FAIRy deep-learning for bioimage analysis

Estibaliz Gómez de Mariscal, EMBO Postdoctoral Fellow

Optical Cell Biology Group (Prof. Ricardo Henriques)

Instituto Gulbenkian de Ciência

@gomez_mariscal

egomez@igc.gulbenkian.pt

Microscopy imaging: be ready to play

Recommendation: Use image processing in the loop & improve the quality of your workflow

Image acquisition	Image preprocessing	Image processing	Quantification	Data analysis

How can we make it accessible?

Open source tools

- Widefield
- Confocal
- SMLM
- DIC
- EM

- High-throughput
- High-content
- Multispectral

- Channel unmixing
- Image classification
- Object detection
- Denoising
- Restoration
- Super-resolution
- Registration

- Segmentation
- Detection
- Classification
- Tracking

- Morphology
- Lineages
- Embedded features
- Data distributions

- Statistics
- Visualisation

All Topics	AGAVE	AICS ImageIO	Aydin	BIAFLOWS	BII
BiofilmQ	Bio-Formats	BoneJ	BrainGlobe	Cellpose	CellProfiler
CLIJ	CytoMAP	Cytomine	DeepLabCut	Fiji	FLIMLib
GerBI	Icy	IDR	ilastik	ImageJ	ImageJ2
ImgLib2	ImJoy	JIPipe	Mars	MCMICRO	MIA
MIB	μManager	MoBIE	MorphoGraphX	MorphoNet	napari
NEUBIAS	OME	OMERO	OmeSliCC	OpenIRIS	OpenSPIM
Orbit	Piximi	PYME	Python-Microscope	QuPath	SCIFIO
scikit-image	sciview	SR-Tesseler	StarDist	starfish	TissUUmaps
vedo	webKnossos	ZeroCostDL4Mic	... Your Icon Here	Zzz Dormant Partners	

KNIME
Open for Innovation

Deep learning playground

Open-source science (technology)   **Reliable**   **User friendly (accessible)**

Open-source science (technology)   **Reliable**   **User friendly (accessible)**

··· 2015   2017   2018

First U-Net development

Papers with code

Our Mission

The mission of Papers with Code is to create a free and open resource with Machine Learning papers, code, datasets, methods and evaluation tables.

CSBDeep

StarDist

DeepCell

README with description of the zip
• C++ and Caffe bianaries

O. Ronneberger & T. Borx, MICCAI 2015

README with instructions for code installation and run

DeepLabCut: a software package for animal pose estimation

E. Gómez-de-Mariscal et al., Sci Reports 2019

4

Open-source science (technology) | Reliable | User friendly (accessible)

... 2015 2017 2018 2019 2020 2021 2022

First U-Net development

Papers with code

README with description of the zip
- C++ and Caffe bianaries

O. Ronneberger & T. Borx, MICCAI 2015

README with instructions for code installation and run

E. Gómez-de-Mariscal et al., Sci Reports 2019

5

Community meeting the challenges for machine learning enabled bioimage analysis

- ImageJ / Fiji plugins
  - CSBDeep
    - Noise2Void
    - DenoiSeg
  - StarDist
  - DeepImageJ
- Ilastik
- Knime
- Icy
- Qupath

- Weka
- CellProfiler
- DeepCell
- Cellpose
- Google Colab*
- ZeroCostDL4Mic
- DeepTrack
- DeepLabCut
- MONAI
- LiveCellMiner
- DeepMIB

Community making it possible

Scientific Community Image fórum: https://forum.image.sc

NEUBIAS Academy@Home: http://eubias.org/NEUBIAS/training-schools/neubias-academy-home/

COBA center (Cimini & Carpenter): https://openbioimageanalysis.org/

iBiology: https://www.ibiology.org/biology-online/

BioImage Informatics Index: https://biii.eu/

CHALLENGES FOR ACCESSIBILITY

Technical Skills

High Computational resources

Data
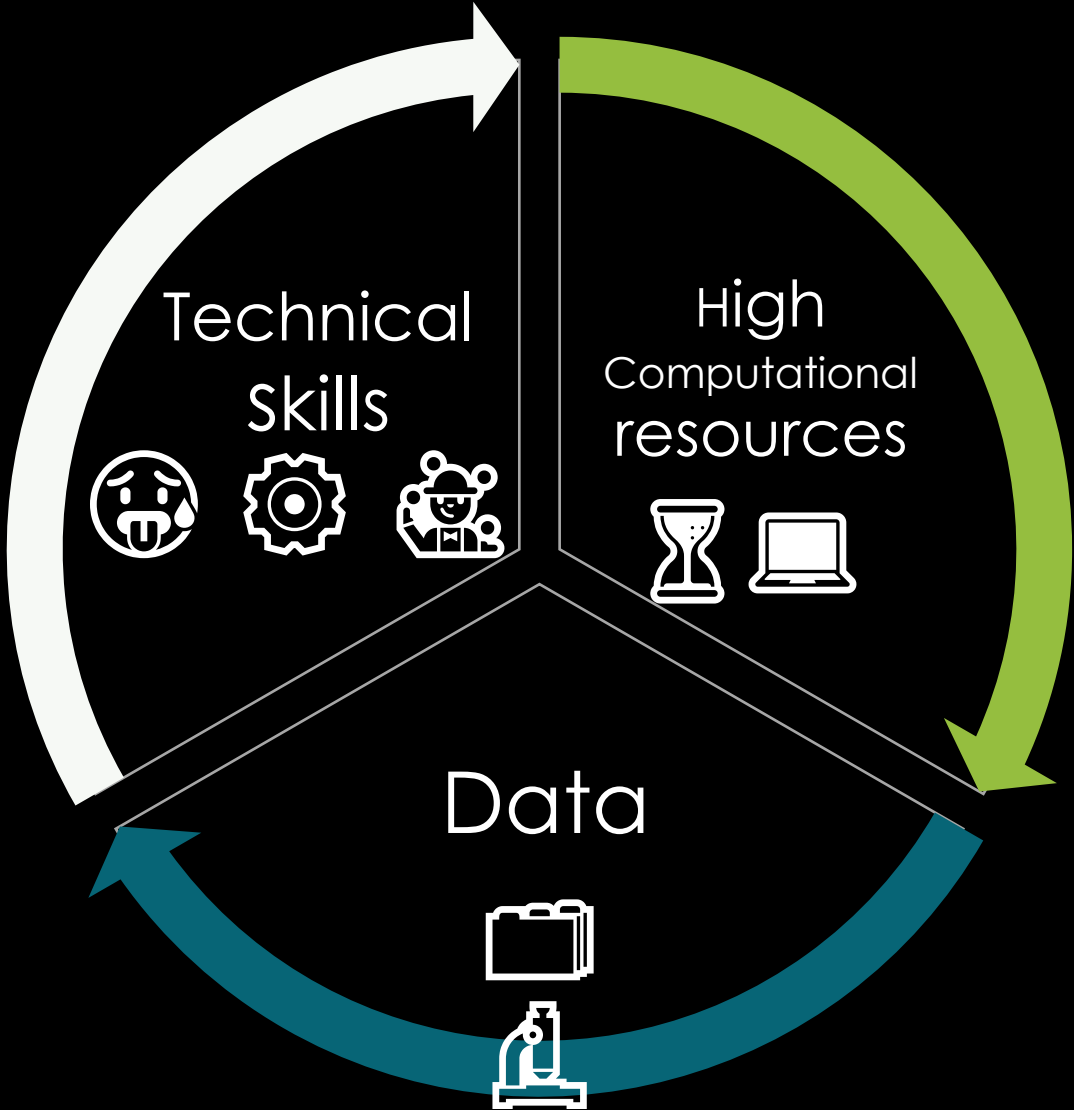
Deep learning:

an extremely hot topic in the field 🔥

... and FAIRy* open-source solutions, the jewel in the crown

*Findable, Accessible, Interoperable and Reusable

FAIRy deep-learning ecosystem

Open-source science (technology)

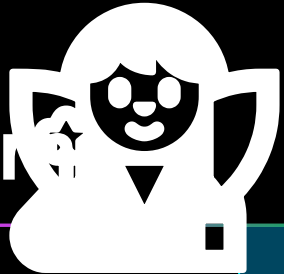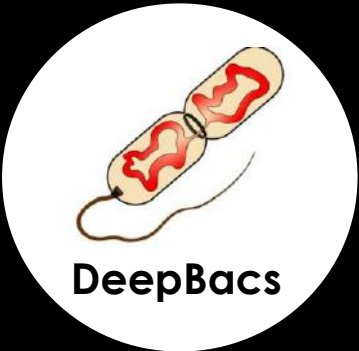|				Standards
deepImageJ	#ZeroCostDL4Mic	DeepBacs	DL4MicEverywhere	BioImage.IO
User friendly frameworks	Resources Computational & Data			Transfer learning

E. Gómez-de-Mariscal*, C. García-López-de-Haro*, et al., Nature Methods 2021
L. von Chamier et al., Nature Communications 2021
O. Wei*, F. Beuttenmueler*, E. Gómez-de-Mariscal*, C. Pape* et al., bioRxiv 2022

C. Spahn, E. Gómez-de-Mariscal, ..., R. Henriques, Communications Biology 2022
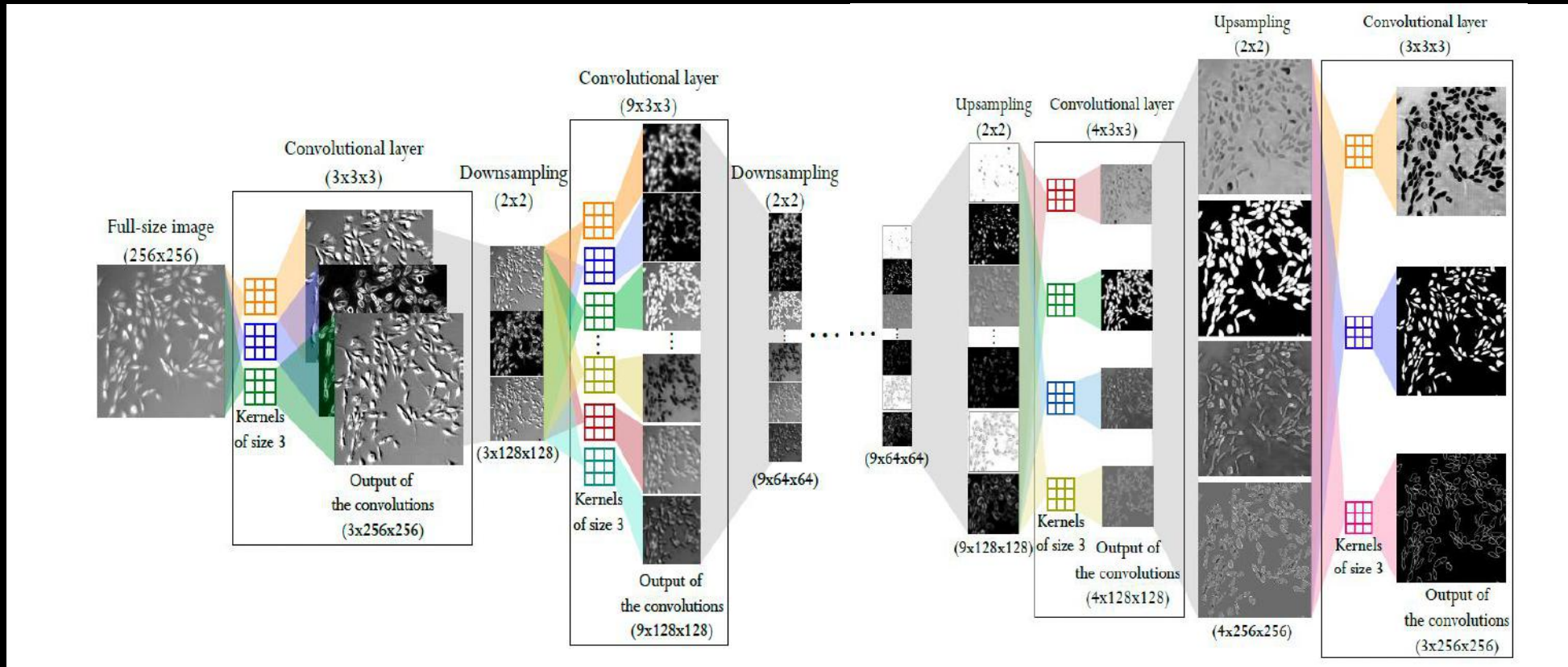I. Hidalgo-Cenalmor et al., bioRxiv 2023

What is a (pre)trained model?

- Architecture
- Trained Weights

> Once trained, the model state is frozen and the inference consists of running a series of static operations (e.g., convolutions, downsampling)
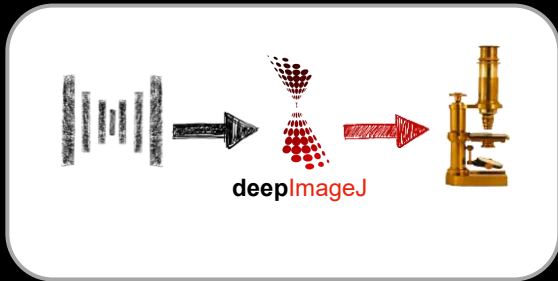
Gómez de Mariscal, E. et al., NEUBIAS Springer, 2022

Back in 2018...

Arrate Muñoz Barrutia

Daniel Sage

deepImageJ

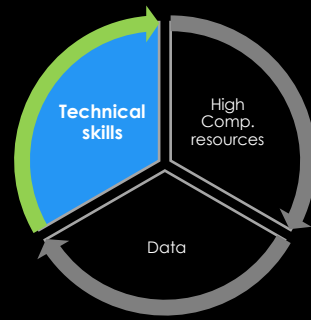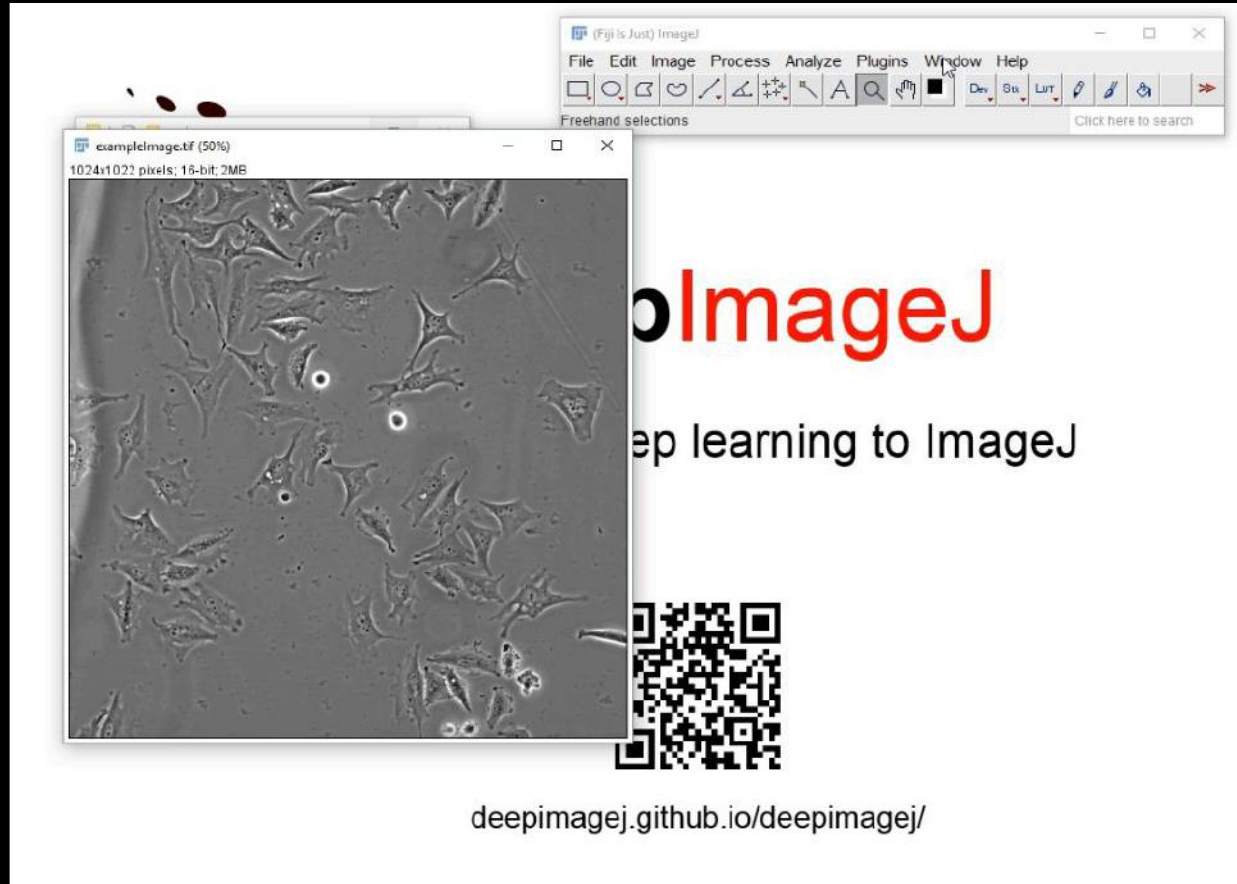*ImageJ is fully equipped for scientific image VISUALIZATION and QUANTIFICATION*

Estibaliz Gómez de Mariscal

Carlos García López de Haro

Wei Ouyang

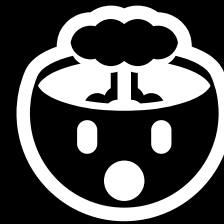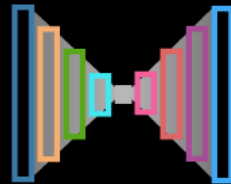Fully equipped for scientific image VISUALIZATION and QUANTIFICATION

10

"How can I apply it to my data?"
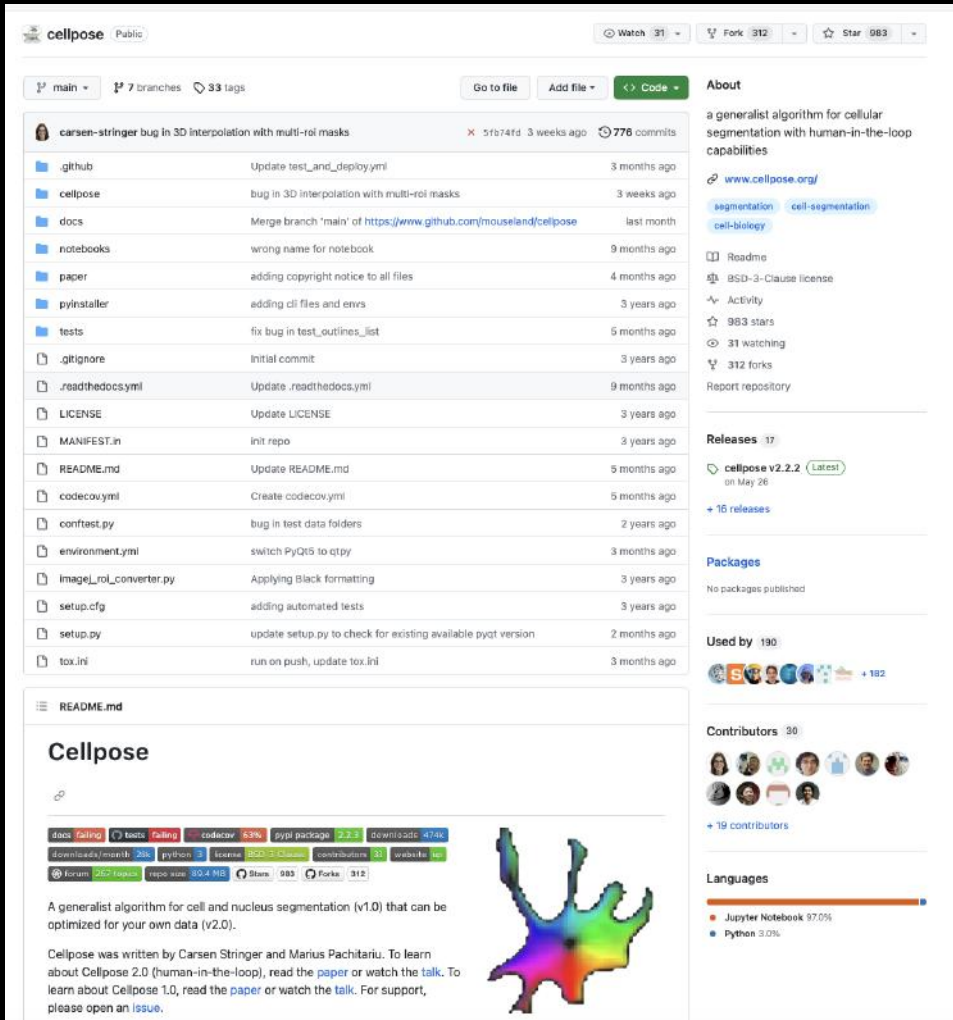
(Re-)training

Fine-tuning

DL-based image processing is mostly about Python code

How easy could it be?

There should be a way to make it easy

ZeroCostDL4Mic: full user-friendly deep learning experience

L. von Chamier et al., Nat Comms 2021
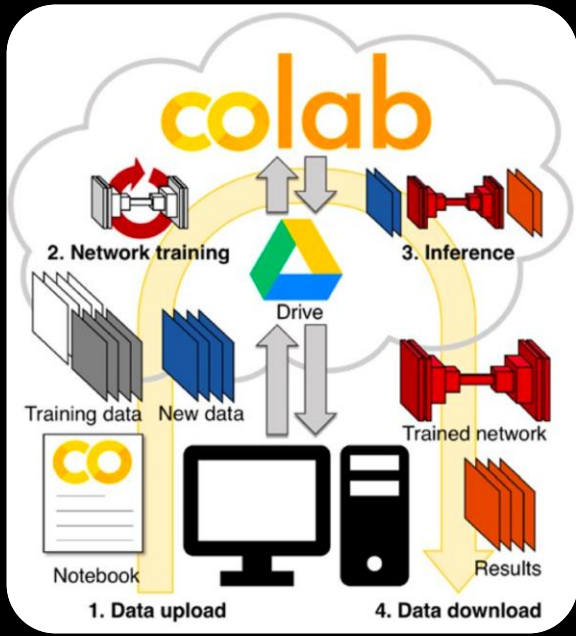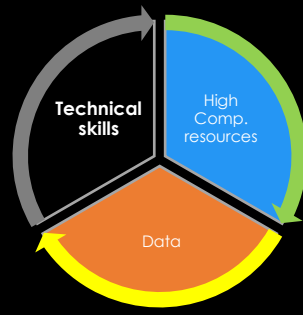
Lucas von Chamier

Romain Laine

Guillaume Jacquemet

Ricardo Henriques

Tasks and networks
implemented by ZeroCostDL4Mic

https://github.com/HenriquesLab/ZeroCostDL4Mic

Unified user-friendly interface

→ No coding

Open in Colab   FREE GPU service

ZeroCostDL4Mic: full user-friendly deep learning experience
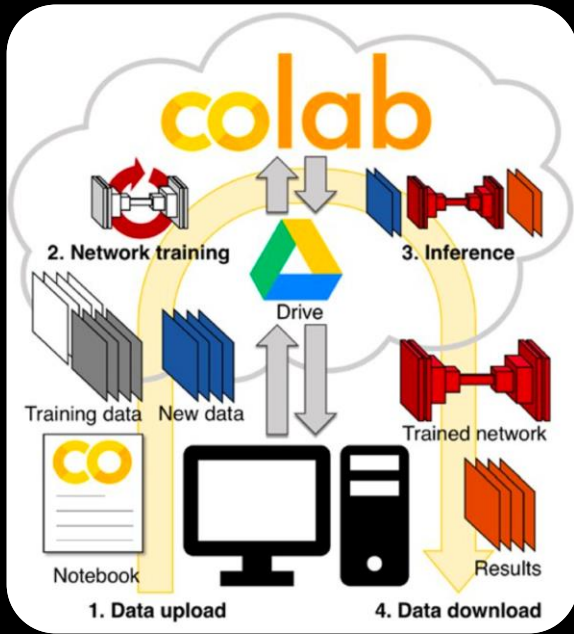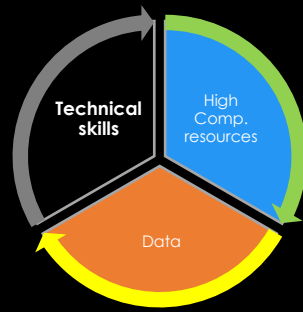
Technical skills

High Comp. resources

Data

Directly open the notebooks from the documentation

2. Network training
3. Inference
Drive
Training data   New data
Trained network
Notebook
Results
1. Data upload
4. Data download

L. von Chamier et al., Nat Comms 2021

Lucas von Chamier

Romain Laine

Guillaume Jacquemet

Ricardo Henriques

Unified user-friendly interface

→ No coding

Open in Colab

FREE GPU service

Segmentation networks

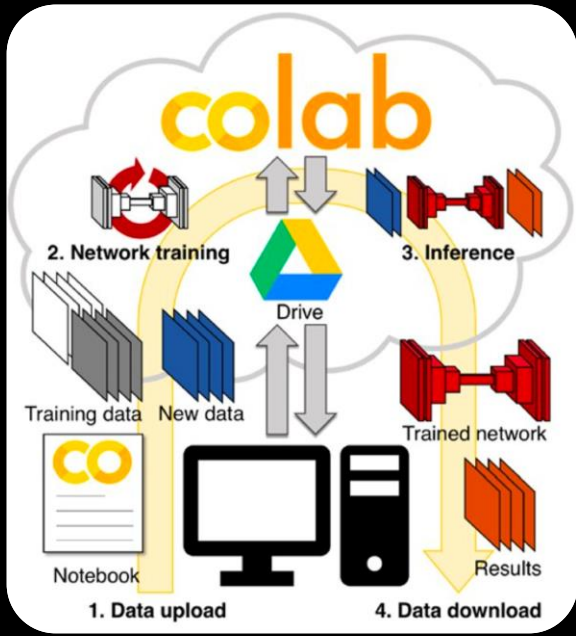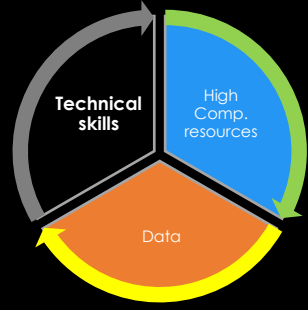| Network	Paper(s)	Tasks	Status	Last test	Link to example training and test dataset	Direct link to the notebook in Colab
U-Net (2D)	here and here	Binary segmentation	Fully supported	27/07/23 ✅ working (IH)	here	Open in Colab
U-Net (3D)	here	Binary segmentation	Fully supported	27/07/23 ✅ working (IH)	EPFL dataset	Open in Colab
U-Net (2D) multilabel	here and here	Semantic segmentation	Under beta-testing	16/07/23 ✅ working (IH)	here	Open in Colab
DenoiSeg	here	Joint denoising and binary segmentation	Fully supported	⚠ broken (no GPU) (GJ)	Available soon	Open in Colab
StarDist (2D)	here and here	Instance segmentation	Fully supported	19/05/23 ✅ working (EGM)	here	Open in Colab
StarDist (3D)	here and here	Instance segmentation	Fully supported	07/10/22 ✅ working (GJ)	from Stardist github	Open in Colab
Cellpose (2D and 3D)	here	Instance segmentation (Cells or Nuclei)	Fully supported	05/09/23 ✅ working	Coming soon!	Open in Colab
SplineDist (2D)	here	Instance segmentation	Fully supported	07/10/22 ✅ working (GJ)	here	Open in Colab
EmbedSeg (2D)	here	Instance segmentation	Under beta-testing	01/01/23 ✅ working (AR)	here	Open in Colab
MaskRCNN (2D)	here	Instance segmentation	Under beta-testing		Coming soon!	Open in Colab
Interactive Segmentation - Kaibu (2D)	here	Interactive instance segmentation	Under beta-testing		Coming soon!	Open in Colab

Denoising and image restoration networks

Network	Paper(s)	Tasks	Status	Last test	Link to example training and test dataset	Direct link to the notebook in Colab
Noise2Void (2D)	here	Self-supervised denoising	Fully supported	14/07/23 ✅ working (IH)	here or here	Open in Colab
Noise2Void (3D)	here	Self-supervised denoising	Fully supported	14/07/23 ✅ working (IH)	here	Open in Colab

Cellpose_2D_ZeroCostDL4Mic.ipynb

Training set up

Path to training images:

training_source: * Insertar texto aquí

training_target: * Insertar texto aquí

Name of the model and path to model folder:

model_name: * Insertar texto aquí

model_path: * Insertar texto aquí

Training Parameters:

Number of epochs:

number_of_epochs: 500

Channel_to_use_for_training: Red

If you have a secondary channel that can be used for training, for instance nuclei, choose it here:

Second_training_channel: None

Advanced Parameters

Use_Default_Advanced_Parameters

Train

4.2. Start Training

When playing the cell below you should see updates after each epoch (round). Network training can take some time.

* CRITICAL NOTE: Training is currently done using Torch.

* CRITICAL NOTE: Google Colab has a time limit for processing (to prevent using GPU power for datamining). Training time must be less than 12 hours! If training takes longer than 12 hours, please decrease the number of epochs or number of patches. Another way circumvent this is to save the parameters of the model after training and start training again from this point.

Start training

Evaluate

Choose the folders that contain your Quality Control dataset

Source_QC_folder: * Insertar texto aquí

Target_QC_folder: * Insertar texto aquí

Channel_to_segment: Red

If you chose the model "cytoplasm" indicate if you also have a nuclear channel that can be used to aid the segmentation.

Nuclear_channel: None

Segmentation parameters:

Object_diameter: 0

Flow_threshold:

mask_threshold:

ZeroCostDL4Mic: full user-friendly deep learning experience

L. von Chamier et al., Nat Comms 2021

Lucas von Chamier

Romain Laine

Guillaume Jacquemet

Ricardo Henriques

DeepBacs

Streamlined deep learning for microbiology

Christoph Spahn

(1) Segmentation
Input   NN

(2) Object detection
Mecillinam   CAM
Nalidixate
Nalidixate   CAM
Vesicle
Control

(3) Denoising
Input   NN

(4) Artificial Labeling
Input   NN

(5) Super-resolution
Input   NN
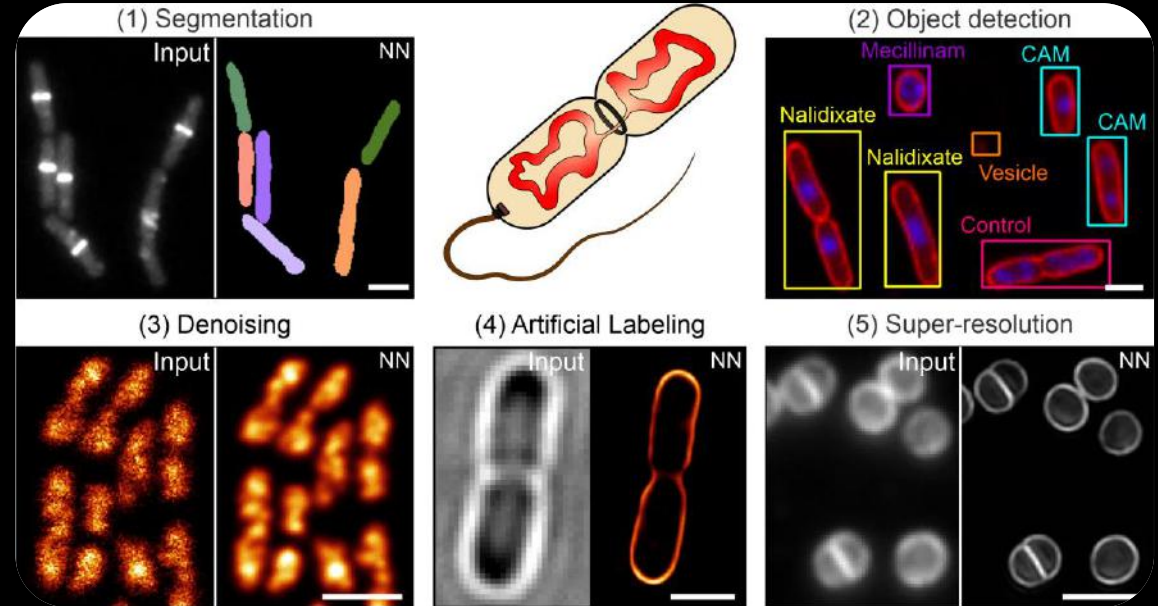
C. Spahn et al., Communications Biology 2022

Unified user-friendly interface

→ No coding

CO Open in Colab   FREE GPU service

DL4MicEverywhere: Deep learning for microscopy made flexible, shareable, and reproducible

I. Hidalgo-Cenalmor   G. Jacquemet   R. Henriques

DL4MicEverywhere

Proprietary platform → Open container initiative

Open source, community driven platforms

Build upon ZeroCostDL4Mic to run DL notebooks everywhere 🌍

Interface → Automated containerization → Safe encapsulation

DL4MicEverywhere

Select:
Notebook
Images folder
Output folder
☐ GPU

Run   Close

IDENTIFY   DEFINE DOCKER IMAGE   RUN CONTAINER

Local   docker hub   Dependencies
🔍 SEARCH   Images   Notebook
Remote   PULL / BUILD   Models

jupyter   local/remote host
▶ Load data
▶ Train model
▶ Quality control
▶ Predict

github.com/HenriquesLab/DL4MicEverywhere/

DEVELOPERS
Easy to deploy, share, showcase, and benchmark

STANDARDS
Reproducible
Transferable
Transparent
FAIR

DL4MicEverywhere

Models
BioImage.IO

USERS   Local / HPC / Cloud
Easy-to-use interface to train, evaluate and use models

DL container images
docker hub

COMMUNITY

Notebooks

Continuous integration

Automatic testing

Bespoke notebooks
Hosted in DL4MicEverywhere
Inspired by ZeroCostDL4Mic

#ZeroCostDL4Mic

External notebooks
DL4MicEverywhere-compatible

I. Hidalgo et al., DL4MicEverywhere: Deep learning for microscopy made flexible, shareable, and reproducible, bioRxiv 2023

FAIRy deep-learning

Findable, Accessible, **Interoperable** and Reusable

BioImage Model Zoo
Advanced AI models in one-click

Wei Ouyang   Florian Jug   Anna Kreshuk

FAIRy deep-learning (Findable, Accessible, Interoperable, Reproducible)

The trained model

The training data

Direct contact with the developer

Spahn C., et al., Comm Biology 2022

The notebook

W. Ouyang *, F. Beuttenmueller*, E Gómez-de-Mariscal*, C. Pape*, et al., bioRxiv 2022

BioImage Model Zoo
Advanced AI models in one-click

Wei Ouyang   Florian Jug   Anna Kreshuk

Interoperability: From the community for the community

Universal deep learning model format

ImJoy #ZeroCostDL4Mic ilastik deepImageJ CSBDeep
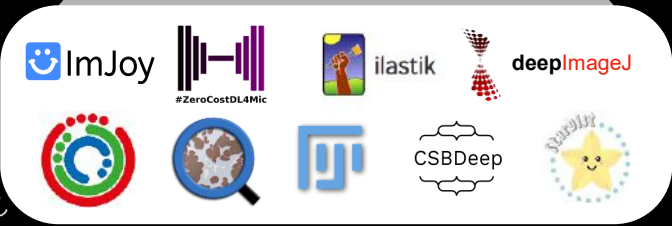
Consumer: **deepImageJ**
segment nuclei in new data

Consumer: **QuPath**
segment, curate, quantify

BioImage.IO
StarDist H&E Nuclei Segmentation
10.5281/zenodo.6338614 chatty-frog
reusable model

programmatic use in notebook
Consumer: **ZeroCostDL4Mic**

programmatic use in python
Consumer: **stardist python**

Technical skills   High Comp. resources   Data

W. Ouyang *, F. Beuttenmueller*, E Gómez-de-Mariscal*, C. Pape*, et al., bioRxiv 2022
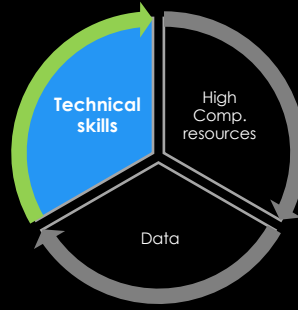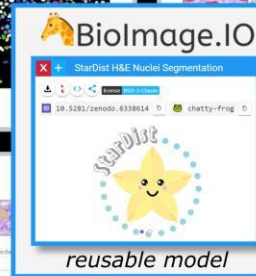
Bringing standards to the FAIRy deep learning ecosystem

Model metadata

Input shape → → Output shape

Normalisation Receptive field Postprocessing

Input tiles — Output tiles

Input image Output image
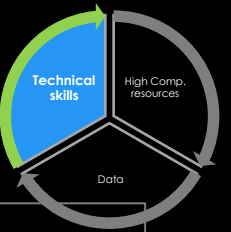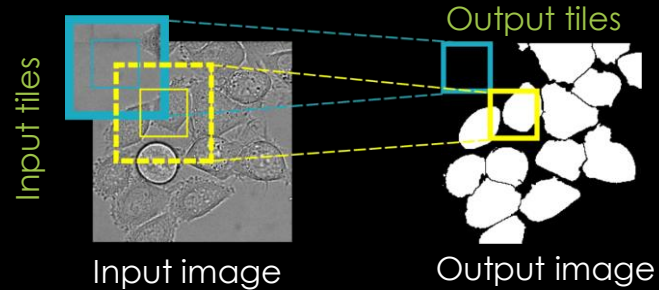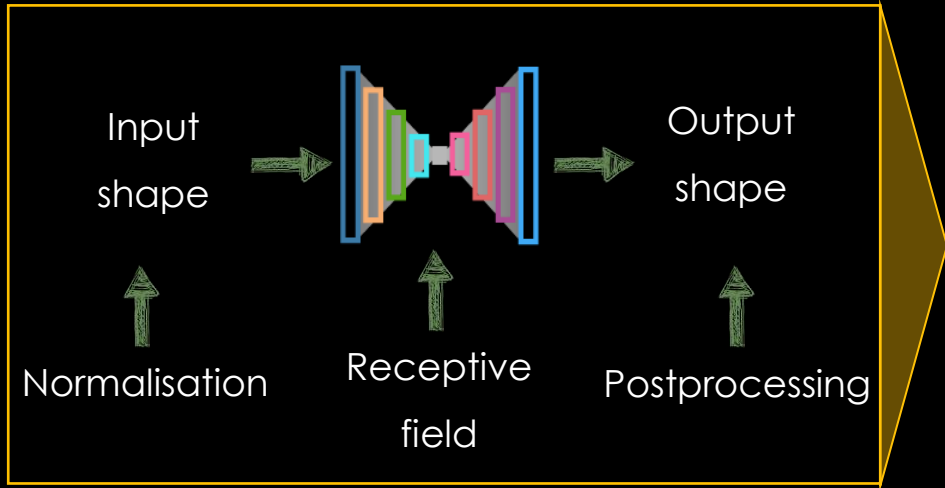
Models tailored for bioimage analysis
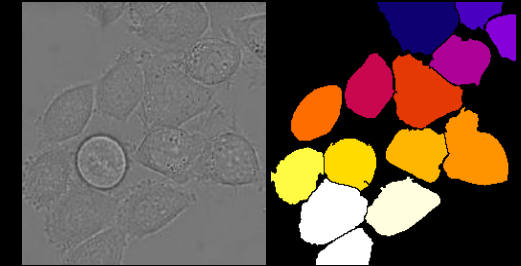
Input example Output example

Metadata

Shape: 256 x 256 pixels

Pixel size: 0.19 μm x 0.19 μm

Common resource specifications

Quick AI feature integrations

Python and Java utilities
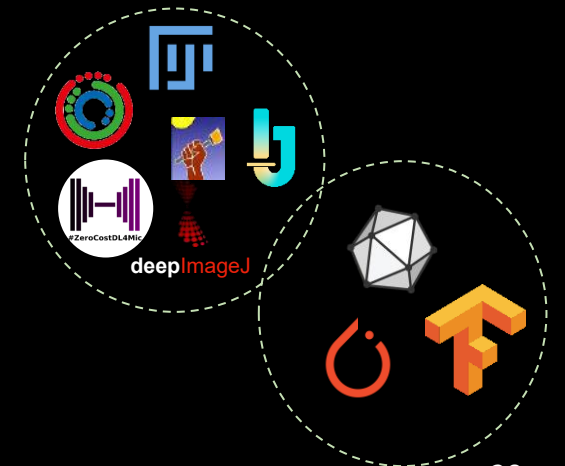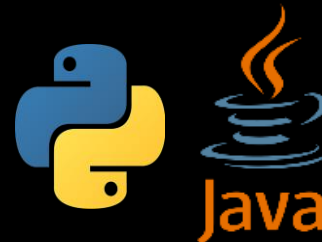
- W. Ouyang *, F. Beuttenmueller*, E Gómez-de-Mariscal*, C. Pape*, et al., bioRxiv 2022
- C. García-López-de-Haro, JDLL: A library to run Deep Learning models on Java bioimage informatics platforms, 2023

Support **reproducibility**

Model fine tuning

GUI for model management

New model contribution

W. Ouyang *, F. Beuttenmueller*, E Gómez-de-Mariscal*, C. Pape*, et al., bioRxiv 2022

Boost data annotation

Download a pretrained model
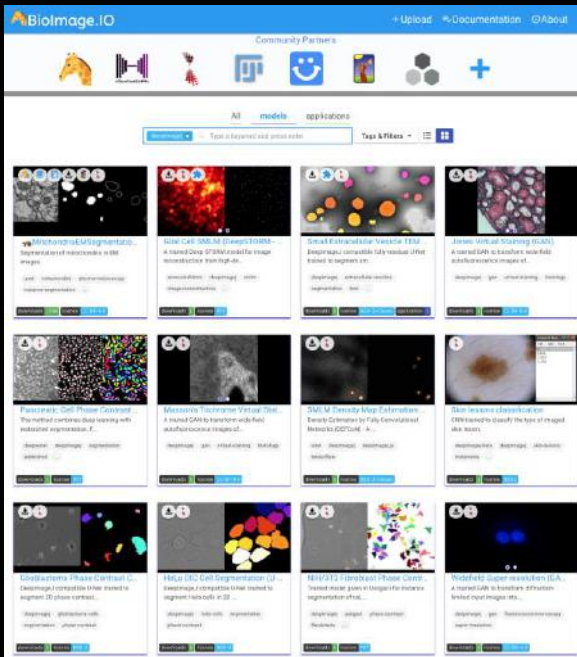
Speed up data annotation

Fine tuning

Analyze

Process an image
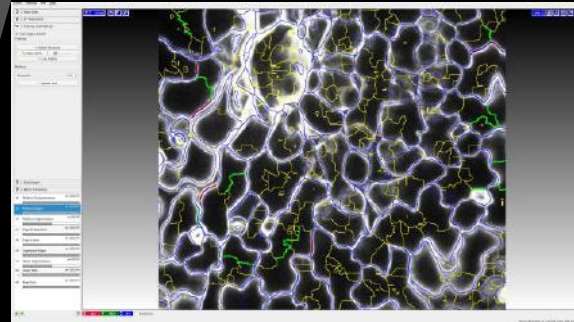
Curate the results

New training data

W. Ouyang *, F. Beuttenmueller*, E Gómez-de-Mariscal*, C. Pape*, et al., bioRxiv 2022
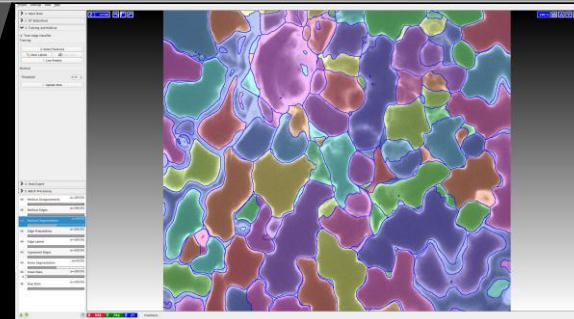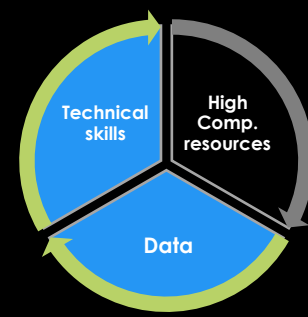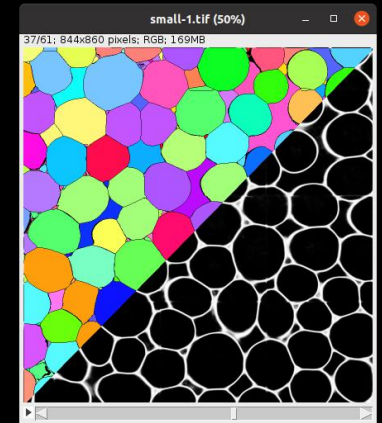
The potential of pre-trained models and interoperability
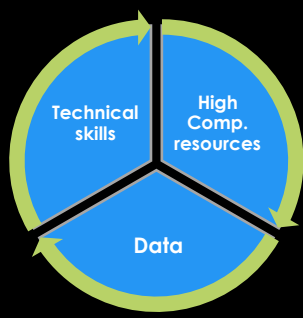
Time lapse movies of 3D volumes
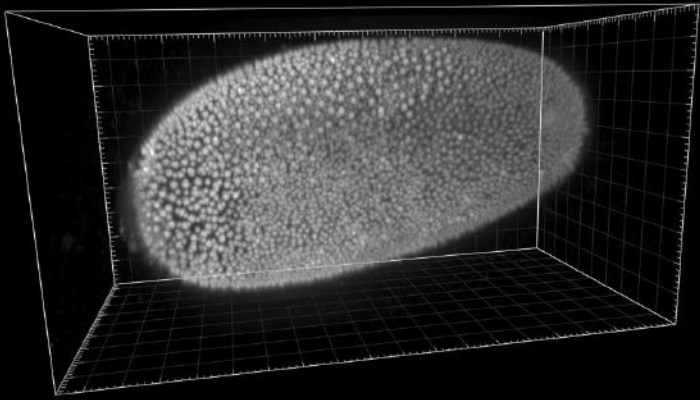
Considerations when annotating data

~300 GB of data

> 300 (3D) nuclei per time frame

2D PARTIAL Annotations (Cell Tracking Challenge)

+

pre-trained model
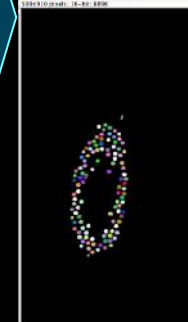
Reconstruct 3D nuclei

DL segmentation + 3D connected components

Ulman, V. et al., Nature Methods 2017
Uwe Schmidt et al., MICCAI 2018
Legland, D., Arganda-Carreras, I., & Andrey, P., Bioinformatics, 2016
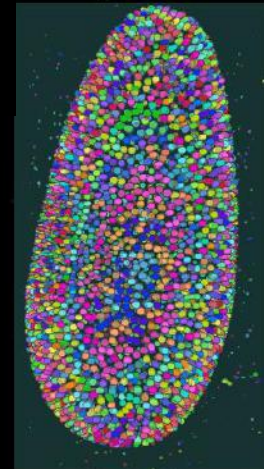
The potential of pre-trained models: domain adaptation
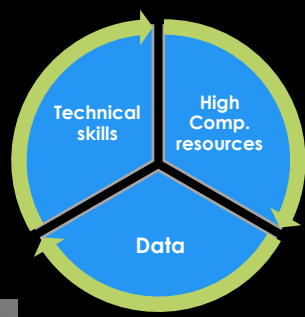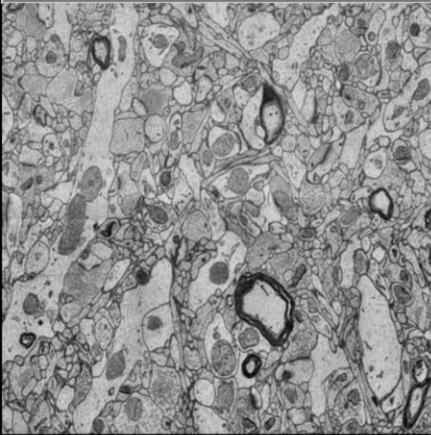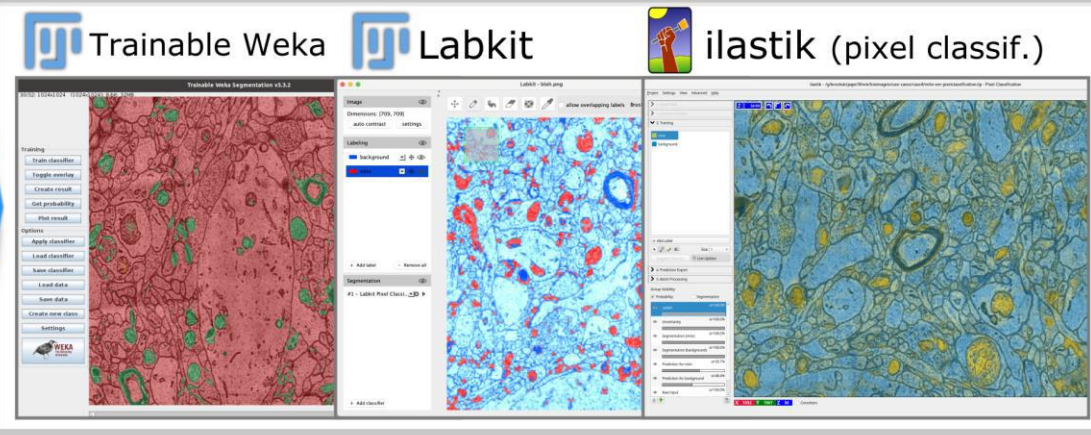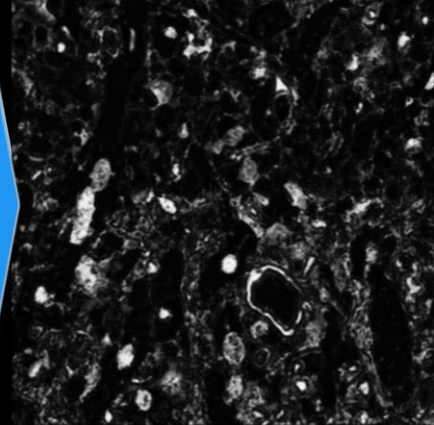
Download a trained enhancer model from the BioImage Model Zoo
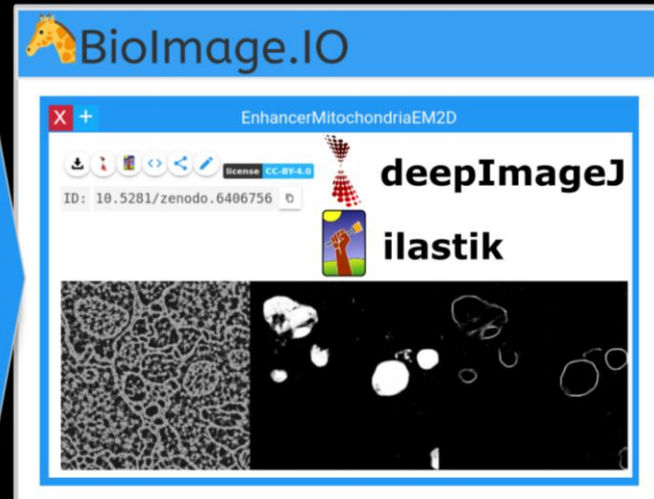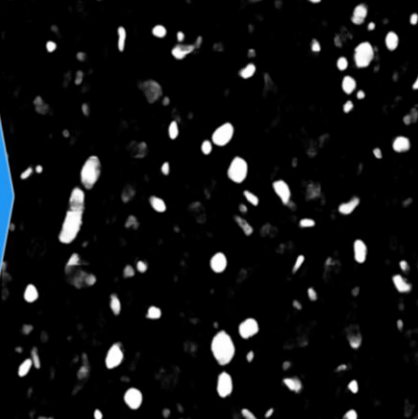
What's the vibe in the community?
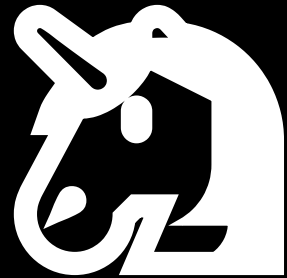
INFRASTRUCTURE FOR LIFE SCIENCES

10 PARTNERS **8** COUNTRIES

3 YEARS **4M** EUROS

AI4Life

AI MODELS AND METHODS FOR THE LIFE SCIENCES

Community Partners

BioImage.IO · #ZeroCostDL4Mic · deepImageJ · ImJoy · ilastik · HPA

Project Partners

EMBL · EMBL-EBI · EMBRC EUROPEAN MARINE BIOLOGICAL RESOURCE CENTRE · EMPHASIS · EURO-BIOIMAGING · eu-openscreen · HUMAN TECHNOPOLE · INSTITUTO GULBENKIAN DE CIÊNCIA · instruct ERIC · KTH VETENSKAP OCH KONST · uc3m Universidad Carlos III de Madrid

LIFE SCIENCE COMMUNITY SOFTWARE AI MODELS STANDARDS AI METHODS COMPUTER SCIENCE COMMUNITY

What's the vibe?

What's the vibe?

Tech-in-the-lab

Seeking for an empowered life-sciences community

Data analysis on the loop

Optimised integration of DL into imaging

Computer Scientists: Awareness of reusability to speed up

- Benchmarking

- Scientific breakthrough

Vladimir Ulman   Martin Maska

nature methods

Analysis   https://doi.org/10.1038/s41592-023-01879-y

The Cell Tracking Challenge: 10 years of objective benchmarking

Martin Maška [1], Vladimír Ulman [1,2], Pablo Delgado-Rodriguez [3,4], Estibaliz Gómez-de-Mariscal [3,4,5], Tereza Nečasová [1], Fidel A. Guerrero Peña [6,7], Tsang Ing Ren [6], Elliot M. Meyerowitz [8], Tim Scherr [9], Katharina Löffler [9], Ralf Mikut [9], Tianqi Guo [10], Yin Wang [10], Jan P. Allebach [10], Rina Bao [11,12], Noor M. Al-Shakarji [12], Gani Rahmon [12], Imad Eddine Toubal [12], Kannappan Palaniappan [12], Filip Lux [1], Petr Matula [1], Ko Sugawara [13,14], Klas E. G. Magnusson [15], Layton Aho [16], Andrew R. Cohen [16], Assaf Arbelle [17], Tal Ben-Haim [17], Tammy Riklin Raviv [17], Fabian Isensee [18,19], Paul F. Jäger [19,20], Klaus H. Maier-Hein [19,21], Yanming Zhu [22,23], Cristina Ederra [24], Ainhoa Urbiola [24], Erik Meijering [22], Alexandre Cunha [7], Arrate Muñoz-Barrutia [3,4], Michal Kozubek [1,25] & Carlos Ortiz-de-Solórzano [24,25]

2021: Reusability guidelines (optional)

ISBI 2022: 11 participants

- All top performing methods

- 100% followed reusability guidelines
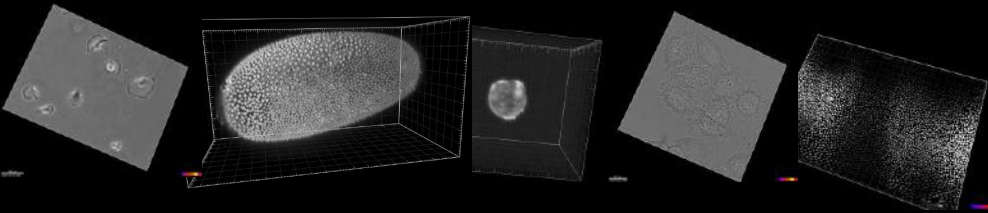
V. Ulman & M. Maska et al., Nat Methods 2017
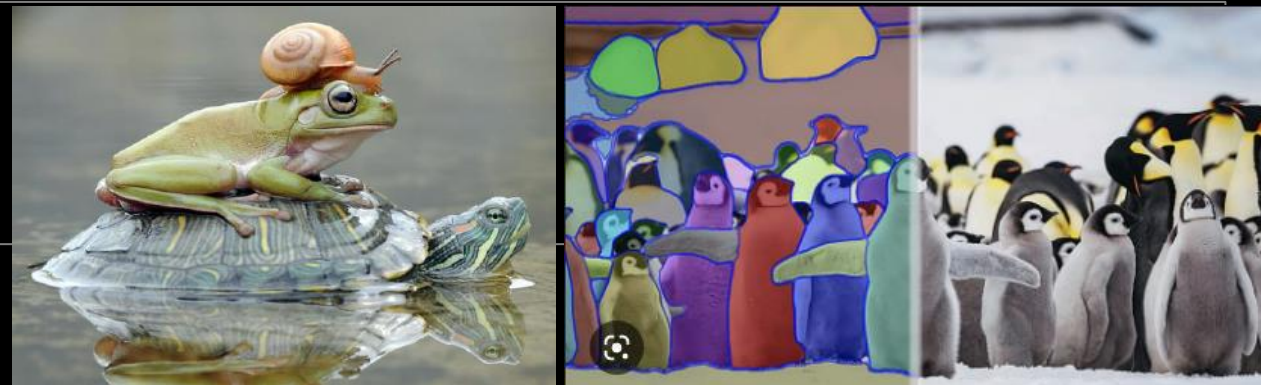
What's the vibe?

Tech-in-the-lab

Seeking for an empowered life-sciences community

Data analysis on the loop

Optimised integration of DL into imaging

Computer Scientists: Awareness of reusability to speed up

- Benchmarking

- Scientific breakthrough

Scientists & researchers: Automatically guided analysis

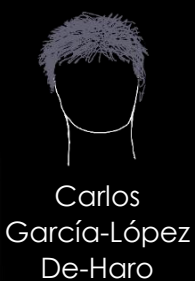- Embedding non-digested scientific knowledge

  (sparse information)

https://github.com/computational-cell-analytics/micro-sam

A. Kirillov et al., Segment Anything, arXiv 2023

deepImageJ team

Universidad Carlos III de Madrid (uc3m)

EPFL

■ EPFL Center for Imaging

KTH VETENSKAP OCH KONST

Carlos García-López De-Haro

Lucía Moya

Wei Ouyang

Arrate Muñoz Barrutia

Daniel Sage

Donati, Laurène, EPFL

Unser, Michael, EPFL

Lundberg, Emma, SciLifeLab, KTH

#ZeroCostDL4Mic

DeepBacs

Prof. R. Henriques

HENRIQUES LAB

INSTITUTO GULBENKIAN DE CIÊNCIA
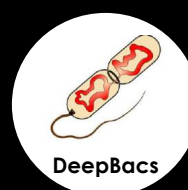
G. Jacquemet

R.F. Laine

C. Spahn

P.M. Pereira

BioImage Model Zoo

- Pengo, Thomas, University of Minnesota
- M. Gordaliza, Pedro, UC3M
- Yair Rivenson, UCLA
- Hongda, Wang, UCLA
- Tosi, Sébastien, IRB Barcelona

- Eglinger, Jan, FMI Basel
- Reuden, Curtis LOCI Lab
- Schmidt, Deborah, MDC
- Olivo-Marin, Jean-Christophe, Institut Pasteur
- Tivenez, Jean-Yves, Institut Pasteur