Deep learning and bioimage analysis

Data Driven Life Sciences (DDLS) course

KTH - 2023

Estibaliz Gómez de Mariscal, EMBO Postdoctoral Fellow

Optical Cell Biology Group (Prof. Ricardo Henriques)

Instituto Gulbenkian de Ciencia

@gomez_mariscal

egomez@igc.gulbenkian.pt

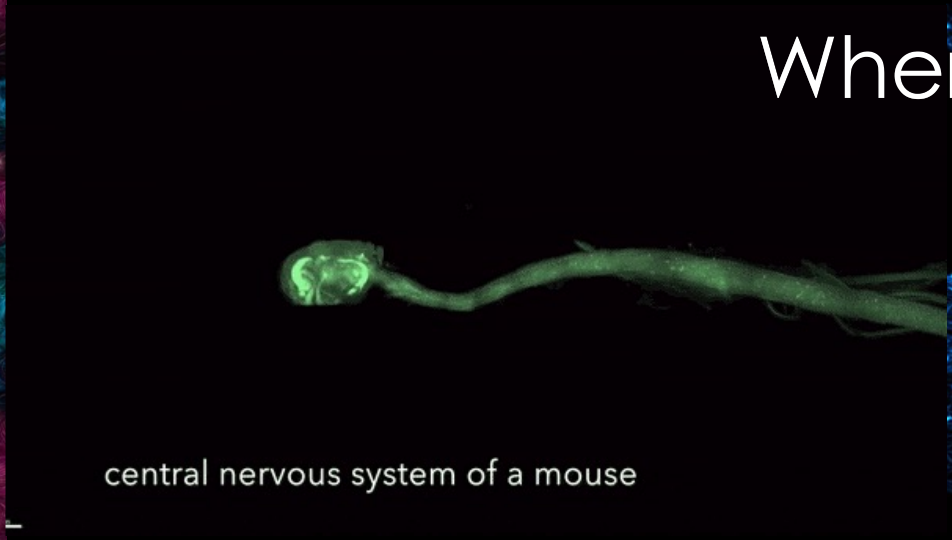
INSTITUTO GULBENKIAN DE CIÊNCIA

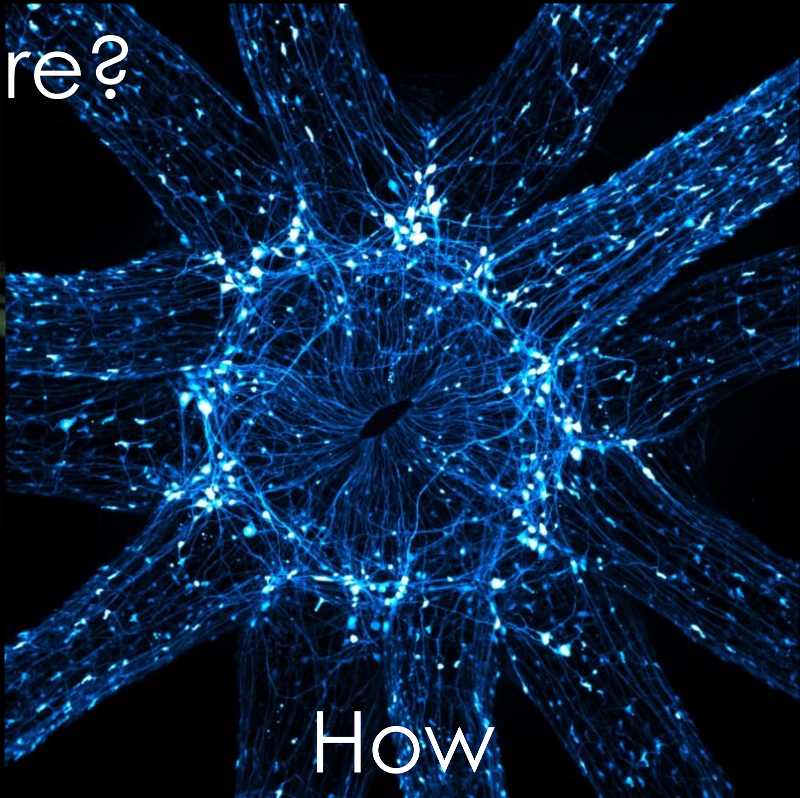
EMBO

HENRIQUES LAB

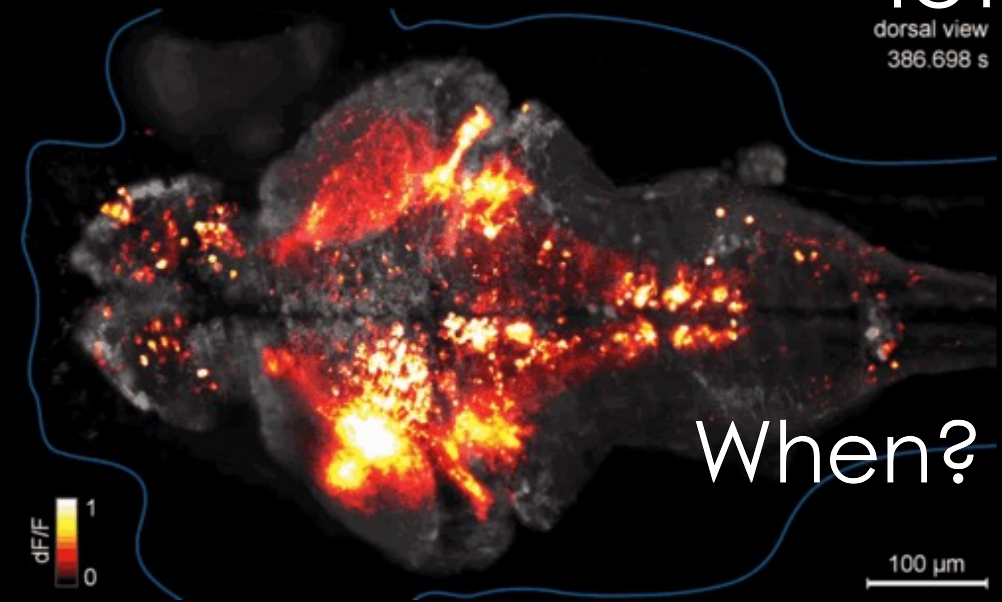
Today’s lecture

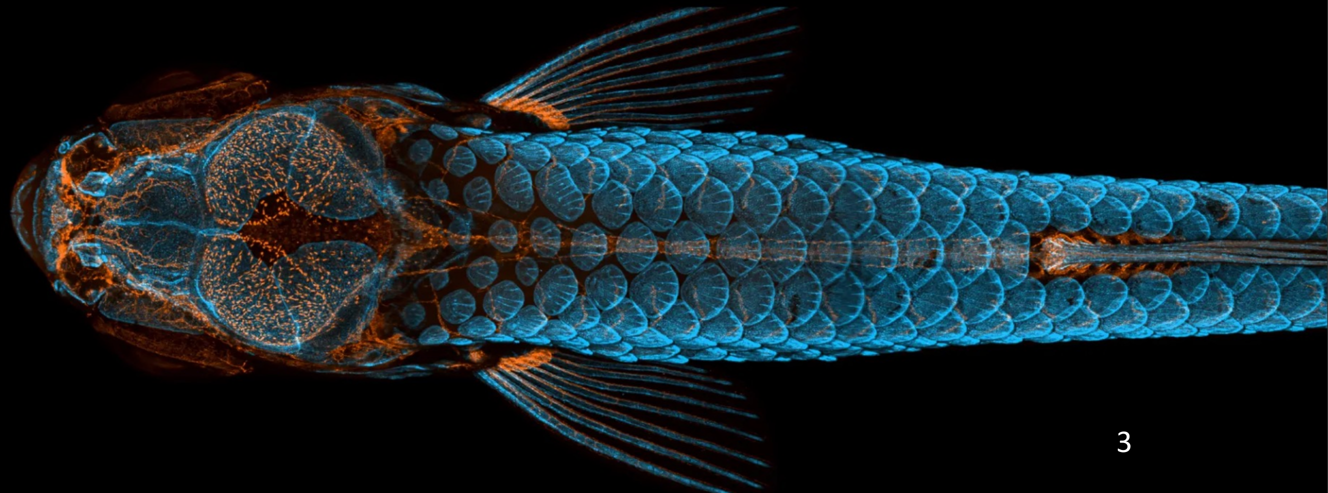
- BioImage analysis: definition
- Deep learning for image processing
- Segmentation
- Considerations about DL





Bioimages contain lot of information





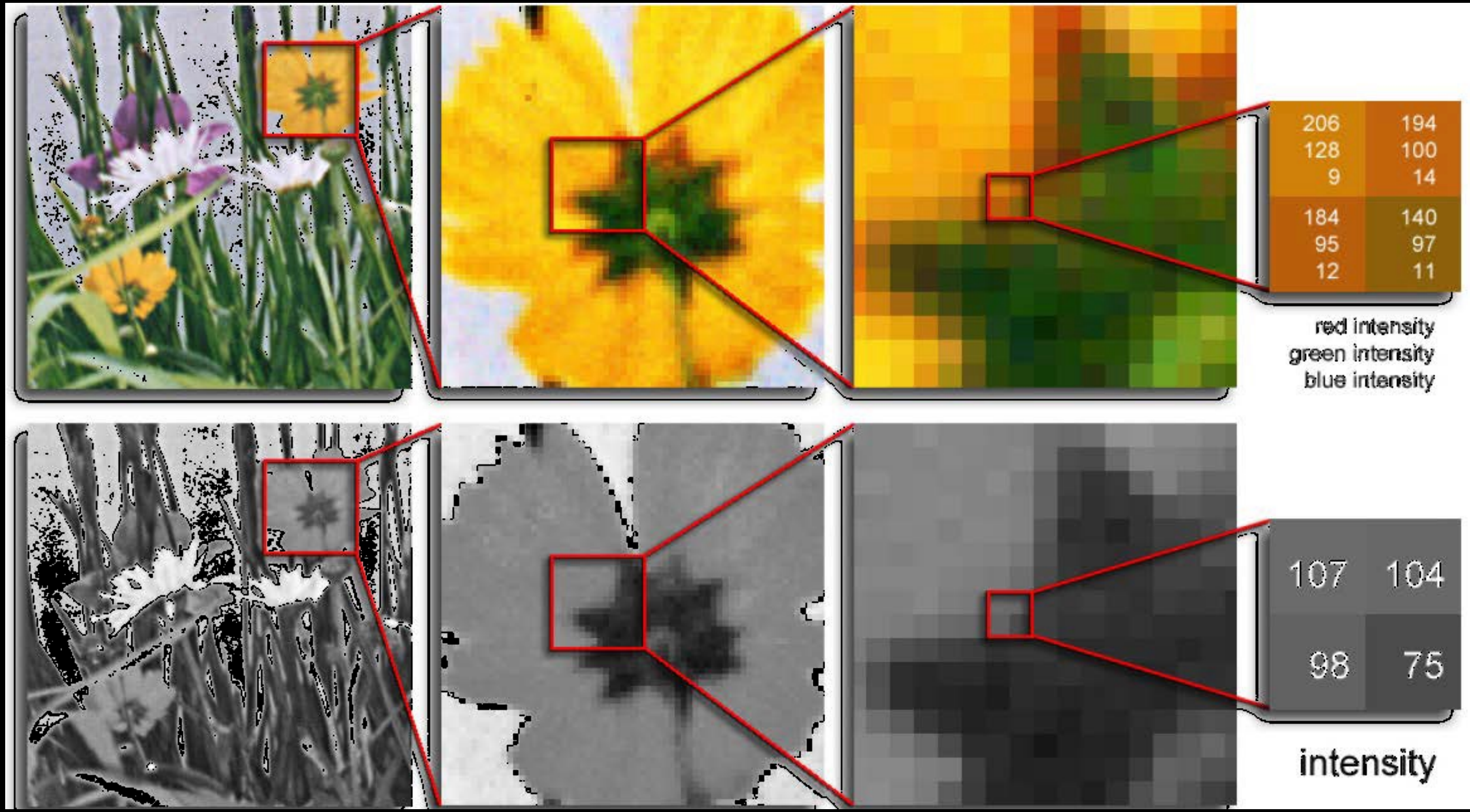
How many cells can we count?

Computational image processing:

- Precise
- Reproducible
- Transferable
- Automatic → FAST

Digital images

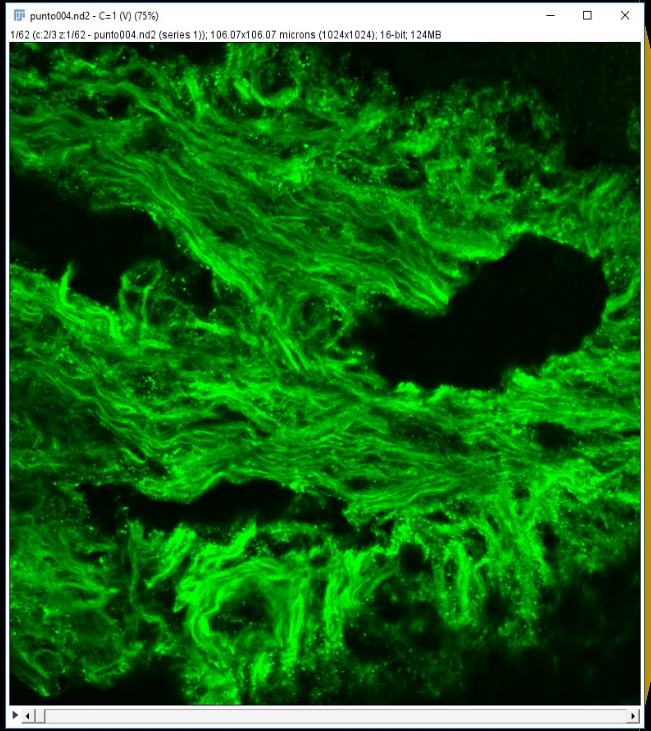
A **digital image** is a mapping of intensities from a 2D grid of (uniformly spaced) discrete points, into a set of numerical values. The grid elements are called **pixels**.

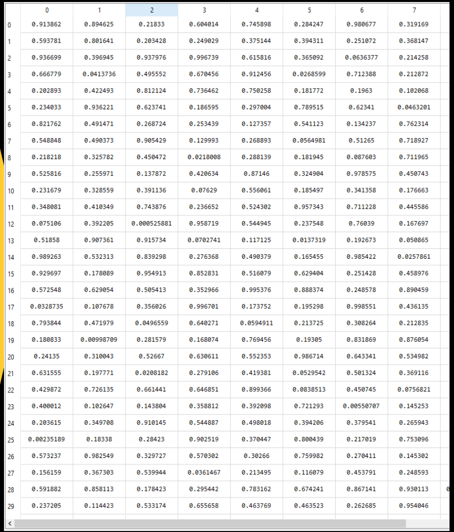


BioImages: biological information given by numbers

Raw data

Codifies the information contained in the image



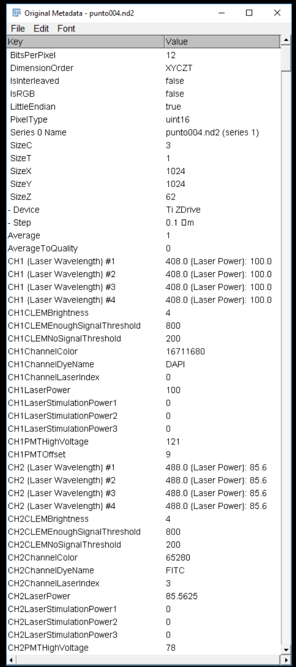


	0	1	2	3	4	5	6	7
0	0.913862	0.894625	0.21833	0.604014	0.745898	0.284247	0.980677	0.319169
1	0.593781	0.801641	0.203428	0.249029	0.375144	0.394311	0.251072	0.368147
2	0.936699	0.396945	0.937976	0.996739	0.615816	0.365092	0.0636377	0.214258
3	0.666779	0.0413736	0.495552	0.670456	0.912456	0.0268599	0.712388	0.212872
4	0.202893	0.422493	0.812124	0.736462	0.750258	0.181772	0.1963	0.102068
5	0.234033	0.936221	0.623741	0.186595	0.297004	0.789515	0.62341	0.0463201
6	0.821762	0.491471	0.268724	0.253439	0.127357	0.541123	0.134237	0.762314
7	0.548848	0.490373	0.905429	0.129993	0.268893	0.0564981	0.51265	0.718927
8	0.218218	0.325782	0.450472	0.0218008	0.288139	0.181945	0.087603	0.711965
9	0.525816	0.255971	0.137872	0.420634	0.87146	0.324904	0.978575	0.450743
10	0.231679	0.328559	0.391136	0.07629	0.556061	0.185497	0.341358	0.176663
11	0.348081	0.410349	0.743876	0.236652	0.524302	0.957343	0.711228	0.445586
12	0.075106	0.392205	0.000525881	0.958719	0.544945	0.237548	0.76039	0.167697
13	0.51858	0.907361	0.915734	0.0702741	0.117125	0.0137319	0.192673	0.050865
14	0.989263	0.532313	0.839298	0.276368	0.490379	0.165455	0.985422	0.0257861
15	0.929697	0.178089	0.954913	0.852831	0.516079	0.629404	0.251428	0.458976
16	0.572548	0.629054	0.505413	0.352966	0.995376	0.888374	0.248578	0.890459
17	0.0328735	0.107678	0.356026	0.996701	0.173752	0.195298	0.998551	0.436135
18	0.793844	0.471979	0.0496559	0.640271	0.0594911	0.213725	0.308264	0.212835
19	0.180833	0.00998709	0.281579	0.168074	0.769456	0.19305	0.831869	0.876054
20	0.24135	0.310043	0.52667	0.630611	0.552353	0.986714	0.643341	0.534982
21	0.631555	0.197771	0.0208182	0.279106	0.419381	0.0529542	0.501324	0.369116
22	0.429872	0.726135	0.661441	0.646851	0.899366	0.0838513	0.450745	0.0756821
23	0.400012	0.102647	0.143804	0.358812	0.392098	0.721293	0.00550707	0.145253
24	0.203615	0.349708	0.910145	0.544887	0.498018	0.394206	0.379541	0.265943
25	0.00235189	0.18338	0.28423	0.902519	0.370447	0.800439	0.217019	0.753096
26	0.573237	0.982549	0.329727	0.570302	0.30266	0.759982	0.270411	0.145302
27	0.156159	0.367303	0.539944	0.0361467	0.213495	0.116079	0.453791	0.248593
28	0.591882	0.858113	0.178423	0.295442	0.783162	0.674241	0.867141	0.930113
29	0.237205	0.114423	0.533174	0.655658	0.463769	0.463523	0.262685	0.954046

Metadata

Set of text data providing additional information about the image.

- Imaging modality
- Objective
- Magnification
- Resolution or Pixel/Voxel size (microns, mm)
- Number of channels
- Excitation spectrum
- Information about the patient



Original Metadata - punto004.nd2

File Edit Font

Key	Value
BitsPerPixel	12
DimensionOrder	XYCZT
IsInterleaved	false
IsRGB	false
LittleEndian	true
PixelType	uint16
Series 0 Name	punto004.nd2 (series 1)
SizeC	3
SizeT	1
SizeX	1024
SizeY	1024
SizeZ	62
- Device	Ti ZDrive
- Step	0.1 □m
Average	1
AverageToQuality	0
CH1 {Laser Wavelength} #1	408.0 {Laser Power}: 100.0
CH1 {Laser Wavelength} #2	408.0 {Laser Power}: 100.0
CH1 {Laser Wavelength} #3	408.0 {Laser Power}: 100.0
CH1 {Laser Wavelength} #4	408.0 {Laser Power}: 100.0
CH1CLEMBrightness	4
CH1CLEMEnoughSignalThreshold	800
CH1CLEMNoSignalThreshold	200
CH1ChannelColor	16711680
CH1ChannelDyeName	DAPI
CH1ChannelLaserIndex	0
CH1LaserPower	100
CH1LaserStimulationPower1	0
CH1LaserStimulationPower2	0
CH1LaserStimulationPower3	0
CH1PMTHighVoltage	121
CH1PMTOffset	9
CH2 {Laser Wavelength} #1	488.0 {Laser Power}: 85.6
CH2 {Laser Wavelength} #2	488.0 {Laser Power}: 85.6
CH2 {Laser Wavelength} #3	488.0 {Laser Power}: 85.6
CH2 {Laser Wavelength} #4	488.0 {Laser Power}: 85.6
CH2CLEMBrightness	4
CH2CLEMEnoughSignalThreshold	800
CH2CLEMNoSignalThreshold	200
CH2ChannelColor	65280
CH2ChannelDyeName	FITC
CH2ChannelLaserIndex	3
CH2LaserPower	85.5625
CH2LaserStimulationPower1	0
CH2LaserStimulationPower2	0
CH2LaserStimulationPower3	0
CH2PMTHighVoltage	78

BIO-FORMATS

The solution for reading proprietary microscopy image data and metadata

Digital (bio)-image analysis

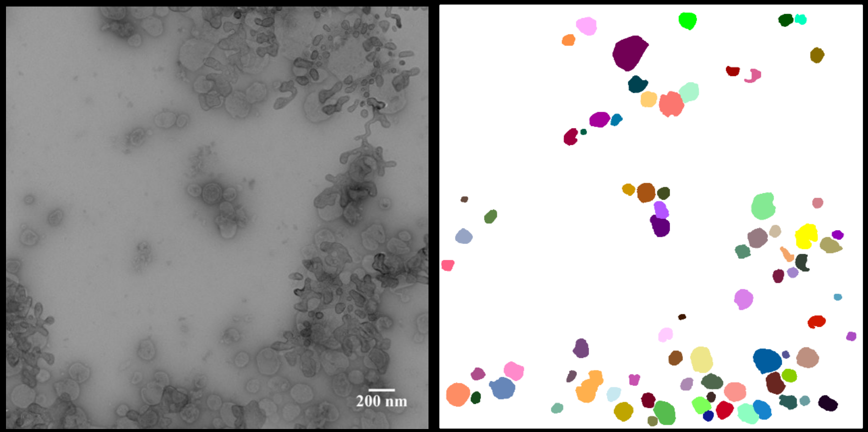
🔬🧫🧪🦠💻📈

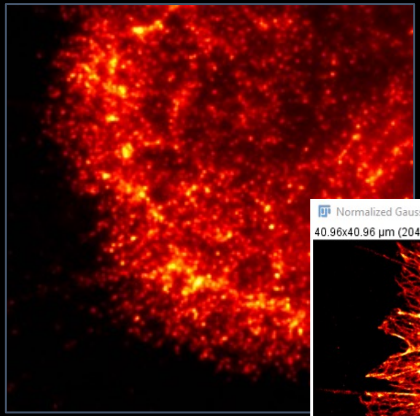
Image processing

Image Processing is any form of data processing for which the input is an image – the output is not necessarily an image.

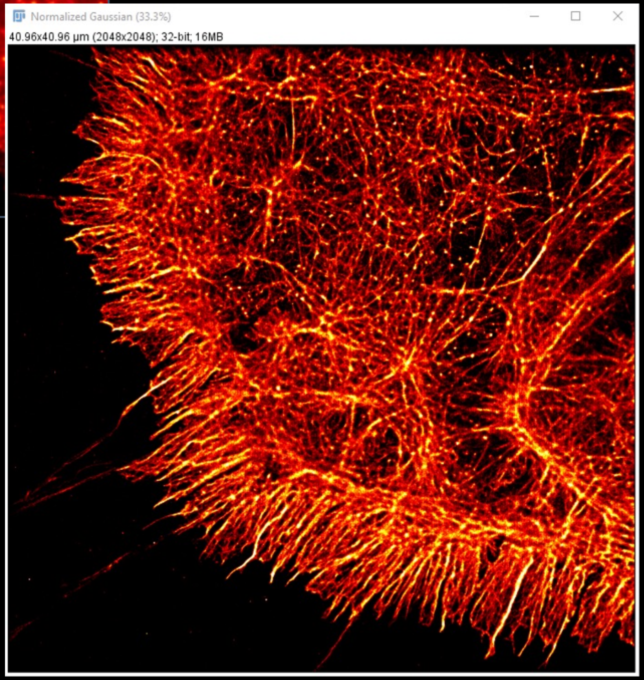
Input: Image - Output: Coordinates (bounding boxes)

Input: Image
Output: Image

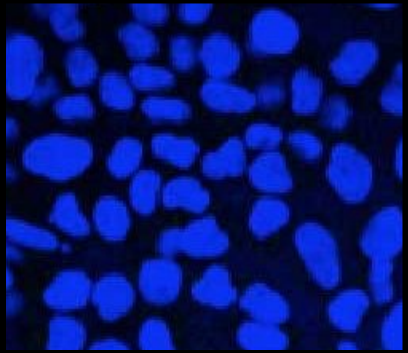




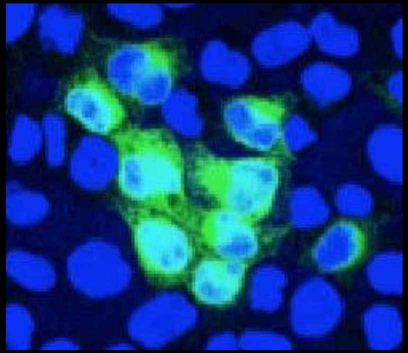
Input: Image
Output: Image



Input: Image
Output: Label



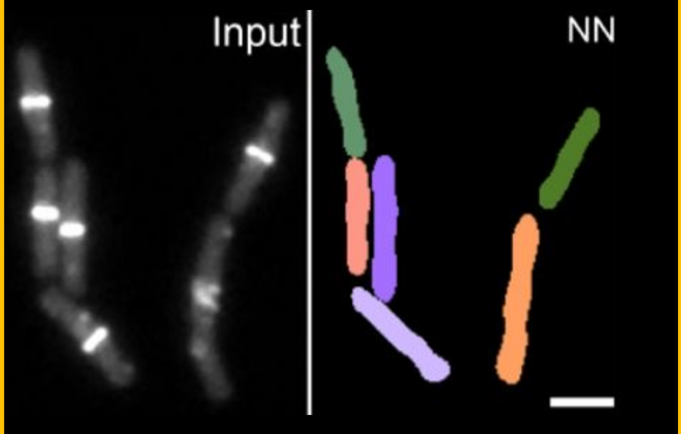
Non infected



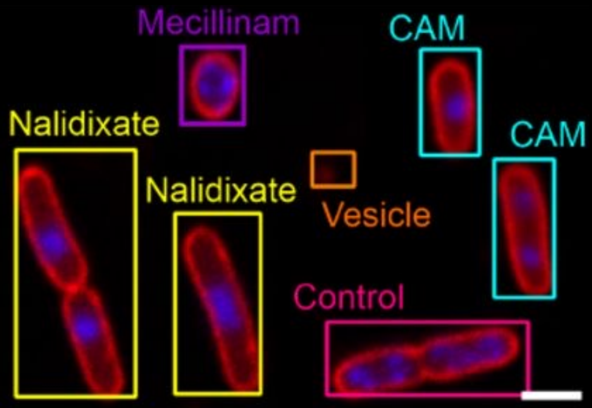
Infected

(*Classical*) image processing tasks

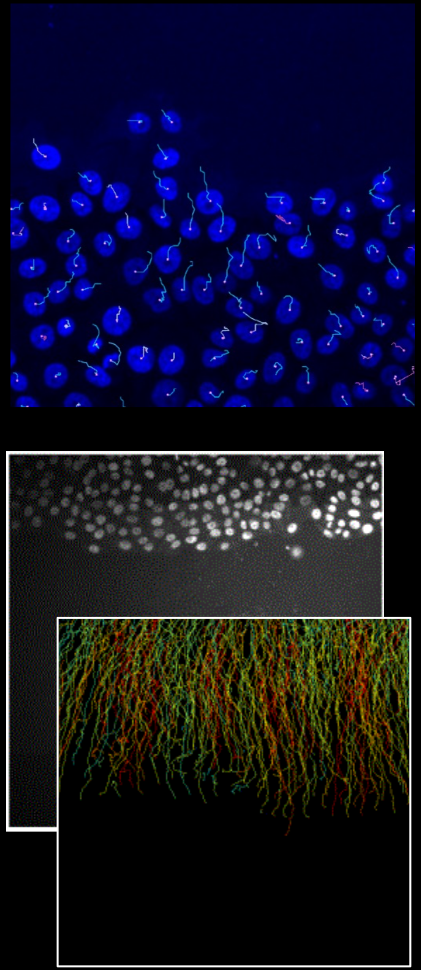
(1) Segmentation



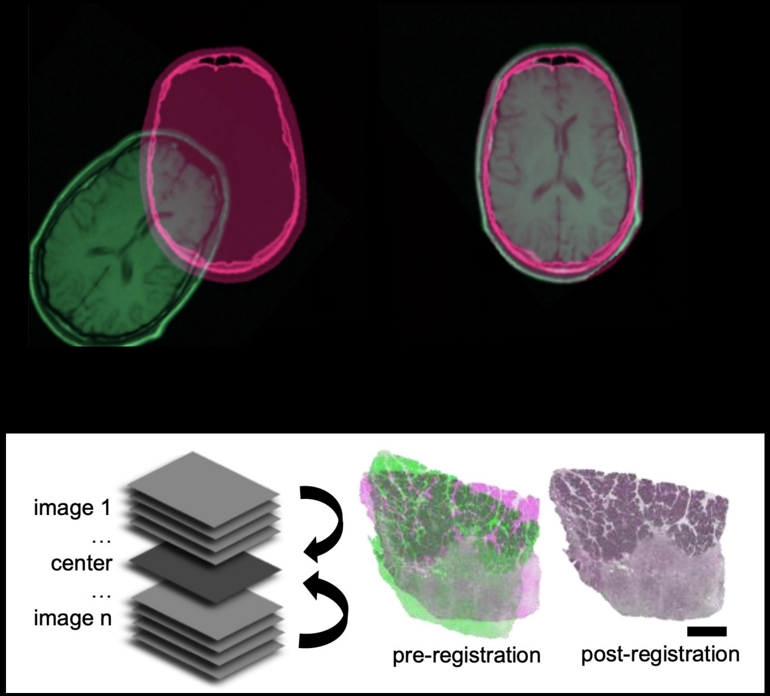
(2) Object detection



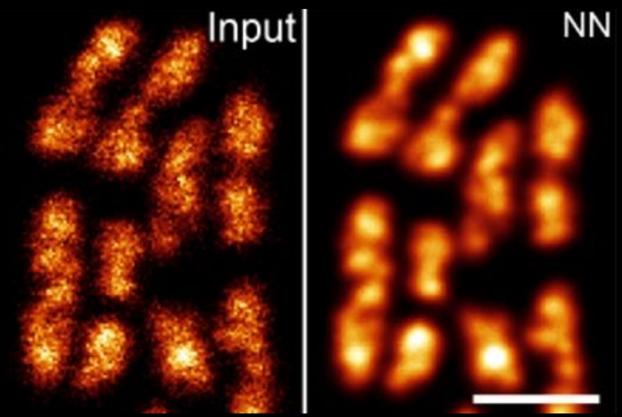
(3) Object tracking



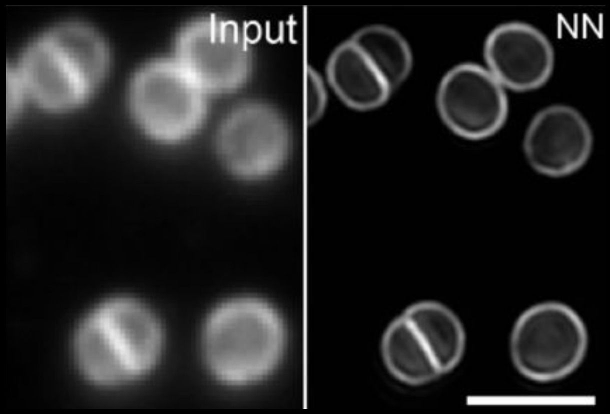
(4) Image registration



(5) Denoising



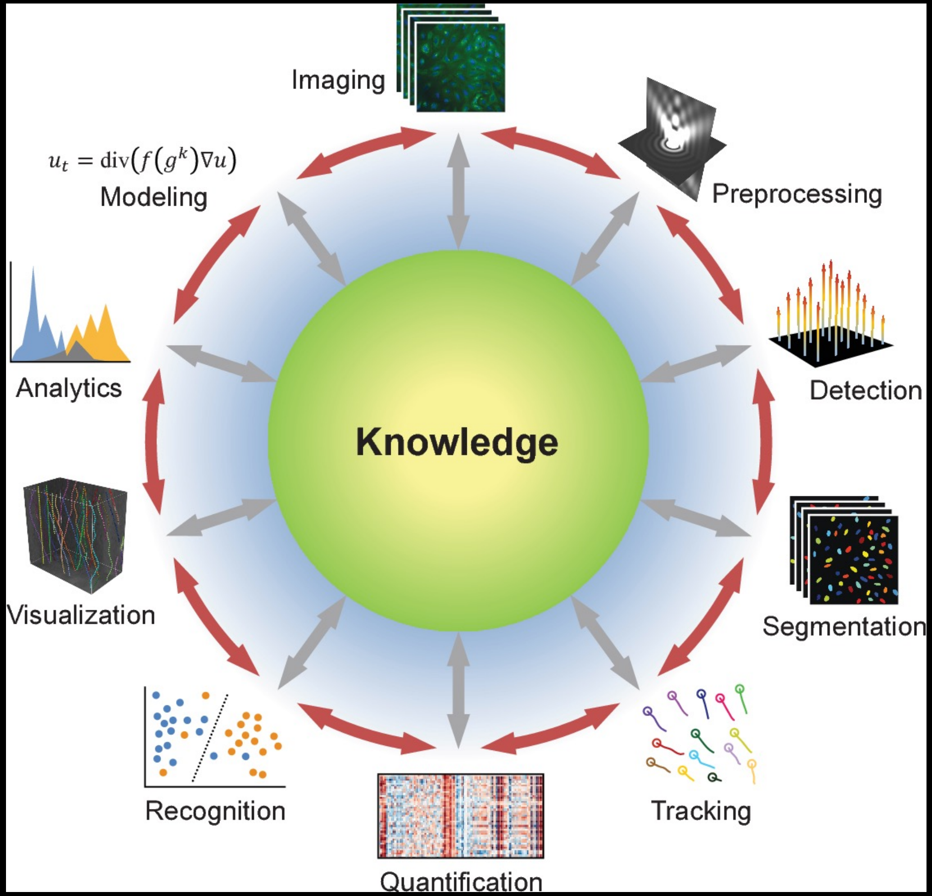
(6) Super-resolution

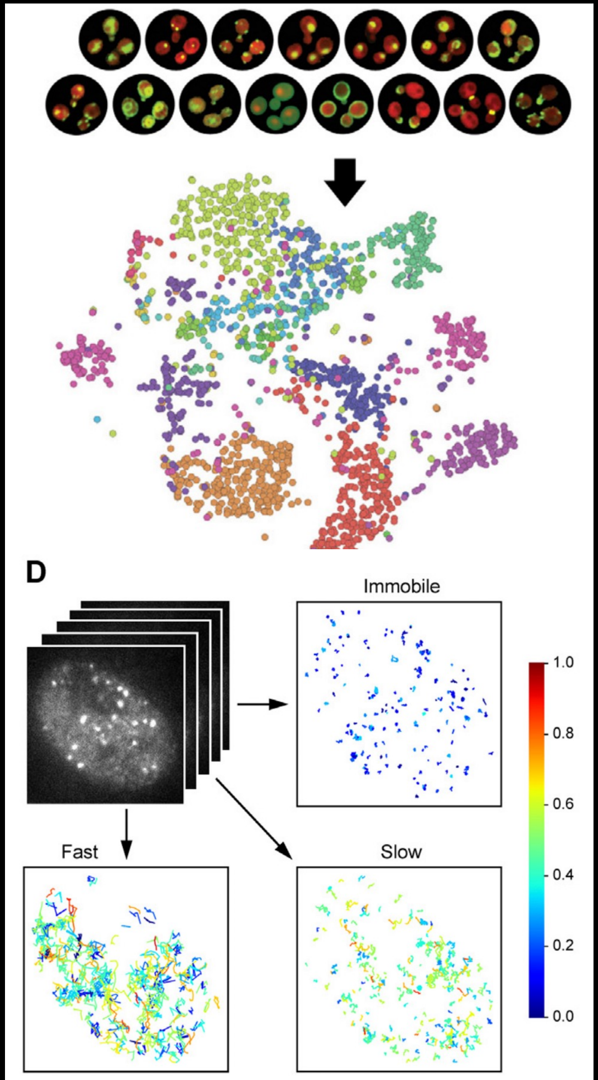


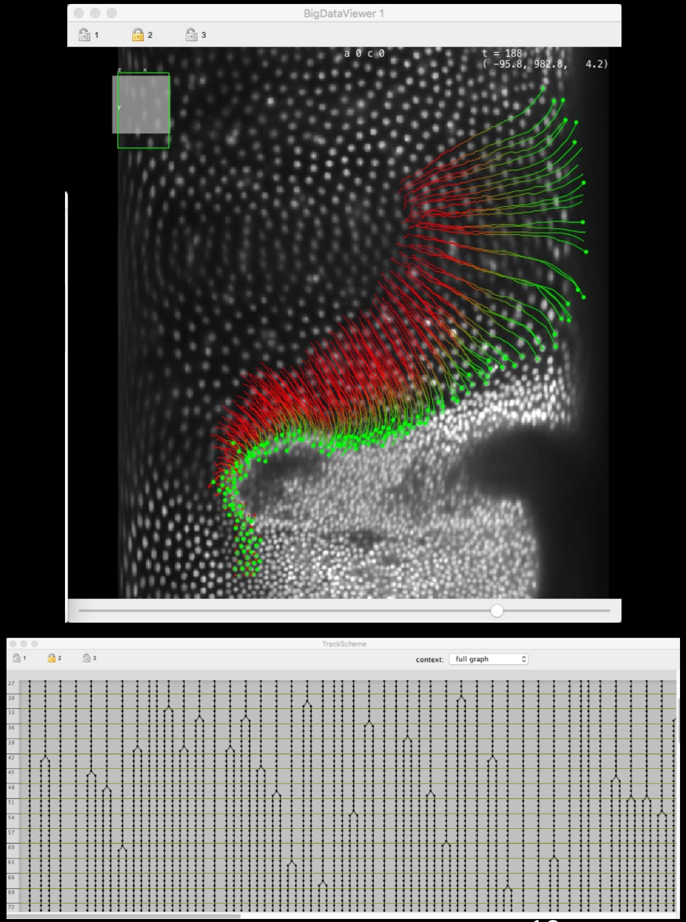
Christoph Spahn, et al., bioRxiv, 2021
Elnaz Fazeli, et al., F1000Research 2020
Ashley Kiemen, et al., bioRxiv 2020

BioImage analysis

Collection of image processing techniques to extract numerical information from scientific images







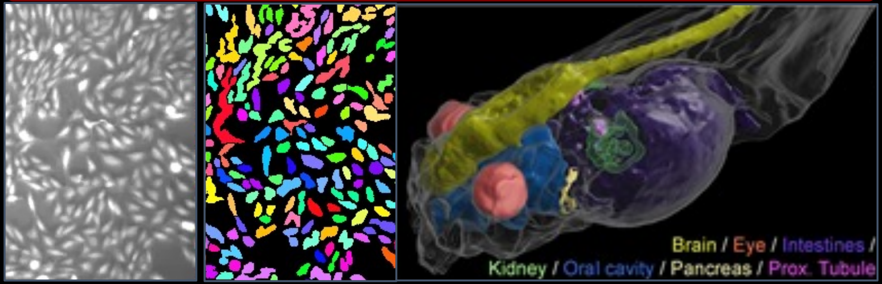
E. Meijering et al., Nature Biotechnology 2016
E. Meijering, 2020
Vladimir Ulman et al., Nature Methods, 2017

Deep learning:

an extremely hot topic in the field 🔥

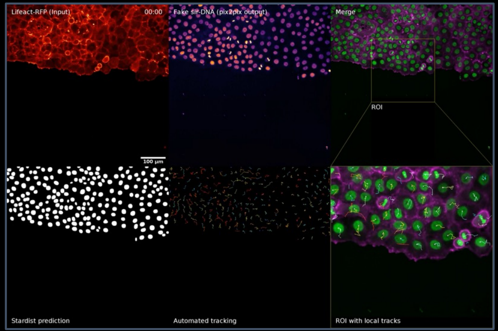
The deep learning landscape for microscopy imaging

Segmentation



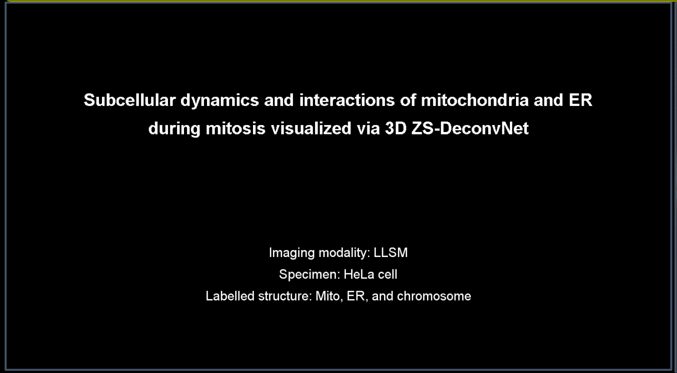
F. Lux & P. Matula, arXiv, 2020 Naert et al., Development 2021

Artificial labelling

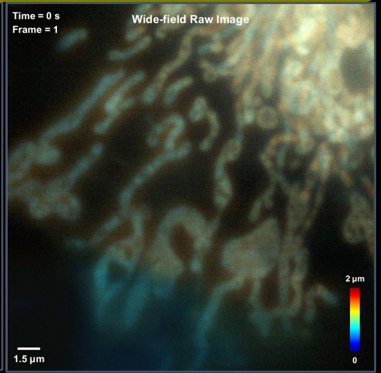


L. Von Chamier, …, R. Henriques, Nat Comms 2021

Resolution enhancement & restoration

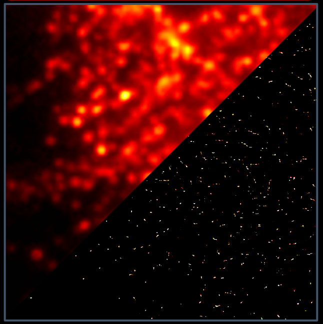


C. Qiao et al., bioRxiv 2023



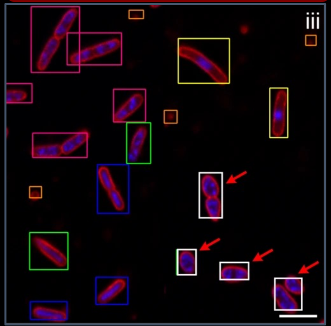
C. Qiao et al., Nat Biotech 2022

SMLM



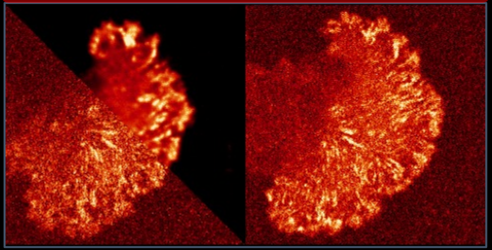
E. Nehme et al., Optica, 2018

Detection



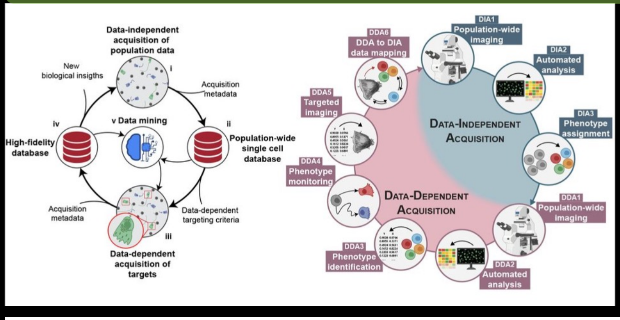
C. Spahn et al., Comm Biology 2022

Denoising

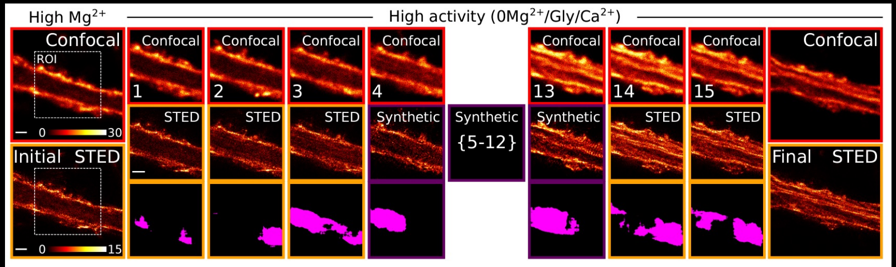


L. Von Chamier, …, R. Henriques, Nature Communications 2021

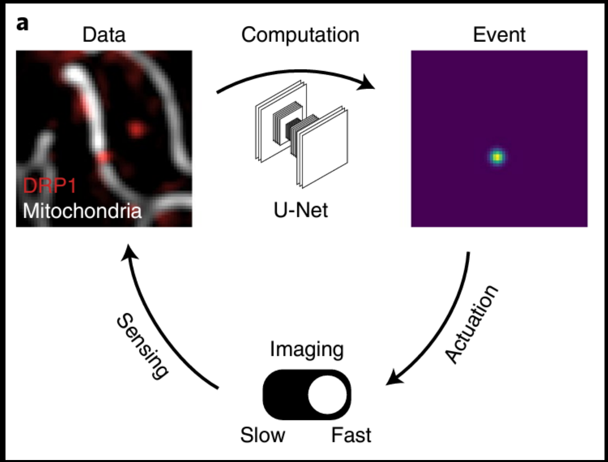
Data driven microscopy



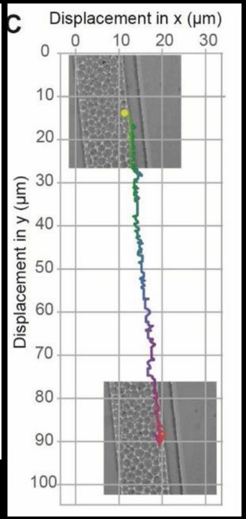
O. André et al., Cell reports 2023



C.Bouchard et al., bioRxiv 2023



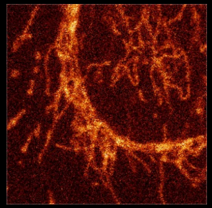
D. Maheic et al., Nat Methods 2022

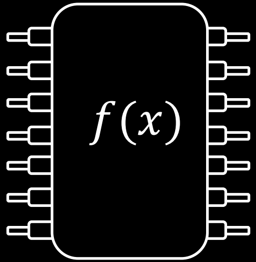


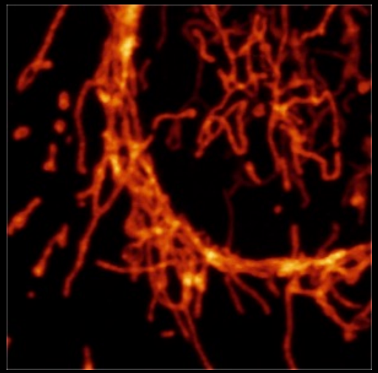
L. Chiron et al., Sci Reports 2022

Why does deep learning pose a new paradigm?

Classical algorithm







Round shape

Red

Smooth texture

Machine learning method

"tomato"

Round shape

Red

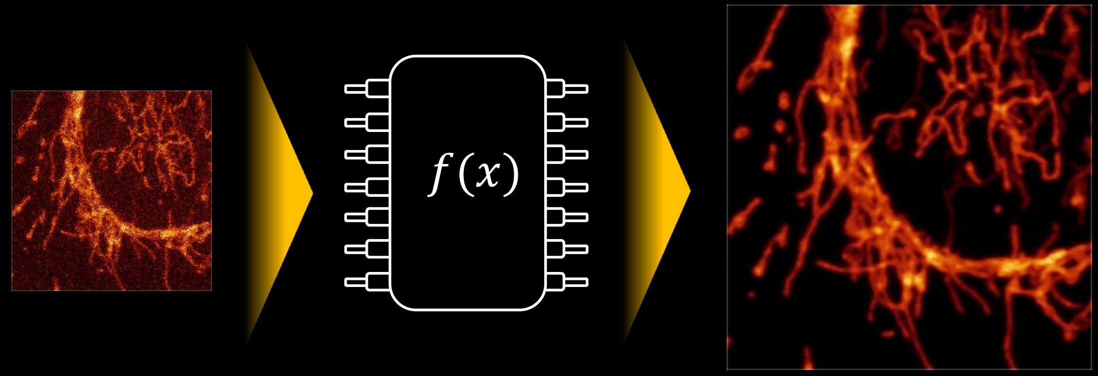
Smooth texture

Machine learning method

"tomato"

Why does deep learning pose a new paradigm?

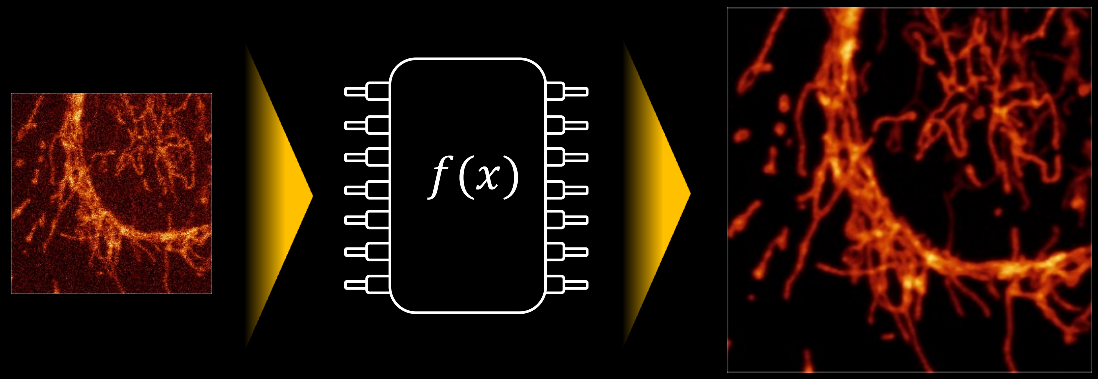
Classical algorithm

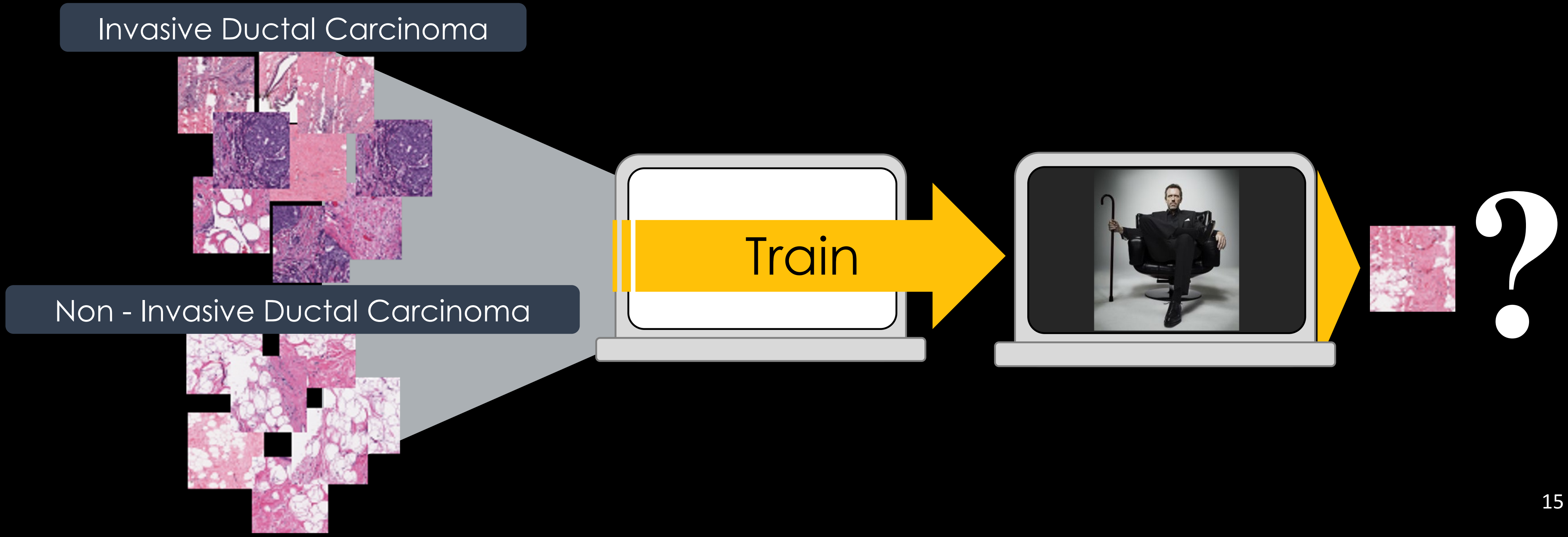


What if the system could learn **automatically** from the **data**?

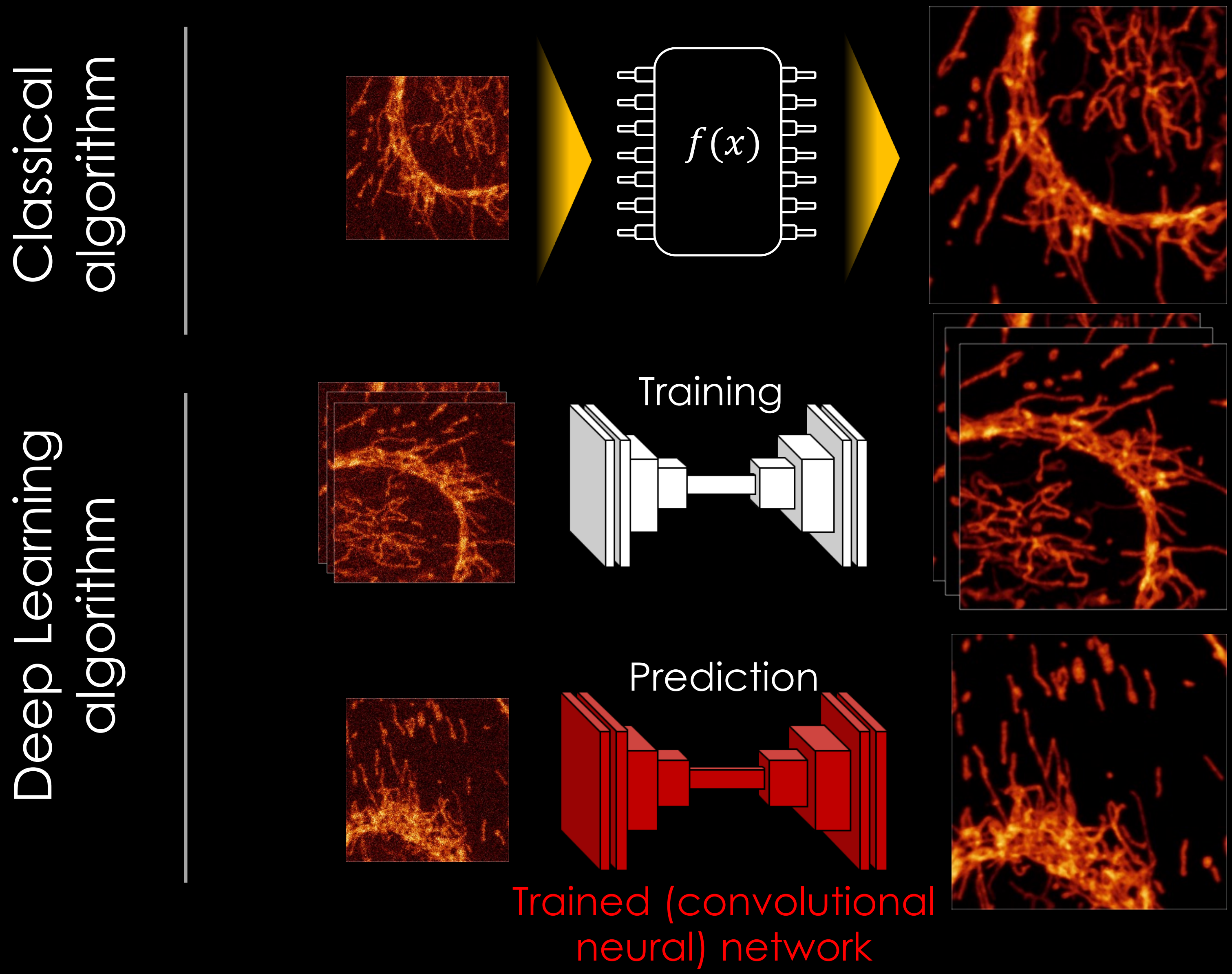
Why does deep learning pose a new paradigm?

Classical algorithm



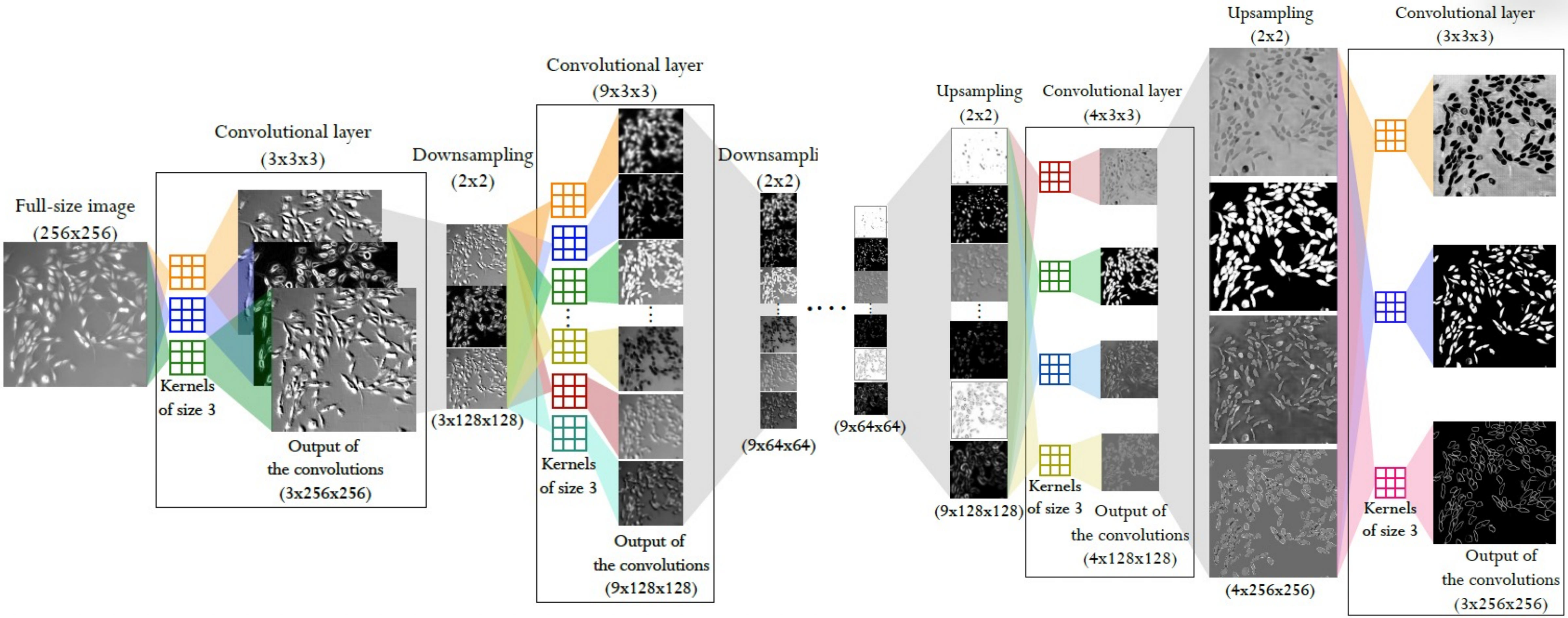


Why does deep learning pose a new paradigm?



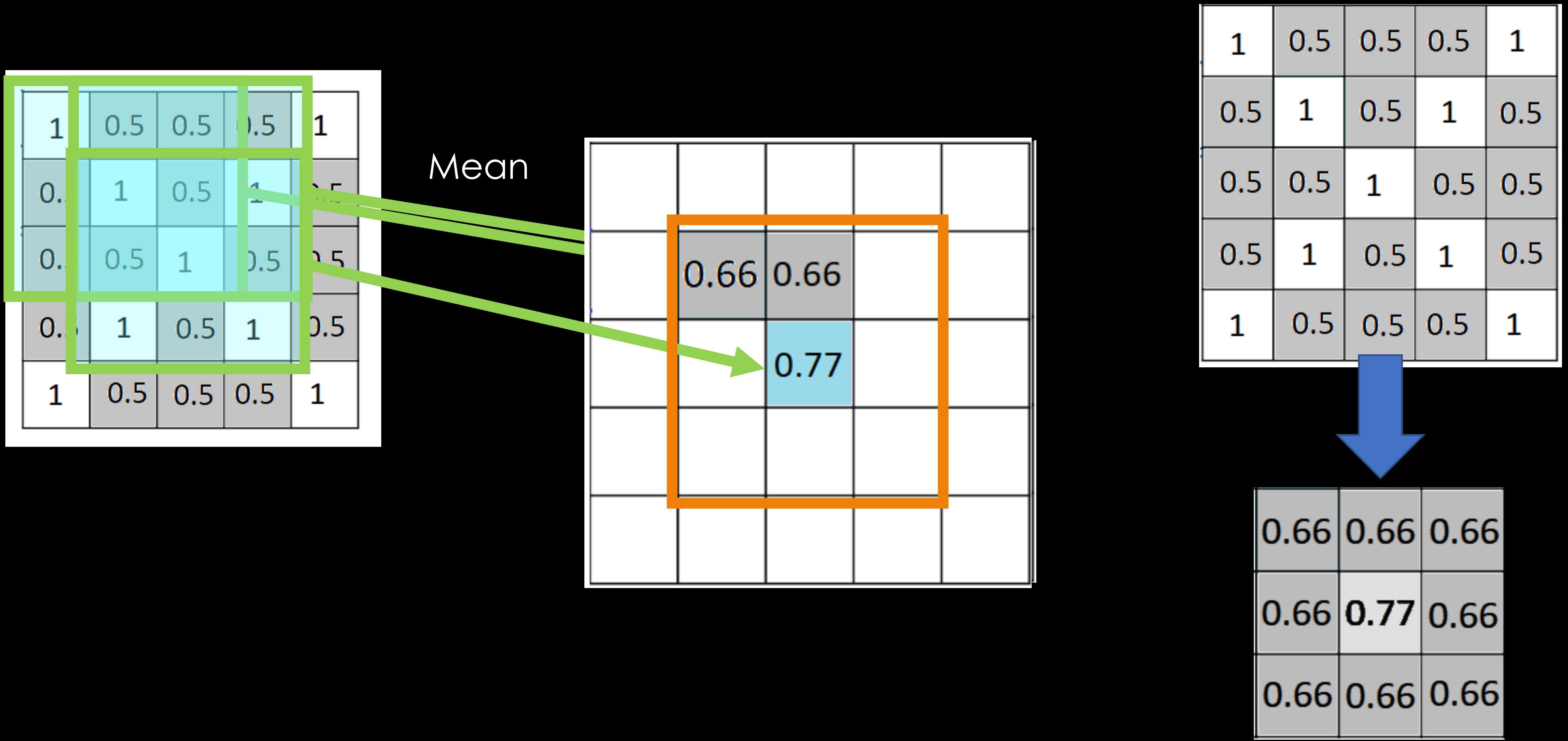
What are (convolutional) neural networks?

Image segmentation convolutional neural network arquitecture (2D U-Net)

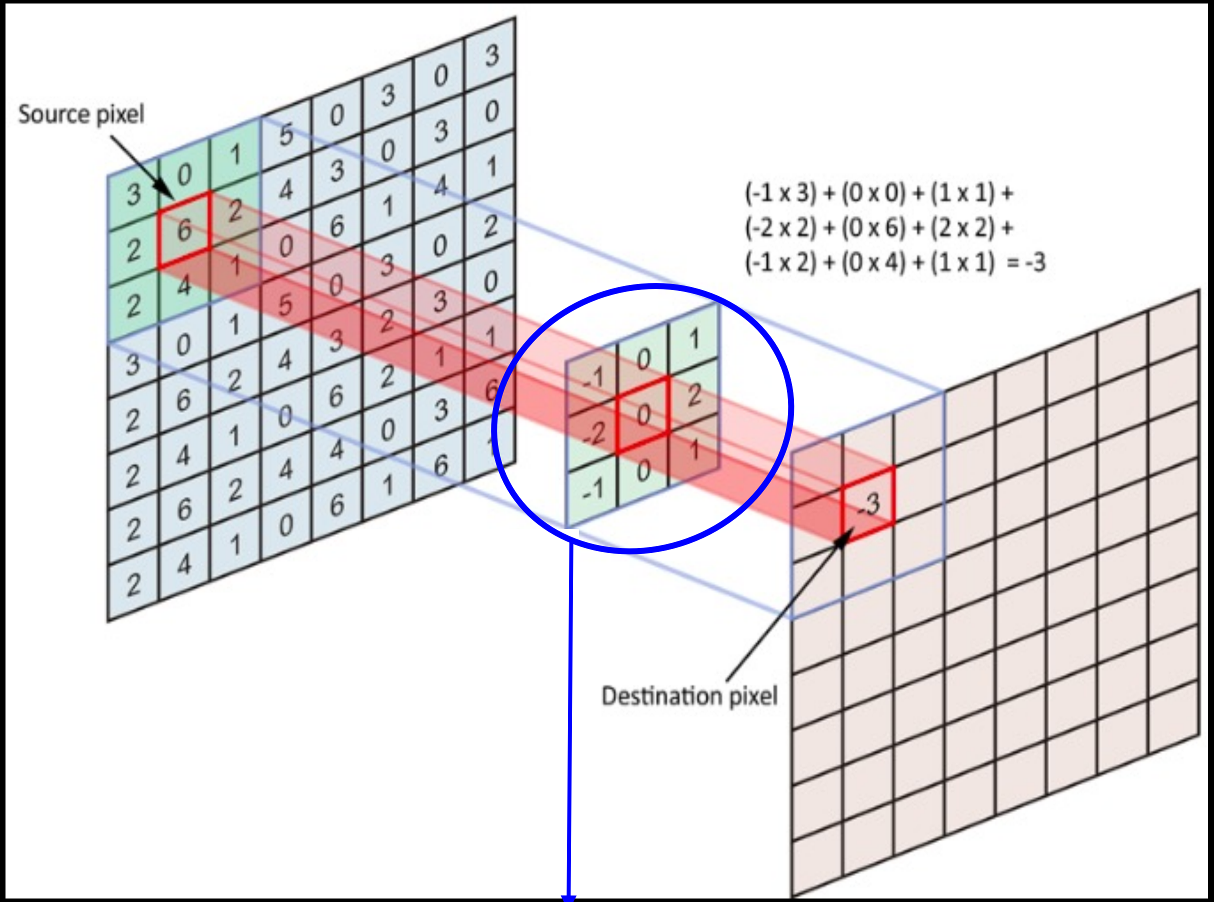


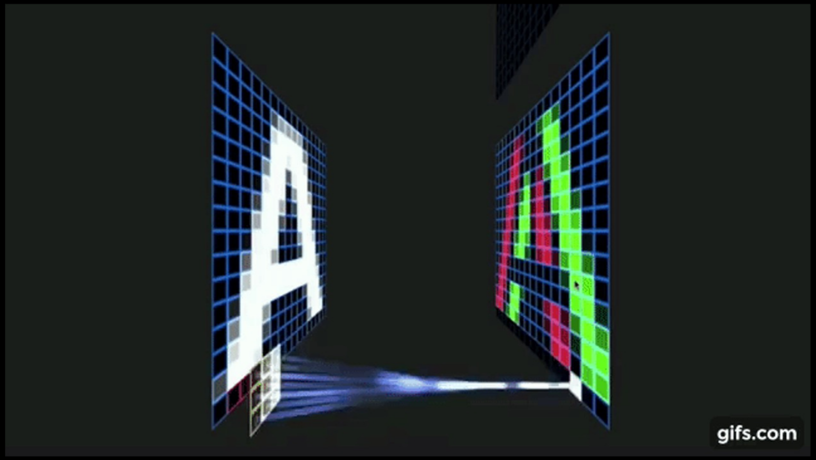
Gómez de Mariscal, E. et al., NEUBIAS Springer, 2022
Olaf Ronneberger, Philipp Fischer, Thomas Brox; arXiv 2015

Convolutions



Convolutions





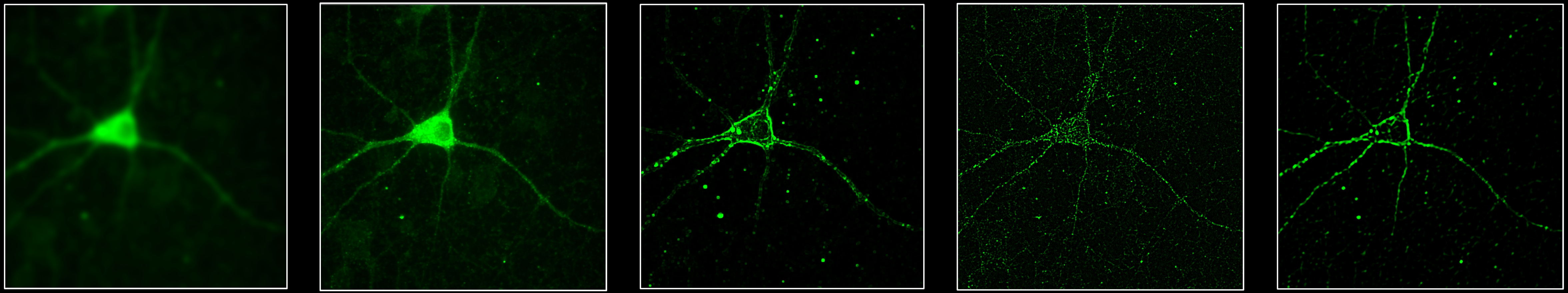
Contextual information around the pixel

Spatial filtering

Convolutional kernel: determines de feature to enhance

Convolutions

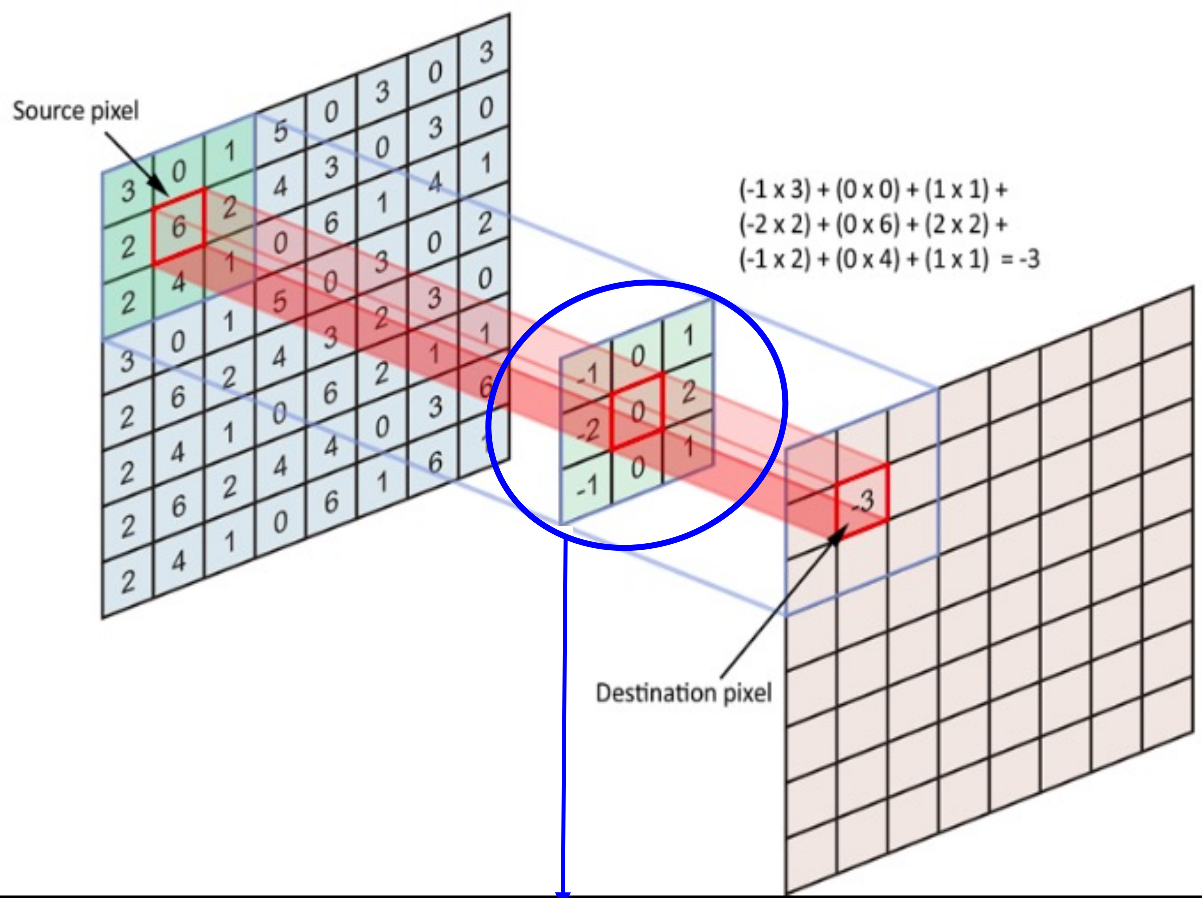
Capacity to quantify and enhance features of interest in the image



Filtered images and image filters can be combined in multiple ways

What if the system could learn the optimal combinations **automatically** from the **data**?

What are (convolutional) neural networks?



Convolutional kernel: determines de feature to enhance

Trainable filter (convolutional layer)

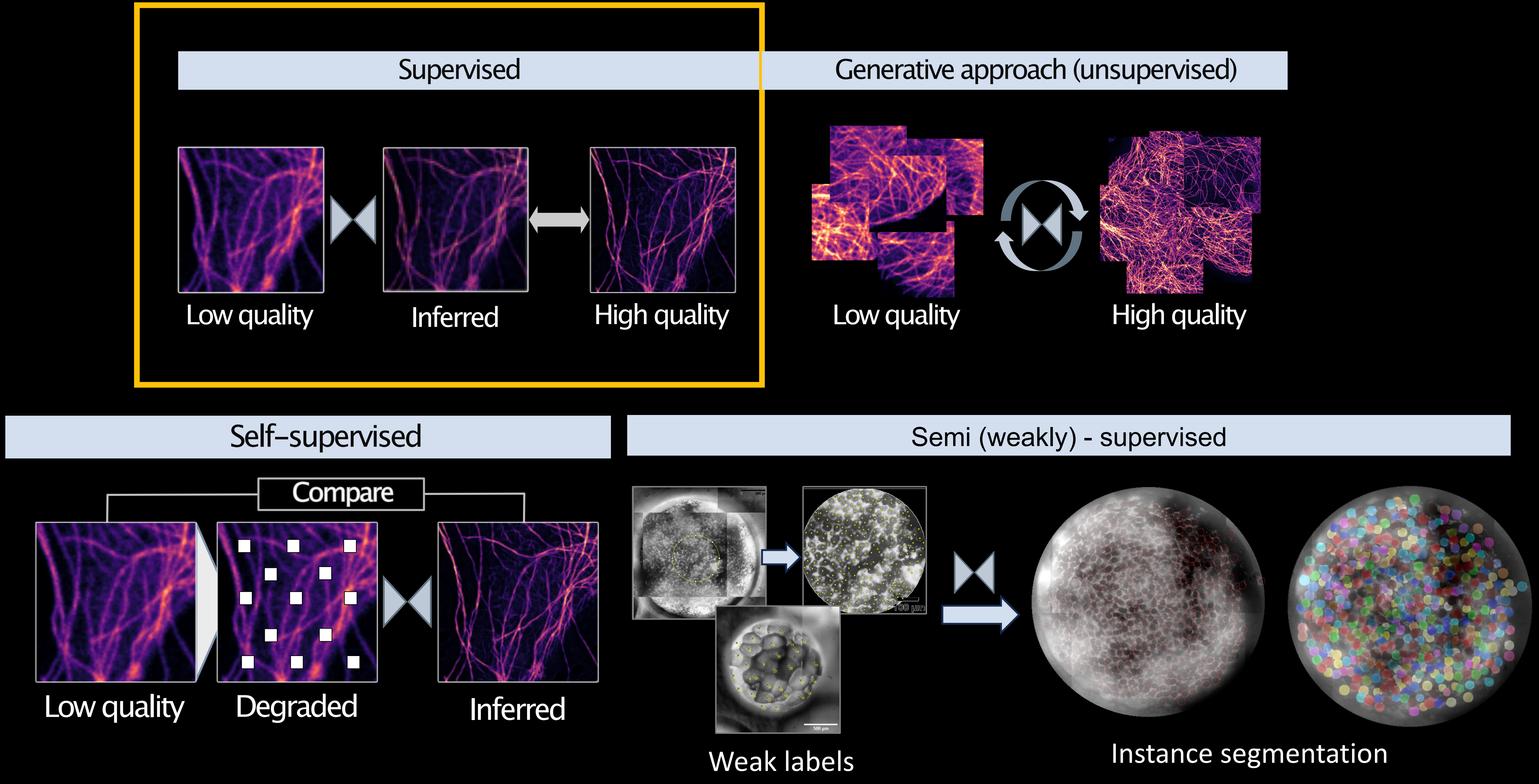
W_{00}	$W_{01}b$	W_{02}
W_{10}	W_{11}	W_{12}
W_{20}	W_{21}	W_{22}

Note: Each unit in the kernel will have its weight, but each convolutional filter will also have a bias:

$C(x) = W \otimes x + b$,

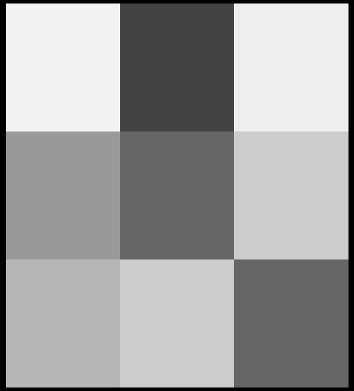
where x is an input image

The learning process

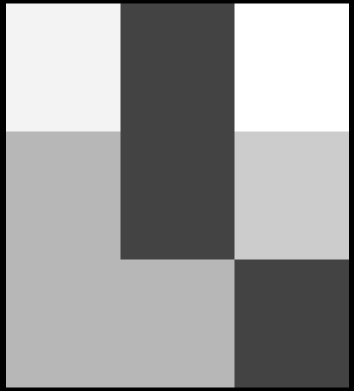


E. Gómez-de-Mariscal, et al. "Harnessing Artificial Intelligence To Reduce Phototoxicity in Live Imaging." *arXiv* (2023).

Supervised CNNs training



$$* \begin{bmatrix} W_{00} & W_{01} & W_{02} \\ W_{10} & W_{11} & W_{12} \\ W_{20} & W_{21} & W_{22} \end{bmatrix} + b =$$



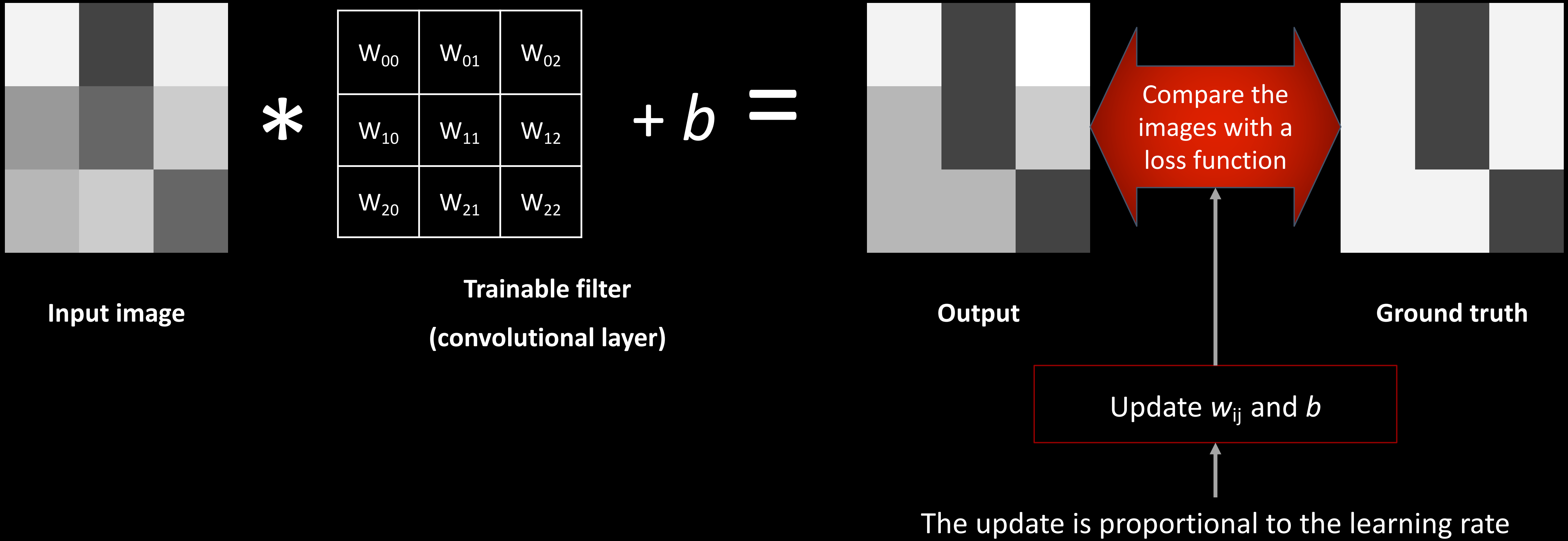
Input image

Trainable filter

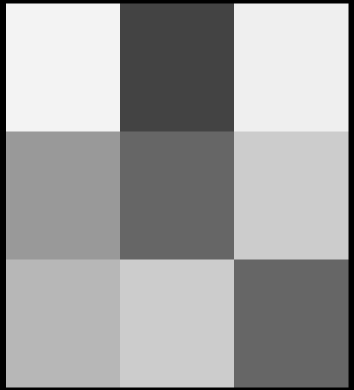
(convolutional layer)

Output

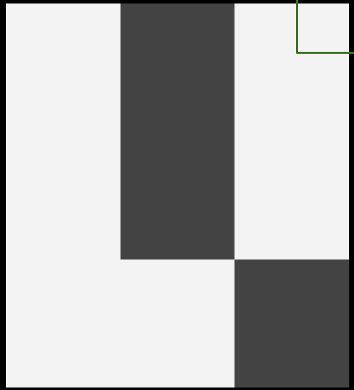
Supervised CNNs training



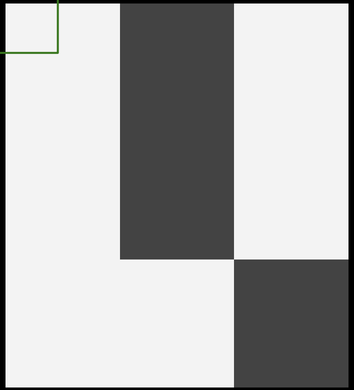
Supervised CNNs training



$$* \begin{bmatrix} W_{00} & W_{01} & W_{02} \\ W_{10} & W_{11} & W_{12} \\ W_{20} & W_{21} & W_{22} \end{bmatrix} + b =$$



Stop training



Input image

Trainable filter

(convolutional layer)

Output

Ground truth

Supervised CNNs training: backpropagation

- Optimization → Gradient descent
- Gradients computation → Backpropagation (use the chain rule for derivatives):

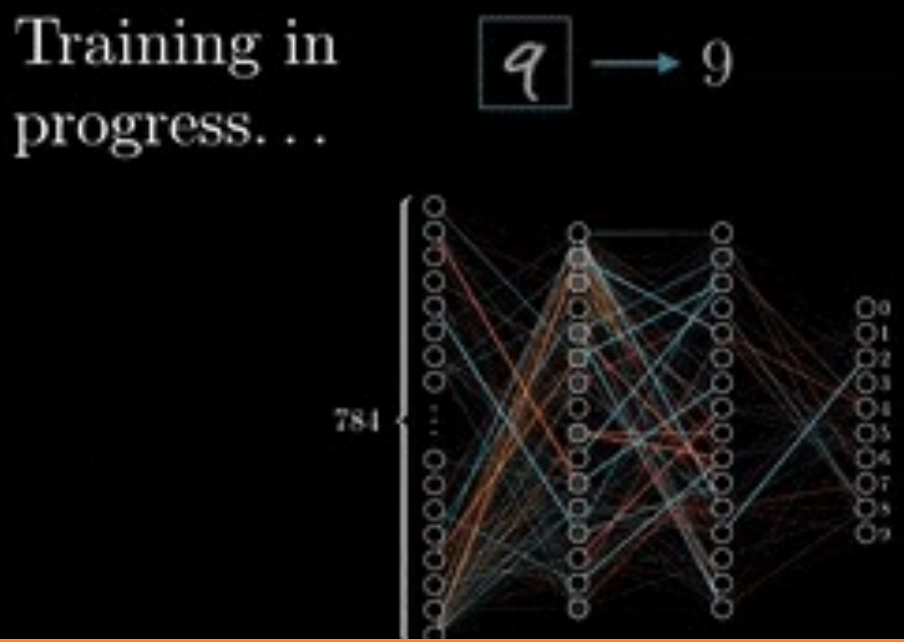
    After each forward pass through the network, a backward pass is performed to adjust the model's parameters (weights and biases) according to the error made by the output of the network.

Forward pass

Loss function: quantitative measure of the error

Learning rate: proportion used to update the parameters on each pass

Most used loss functions:
- Mean Squared Error (MSE or L1)
- Mean Absolute Error (MAE or L2)
- Binary Cross Entropy (Categorical cross entropy)

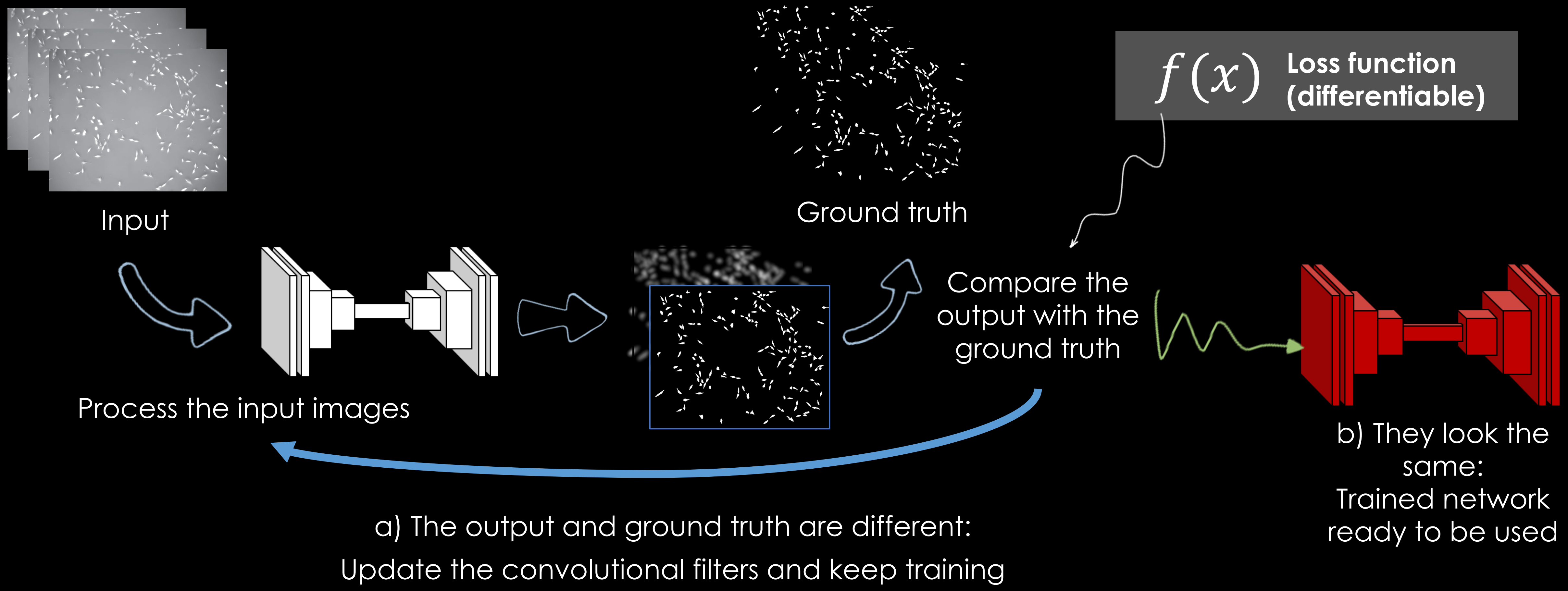


Error backpropagation

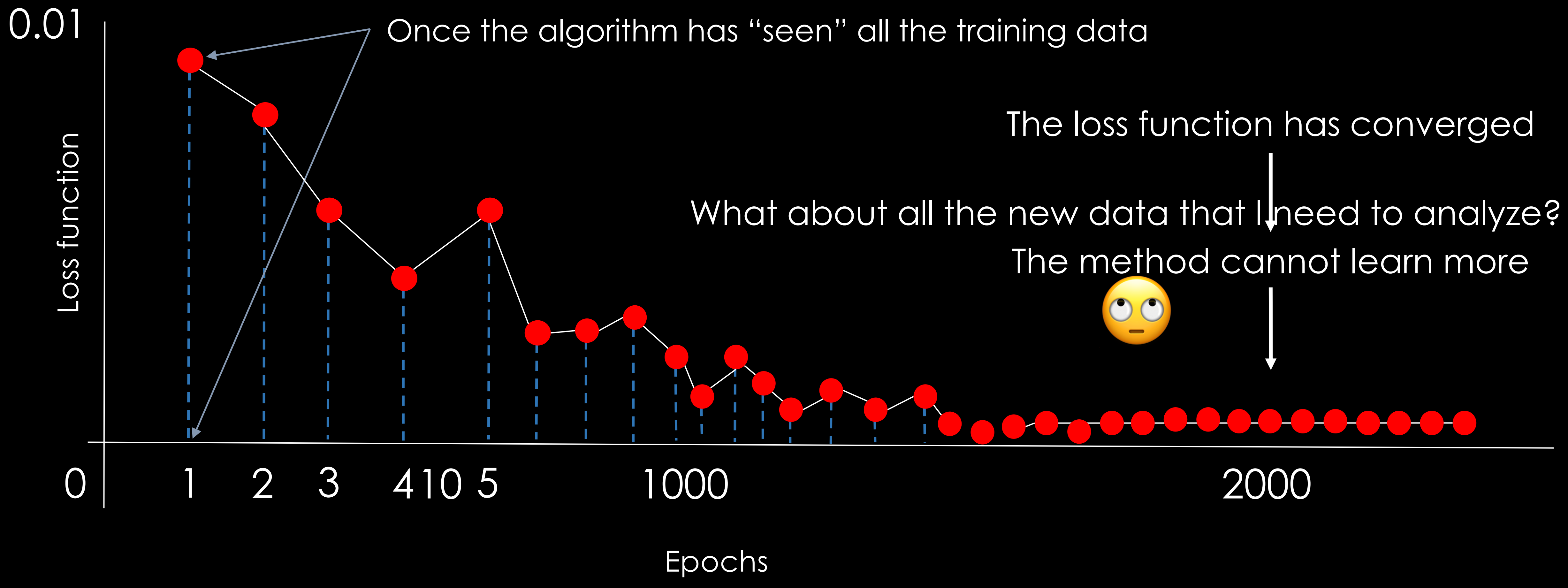
Introduced in 1960s
Popularized by Rumelhart, Hinton and Williams in "_Learning representations by back-propagating errors_", 1989

Supervised CNNs training

Pairs of inputs and desired outputs (i.e., ground truth)



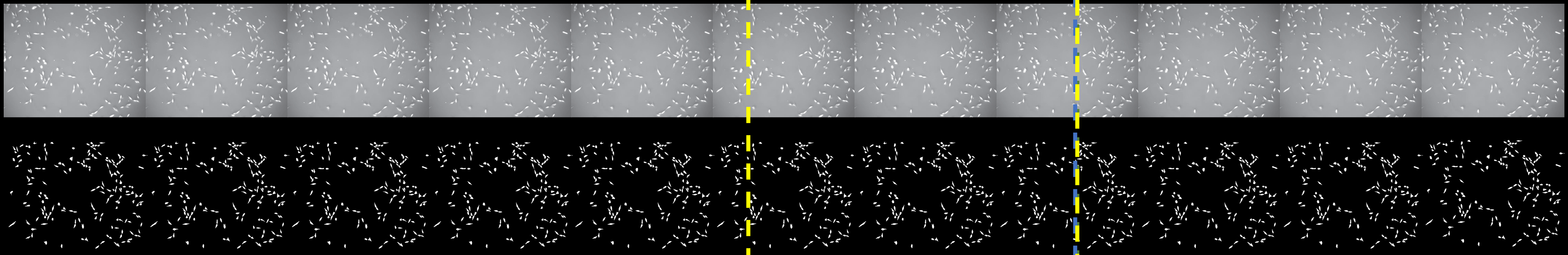
Training a neural network



Training a neural network: data

Input

Output



Training data
Parameter optimisation

Validation data
Model selection

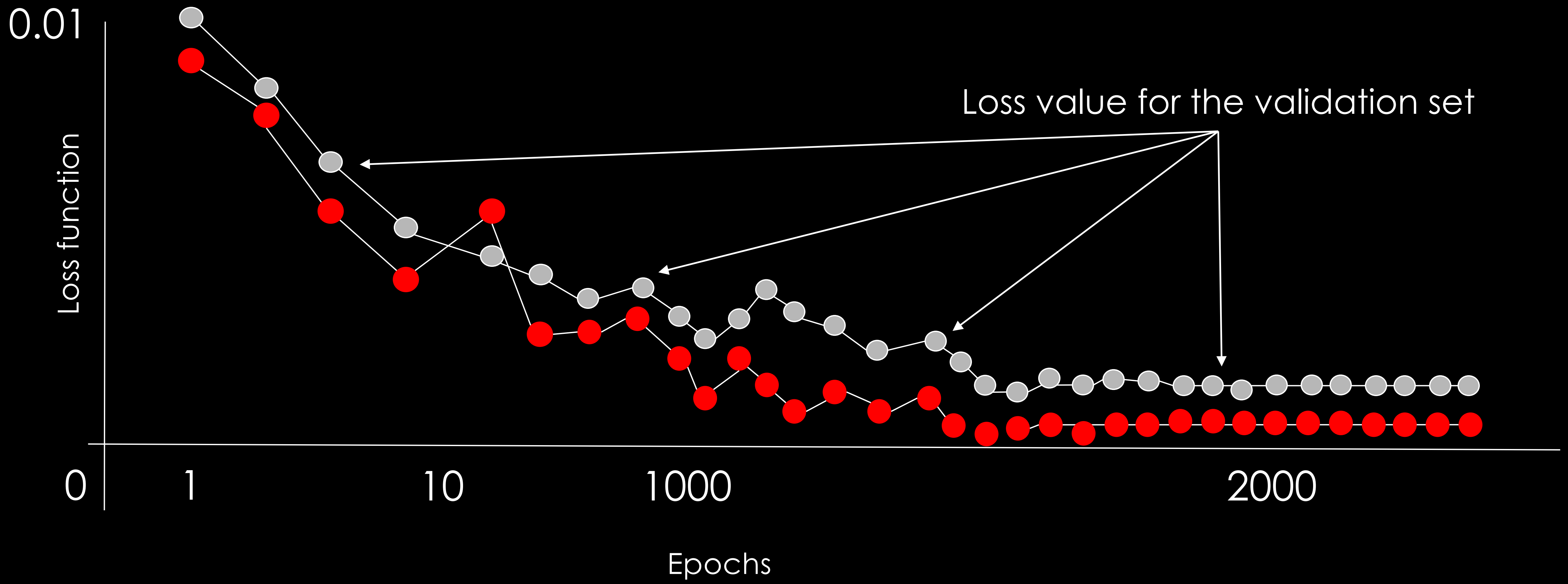
Test data
Final model evaluation
(not seen until the end)

The method learns from this dataset

- Large
- Heterogeneous
- Covers all the possible events in the problem

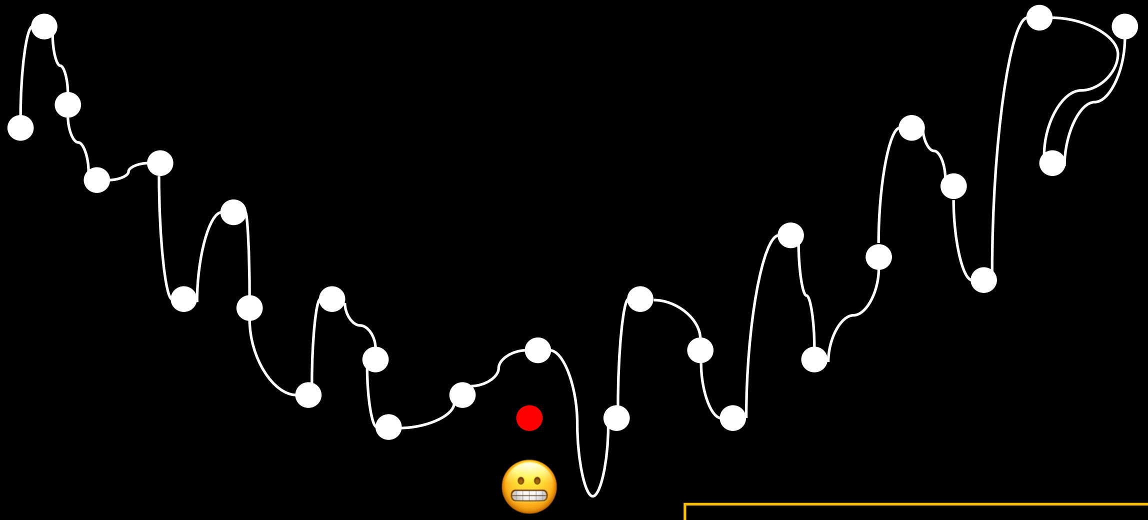
Independent of the training set

Training a neural network: data

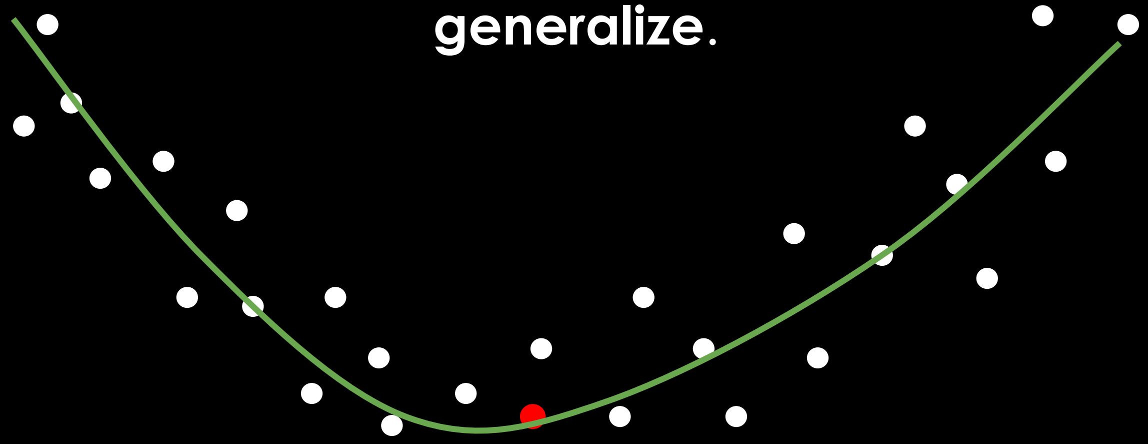


Training a neural network: data

If the method fails in the validation data, then it is called **overfitting**.



If the test processing is as good as in the training data, then we say that it can **generalize**.



You: how do I prevent overfitting?

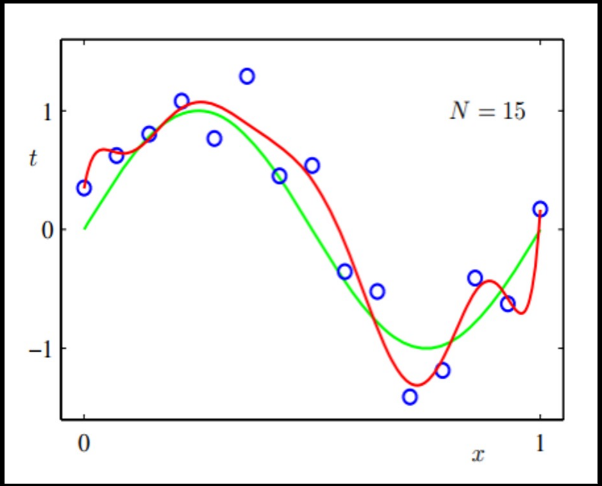
Esti & Wei: more data 😊

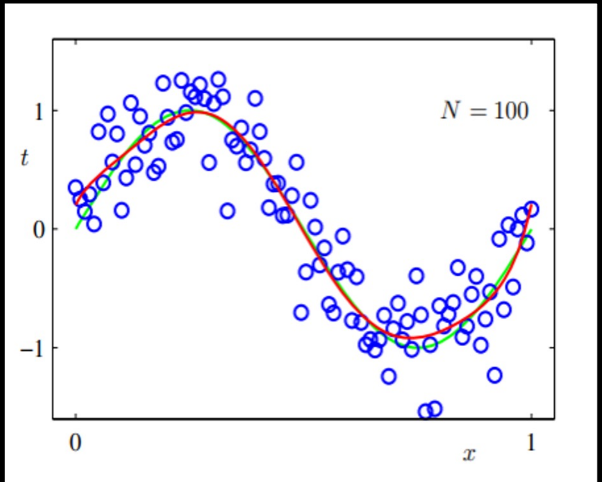
You: 😵‍💫🫠

Training a neural network: data

The bigger the data the better

→ Cover a real scenario

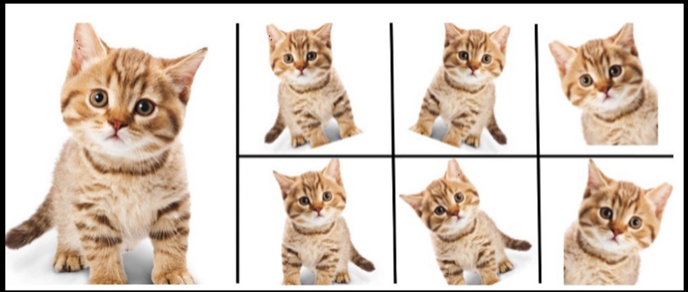




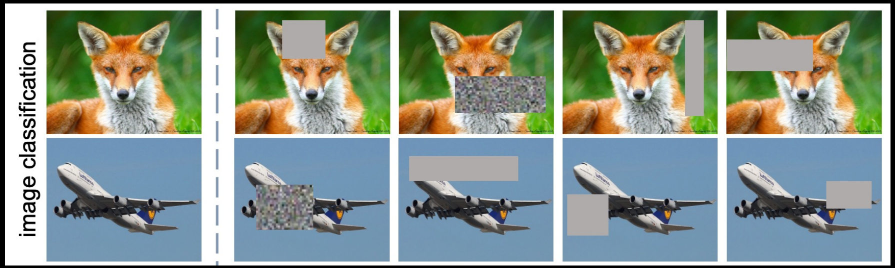
Christopher M. Bishop, Pattern Recognition and Machine Learning

Data augmentation

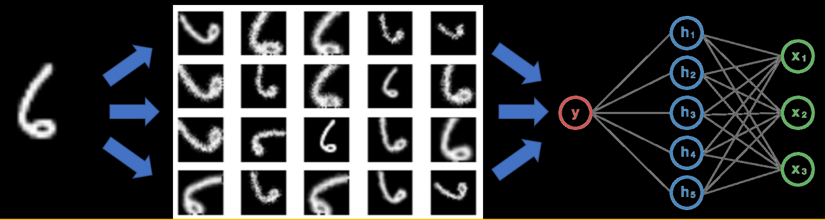
All of them are cats, indeed, the same cat



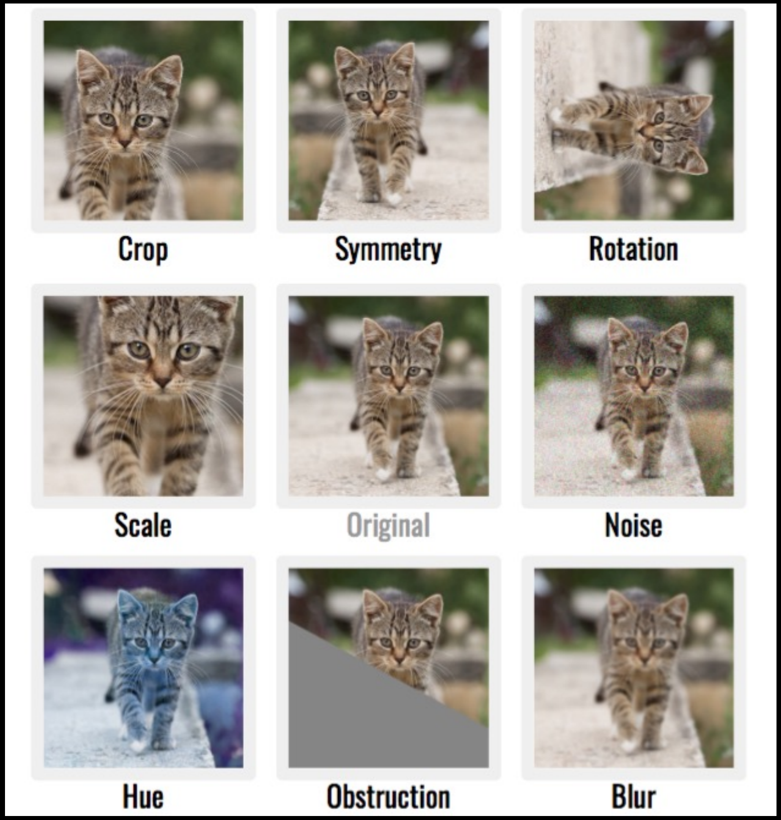
No doubt, it is a fox or an airplane



It is always the number 6



Most common strategies to augment data in image classification



Training a neural network: data augmentation in microscopy

I. The ground truth also needs to be augmented with the same transformations.

II. Image transformations need to preserve the meaning and biophysical properties of the data.

Geometrical transformations

Linear transformations (preserve shape)

- Rotation
- Translation

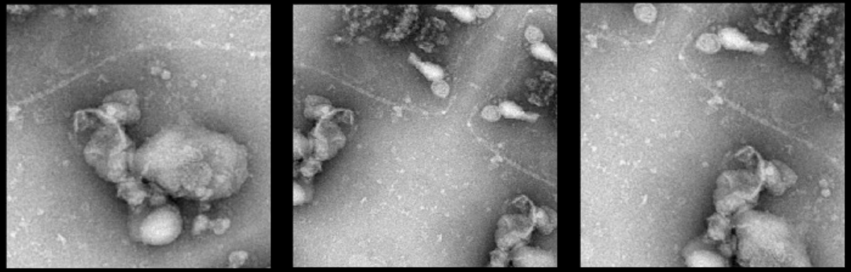
Non-linear (elastic) transformations (shape changes)

- Zooming
- Shearing

Signal artifacts:

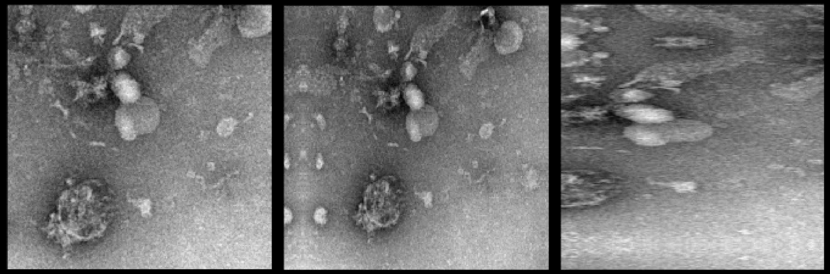
- Noise
- Contrast
- Blurring

Linear transformations



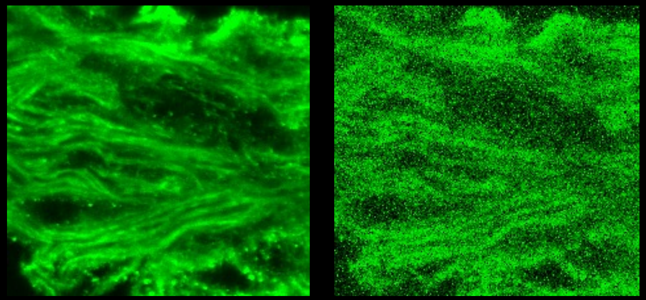
Original patch — Rotation + Shift — Rotation

Non-linear transformations



Original patch — Zoom — Shearing

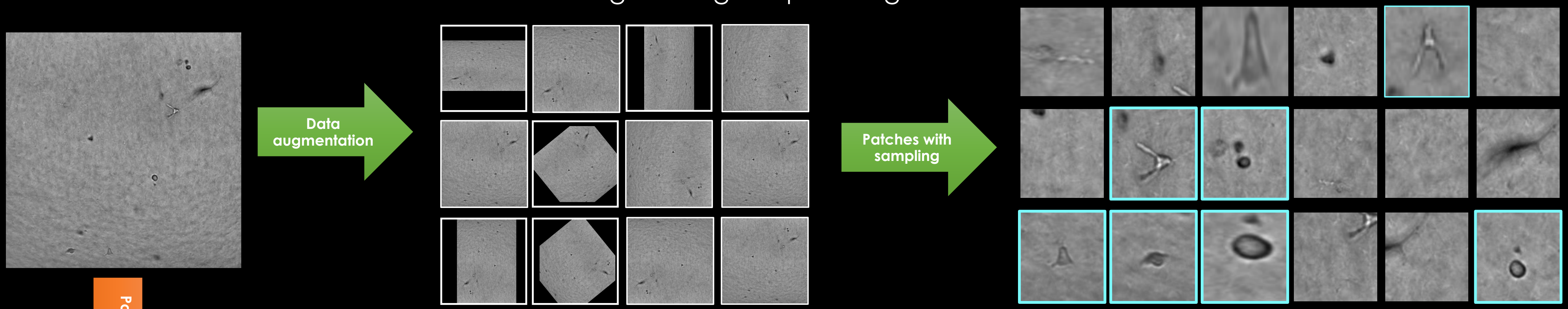
Adding noise

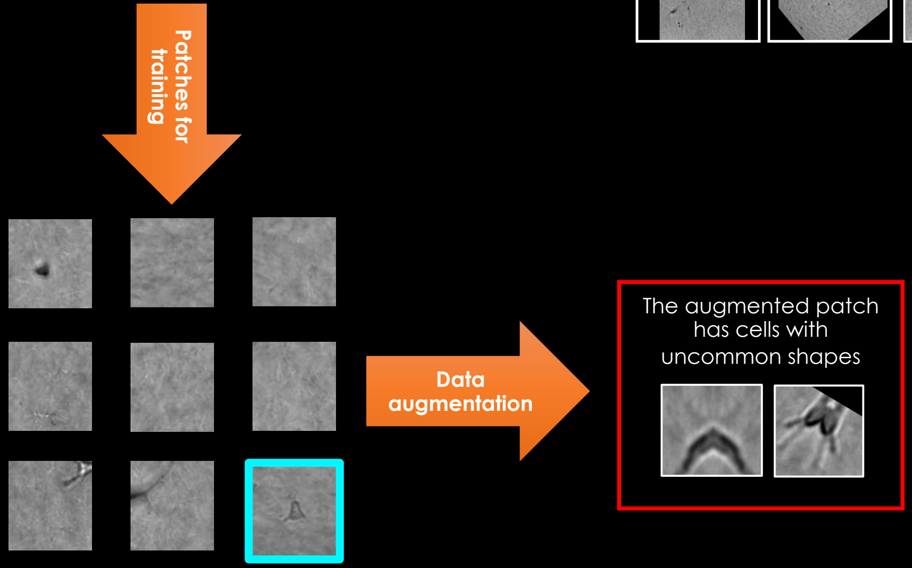


Original image — Noisy

Segmentation with CNNS: patches and data augmentation

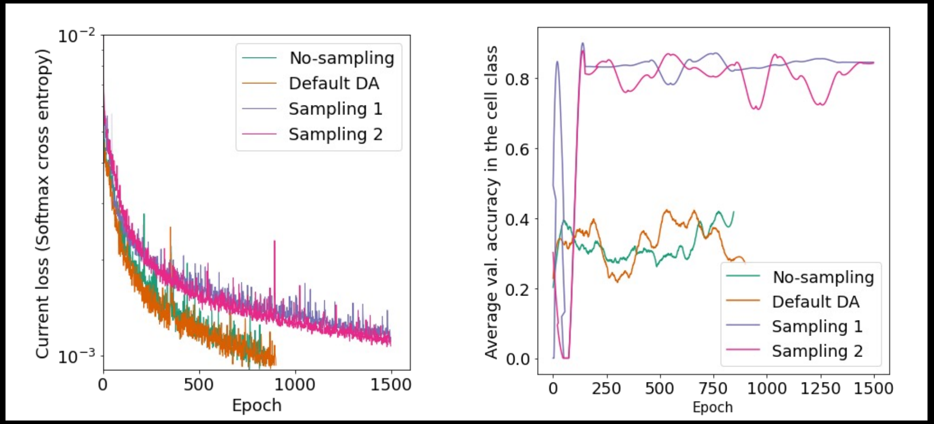
Make sure that artifacts are not introduced when augmenting the patching





Only one patch has a cell

Results on the learning process with different strategies for data augmentation



E. Gómez de Mariscal et al., arxiv 2021

Training a neural network: data

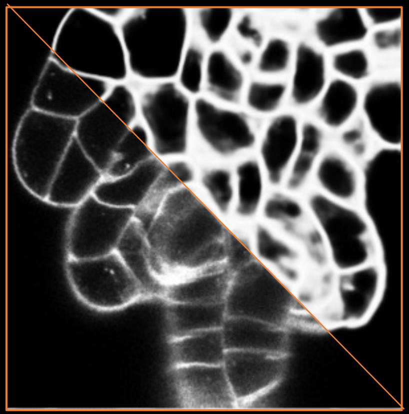
Data collection (&curation) is expensive. 

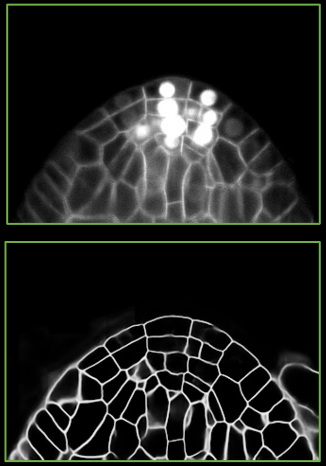
Small datasets for bioimage analysis

Fine tuning

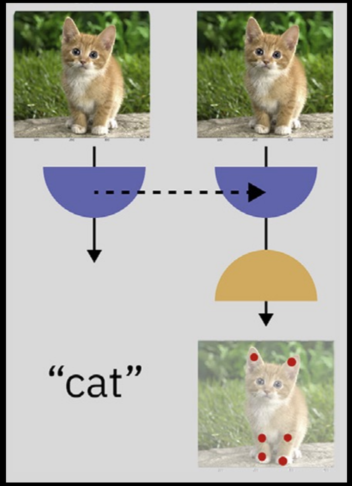
Pretrained model for boundary segmentation

Same task, similar features, different data distribution





Transfer learning



Classification

Segmentation

Classification

Pose estimation

A. Mathis, et al., Neuron 2020
A. Wolny et al., eLife 2020
W. Ouyang et al., bioRxiv 2022

Deep learning systems

Model architecture

- U-Net
- ResNet
- MobileNetV2
- (cycle)GANS
- DenseNets

Loss function

- Mean Squared Error
- Mean Absolute Error
- (binary)cross-entropy
- Focal loss
- Dice loss

Optimisation

- Stochastic gradient descent
- ADAM

Data (curation)

- Labelled masks
- Keypoints&landmarks
- Paired images (high&low SNR)
- Bounding boxes
- Tracks

Influence the final performance and behaviour of your system

A. Mathis, et al., Neuron 2020

Deep learning systems

Model architecture **Loss function** **Optimisation** **Data**

Each configuration affects the inference speed, training data requirements, memory demands.

Task specific:

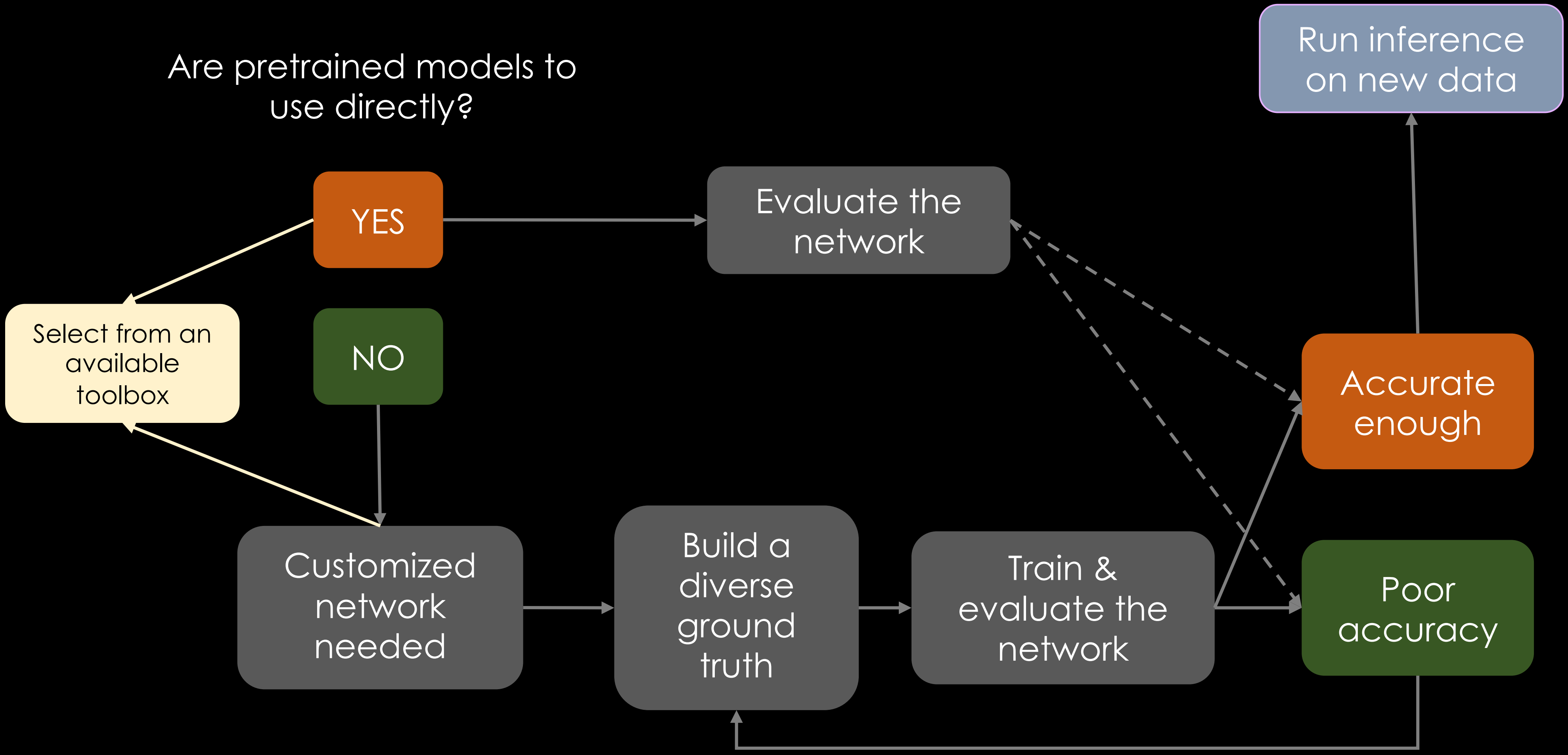
- 2D, 3D, time lapse, multichannel.
- Image-to-image vs.image-to-vector processing.
- Criterion needed to learn: regression vs classification.

A. Mathis, et al., Neuron 2020

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?



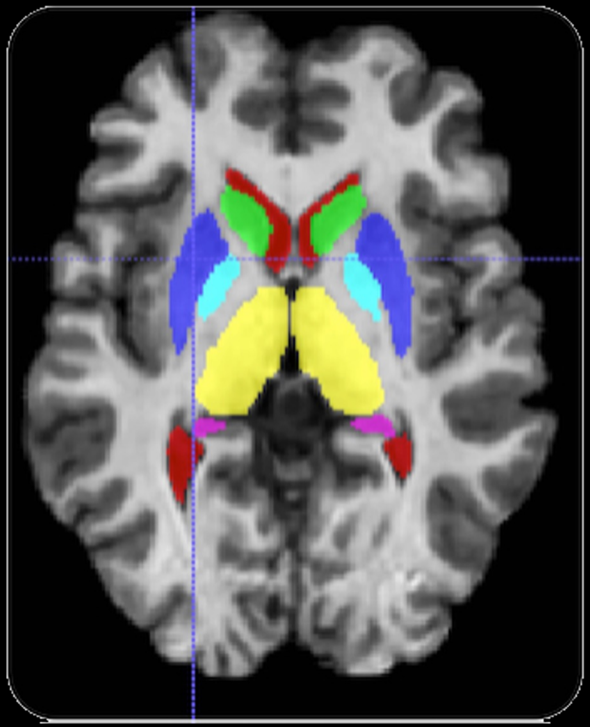
A. Mathis, et al., Neuron 2020

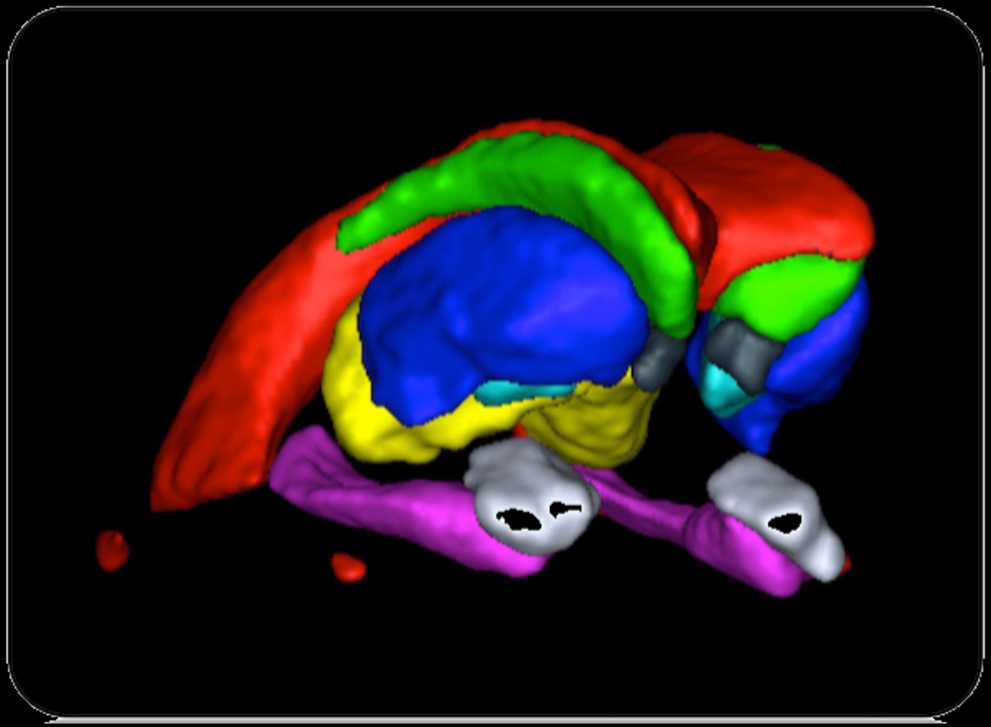
Segmentation

Segmentation

Formal definition: partitioning of the image domain Ω into several (usually disjoint) regions Ω_i

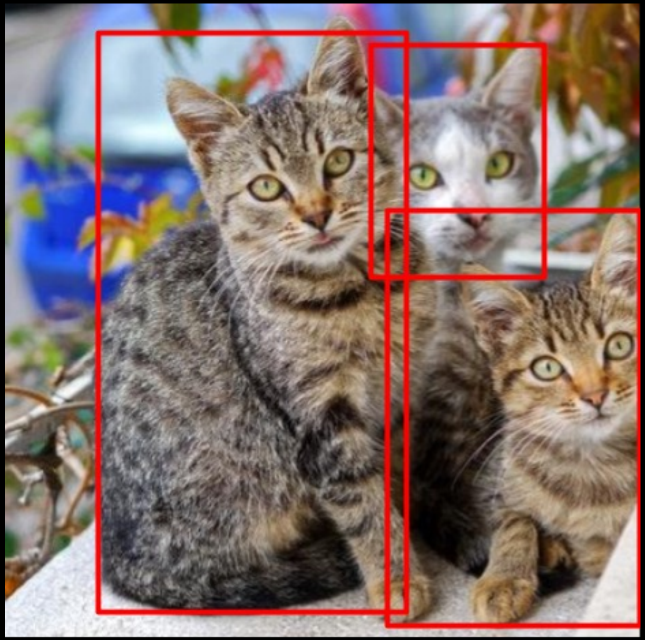
$$\Omega = \bigcup_i \Omega_i, \qquad \Omega_i \cap \Omega_j = \emptyset, \forall i \neq j$$



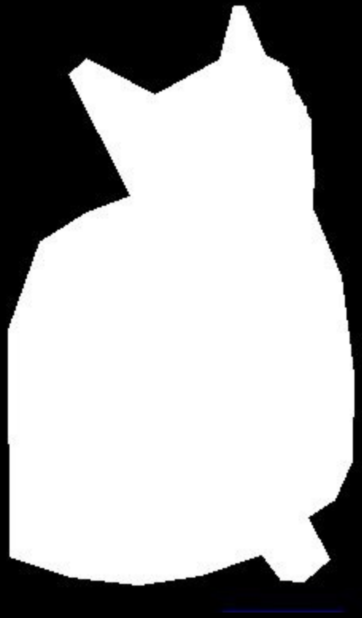


P. Coupé et al., NeuroImage 2011

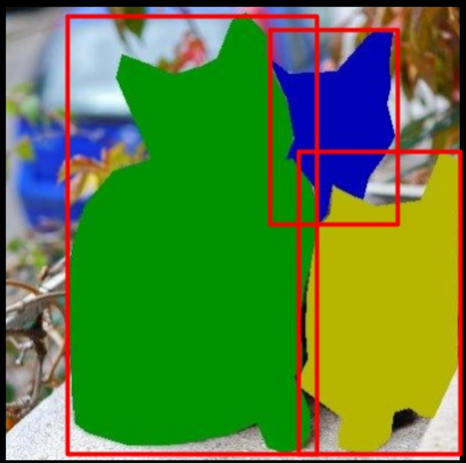
Segmentation



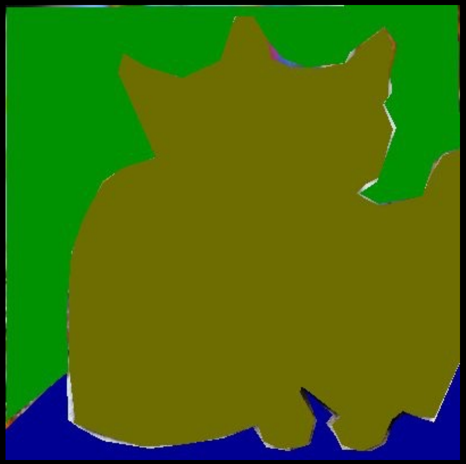
Detection



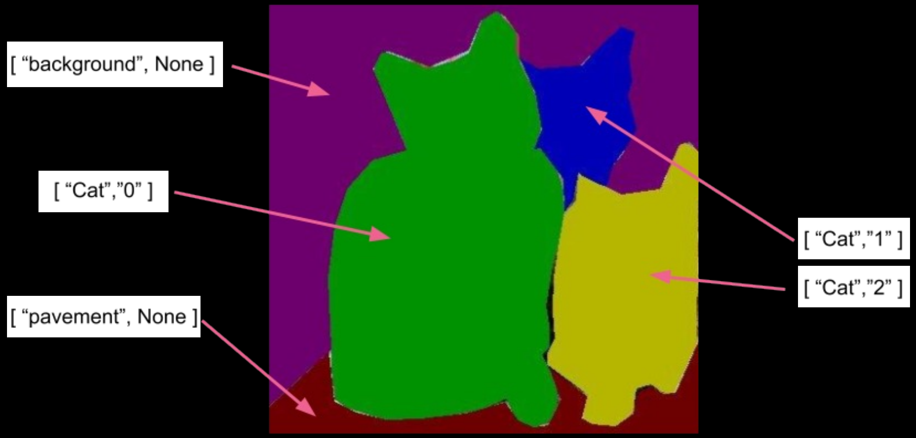
Binary segmentation



Instance segmentation
(Detection + segmentation)



Semantic segmentation



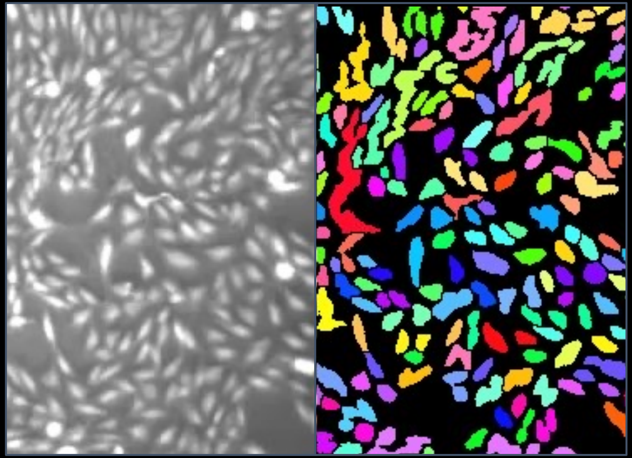
Panoptic segmentation
(Instance + Semantic segmentation)

https://analyticsindiamag.com/semantic-vs-instance-vs-panoptic-which-image-segmentation-technique-to-choose/

Segmentation

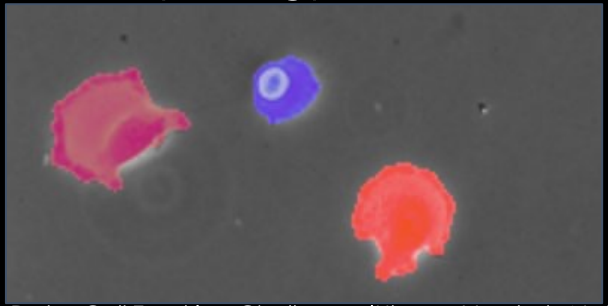
The information is partitioned in different segments to simplify its representation into something that is easier to analyze

Cell counting



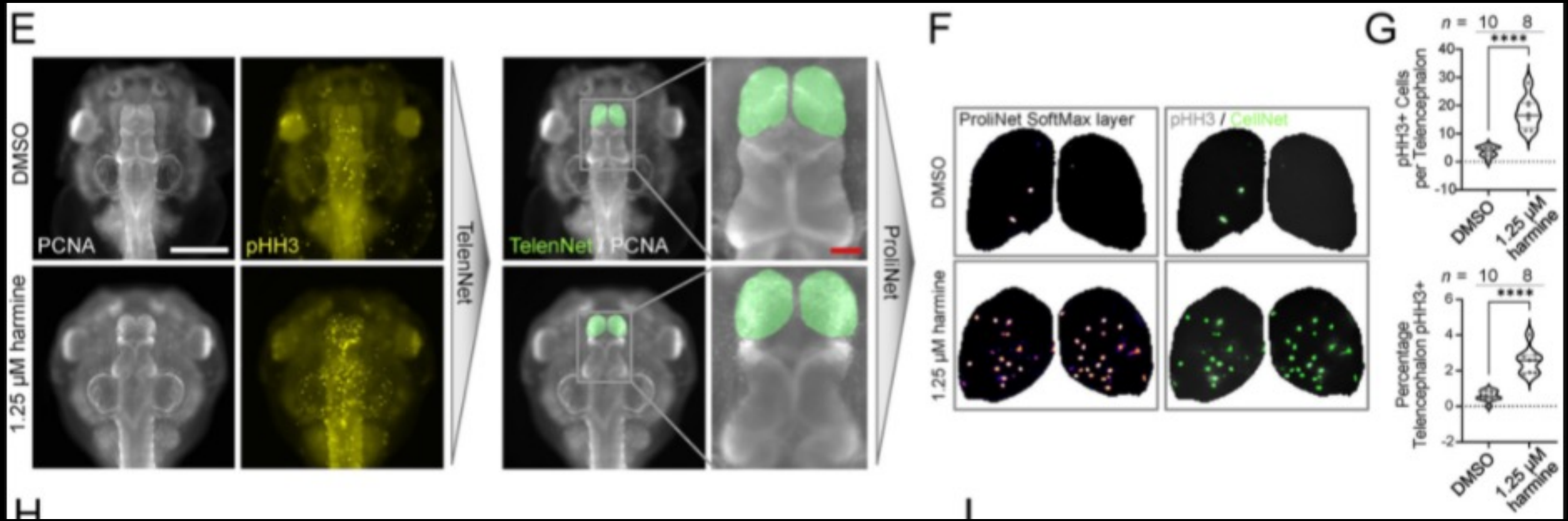
F. Lux & P. Matula, arXiv, 2020

Morphology assessment



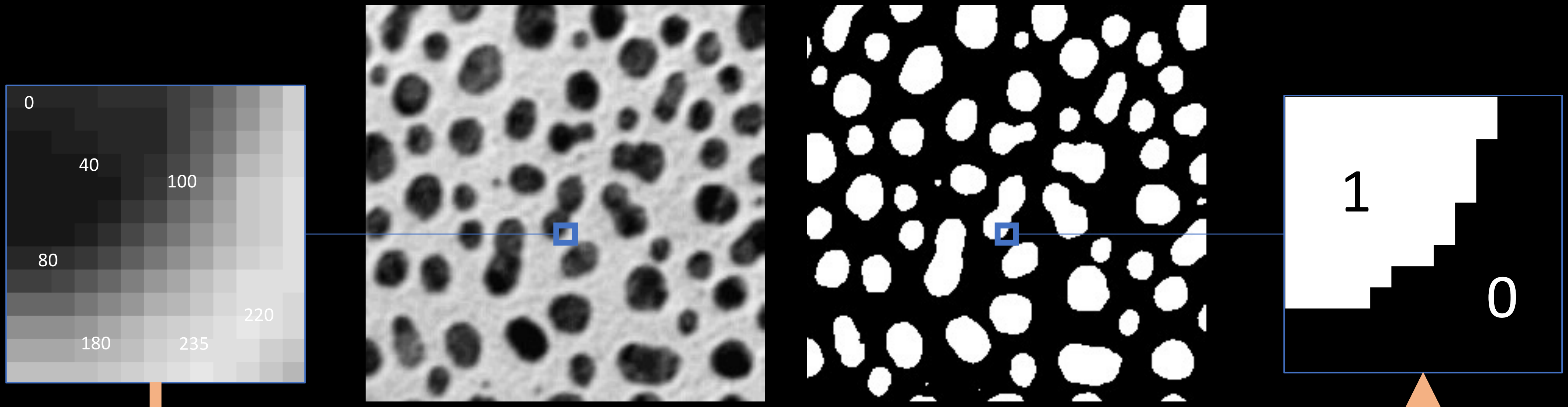
Data: Cell Tracking Challenge (Ulman, V., et al., Nat Methods 2017), Traning: João Luis Soares Lopes (EPFL)

Determine anatomical regions (telencephalons) to measure cell activity (proliferating pHH3+ cells)



Thomas Naert, Development, 2021

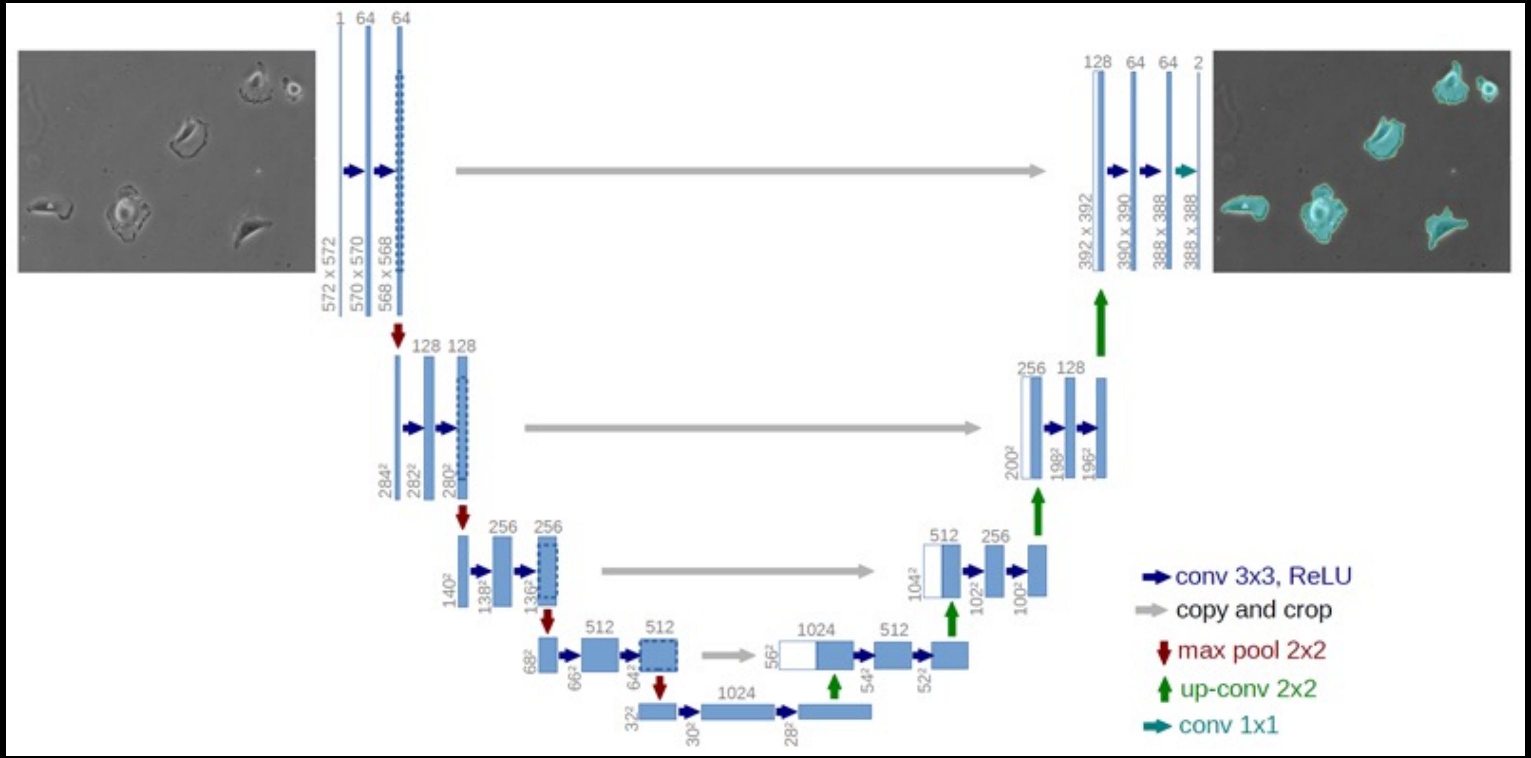
Segmentation: Thresholding is the most basic form of obtaining binary images



Threshold all the values smaller than 128

Segmentation with CNNS: U-Net encoder-decoder for binary segmentations

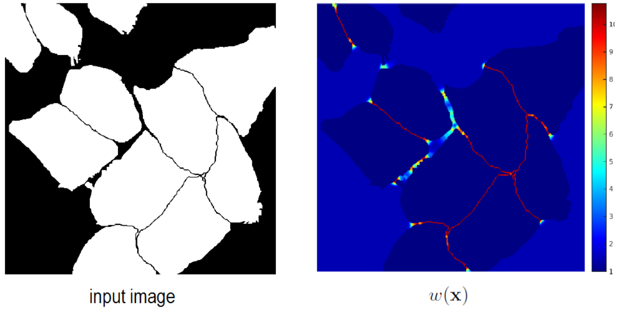
Skip connections:
Take the output of each level in the encoder path and copy it with the input of the decoder path.
It helps preserving high resolution details during decoder process.



Olaf Ronneberger, Philipp Fischer, Thomas Brox; U-Net: Convolutional Networks for Biomedical Image Segmentation; arXiv 2015

Segmentation: alternative strategies

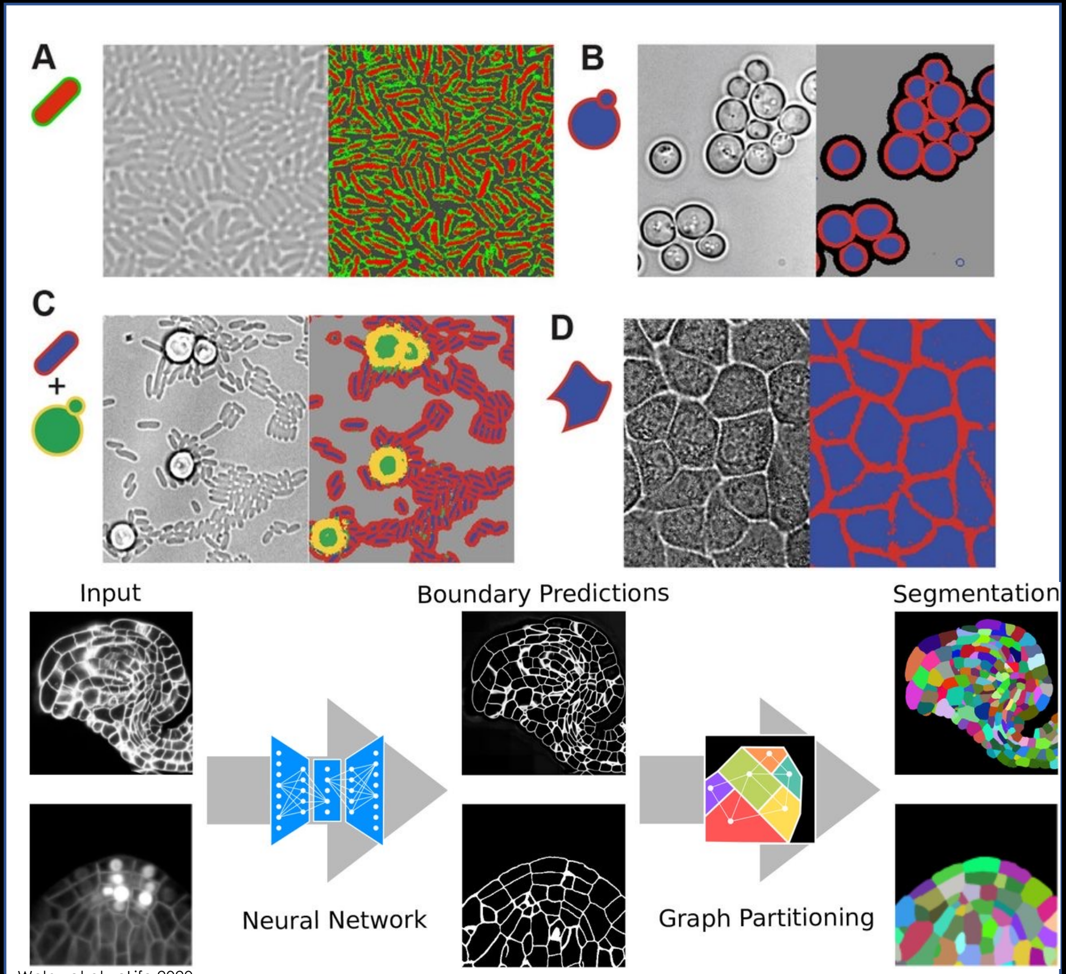
Weighted loss functions



$$w(\mathbf{x}) = w_0 \cdot \exp\left(-\frac{(d_1(\mathbf{x}) + d_2(\mathbf{x}))^2}{2\sigma^2}\right)$$

Olaf Ronneberger, et al., arXiv 2015

Use different labels



A. Wolny et al., eLife 2020

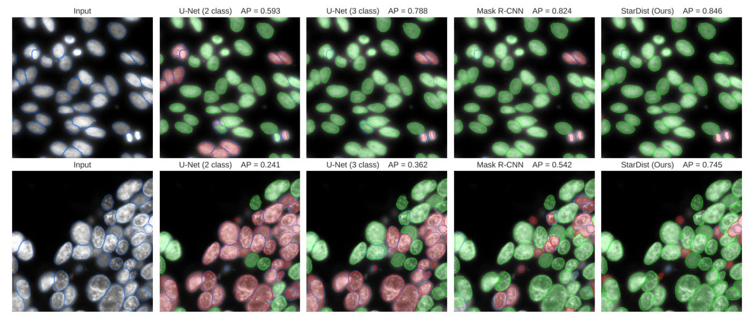
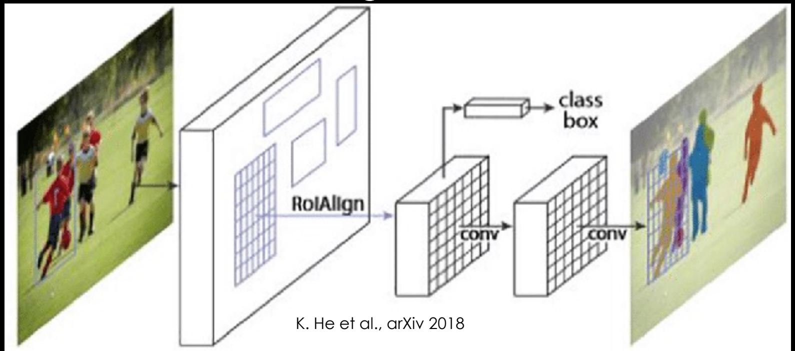


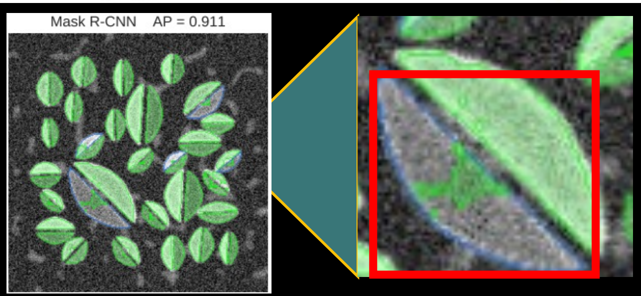
Fig. 4: Two segmentation results ($\tau = 0.5$) for DSB2018. See Fig. 2 caption for legend.

https://github.com/maweigert/neubias_academy_stardist
Schmidt, Weigert et al 2018

Instance segmentation with ROI classification and segmentation



K. He et al., arXiv 2018

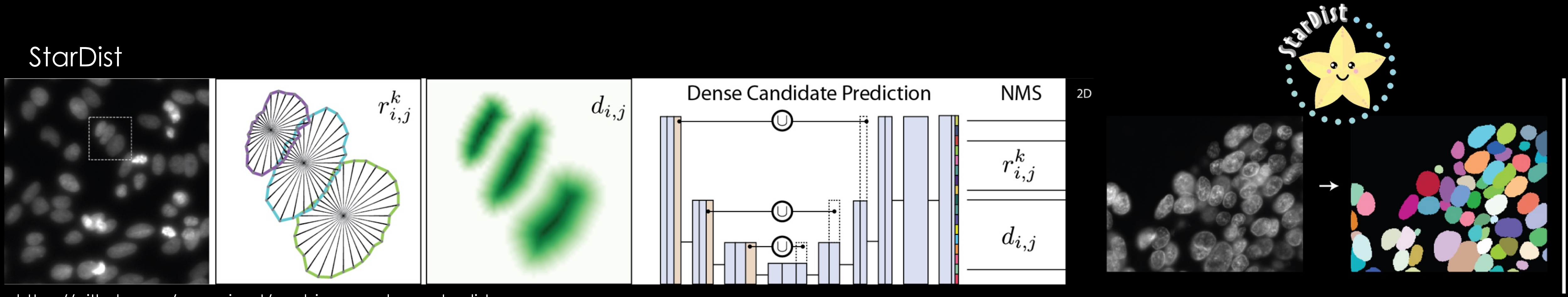


Limitations of binary image segmentation

- Will not work with dense, packed or clustered objects.
  → Additional labels to split indpendent objects
- Overlapping objects cannot be represented in one single mask
- ROIs need to be predefined and do not scale well to cellular shapes

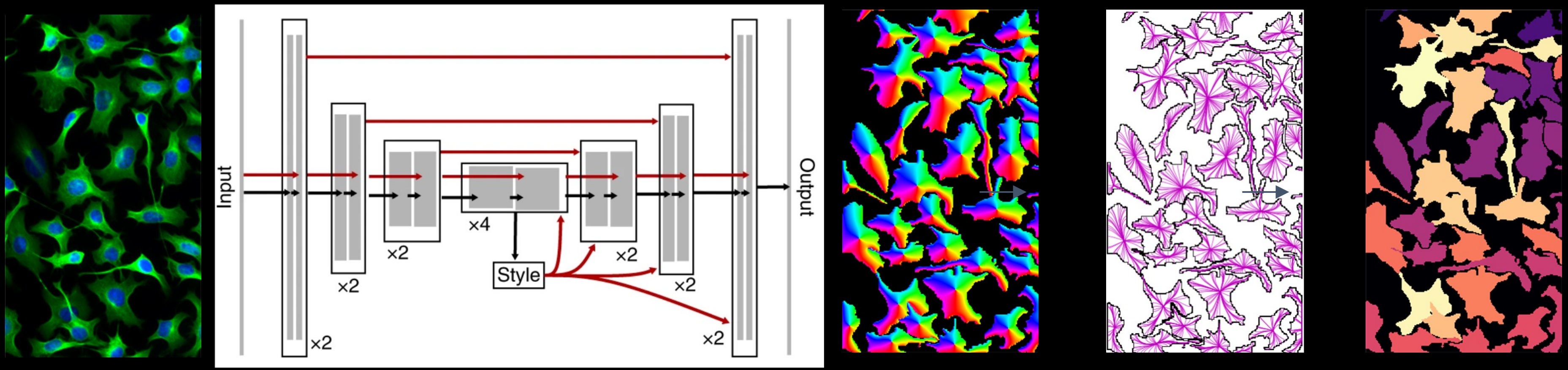
Segmentation: learn deterministic features rather than discrete labels

StarDist



https://github.com/maweigert/neubias_academy_stardist
Schmidt, Weigert et al 2018

Cellpose



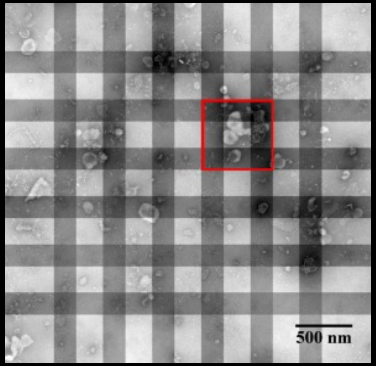
https://www.cellpose.org

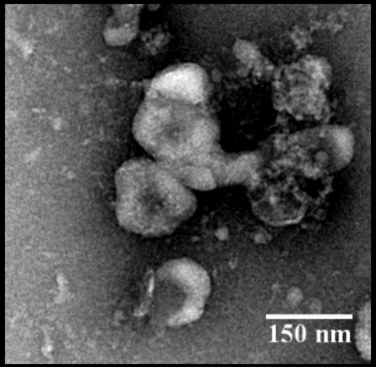
Stringer, Wang, Michaelos, Pachitariu, Nature Methods, 2021

Segmentation with CNNS: image preparation and features

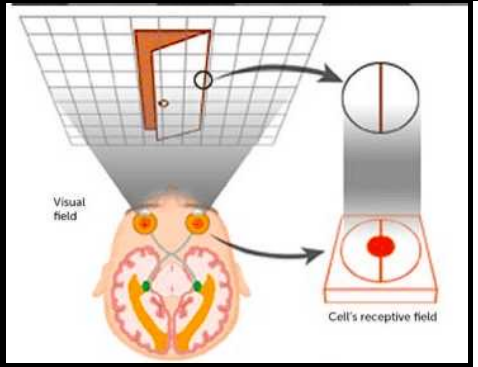
Biomedical images can get really large (up to TB for electron microscopy) → GPU memory a major limitation

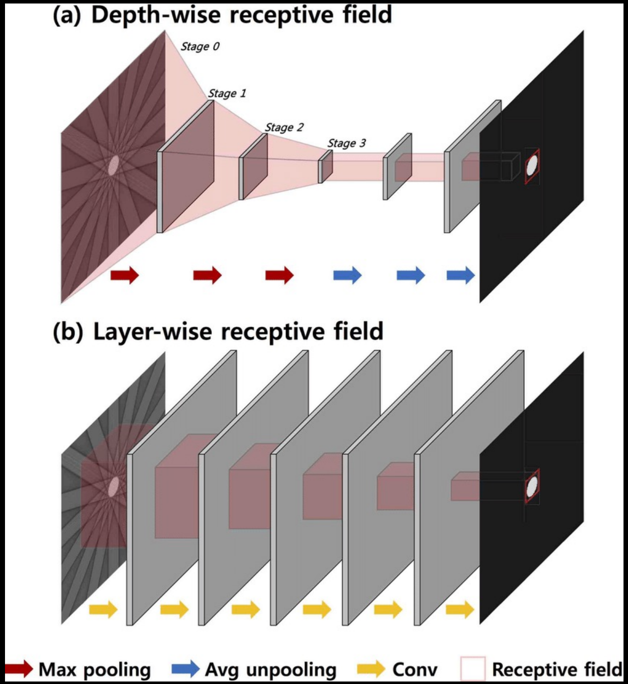
Divide images into patches →increase the training data variability

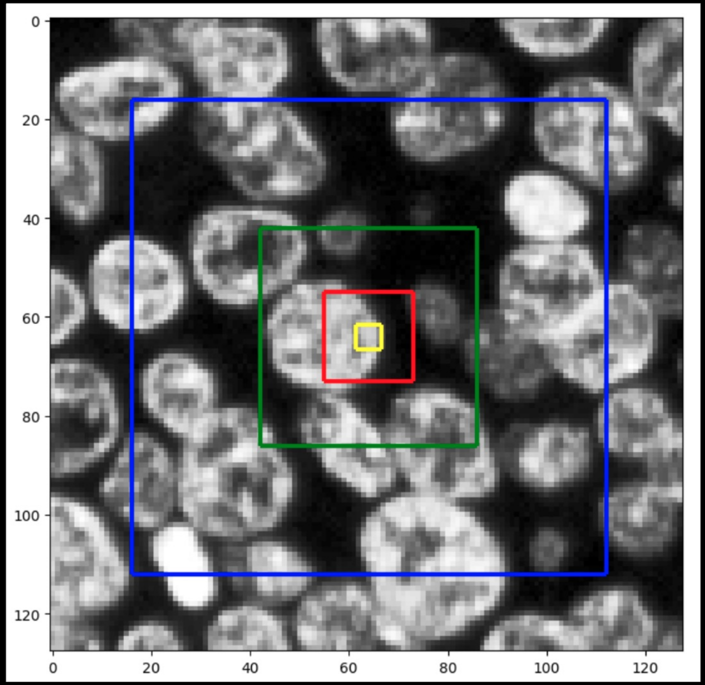




How big? → Receptive field of the network → it needs to have enough information to learn and discriminate





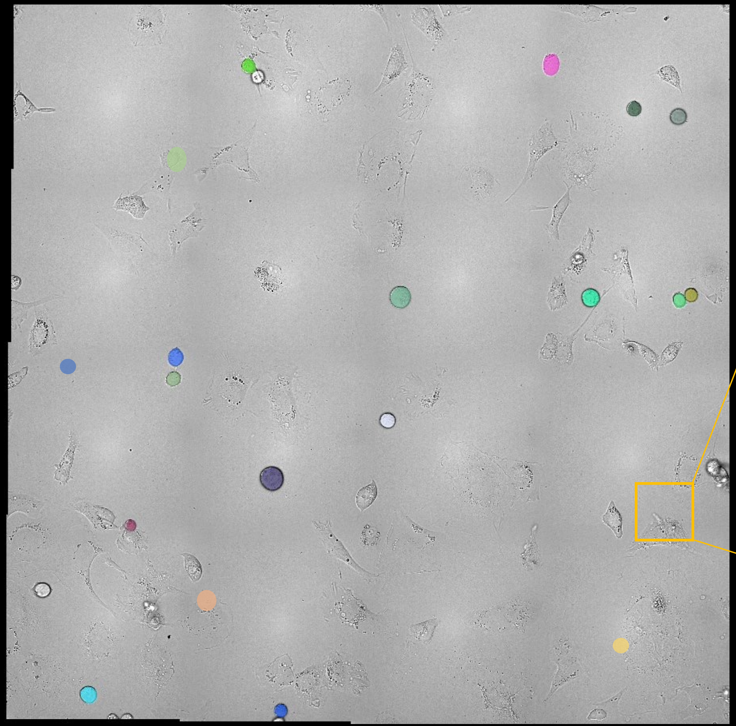


TissueNet,
Greenwald, Miller et al 2021

Segmentation with CNNS: image preparation and features

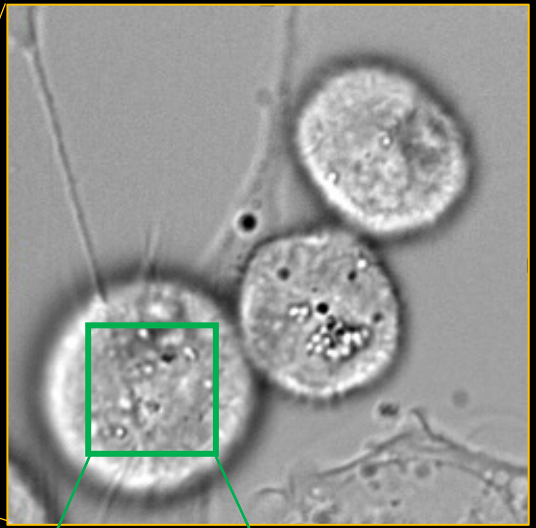
Image processing task: segment cells in mitosis

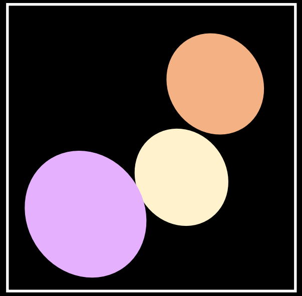
7271 pixels



7271 pixels

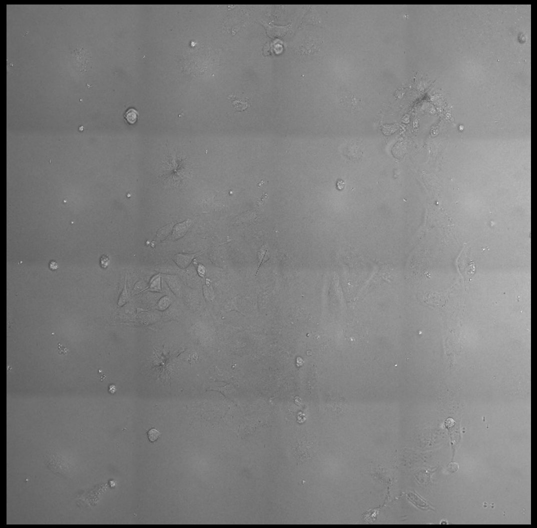
512 x 512 pixels



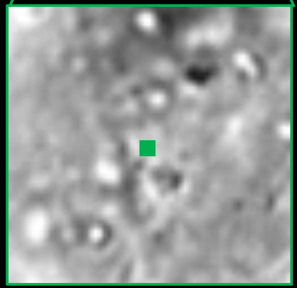


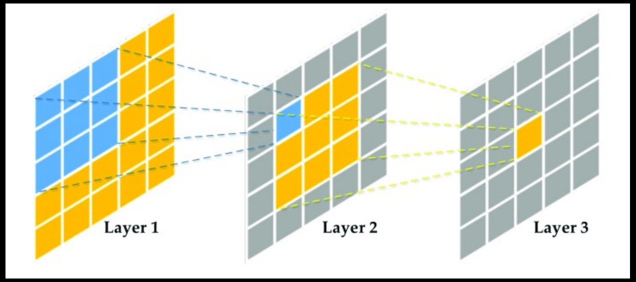
Training patch size for StarDist

Result: empty masks



Receptive field of StarDist: “What the network sees to determine the value of one pixel”





Schmidt, Weigert et al 2018

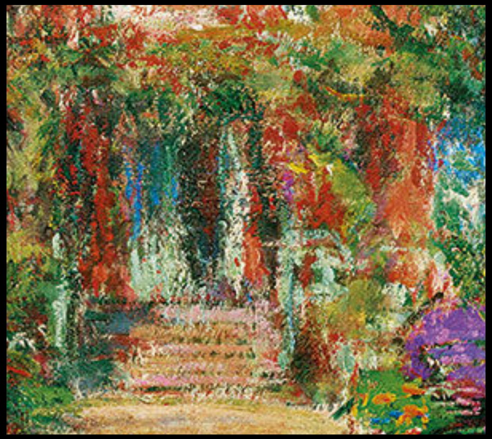
Original vídeos: 4x4 field of views (x63)

Pixel size = 0.108 um/pixel

Segmentation with CNNS: image preparation and features

A way to understand could be… asking what is the perfect distance to decipher the scenes of Claude Monet's art

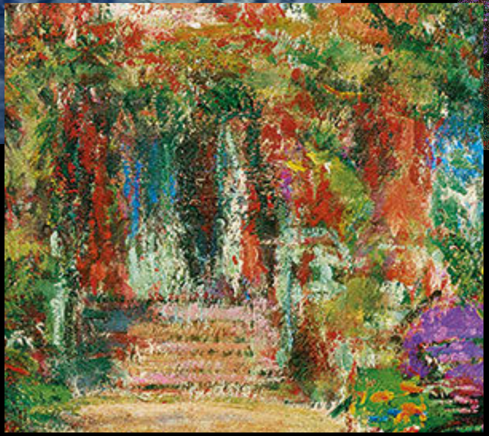
→ enough as to get the context with still meaningful details

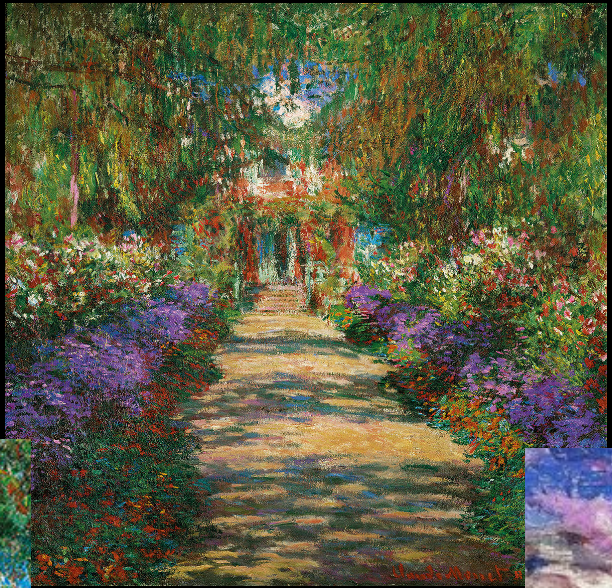


Segmentation with CNNS: image preparation and features

A way to understand could be… asking what is the perfect distance to decipher the scenes of Claude Monet's art

→ enough as to get the context with still meaningful details

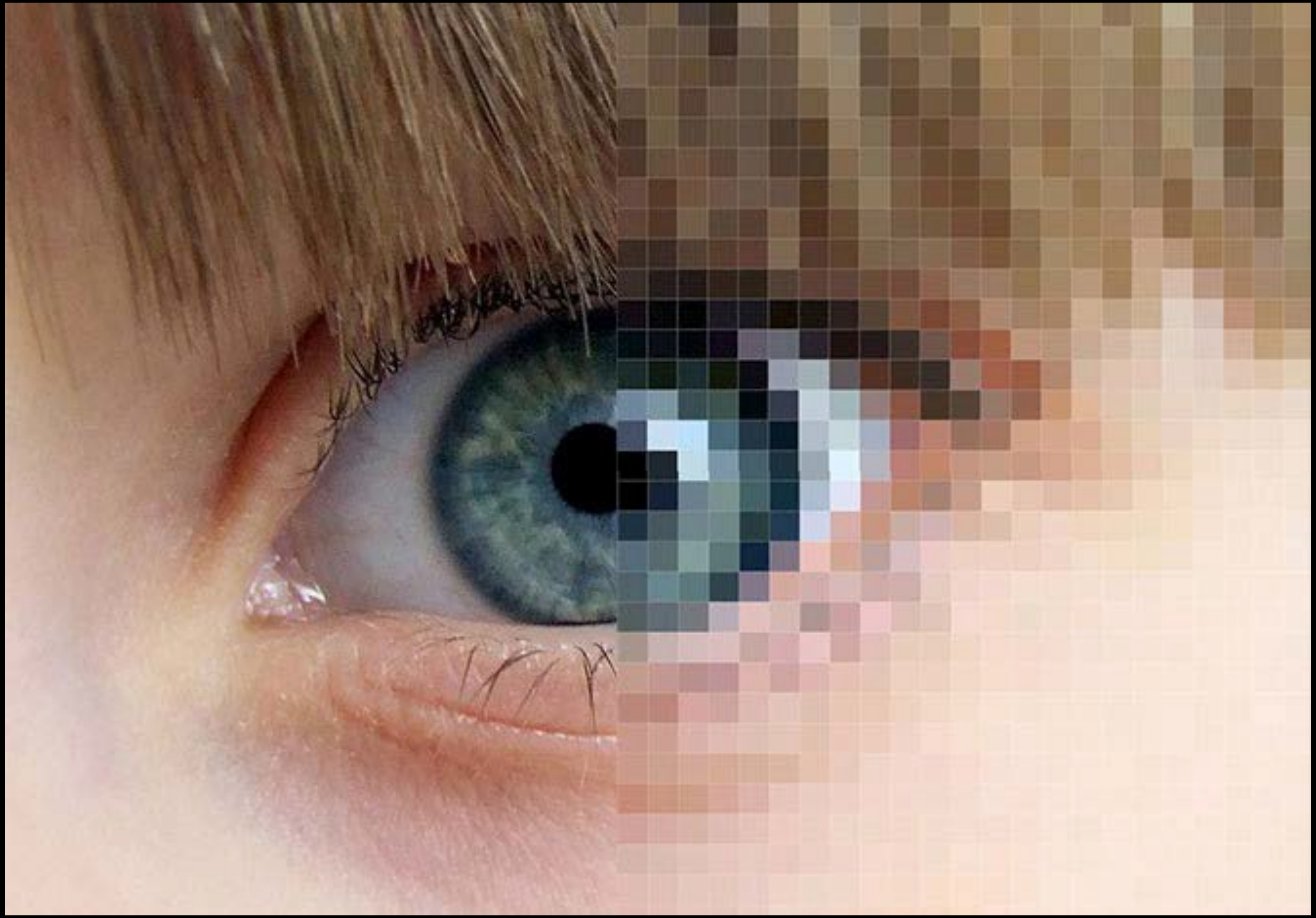




Garden path at Giverny

The house among roses

Segmentation with CNNS: image resolution and its effect in segmentation

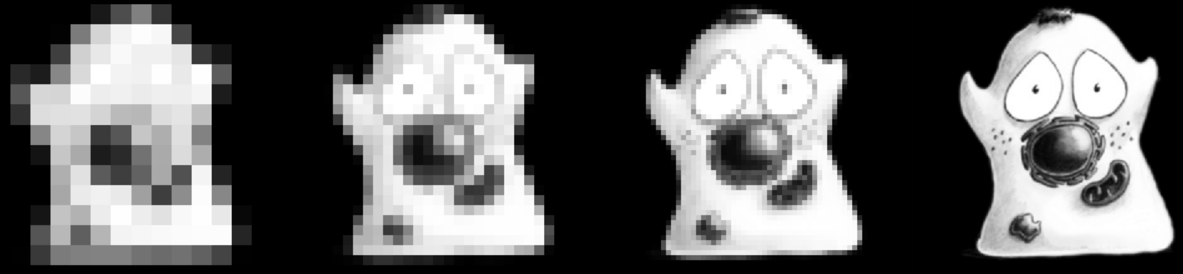


Segmentation with CNNS: image resolution and its effect in segmentation

Basic definition: Number of pixels in an image ➡ Pixel size: physical size (length and width) covered by one pixel

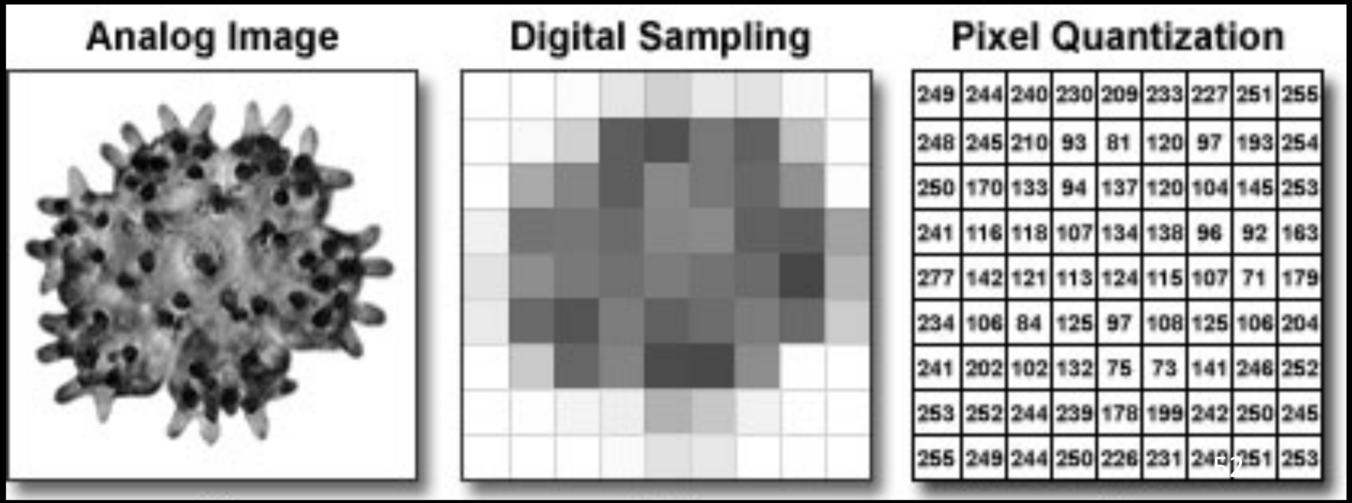
16 cm

16 cm



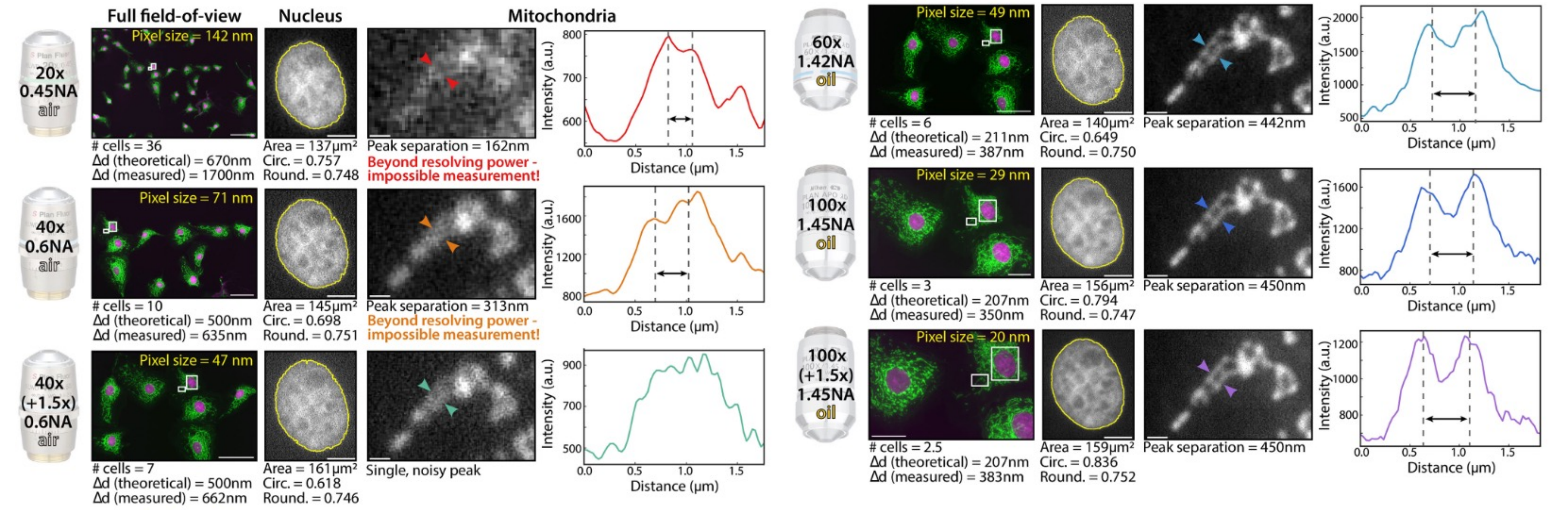
8x8 64x64 128x128 512x512

- The pixel size influence the amount of details for a given field of view.
- Detectors also limit the size of the finniest detail that we can acquire



249	244	240	230	209	233	227	251	255
248	245	210	93	81	120	97	193	254
250	170	133	94	137	120	104	145	253
241	116	118	107	134	138	96	92	163
277	142	121	113	124	115	107	71	179
234	106	84	125	97	108	125	106	204
241	202	102	132	75	73	141	246	252
253	252	244	239	178	199	242	250	245
255	249	244	250	226	231	[illegible]	251	253

Segmentation with CNNS: image resolution and its effect in quantification

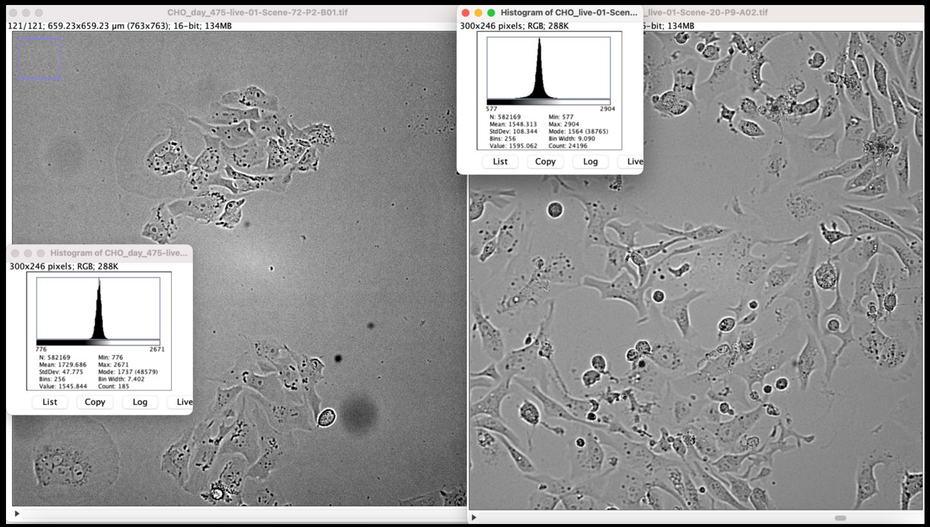


Segmentation with CNNS: image preparation and features

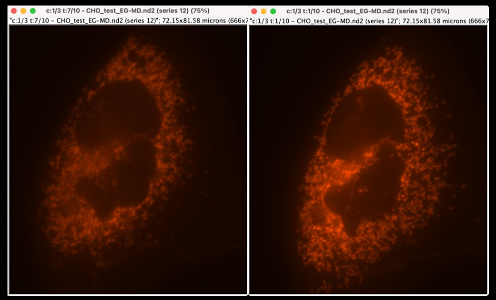
Intensity values vary with the physical properties of the data, the calibration of imaging devices or the natural variability of the sample

MNIST data: black and white pixels

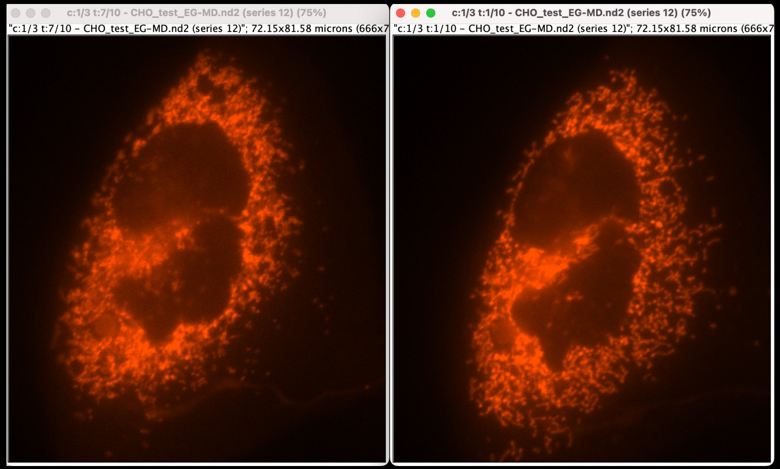
Non-normalized images



Before normalization



After percentile normalization

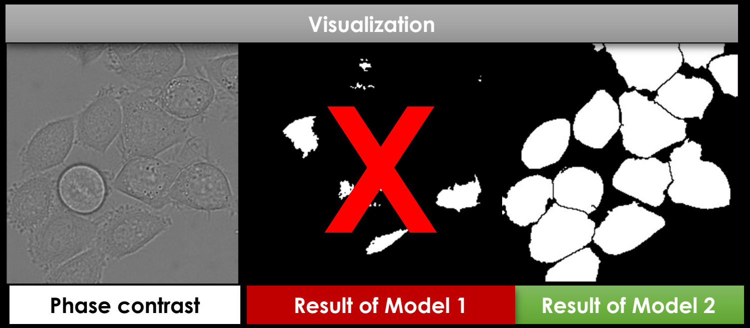


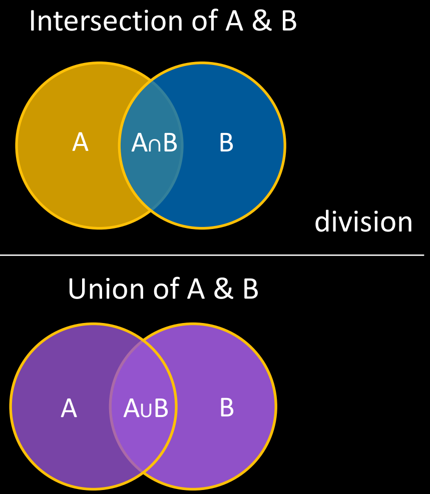
Common strategies:

- Intensity projection: Clip the dynamic range of values to the [0, 1] range
- Standardization with the mean and standard deviation
- Percentile projection (common in fluorescence): remove outliers (i.e., noise and artifacts) from the intensity distribution (extremes in the tails) and clip to the [0, 1] range.
- Normalize w.r.t. the entire population (training data)

Evaluation of the model performance → Accuracy 👀

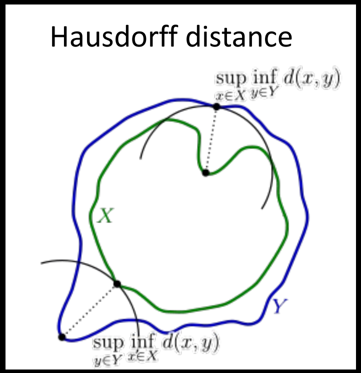
You need to verify that the model is doing precisely what you want

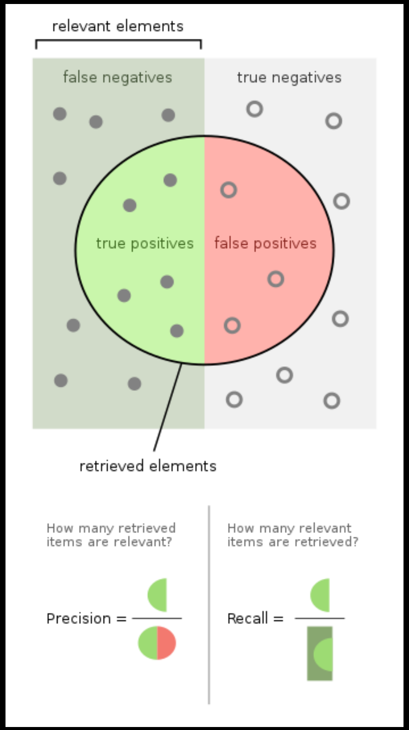




Jaccard (A, B) = A∩B / A∪B =

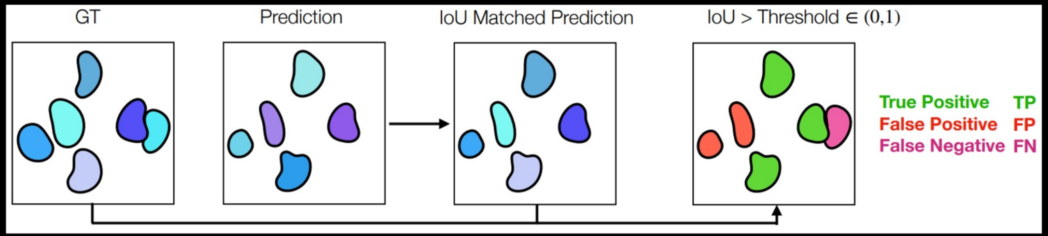
Dice (A, B) = 2*A∩B / (A+B)z





Quantify the accuracy

- Precision, recall, F1 (= (2*precision*recall) / (precision + recall))
- Jaccard index / Dice coefficient
- Hausdorff distance
- Mean Squared Error (L2)
- Structural similarity index (SSIM)
- Biologically relevant measures (cell densities, fluorescence intensities, diameters)



https://github.com/maweigert/neubias_academy_stardist
Schmidt, Weigert et al 2018

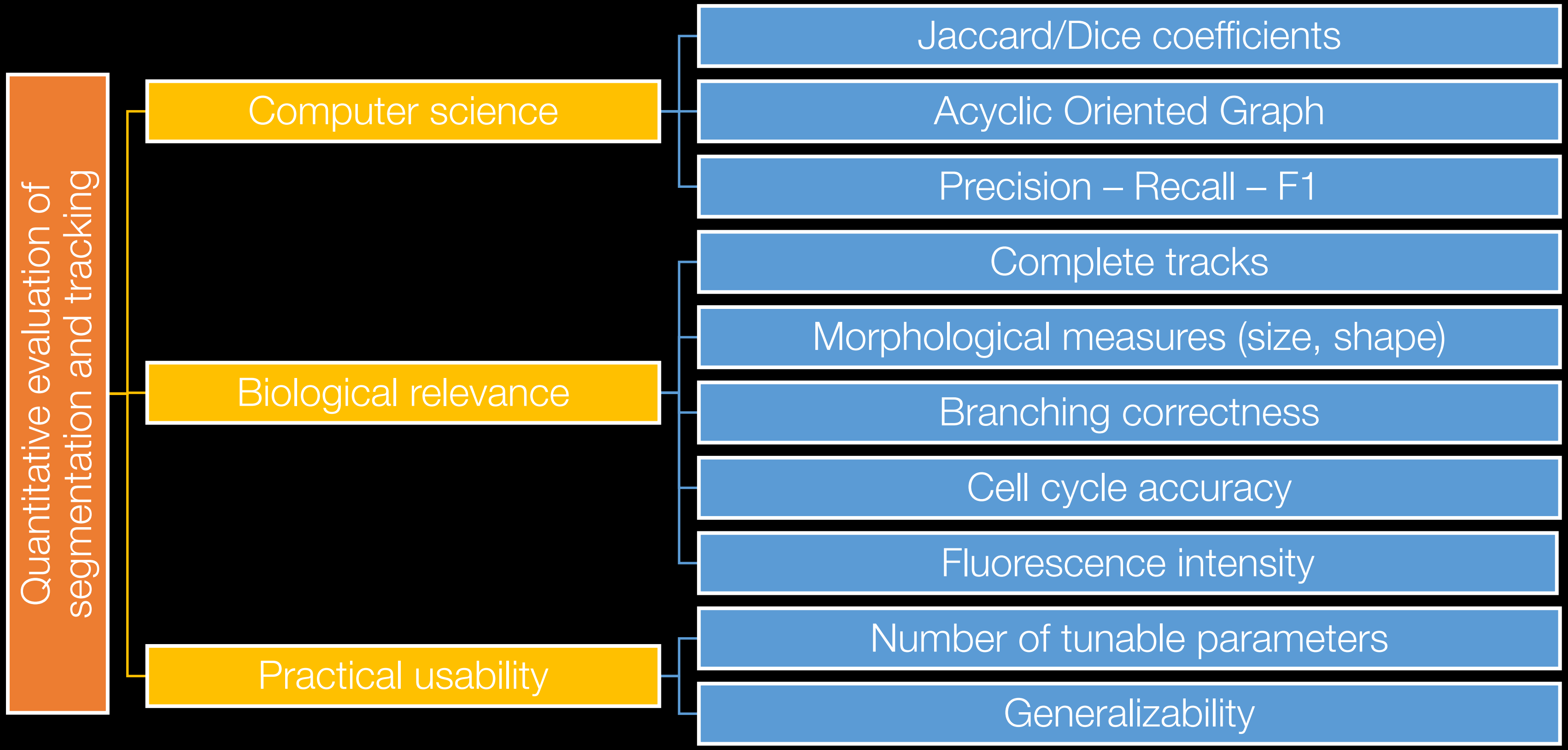
Biological relevance of the segmentation results

Segmentation result

Real object

✅ Accuracy (overlap)

❌ Length, curvature, diameter



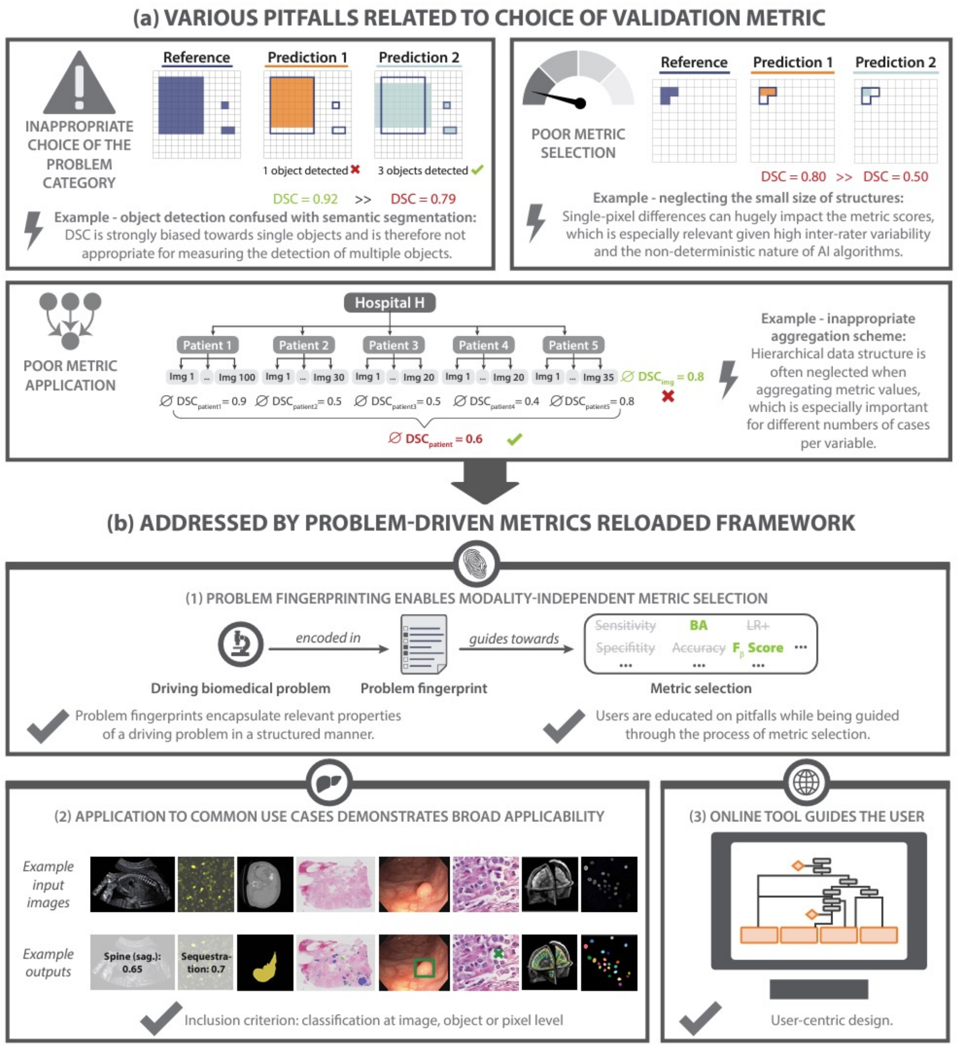
Lena Maier-Hein, Metrics reloaded: Recommendations for image analysis validation, https://arxiv.org/abs/2206.01653, arXiv 2022 (last update June 2023)

Computer Science > Computer Vision and Pattern Recognition

[Submitted on 3 Jun 2022 (v1), last revised 30 Jun 2023 (this version, v6)]

Metrics reloaded: Recommendations for image analysis validation

Lena Maier-Hein, Annika Reinke, Patrick Godau, Minu D. Tizabi, Florian Büttner, Evangelia Christodoulou, Ben Glocker, Fabian Isensee, Jens Kleesiek, Michal Kozubek, Mauricio Reyes, Michael A. Riegler, Manuel Wiesenfarth, A. Emre Kavur, Carole H. Sudre, Michael Baumgartner, Matthias Eisenmann, Doreen Heckmann-Nötzel, A. Tim Rädsch, Laura Acion, Michela Antonelli, Tal Arbel, Spyridon Bakas, Arriel Benis, Matthew Blaschko, M. Jorge Cardoso, Veronika Cheplygina, Beth A. Cimini, Gary S. Collins, Keyvan Farahani, Luciana Ferrer, Adrian Galdran, Bram van Ginneken, Robert Haase, Daniel A. Hashimoto, Michael M. Hoffman, Merel Huisman, Pierre Jannin, Charles E. Kahn, Dagmar Kainmueller, Bernhard Kainz, Alexandros Karargyris, Alan Karthikesalingam, Hannes Kenngott, Florian Kofler, Annette Kopp-Schneider, Anna Kreshuk, Tahsin Kurc, Bennett A. Landman, Geert Litjens, Amin Madani, Klaus Maier-Hein, Anne L. Martel, Peter Mattson, Erik Meijering, Bjoern Menze, Karel G.M. Moons, Henning Müller, Brennan Nichyporuk, Felix Nickel, Jens Petersen, Nasir Rajpoot, Nicola Rieke, Julio Saez-Rodriguez, Clara I. Sánchez, Shravya Shetty, Maarten van Smeden, Ronald M. Summers, Abdel A. Taha, Aleksei Tiulpin, Sotirios A. Tsaftaris, Ben Van Calster, Gaël Varoquaux, Paul F. Jäger



Fingerprint name	Fingerprint illustration	Fingerprint description
Image processing category identified by category mapping		Semantic segmentation (SemS): assignment of one or multiple category labels to each pixel.
Domain interest-related properties (selection)		
Particular importance of structure boundaries		The biomedical application requires exact structure boundaries. *Example: segmentation for radiotherapy planning; knowledge of exact structure boundaries is crucial to destroy the tumor while sparing healthy tissue.* Important: Overlap-based metrics do not measure shape agreement. In the case of complex shapes (high boundary-to-volume ratio) it is therefore typically advisable to set this property to TRUE.
Particular importance of structure center (e.g., in cells, vessels)		The biomedical application requires accurate knowledge of structure centers. *Example: cell centers are subsequently used for cell tracking and cell motion characterization, so false center movement should be suppressed.*
Compensation for annotation imprecisions requested	Ref 1 Ref 2 / Ref 1 Ref 2	The reference annotation is typically only an approximation of the (forever unknown) ground truth. It may be desirable to compensate for known uncertainties, such as intra-rater or inter-rater variability, by configuring the metric accordingly. This is only possible for some metrics.
•••	••• •••	•••
Target structure-related properties (selection)		
Small size of structures relative to pixel size		Structures of the provided class are only a few pixels in size. *Example: multiple sclerosis lesions in magnetic resonance imaging (MRI) scans.*
High variability of structure sizes (within an image and/or across images)		The target structures vary substantially in size, such that some structures are several times the size of others. *Example: polyps in colonoscopy screening, where some polys are several times the size of others. Counterexample: large organs, such as the liver or the kidneys, which are relatively comparable in size across individuals.*
•••	••• •••	•••
Data set-related properties (selection)		
Presence of class imbalance		The class prevalences differ substantially. *Example: In a screening application, the positive class (e.g., cancer) may occur extremely rarely. In this case, prevalence-dependent metrics, such as Accuracy, may be extremely misleading.*
Non-independence of test cases		The test cases are hierarchically structured, indicating non-independence of test cases. *Examples: multiple images of the same patient, hospital or video.*
•••	••• •••	•••
Algorithm output-related properties (selection)		
Possibility of algorithm output not containing the target structure(s)	Pred / Pred	The algorithm may yield outputs in which not all classes are present.
•••	••• •••	•••

Considerations: The objective

Accuracy versus validity

Example:

High segmentation accuracy but

poor temporal consistency

→ Limit object tracking

Segmentation

Real object

✅ Fluorescence average value

❌ Length, curvature, diameter, shape

Mistaken objective

Example:

Diagnosis of diabetic retinopathy.

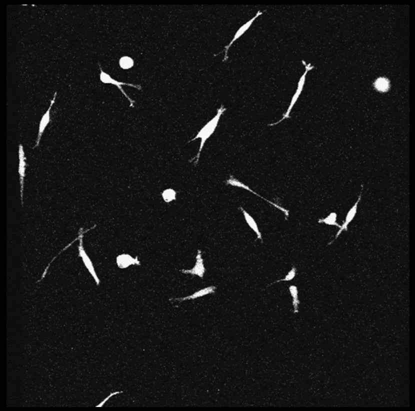
Issues:

1. Discrepancy among doctors and non-valid majority voting
2. Hiden real objective → *"Should this patient see a doctor?"*

Patric Riley, Three pitfalls to avoid in machine learning, Nature 2019

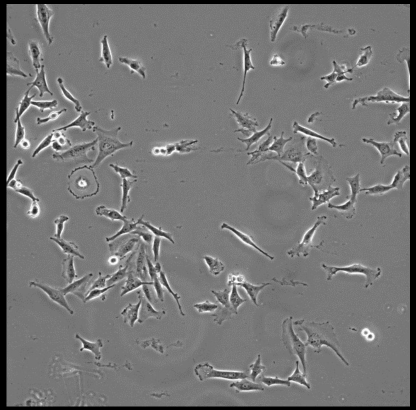
Considerations: Generalizability

DL models are extremely sensitive to pixel sizes (object size), imaging modalities, morphologies, cell types, fluorescence channels…



FluoC3DLMDA231

PhC-C2DL-PSC



Usiigaci

Generalizability is still an active and hot field of study with important open questions:

- Is it because we lack enough training data?
- Should we get deeper models?
- Is it possible to have one single model for a specific task regardless the data?

V Ulman et al., An objective comparison of cell-tracking algorithms, Nature Methods 2017
Hsieh-Fu Tsai et al., Usiigaci: Instance-aware cell tracking in stain-free phase contrast microscopy enabled by machine learning, SoftwareX 2019