

Bio-Image Data Science Training School 2024

Anja Neumann, Christian Martin, Dušan Praščević, Jan Ewald, Laura Žigutytė,
Marie-Sophie von Braun, Matthias Täschner, Robert Haase

GEFÖRDERT VOM



Bundesministerium
für Bildung
und Forschung



Diese Maßnahme wird gefördert durch die Bundesregierung aufgrund eines Beschlusses des Deutschen Bundestages. Diese Maßnahme wird mitfinanziert durch Steuermittel auf der Grundlage des von den Abgeordneten des Sächsischen Landtags beschlossenen Haushaltes.

Safety first



Welcome @ ScaDS.AI

Center for Scalable
Data Analytics and
Artificial Intelligence



UNIVERSITÄT
LEIPZIG



TECHNISCHE
UNIVERSITÄT
DRESDEN

Missions:

Excellent Research scalable Data Science and Artificial Intelligence

Responsible, Trustworthy, Efficient

Transfer of AI-methods in applications

Training of the next generations of AI experts

Facts

- one of 5 new AI research centers in Germany
- 50/50 financed by Federal and State Government
- About >200 Employees in various research areas and disciplines
- **Localer / regional partner for Research**
 - Basic science / expertise
 - AI & Data Science in real-world applications
- **History**
 - since Oct. 2014 Big-Data-Centre ScaDS Dresden/Leipzig
 - since Nov. 2019: AI- and Data-Science-Centre ScaDS.AI
 - since 2020 financial support through the State of Saxony

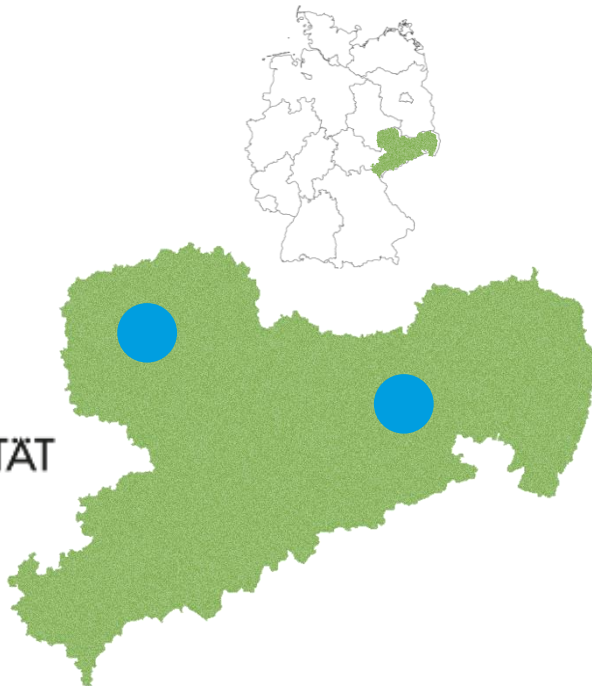


www.humboldt-foundation.de

Partners



UNIVERSITÄT
LEIPZIG



TECHNISCHE
UNIVERSITÄT
DRESDEN



MAX-PLANCK-INSTITUT
FÜR KOGNITIONS- UND NEUROWISSENSCHAFTEN



InfAI[®]
Institut für Angewandte Informatik



Fraunhofer
IAIS



Institute of Software Methods
for Product Virtualization



Leibniz-Institut
für ökologische
Raumentwicklung



CBG
Max Planck Institute
of Molecular Cell Biology
and Genetics

ScaDS.AI
DRESDEN LEIPZIG



TECHNISCHE
UNIVERSITÄT
DRESDEN



UNIVERSITÄT
LEIPZIG

Research Areas

Applied AI &
Big Data

AI Algorithms
& Methods

Big Data Analysis

Topics

Life-Science and Medicine

Earth & Environmental Science

Software Engineering

Physics / Chemistry

Engineering / Business

Understanding Language

Methods and Hardware for Neuro-Inspired Computing

Graph-based Artificial Intelligence

Knowledge Representation

Scalable Visual Computing

Distributed, Efficient Learning

Mathematical Foundations & Statistical Learning

Big Data Analysis

Open Data & Open Models

Data Quality & Data Integration

Cross-cutting topics

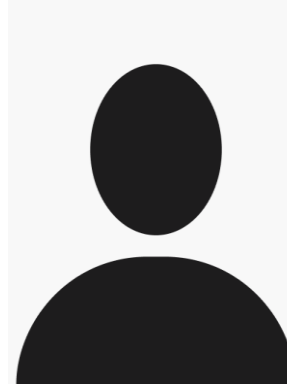
Responsible AI: Ethical and social Dimensions

Architectures / Scalability / Security

Trainers & Supporters



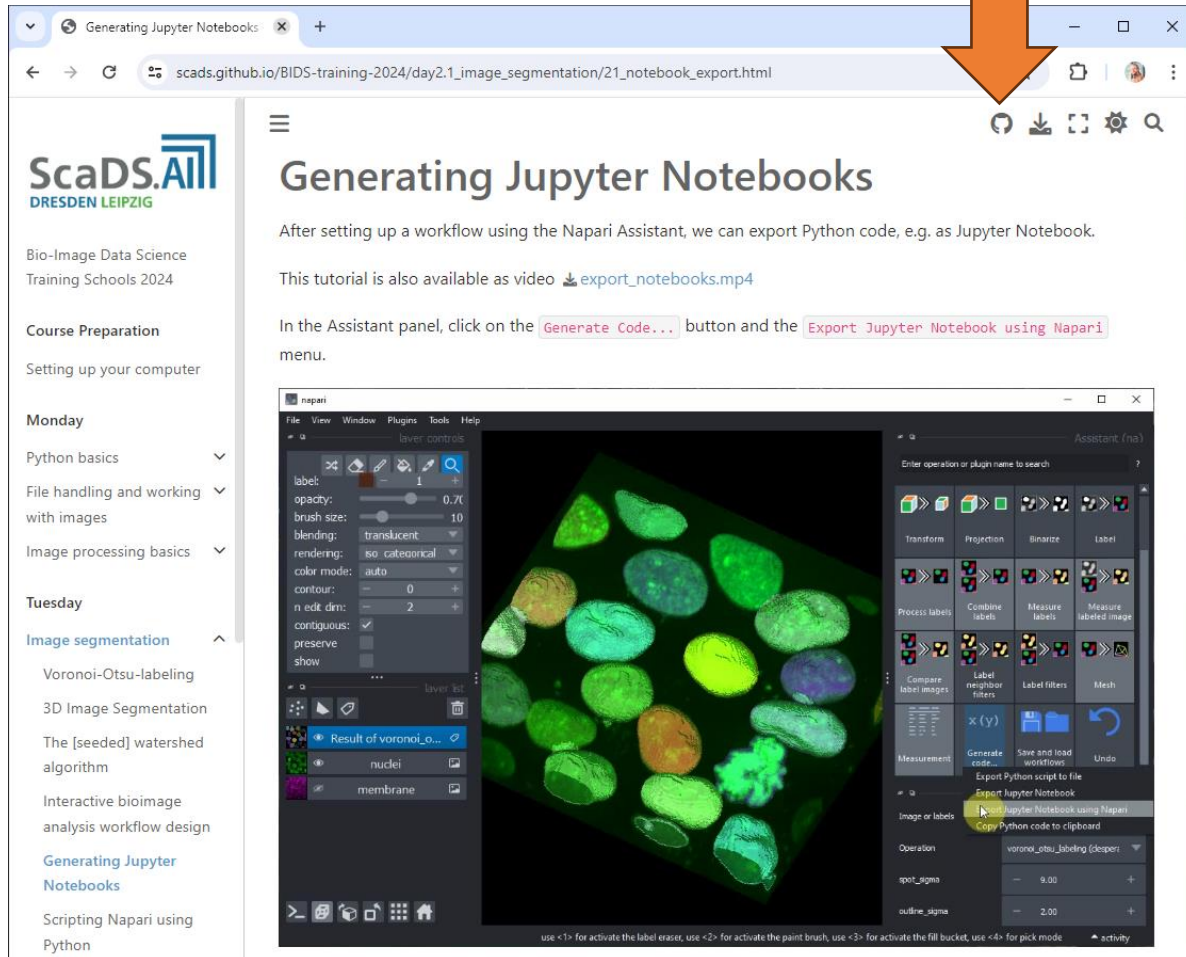
Anja Neumann, Christian Martin, Dušan Prašević, Jan Ewald, Laura Žigutytė, Marie-Sophie von Braun, Matthias Täschner, Robert Haase



Mara Lampert, Susann Tilger, Nicole Spieß, Marvin Rosé

Training Materials

Download the materials to
a place where you can
easily find them



ScaDS.ALI
DRESDEN LEIPZIG

Bio-Image Data Science
Training Schools 2024

Course Preparation
Setting up your computer

Monday

- Python basics
- File handling and working with images
- Image processing basics

Tuesday

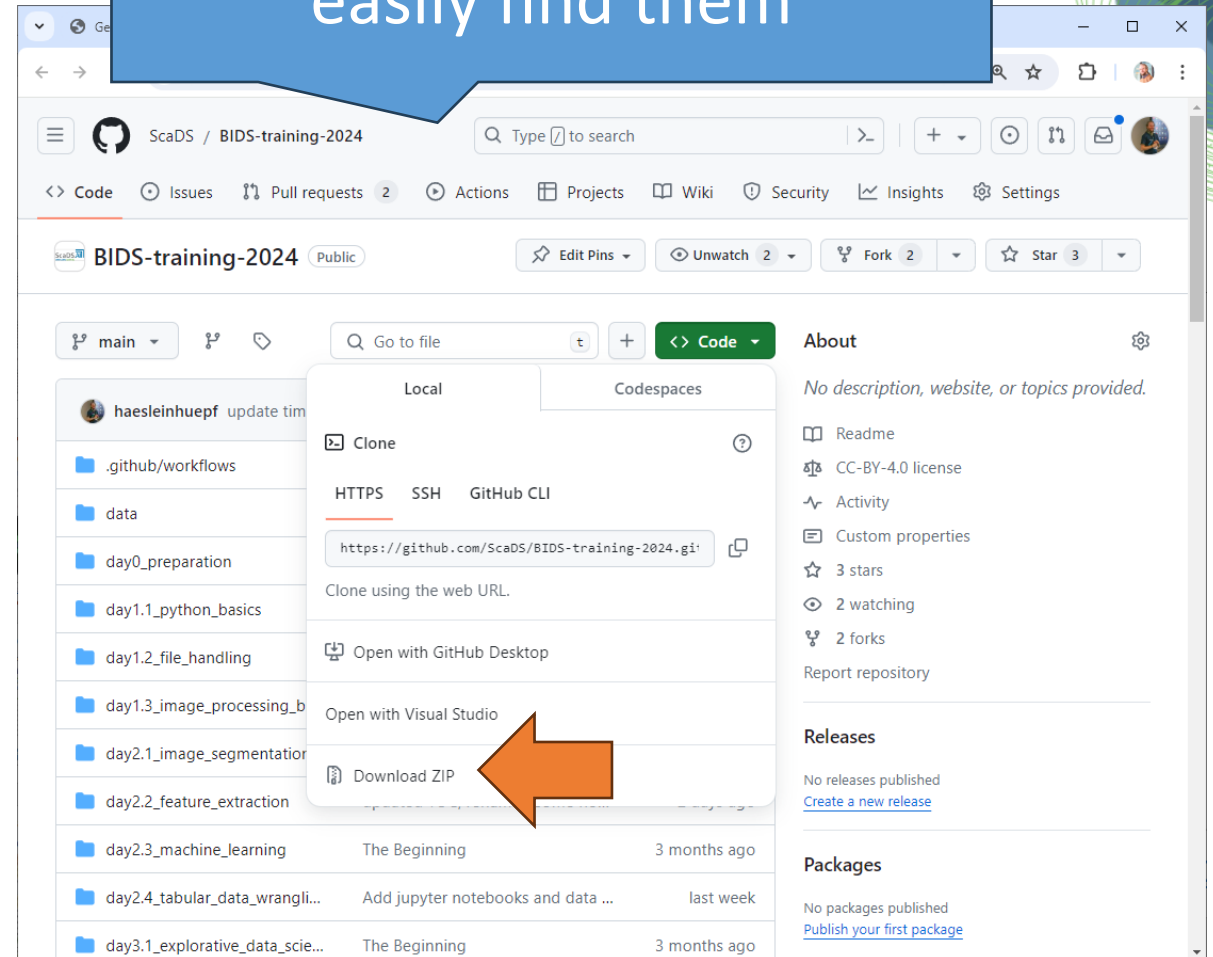
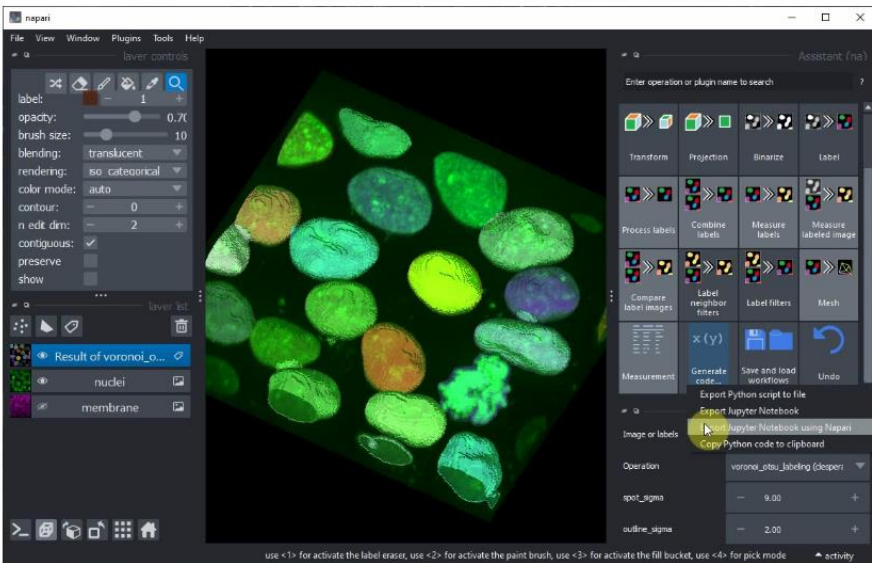
- Image segmentation
 - Voronoi-Otsu-labeling
 - 3D Image Segmentation
 - The [seeded] watershed algorithm
 - Interactive bioimage analysis workflow design
 - Generating Jupyter Notebooks
 - Scripting Napari using Python

Generating Jupyter Notebooks

After setting up a workflow using the Napari Assistant, we can export Python code, e.g. as Jupyter Notebook.

This tutorial is also available as video [export_notebooks.mp4](#)

In the Assistant panel, click on the **Generate Code...** button and the **Export Jupyter Notebook using Napari** menu.



ScaDS / BIDS-training-2024

Type to search

<> Code Issues Pull requests 2 Actions Projects Wiki Security Insights Settings

BIDS-training-2024 Public

Edit Pins Unwatch 2 Fork 2 Star 3

main Go to file Code

haesleinhuepf update time

- .github/workflows
- data
- day0_preparation
- day1.1_python_basics
- day1.2_file_handling
- day1.3_image_processing_b
- day2.1_image_segmentation
- day2.2_feature_extraction
- day2.3_machine_learning
- day2.4_tabular_data_wrangli...
- day3.1_explorative_data_scie...

Local Codespaces

Clone

HTTPS SSH GitHub CLI

<https://github.com/ScaDS/BIDS-training-2024.git>

Clone using the web URL.

Open with GitHub Desktop

Open with Visual Studio

Download ZIP

About

No description, website, or topics provided.

Readme

CC-BY-4.0 license

Activity

Custom properties

3 stars

2 watching

2 forks

Report repository

Releases

No releases published

[Create a new release](#)

Packages

No packages published

[Publish your first package](#)

Schedule

This is a
crash course!



- Monday, May 13th

9:00-10:30	Arrival	
10:30-11:00	Welcome, organizational (Robert)	
11:00-12:30	Python basics (Matthias, Robert)	
12:45-13:45	Lunch	
13:45-15:15	File handling + Image loading/visualization	(Robert, Matthias)
15:15-15:45	Coffee	
15:15-17:15	Image Processing basics (Anja, Marie-Sophie)	
17:15-18:00	Work on your own data / discuss with experts	

- Tuesday, May 14th

9:00-10:30	Image Segmentation (Robert, Matthias)	
10:30-11:00	Coffee	
11:00-12:30	Feature extraction (Robert, Christian)	
12:45-13:45	Lunch	
13:45-15:15	Machine learning (Pixel + Object Classification) (Christian, Anja)	
15:15-15:45	Coffee	
15:45-17:15	Tabular data wrangling using Pandas	(Matthias, Christian)
17:15-18:00	Wrap up & work on your own data	
19:00-21:00	Scientific Networking Event	(Matthias, Robert)

- Wednesday, May 15th

9:00-10:30	Explorative data science, unsupervised machine learning (Laura, Matthias)	
10:30-11:00	Coffee	
11:00-12:45	Data Visualization theory (Jan, Laura)	
12:45-13:45	Lunch	
13:45-15:15	Plotting using Seaborn (Marie-Sophie, Jan)	
15:15-15:45	Coffee	
15:15-17:15	Wrap up & work on your own data	
17:15-18:00	Departure	

Quick survey

- How often do you use Python for data analysis / plotting / statistics?

Never



Monthly

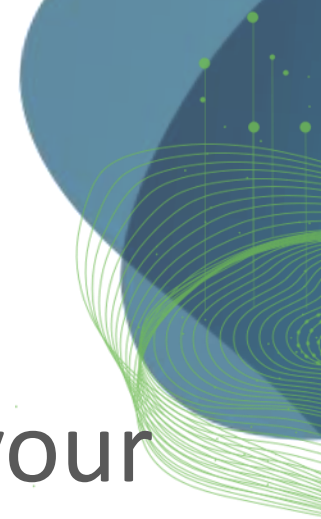


Weekly



Daily



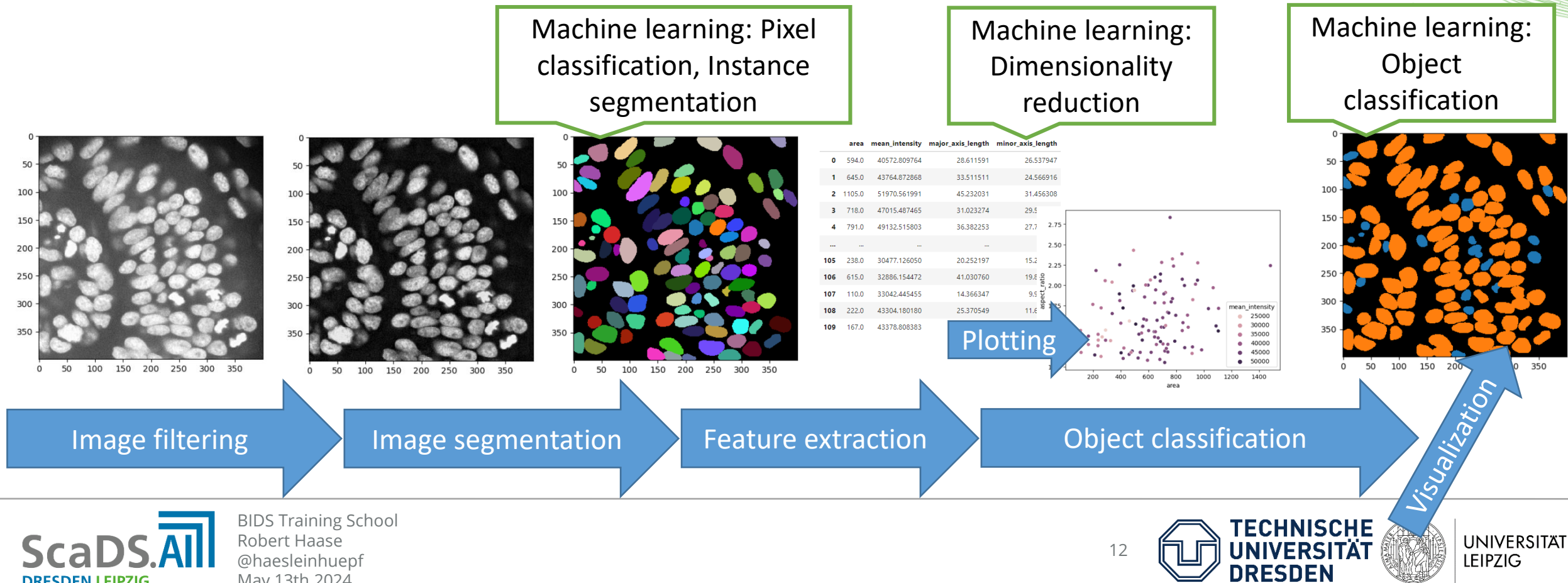


Open points

- Photos will be taken during the event – for PR purposes. Let us know if you don't want to see your face on the internet.
- Dinner tomorrow: Who joins NOT ?

Overview: Bio-image Data Science

- [Image] Data Analysis workflows
- Goal: **Quantify observations, substantiate conclusions with numbers**

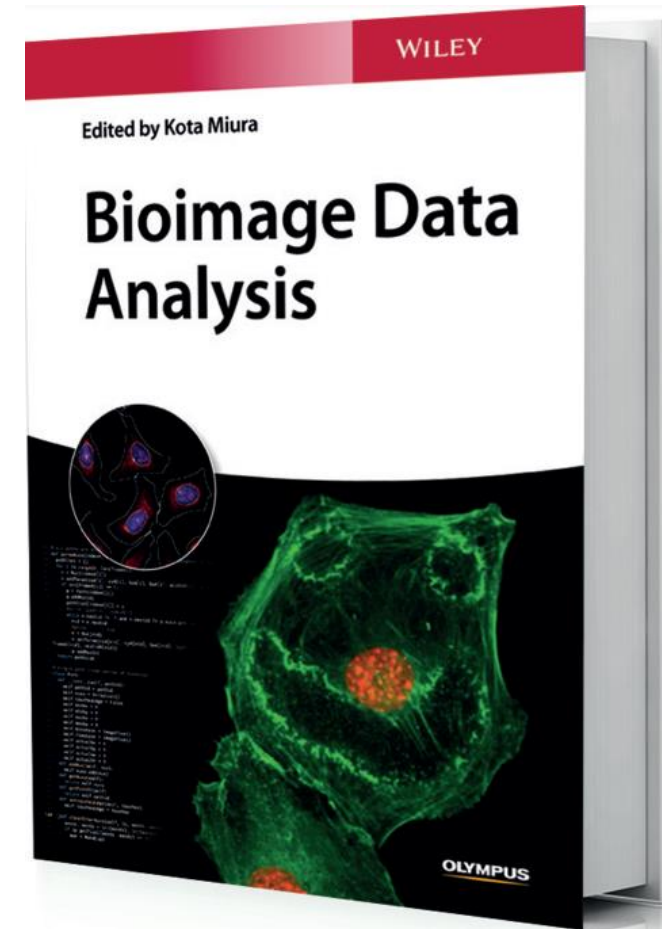


Bioimage Analysis

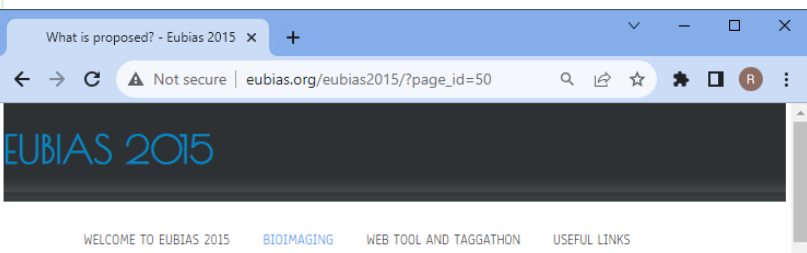
- Kota Miura & Sebastien Tosi 2015:

In the light of this definition, image analysis, which is also called “computer vision,” aims at mimicking the way we see the world and how we identify its visible structures. Image analysis in biology does undeniably also hold this element, but more importantly, its main goal is to *measure* biological structures and phenomena in order to study and understand biological systems in a quantitative way.

To achieve this task, we in fact do not have to be bothered with similarity to the human recognition – we have more emphasis on the objectivity of quantitative measurement, rather than how that computer-based recognition becomes in agreement with human recognition. Therefore, in biology, image analysis is a process of identifying spatial distribution of biological components in images and measuring their characteristics to study their underlying mechanisms in an unbiased way. To underline this difference in the goals of image analysis in the two fields and to distinguish them from each other, we will now on refer to image analysis in biology as *bioimage analysis*.



BIAS Symposia & Training Schools



WHAT IS PROPOSED?

Increasing the Accessibility to Open Software

How could we increase the accessibility and diffusion of image processing / analysis infrastructures, especially in terms of software packages? Existing software packages already offer a rich and wide range of tools for processing and analysis of microscope images. They are constantly evolving and the number of new functions is steadily growing. With so many options offered to users, the problem of accessibility is already excellent, but the enhancement and strengthening of a collaborations / consultations / teaching between the availability of software from multiple angles. At these reasons, we propose the following:

A Web Tool

A list of software packages is already in actual research scenarios we want to it is contained. Many software packages associated documentations should become a web platform that integrates all "do wheel" and may at the same time change functional and highly interactive.

- One page per Function

Meetings

Meeting will be a place to collectively spread and absorb image processing and analysis activities. In terms of increasing the software packages / tools accessibility, three different types of meetings should be organized with clearly defined target groups (developers, analysts and users) and directed flow of information between these groups. As a byproduct meetings strengthen human networks that trigger individual communication to be smoother and more informal, which allows specific solutions to problems by connecting developers, analysts and users. We include teaching oriented activities, or courses, as meetings since although the format differs, we think that the functionality is common in terms of increasing the accessibility to software packages.

"Show Case" meetings (Developers and Analysts)

Software packages upgrade on daily basis and it is quite an effort to follow changes and new functions. An annual meeting with presentations explaining upgrades given by a developer from each software packages would be a great benefit for analysts to be updated with the latest changes. We call this a "Show Case" meeting. In the same meeting, analysts would provide feedback by showing their usages, explaining merits and demerits of the software package and also to request missing functionalities. This meeting also invites prominent developers who are creating cutting edge algorithms so that the progress in the front line of image processing could also be known to developers and to analysts.

Image Analysis Courses (Analysts teaching Users)

In image analysis courses, analysts teach users on how to combine various tools to tackle practical problems, in order to narrow the gap between software packages and user's demands by increasing the image analysis literacy of users. The BIAS 2013 course could be a template for such courses.

Several outcomes will be expected from this activity. The first is that users become more fluent with image analysis, which heightens their ability to choose and assemble appropriate tools. The second is that analysts learn from other analysts how they solved some specific problems and also possibly get aware of new problems they might be expected to solve in the future. The third is that by working together the community of analysts becomes strengthened, leading to an increase in lateral communications.

Courses could target two different types of user group: the first group is graduate students, postdocs and research scientists. The teaching aims that they could reach a level to do basic analysis on their own. The second group is microscope facility staffs. As analysts are still missing in many places, microscope facility staffs are taking the role of analysts beside the maintenance of microscopes. For this group of people the course teaches more advanced contents and also provides and assists teaching in their own institutes (Meta-teaching).

General Assembly (Developers, Analysts, Users)

A general assembly is a gathering of all developers, analysts and users. During this gathering, hackathon presentations



Luxembourg 2019



Goth

- EuBIAS 2013 -

Barcelona, October 7-12, 2013

European BioImage Analysis Symposium

EURO-BIOIMAGING

Welcome!

3613 days since EuBIAS 2013

Navigation

- Home
- Introduction
- BioImage Community
- What is proposed?
- Community Meeting EuBIAS 2013
 - Meeting REPORT
 - Program
 - Registration
 - Invited Speakers/Participants
- Webtool: Taggathon
 - Protocol
 - Schedule
 - Forum
- BioImage Data Analysis Course
 - Planning
 - Registration
 - Instructors
- Venue & Contact
- Gold Sponsors
 - EURO-BIOIMAGING
 - OME

- EuBIAS 2013 -

European BioImage Analysis Symposium

Barcelona, October 7-12, 2013

A Multi-purpose Symposium for the BioImage Analysis Community

The Symposium at a Glance

Image analysis is powerful and essential in modern biology. However, many people involved in bioimage analysis might share the following problems:

- Which tool is appropriate to address my question?
- Who should I ask my very specific question about image analysis?
- I want to increase my skill, but I do not know how.
- As a developer I want to know more about the demand of biologists.
- I became an image analysts specialist. Is there a community for me?

To overcome these problems, we decided in June 2013 to organize this meeting in Barcelona to strengthen the network among those involved in bioimage analysis and provide direct solutions, the meeting is organized in 3 events:

1. Open Community Meeting (2 days): Top-Developers, leading analysts and biologists sit together to address the previous issues and exchange information to figure out possible solutions. Anyone is welcome to join to get solid information on the current scene of bioimage analysis and share point of views.
2. The open community meeting is followed by a course targeting microscopy facility staffs to disseminate the knowledge and techniques of image analysts to the scientific community.
3. In parallel to the course, we will start up to build a public webtool that is expected to evolve into a practical solution centralizing information on image analysis tools and pipelines. The tools will be initially tagged by the

12th of October:

The Symposium is Over !!

An incredible event and widely acknowledged success!

More info, reports, textbooks, Tag clouds, Webtool to come soon !!!

Before that, we truly thank all speakers, analysts, students and taggers for making this meeting a thrilling event:

Leipzig 2024

Heidelberg EUBIAS

Barcelona 2016

Barcelona EUBIAS 2013



Funded by:



<http://simplemaps.com/resources/svg-europe>
<http://eubias.org/eubias2013>
<http://eubias2013.irbbarcelona.org>
https://twitter.com/gomez_mariscal/status/1656428422011187200
<https://twitter.com/StRigaud/status/1664693461310754827>



Sources:

BIAS Symposia & Training Schools

- Training schools around the globe



Paula Sampaio @pm_sampaio · May 9

2nd day TS2 defragmentation starts with GPU-accelerated image processing by Marcelo Zoccoler
[@NEUBIAS_](#) [@Eosclife](#) [@EuroBioImaging](#) [@AI4LifeTeam](#)



9



32



5,491



Jean-Yves Tinevez @jytinevez · Jun 6

One week ago our first Pasteur NEUBIAS course started and finished last Friday. As we are wrapping up and tidying the place, it's time to say thanks!



Stephane Rigaud @StRigaud · Jun 2

Final clap for the #NeubiasPasteur23.
It was a blast 🙌



1



6



34

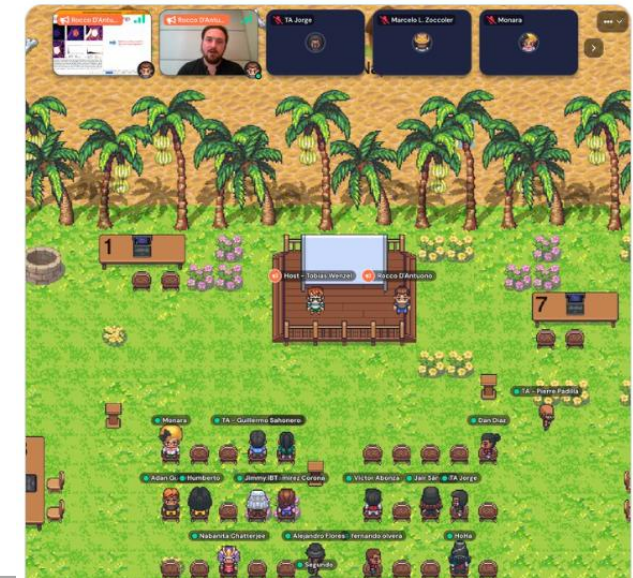


2,421



Tobias Wenzel @MakerTobey

Soooo excited about our ongoing python [@napari_imaging](#) image data science workshop for [@LatamBioimaging](#) [#LIBREhub](#).
Today we learn again from the best! [@zoccolermarcelo](#), [@RogerDantuono](#) & [@jnuneziglesias](#)
All free and open source! [librehub.github.io/napari-LatAm-w...](#)
[@czscience](#) [#bioimaging](#)

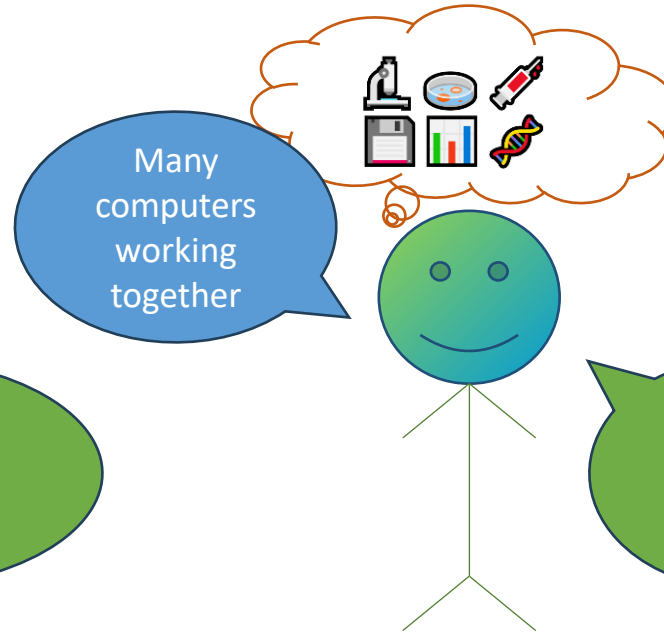


Biolmage Analysis



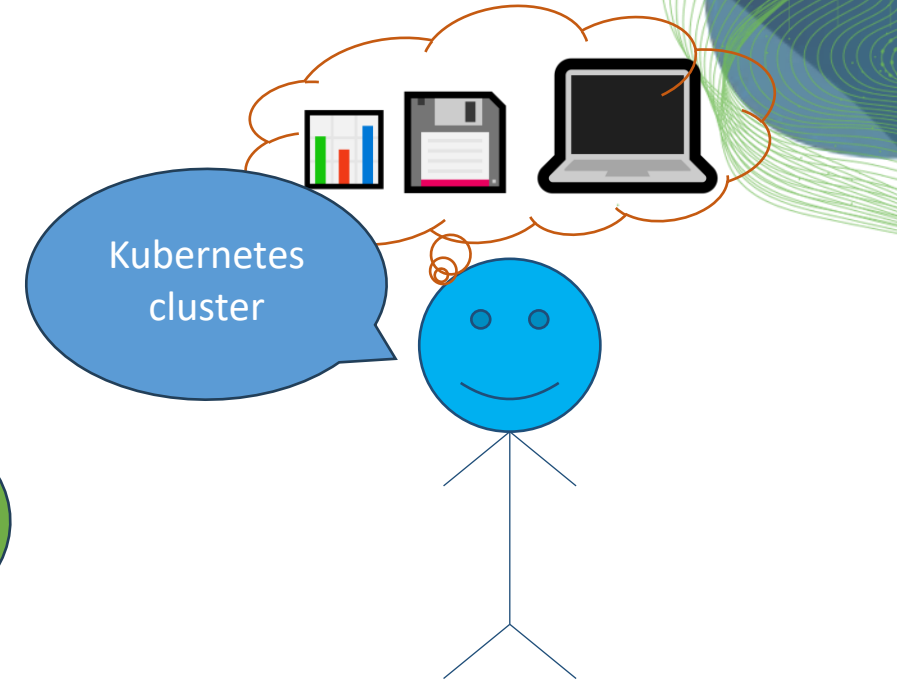
Biologist

*Domain-specialist
(focused on
real-world problems)*



Bio-image Analyst

*← Generalist →
(data-driven,
service-oriented)*



Computer Scientist

*Method + infrastructure specialist
(algorithm-centered)*

Quantitative bio-image analysis

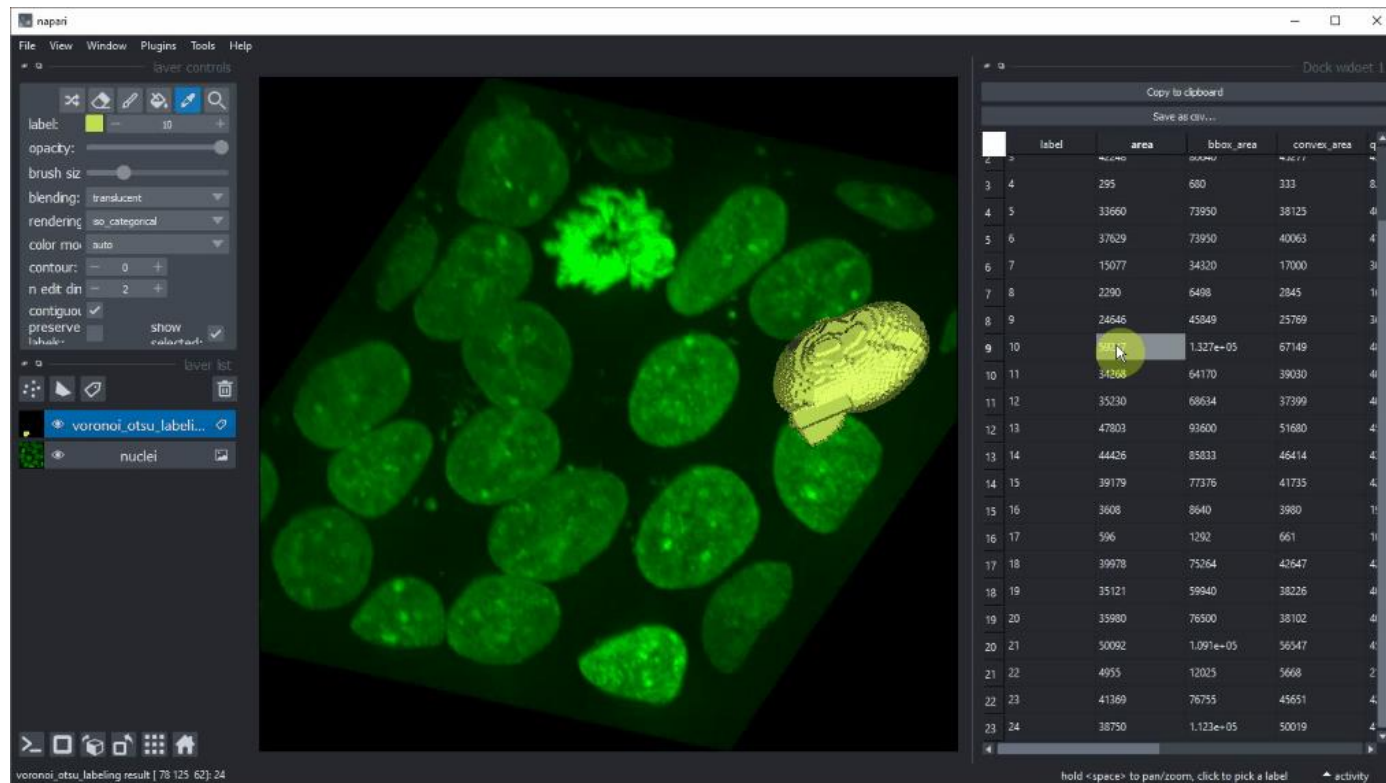
- Deriving quantitative information from images of biological samples taken with microscopes



$$\text{cat height} = \underline{1.5} \times \text{microscope height}$$

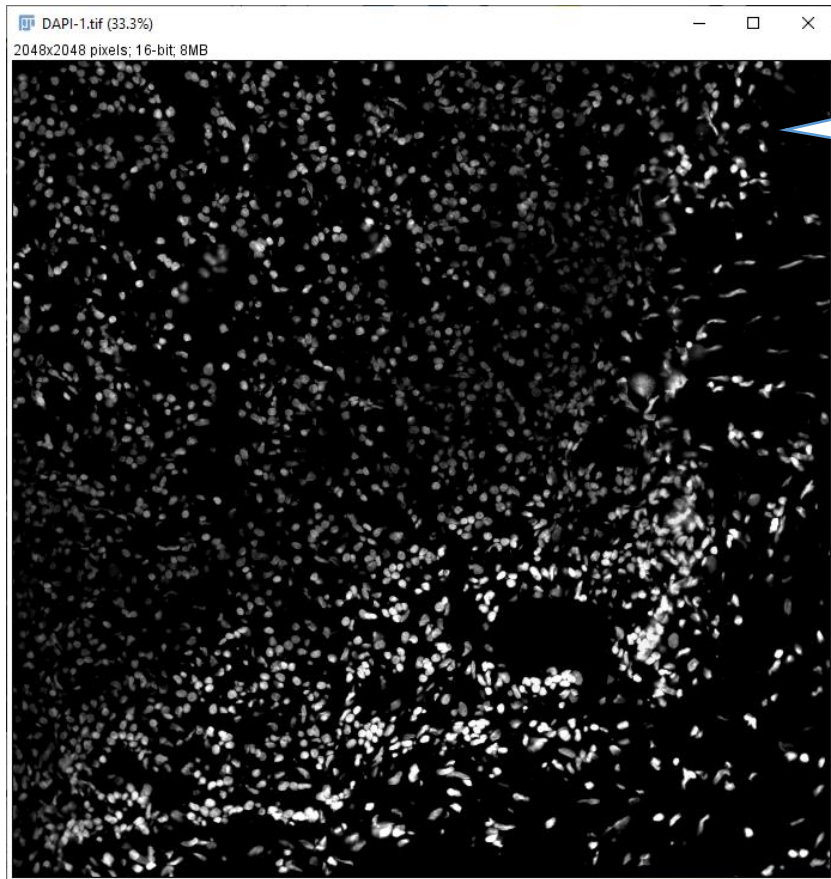
Quantitative bio-image analysis

- Deriving quantitative information from images of biological samples taken with microscopes + visualization



Objective bio-image analysis

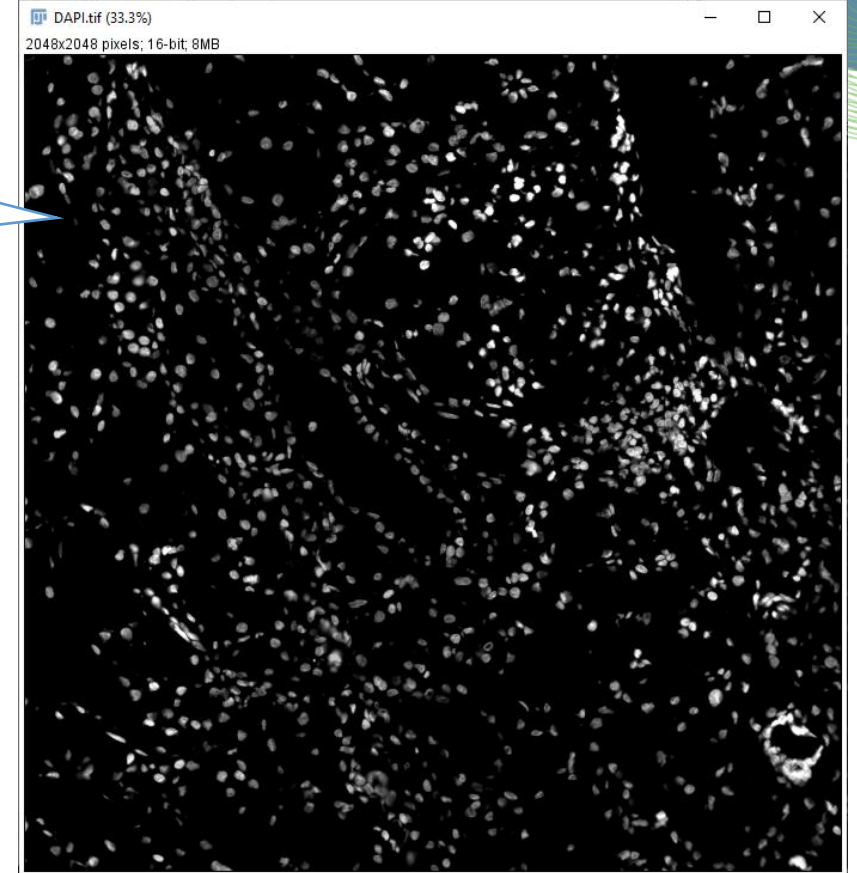
- Measurements should be objective, not influenced by human interpretation



Nuclei in this
image are ...

... more dense
than in this image.

Use automation for
less subjective
analysis.



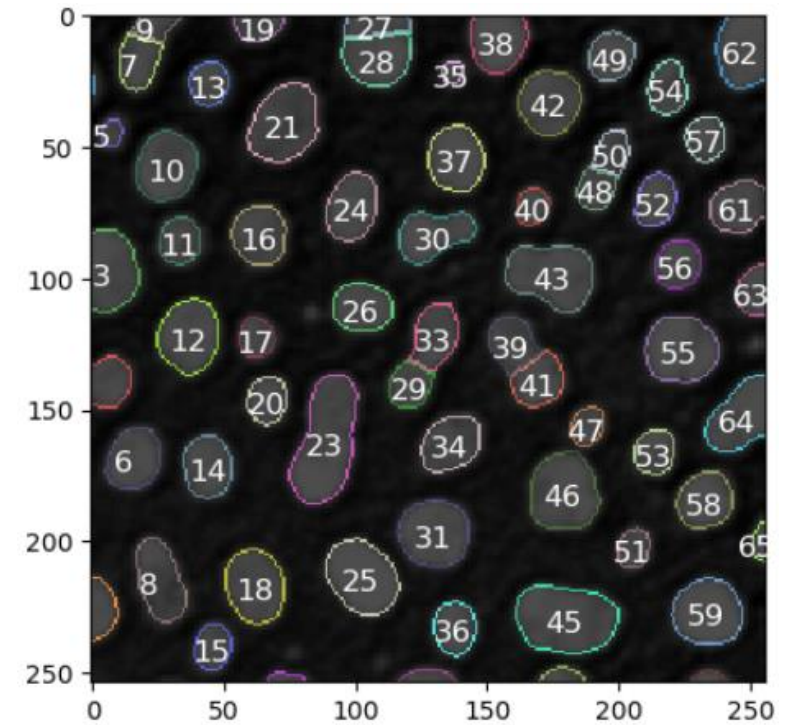
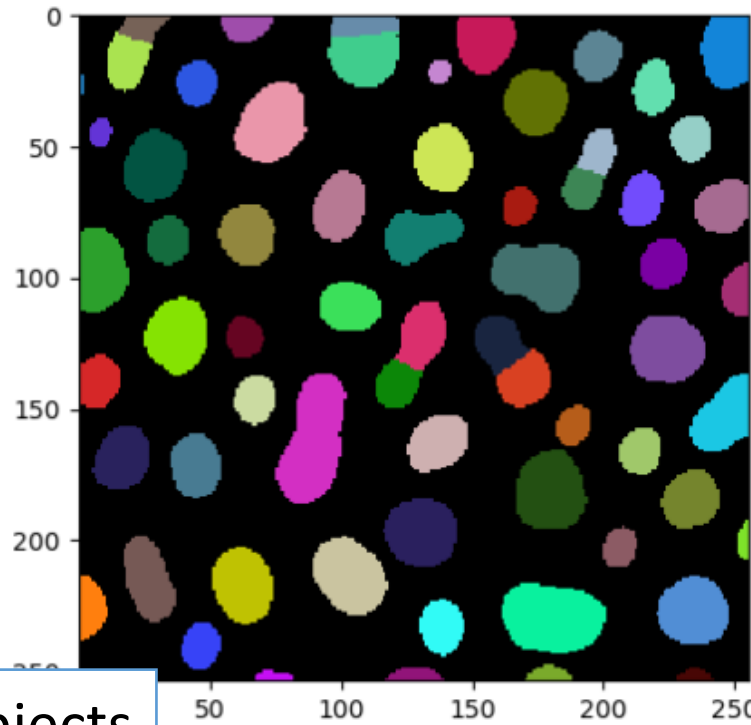
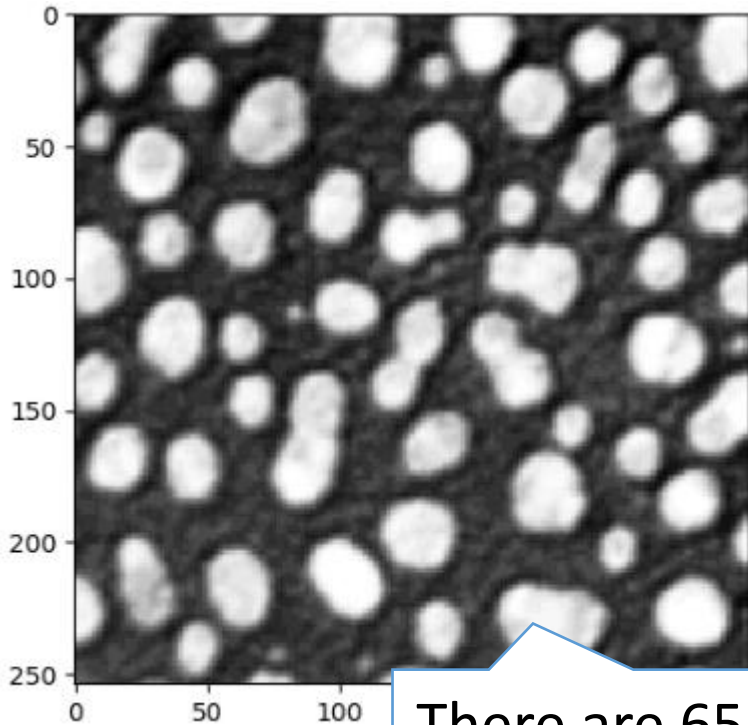
Reliable bio-image analysis

- Algorithms must be reliable (trustworthy).
- Visualization helps gaining trust in automated methods.

Original image

Label image

Overlay

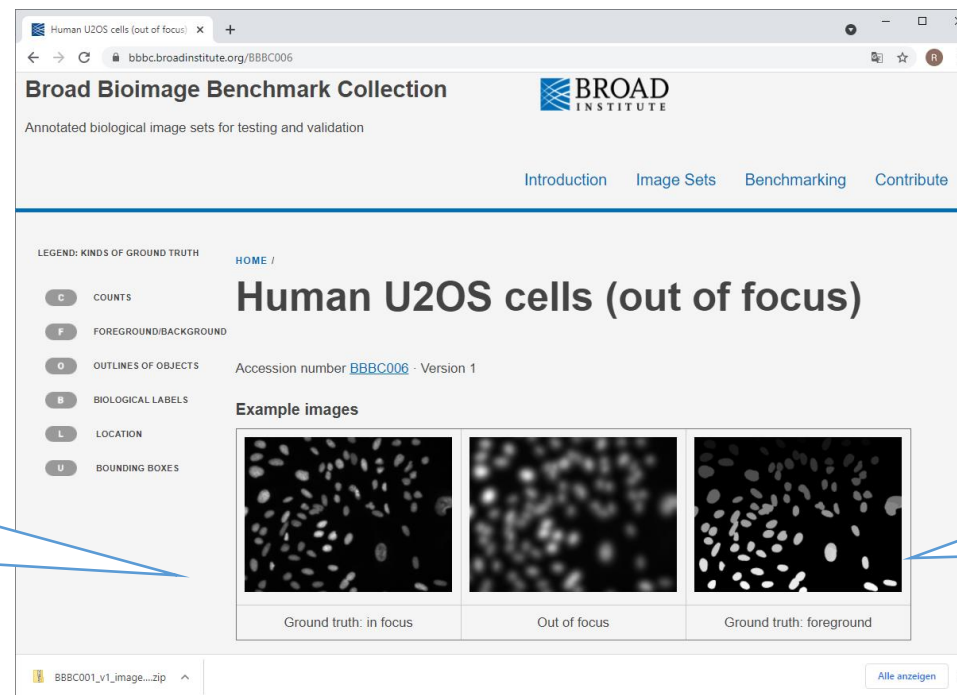


There are 65 objects
in this image.

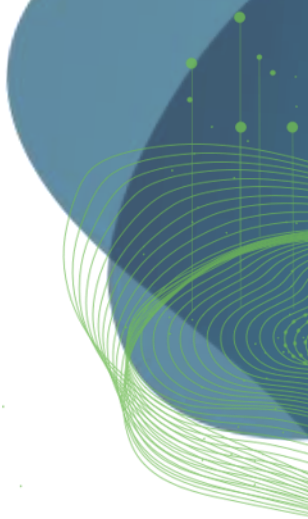
Reliable bio-image analysis

- Algorithms must be reliable (validated methods).
- Publicly available benchmark data sets allow to compare algorithms on common data.

Original image
data



“Ground truth”
label images



Reproducible data analysis

- Allowing others to do your experiment again.
- “The image data was analyzed with Python.”

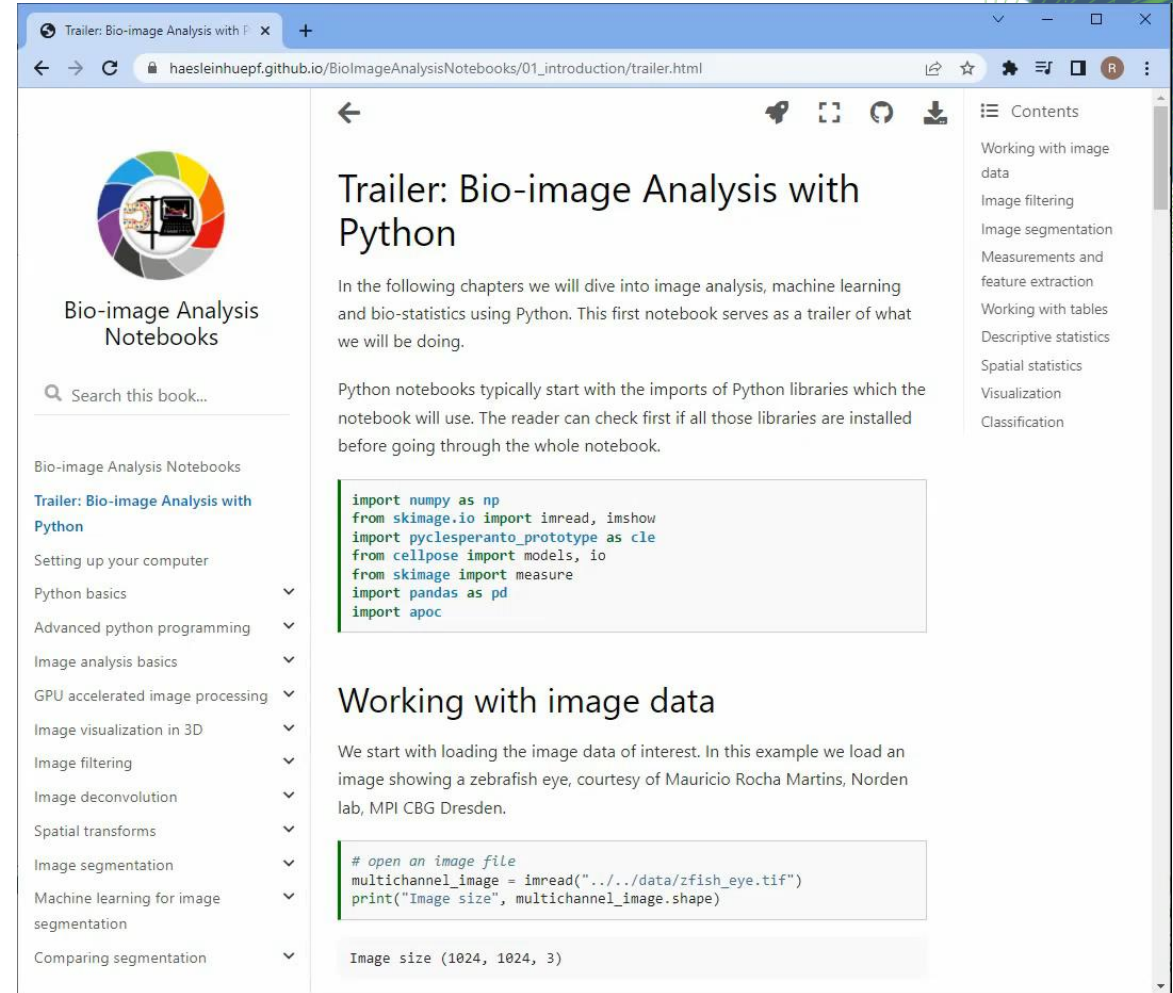
Can you reproduce
what they did?

Reproducible bio-image analysis

- Allowing others to do your experiment again.
- “The image data was analyzed with Python.”

Can you reproduce what they did?

Can you reproduce what they did?



The screenshot shows a web browser window displaying the 'Trailer: Bio-image Analysis with Python' page. The page is part of the 'Bio-image Analysis Notebooks' repository on GitHub. The left sidebar contains a table of contents with links to various notebooks, including 'Trailer: Bio-image Analysis with Python', 'Setting up your computer', 'Python basics', 'Advanced python programming', 'Image analysis basics', 'GPU accelerated image processing', 'Image visualization in 3D', 'Image filtering', 'Image deconvolution', 'Spatial transforms', 'Image segmentation', 'Machine learning for image segmentation', and 'Comparing segmentation'. The main content area shows the title 'Trailer: Bio-image Analysis with Python' and a brief introduction. Below the introduction, there is a code block showing the imports for the notebook:

```
import numpy as np
from skimage.io import imread, imshow
import pyclesperanto_prototype as cle
from cellpose import models, io
from skimage import measure
import pandas as pd
import apoc
```

The code block is followed by the text 'Working with image data' and a paragraph explaining that the notebook starts with loading the image data of interest. Below this, there is another code block showing the code to open an image file and print its size:

```
# open an image file
multichannel_image = imread("../data/zfish_eye.tif")
print("Image size", multichannel_image.shape)
```

The output of the code is displayed as 'Image size (1024, 1024, 3)'.

Replicable bio-image analysis

- Others run the same analysis on their data and have consistent results / same conclusions.
- Can only be achieved if data analysis protocol was documented reproducibly.
- See also: *Replication crisis*
 - In Psychology (surveys)
 - In Medicine (clinical trials)
 - In Computer Science (executable code)
 - ...

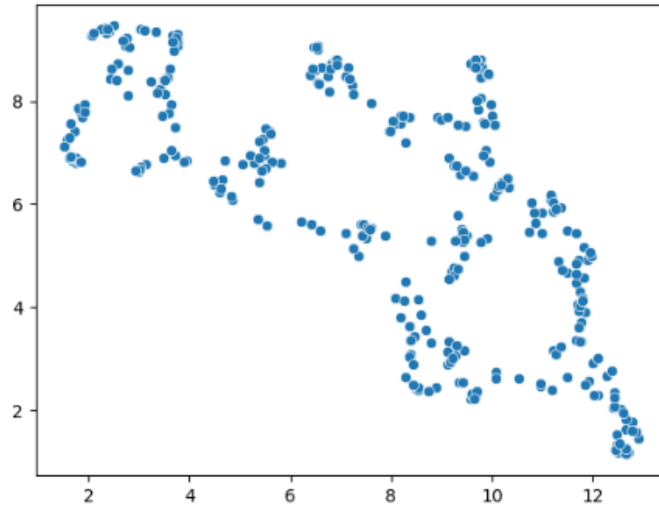


Repeatable data analysis

- In wet-lab experiments, samples may get destroyed while executing the experiment.
- Repeatability is a property of the experiment / algorithm. You cannot improve repeatability by better documentation.

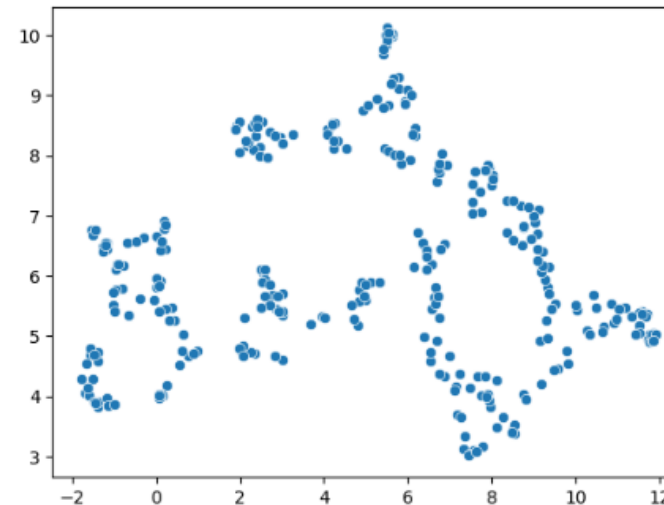
```
[11]: reducer = umap.UMAP()  
      embedding2 = reducer.fit_transform(scaled_statistics)  
  
      seaborn.scatterplot(x=embedding2[:, 0],  
                          y=embedding2[:, 1])
```

[11]: <AxesSubplot: >

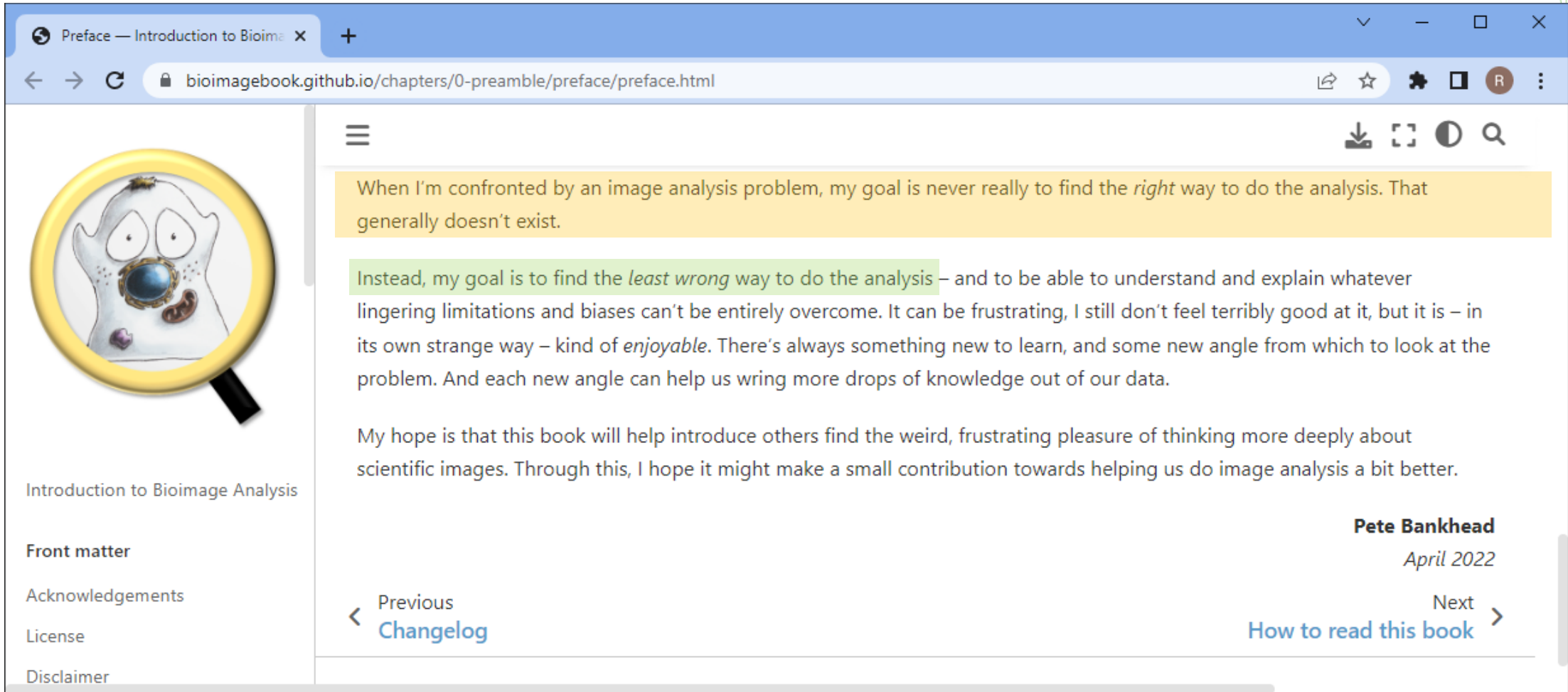


```
[12]: reducer = umap.UMAP()  
      embedding2 = reducer.fit_transform(scaled_statistics)  
  
      seaborn.scatterplot(x=embedding2[:, 0],  
                          y=embedding2[:, 1])
```

[12]: <AxesSubplot: >




Bio-image Analysis: good scientific practice



Preface — Introduction to Bioimage x +

bioimagebook.github.io/chapters/0-preamble/preface/preface.html



Introduction to Bioimage Analysis

Front matter

Acknowledgements

License

Disclaimer

When I'm confronted by an image analysis problem, my goal is never really to find the *right* way to do the analysis. That generally doesn't exist.

Instead, my goal is to find the *least wrong* way to do the analysis – and to be able to understand and explain whatever lingering limitations and biases can't be entirely overcome. It can be frustrating, I still don't feel terribly good at it, but it is – in its own strange way – kind of *enjoyable*. There's always something new to learn, and some new angle from which to look at the problem. And each new angle can help us wring more drops of knowledge out of our data.

My hope is that this book will help introduce others find the weird, frustrating pleasure of thinking more deeply about scientific images. Through this, I hope it might make a small contribution towards helping us do image analysis a bit better.

Pete Bankhead
April 2022

< Previous
Changelog

Next
How to read this book >

Bio-image analysis is supposed to be

- **Quantitative**
 - We derive numbers from images which describe physical properties of the observed sample.
- **Objective**
 - The derived measurement does not depend on who did the measurement. The measurement is free of interpretation.
- **Reliable (trustworthy / validated)**
 - We are confident that the measurement is describing what it is supposed to describe.
- **Reproducible**
 - Enabling others to re-do the experiment. For this, documentation is crucial!
- **Replicability**
 - Others *do* execute the same analysis, potentially on other data, and see consistent results.
- **Repeatable**
 - We can do the same experiment twice under the *same conditions* and get the same measurements.

The Image Science Community

- Ask your question online and an expert will likely reply the same day 😊

The screenshot shows a forum post on the image.sc website. The post is titled "Install issues with py-clesperanto-assistant" and is categorized under "Usage & Issues". The user William Giang (@haesleinhuepf) is asking for help with installing the napari-pyclesperanto-assistant plugin (v0.9.3) for napari 0.4.10 on Windows 10. He provides a list of installed tools: MSVC v142 - VS 2019 C++ x64/x86 build tools (Latest), Windows 10 SDK (10.0.19041.0), C++ CMake tools for Windows, Testing tools core features - Build Tools, and C++ AddressSanitizer. He also mentions that he can install "napari[all]" without issue after creating and activating a conda environment with python=3.8. The post has 1 reply and 1 like.

The screenshot shows a reply to the forum post by Robert Haase (@haesleinhuepf), who is the maintainer of the plugin. He thanks the user for reporting the issue and explains that he recently hit issues with recent pyopenc1 > 2021.2.1 and thus limited that dependency. He is considering removing this with the next minor release next week. He suggests trying to install using conda and provides the following commands:

```
conda create --name bio11 python==3.8.5
conda activate bio11
conda install -c conda-forge pyopenc1==2021.2.1
pip install napari-pyclesperanto-assistant
pip install napari[all]
```

He asks if this helps and signs off with "Cheers, Robert". The reply has 2 likes and is marked as a solution. Below the reply, there is a link to "Setup.py for plugin that depends on pyclesperanto".

Acknowledgements

Big thanks to our funders

GEFÖRDERT VOM



Bundesministerium
für Bildung
und Forschung



Diese Maßnahme wird gefördert durch die Bundesregierung aufgrund eines Beschlusses des Deutschen Bundestages. Diese Maßnahme wird mitfinanziert durch Steuermittel auf der Grundlage des von den Abgeordneten des Sächsischen Landtags beschlossenen Haushaltes.