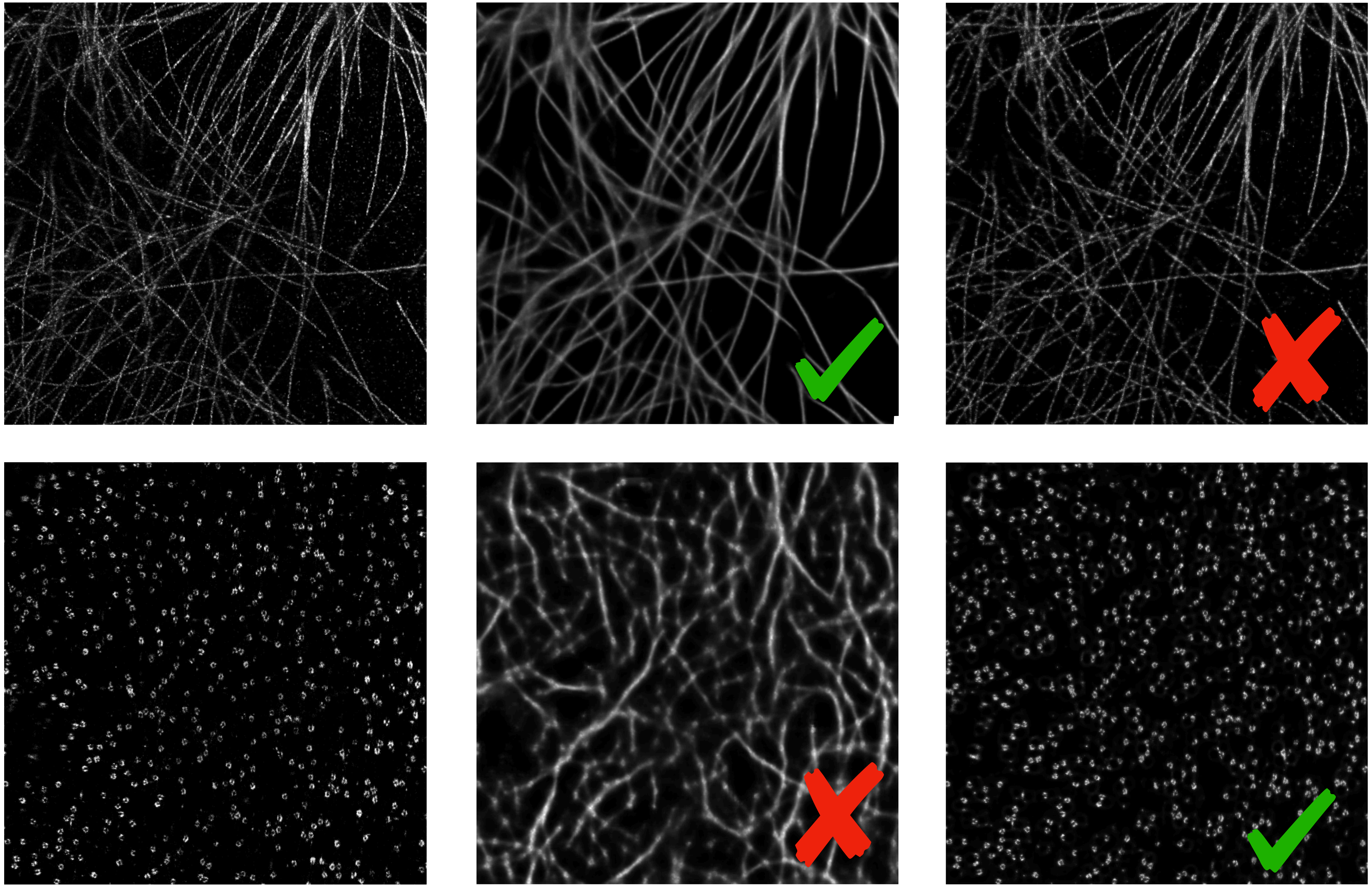
deepImageJ

ANNA-PALM
Wei Ouyang et al.
Nature Biotechnology, 2018.

Image of the dataset | Prediction NN Tubulins | Prediction NN NPC

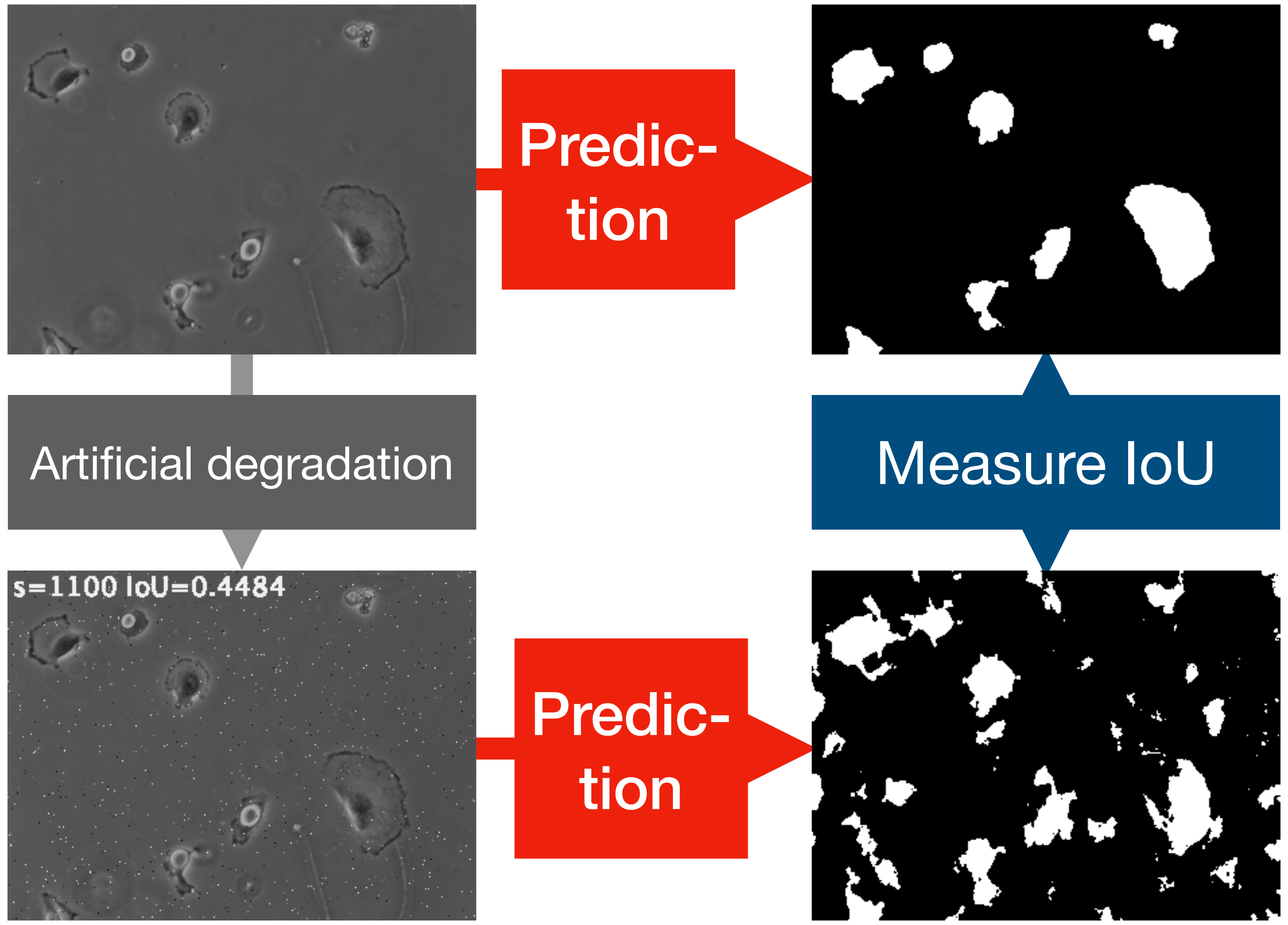
Image used to
train the network
Super-resolution Tubulins

Image used to
train the network
Super-resolution NPC



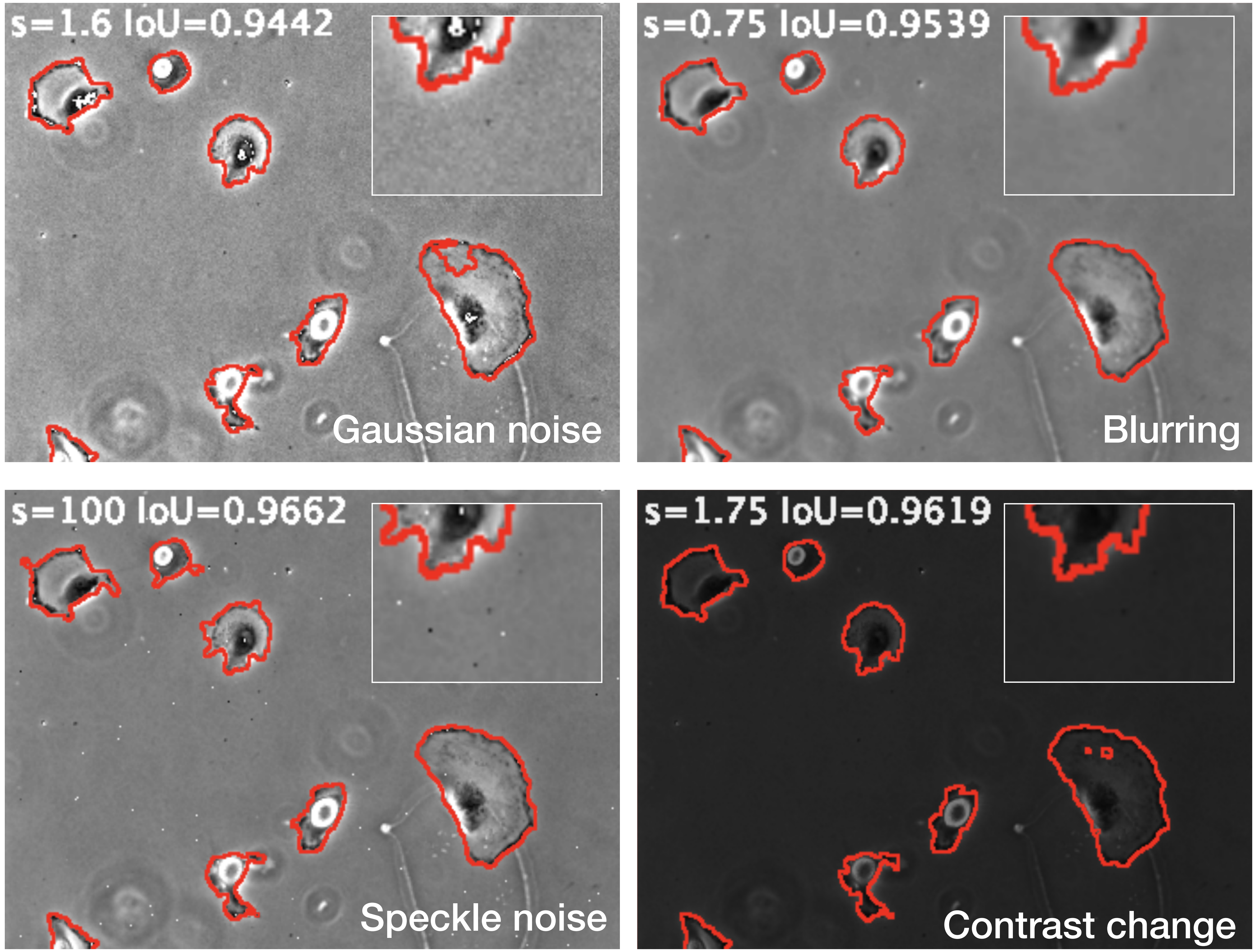
Adversarial Attacks

deepImageJ



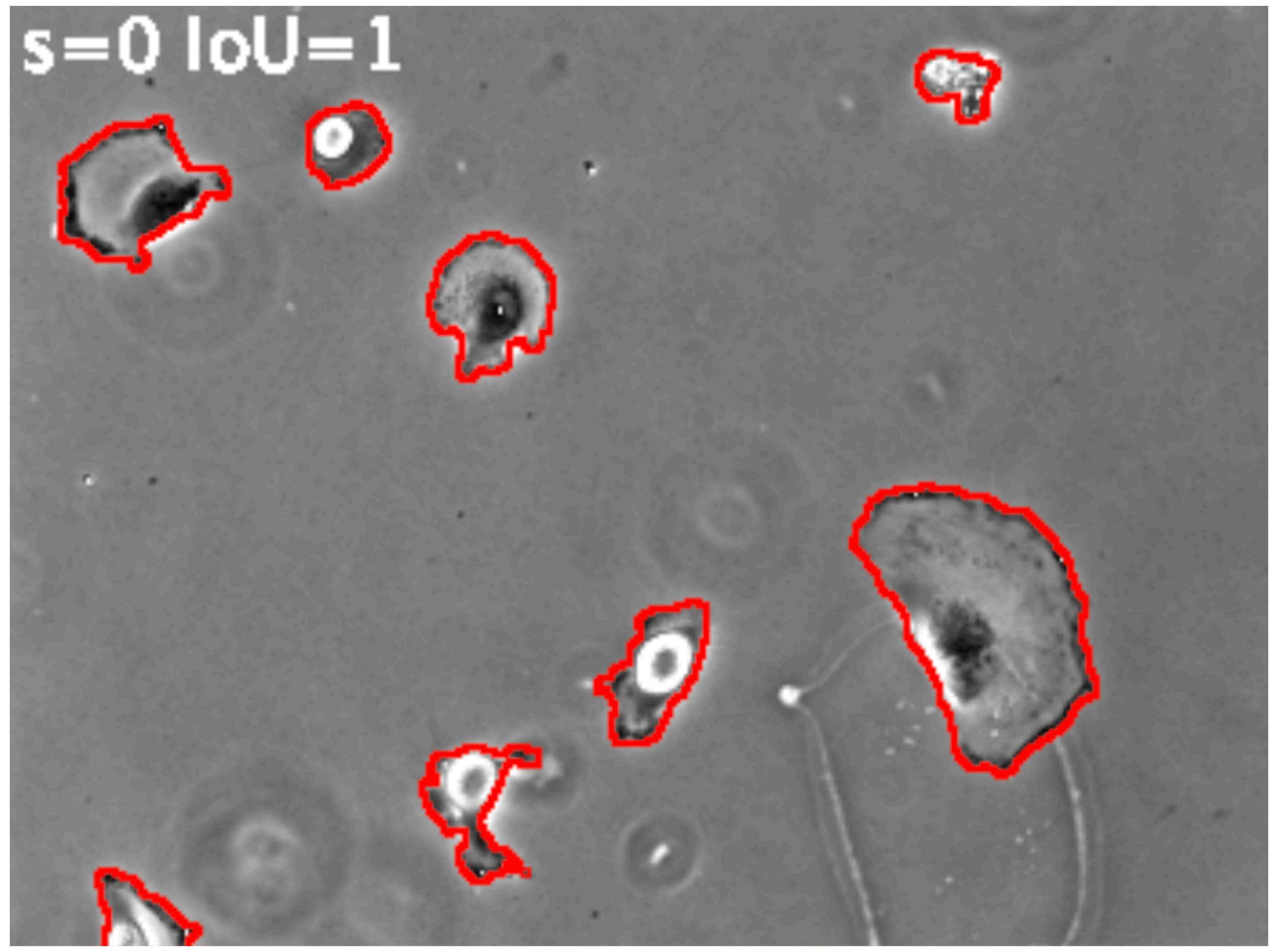
Robustness

Largely studied in computer vision but little attention in microscopy!



Data

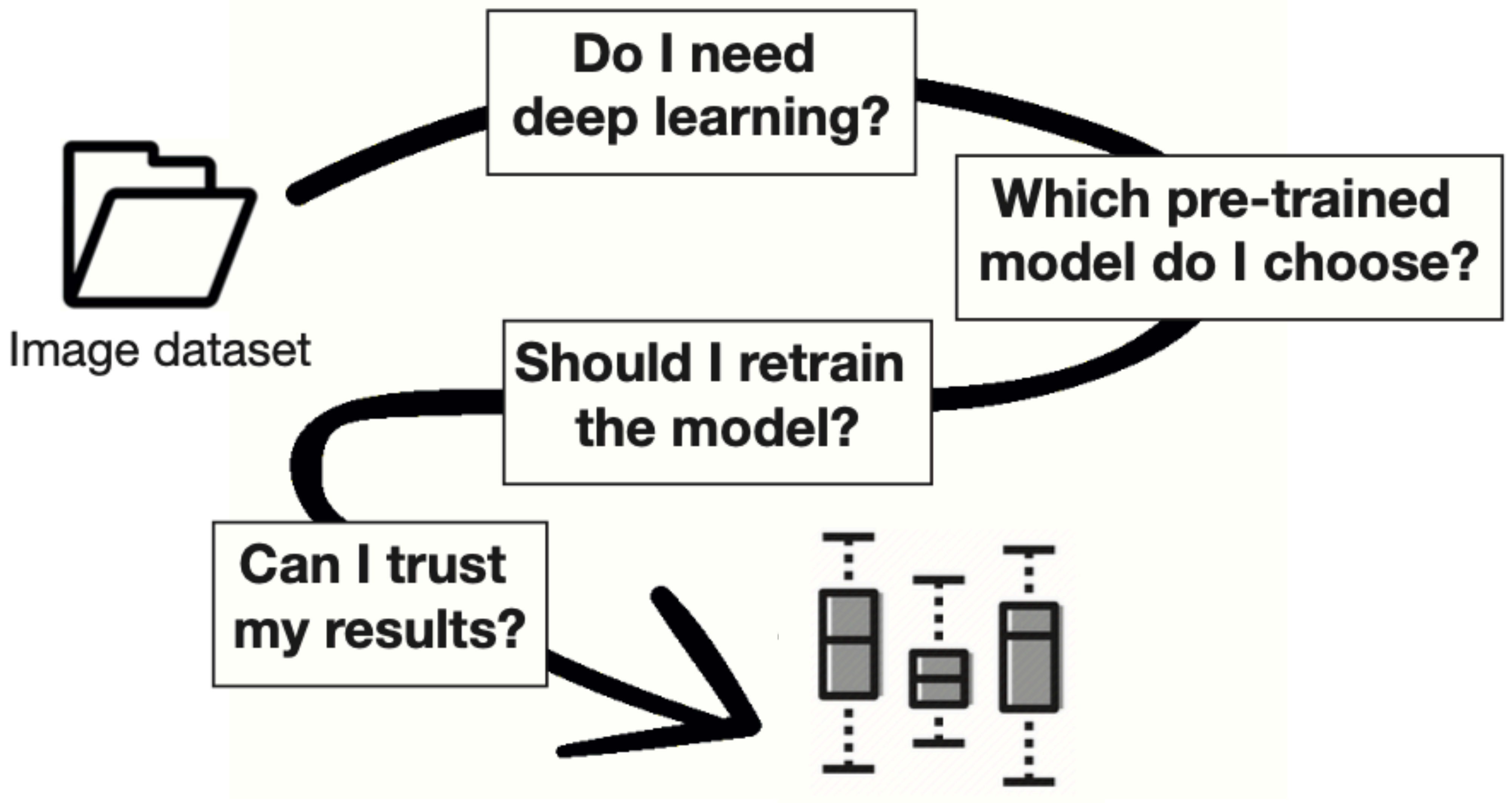
Cell tracking challenge



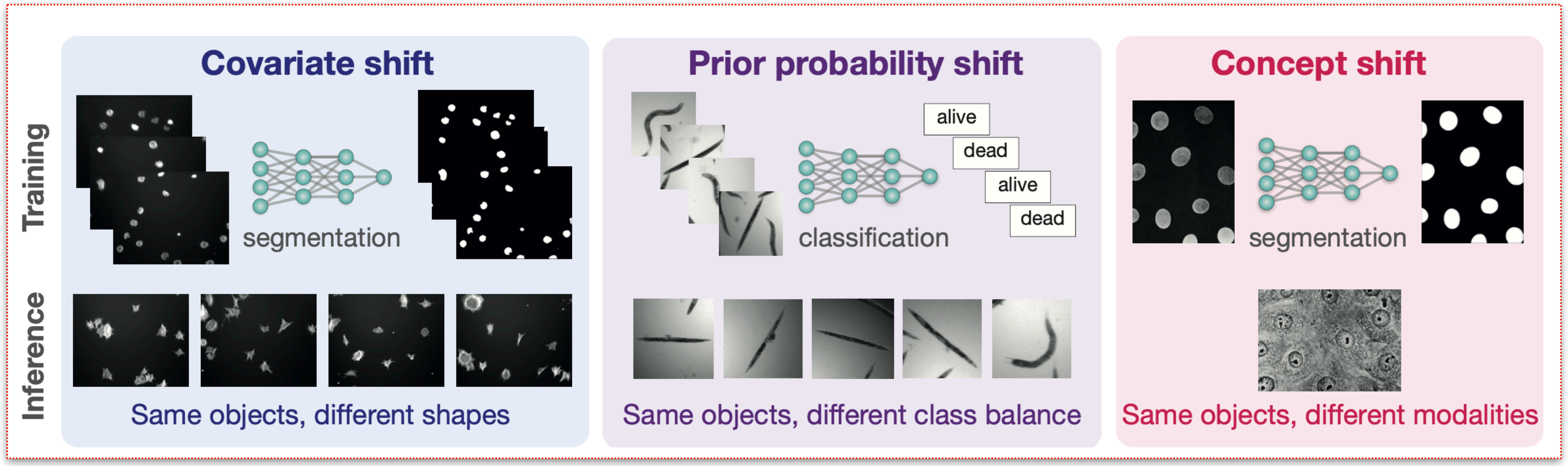
Risks and Challenges

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

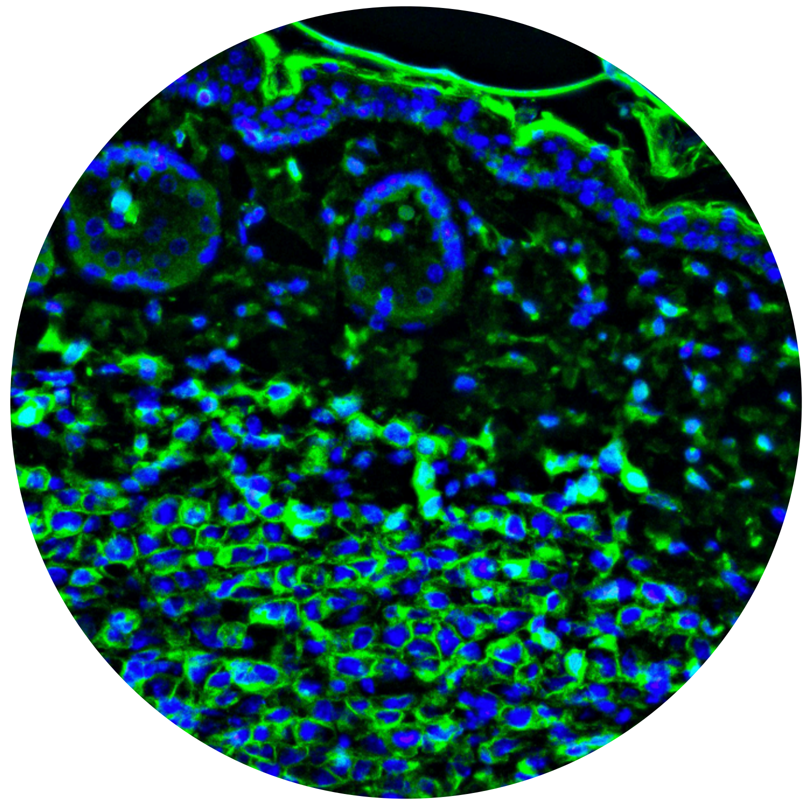
→ Validation / Interpretation

Conclusions

■ Responsibility

Scientific design of the your train dataset

Preparation, curation, annotation, validation, integrity, open



■ Education

Knowledge and skills in deep learning and image analysis

Integration of knowledge, reuse, engineering image analysis skills

■ Ethics

Conscience of impacts and limits

Sample, annotators, bias, environmental footprint, FAIR rules

AI4Life in a nutshell

Goals

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

Outcome

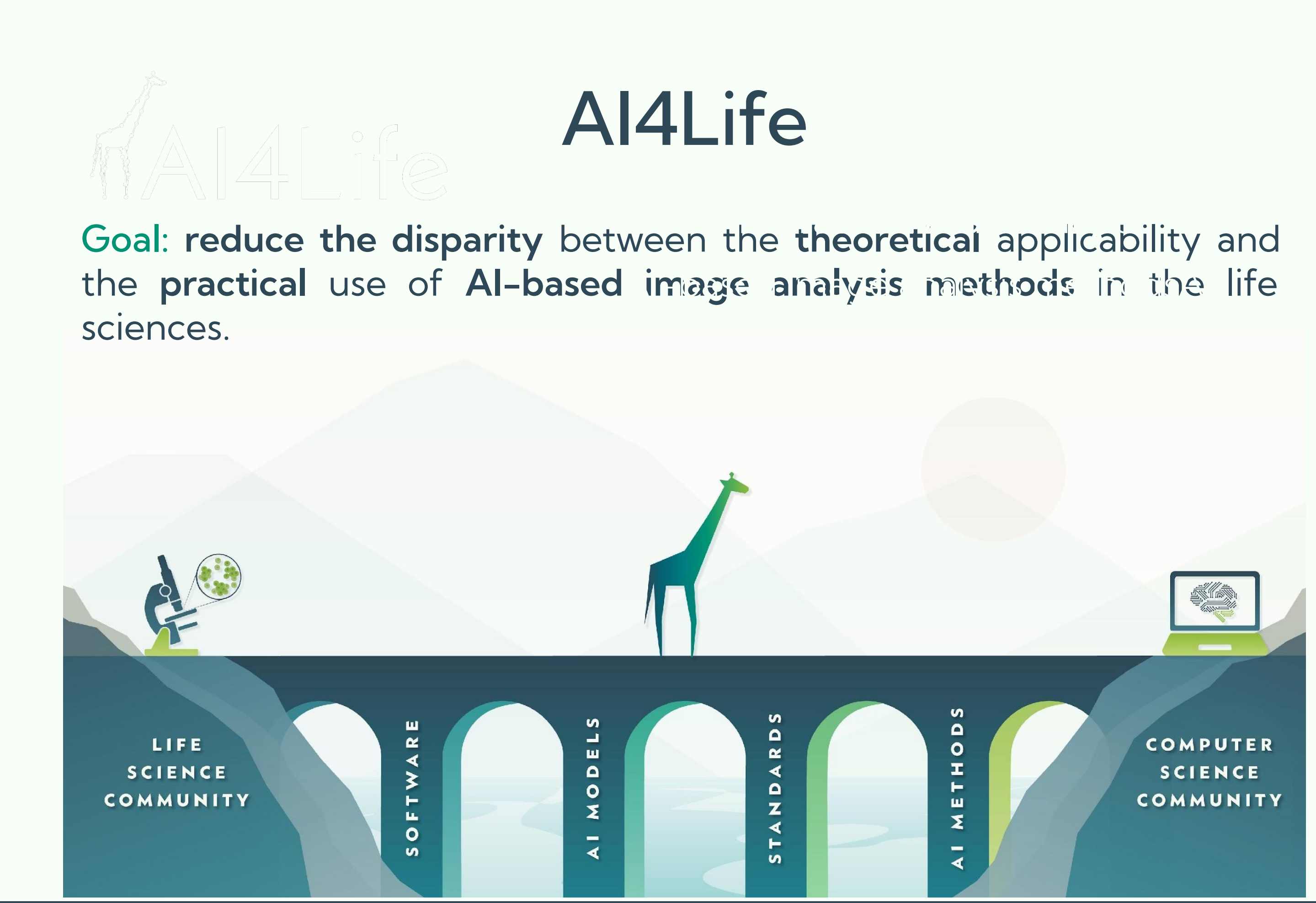
Services provided and solution developed to solve microscopy image analysis problems

Facets

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

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

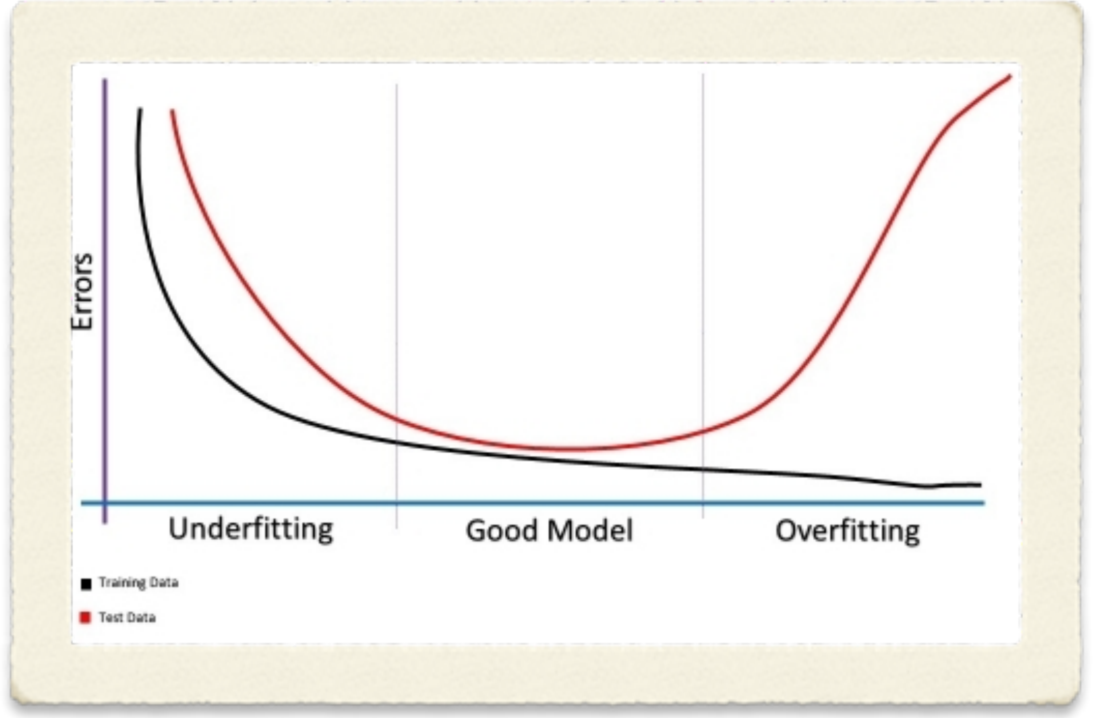
Organize **Open Calls and Challenges** for image analysis problems

Establish standards for the submission, storage and FAIR access

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

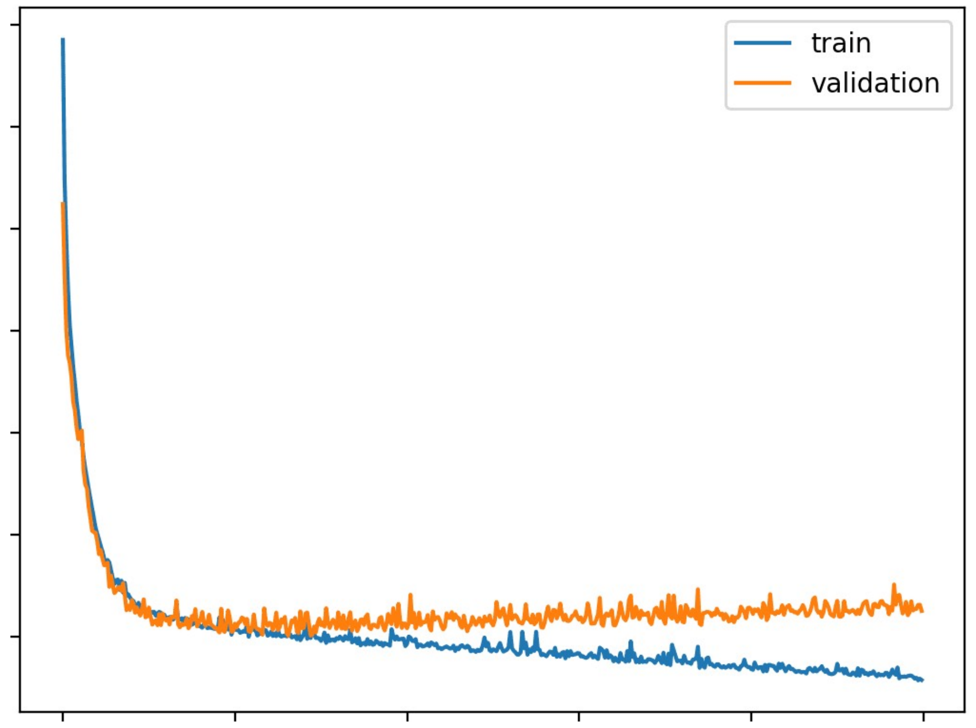
Organizing outreach and training events courses/workshops and participation in conferences

Diagnosis of the Training



Overfitting

Overfitting refers to a model that has learned the training dataset too well.

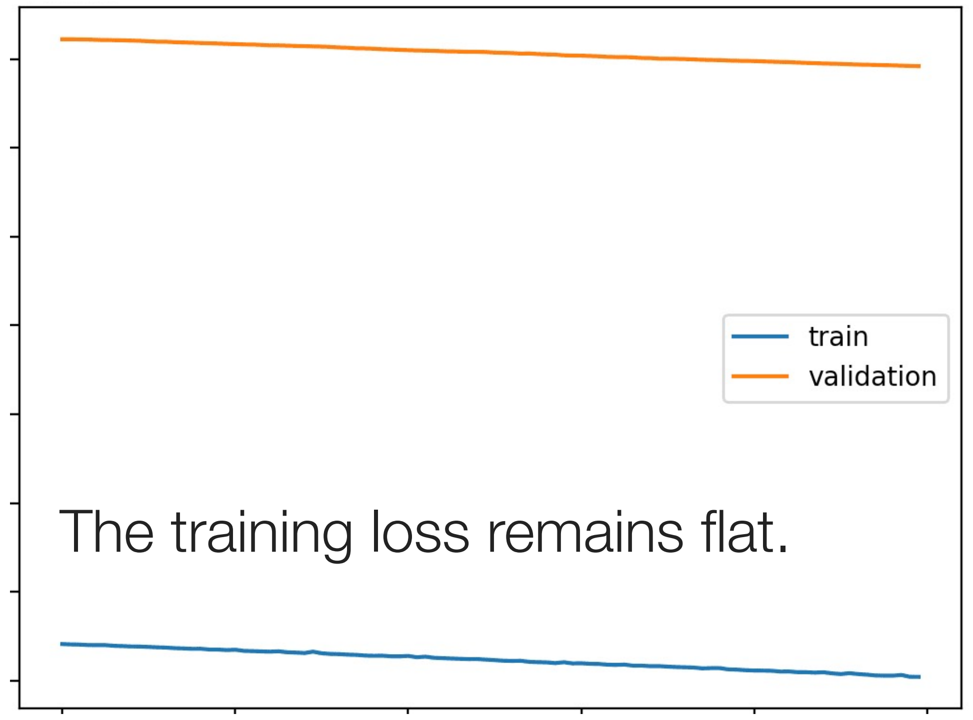


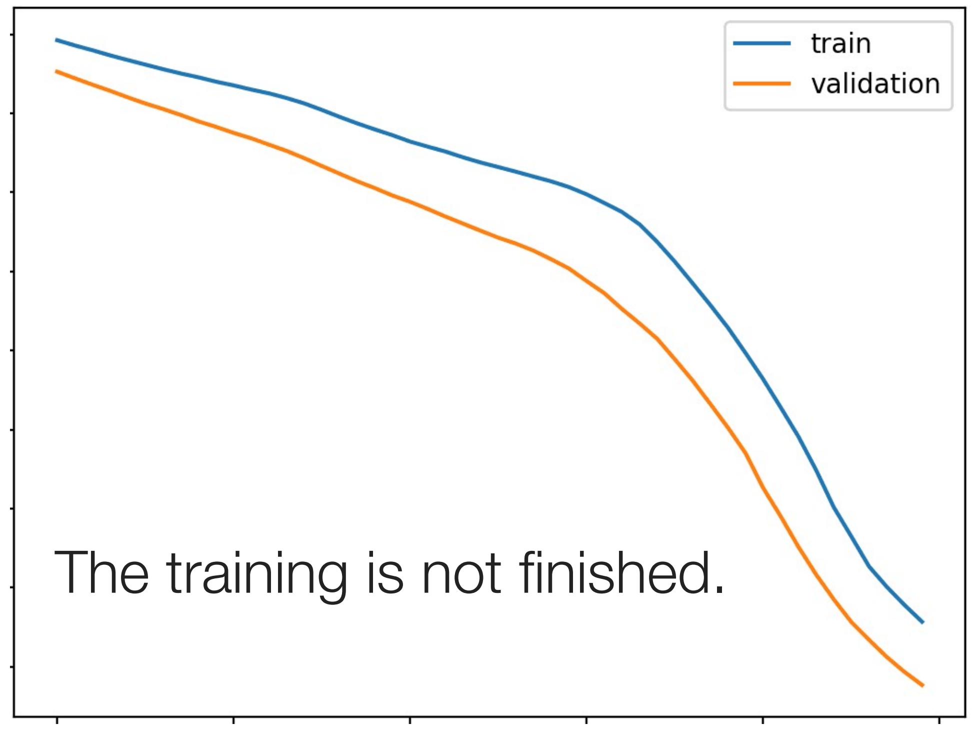
The validation loss decreases to a point and increases later.

The training loss continues to decrease with experience.

Underfitting

The model does not have a suitable capacity for the complexity of the dataset.

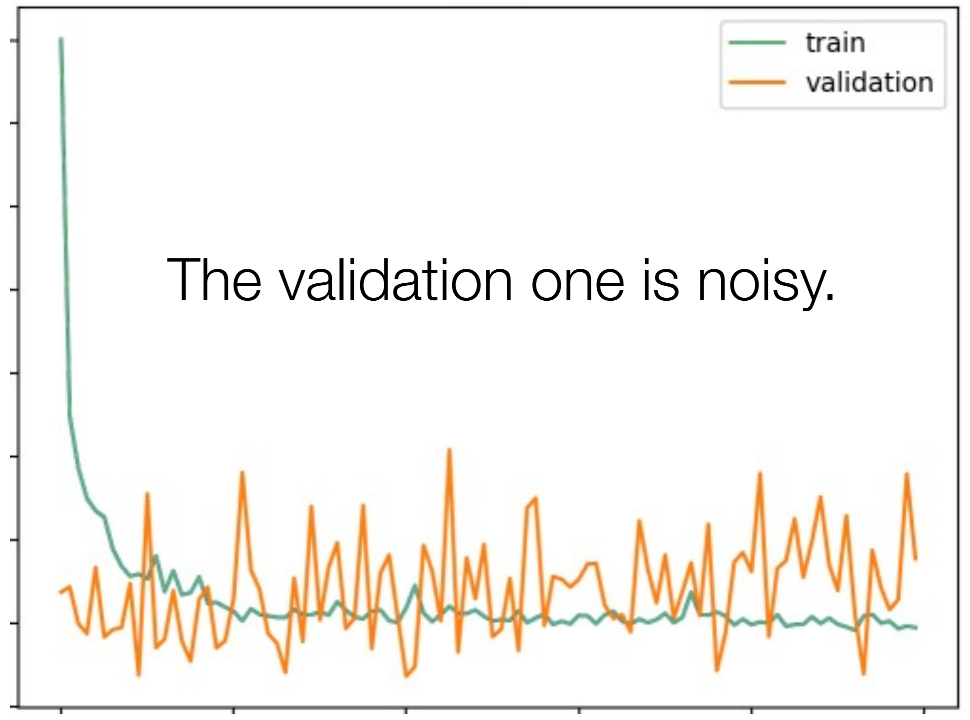


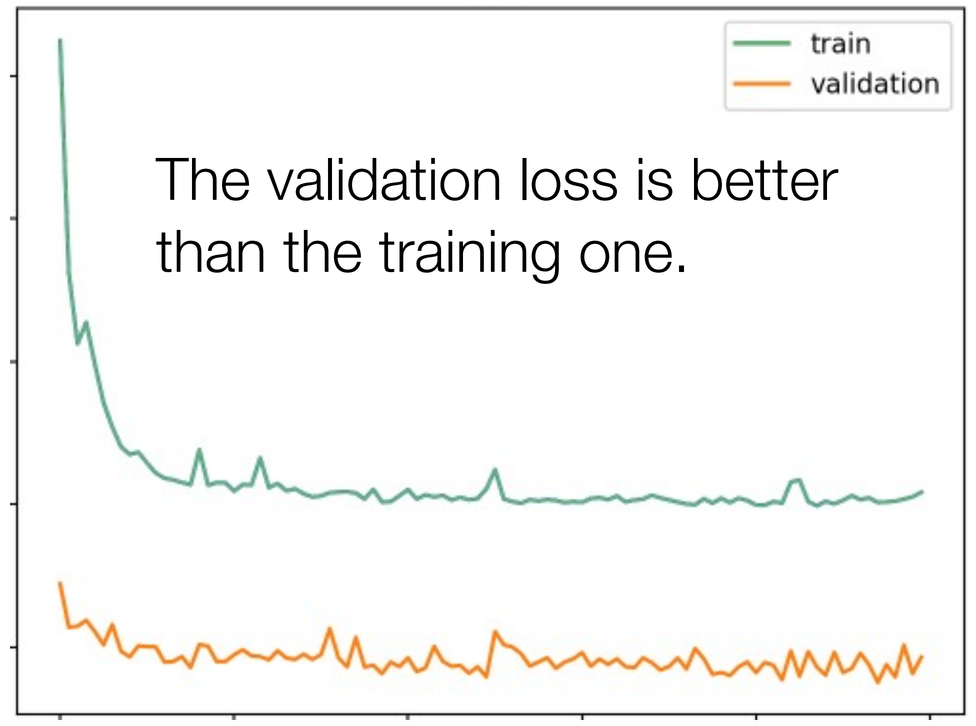


Observation of the training curve.

Unrepresentative validation dataset

The validation dataset does not provide enough information to be able to generalize.



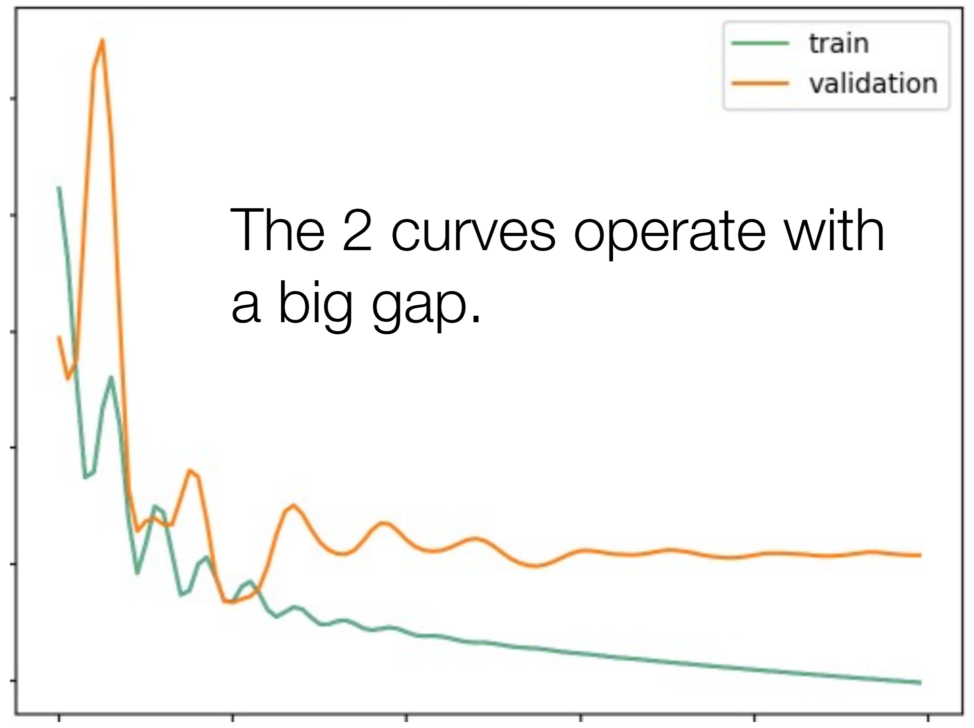


Different behaviour between training and validation.

the validation dataset is easier than the training.

Unrepresentative training dataset

The data available is not enough to capture the model, relative to the validation dataset.



Good Fit