Machine learning based analysis of biomedical microscopy images

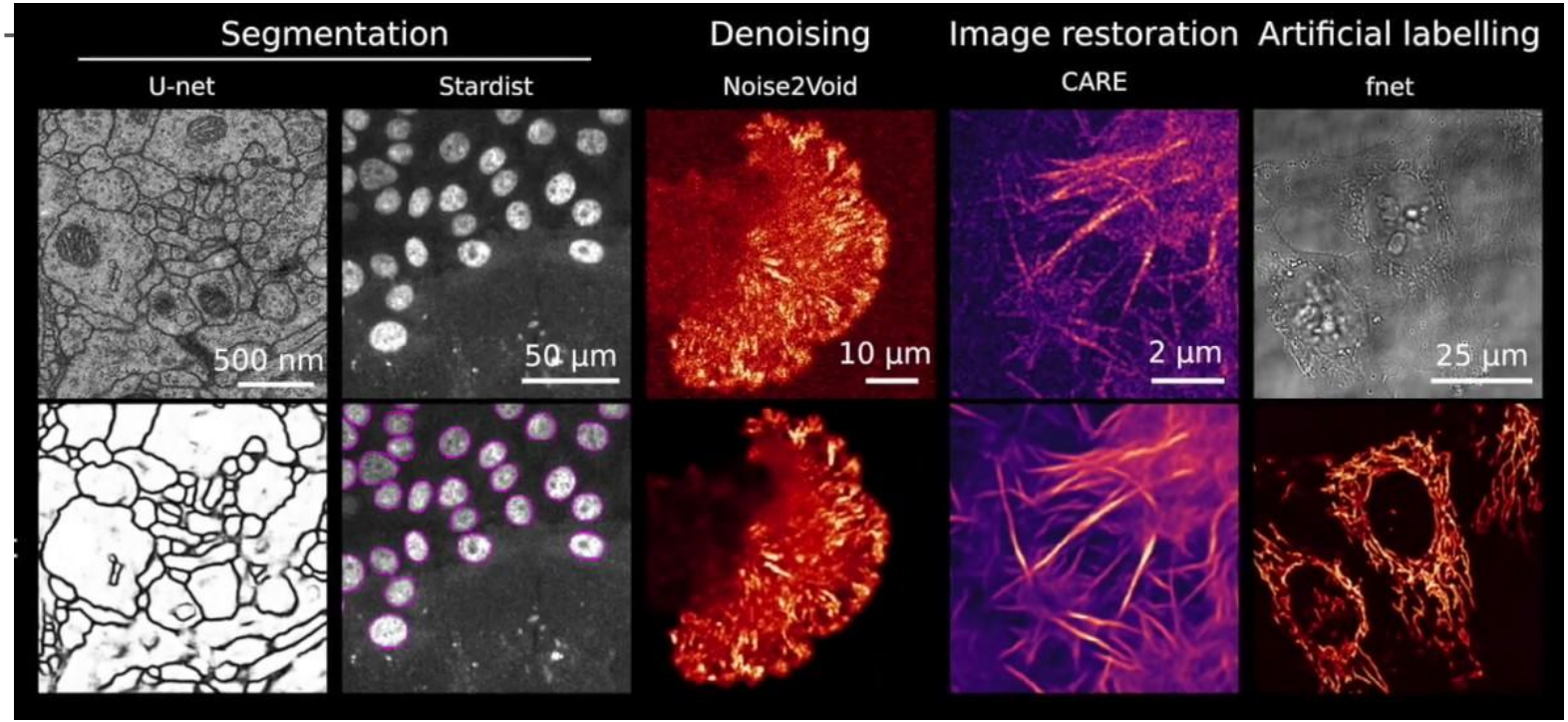
Simon F. Nørrelykke

Image and Data Analysis group (IDA)

Scientific Center for Optical and Electron Microscopy (ScopeM

ETH Zurich

Example tasks and their ML solution



Who we are (IDA)

IDA
IMAGE & DATA ANALYSIS

- Simon F. Nørrelykke (staff)
  - Physicist
- Szymon Stoma (staff)
  - Computer Scientist
- Andrzej Rzepiela (staff)
  - Bioinformatician

Where we are

- Part of ScopeM:
  - ~40 permanent staff members (~30 FTE)
  - Electron and light microscopy
  - Lab automation
  - Sample preparation
  - ~700 active users
  - ~20 EMs, ~30 LMs, ~40 misc tools
  - Services all of ETH Zurich + industry
- Part of ETH Zurich
  - 6’500 academic staff
  - ~20’000 students

What we do

- User support for image and data analysis projects (charged and free)
  - ~20%
- Internal support in ScopeM
  - ~10%
- Teach, internally and external
  - ~20%
- Build and maintain IT infrastructure
  - ~10%
- Overhead (meetings, reporting, grant writing, ...)
  - ~20%
- R&D (further education: conferences, reading papers)
  - ~20%

Project types

Almost exclusively in biology. Mainly LM, increasing number of EM projects.

- High content screening campaigns
  - Run over multiple years
  - More data wrangling and analysis than actually image analysis
  - Considerable infrastructure overhead (computing, storage, access)
- Light and electron microscopy projects
  - Counting objects, co-localisation, measuring intensities, quantifying structures
  - Using open source software and custom written code
- Deep Learning approach
  - HCS, LM, EM: image restoration, segmentation, classification
  - Focus is on preparing the data and managing user expectations
  - Using existing/published architectures and models

Project types

By duration and charge

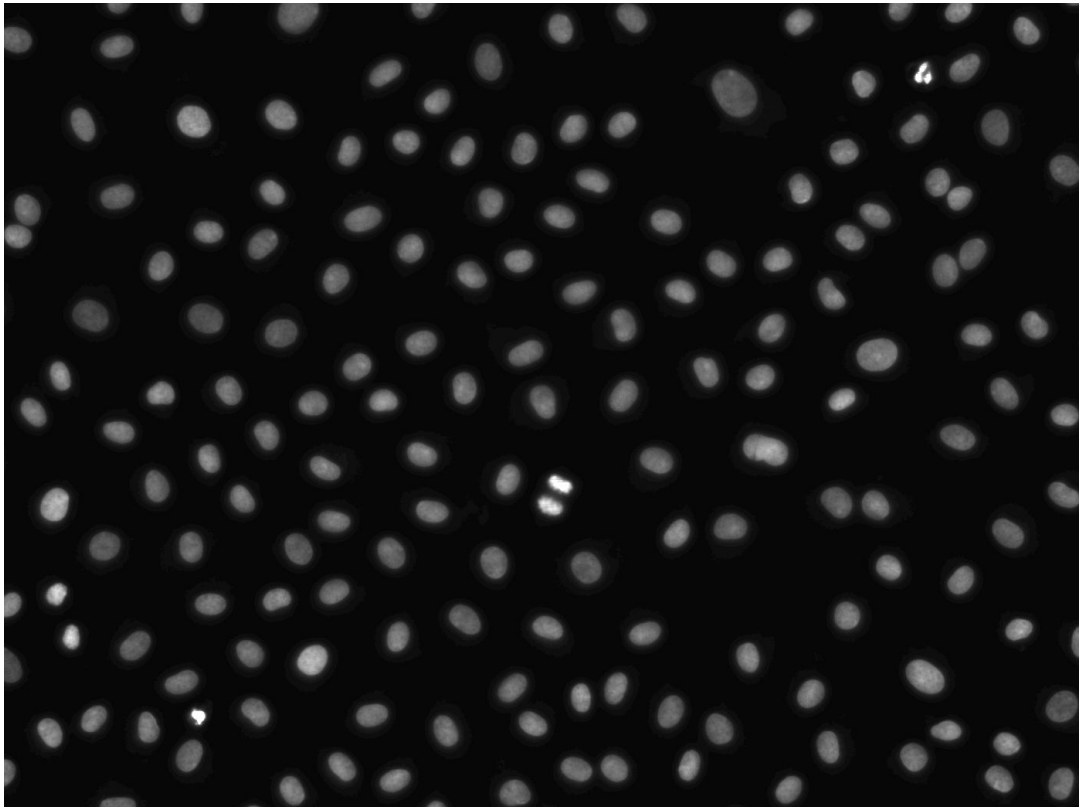
- Image Clinics
  - Free 1h consults on a weekly basis, 1-on-1
  - About 50 consults per year
- Free trials
  - 4-8 hours of feasibility study or complete solution
  - 10-20 per year
- Paid projects
  - Charged by the hour, currently 80CHF/hour (200 CHF/hour industry)
  - 10-20 per year
- Larger collaborations
  - Lump-sum payment for 200-300h, typically high-content screening campaigns
  - 1-3 per year

Teaching

We teach image analysis to life scientist

- Recurrent 3-day introductions based on Fiji
- Occasional 1-day trainings in PhD schools
- Yearly summer school (ZIDAS) since 2017- now with EPFL, Basel, Bern
- Participate in various schools
  - EXCITE: 2-week ETH course in imaging
  - EMBL: “Deep Learning for Image Analysis” and “Advanced Methods in BioImage Analysis”
  - NEUBIAS - training schools, started with Kota Miura’s BIAS @ EMBL
- Why do we teach?
  - We have to: Part of job description; few others bridge biology to computer-science
  - Scales: Give user a fishing-rod, not just a fish
  - Fun and rewarding for all involved

Example screening data

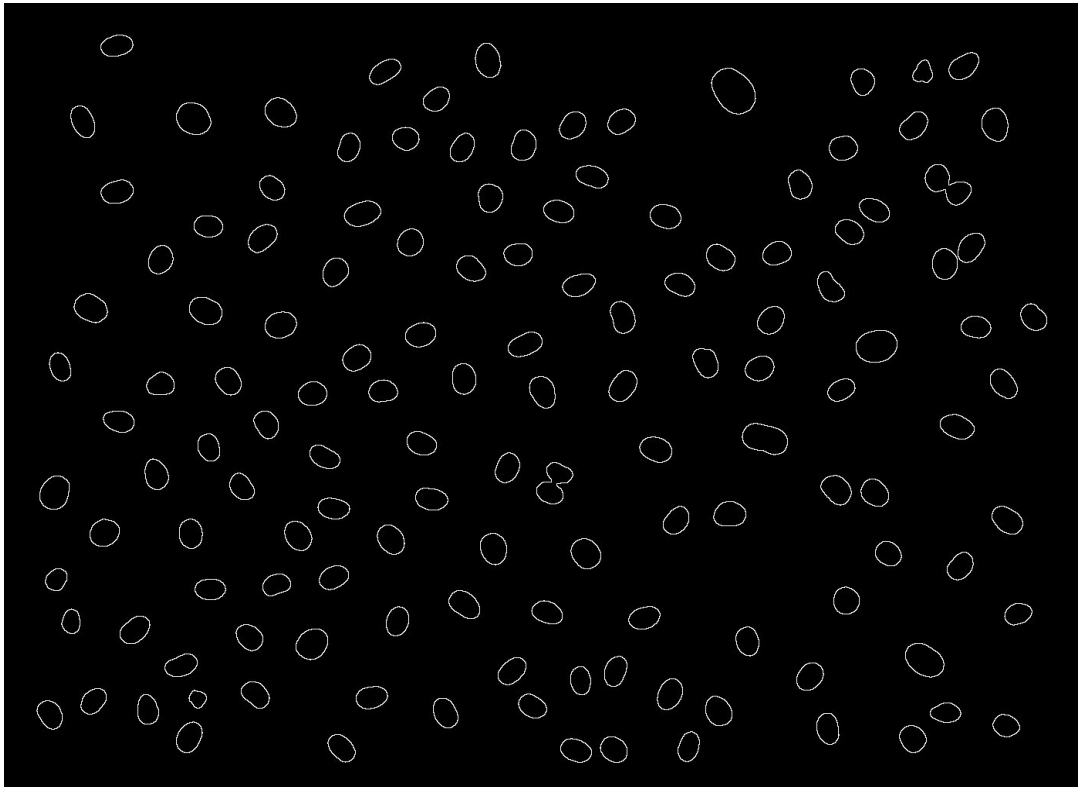


DAPI stained nuclei

You can segment these blindfolded

Now do it in 100’000 images

Example data



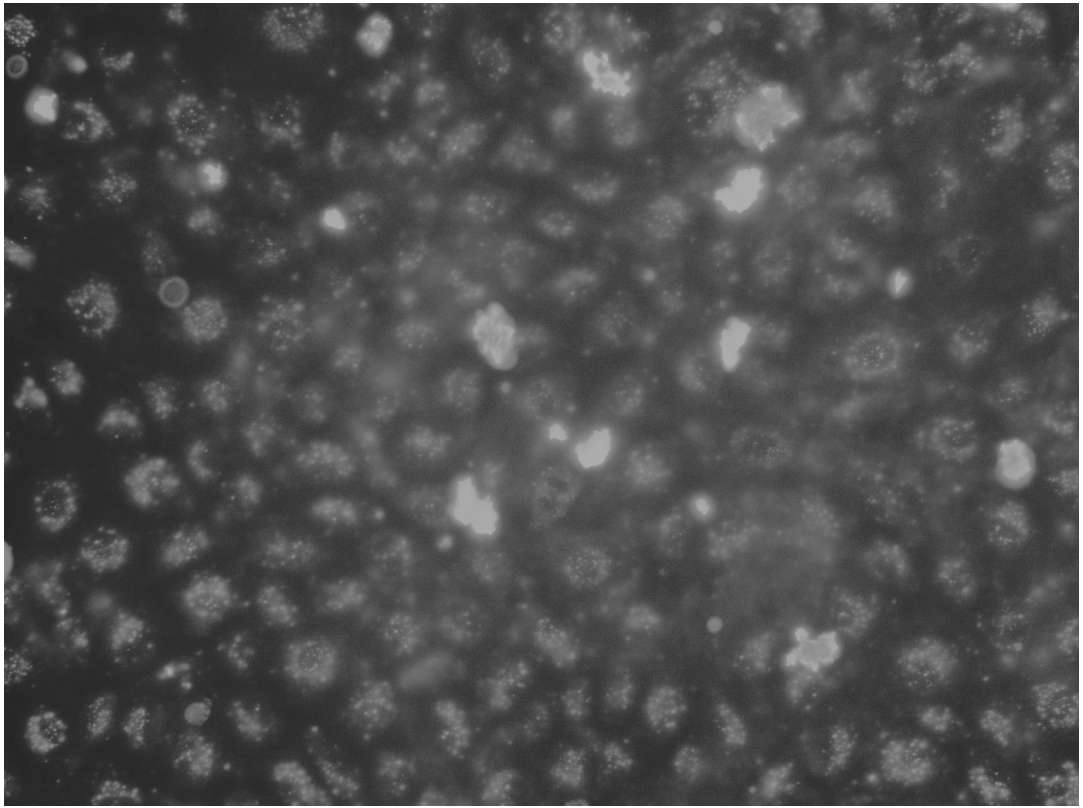
Segmentation result

Not very interesting

Classic methods fine

Bread and butter IA

Example data



RFP channel, same field of view as previous image

One particular protein is stained and we want to “quantify it”

Segment or approximate? What is good enough?

Projects and constraints

- The question is not if we *can* do the project, but if we can do it in *days/hours*
- Unlike researchers we don't have just a few projects we work on for years
- We have tens of projects that vary in size and duration
- Users (their PIs) don't want to pay for multiple weeks, let alone months
- This means we rarely have the time to develop own tools
- It also means we often operate with "good enough"
    - *If the signal is strong the method can be weak*

Open source to the rescue

- Not having time to develop tools, we use and adapt existing ones
- Unlike 10-20 years ago we are now flooded with image analysis software
- We have the luxury problem of picking among existing solutions
- Means we have to stay informed of what comes out

Lessons learned (things to keep in mind) re. tools

- Good enough *is* good enough (*perfect* is the enemy of *done*)
- Rarely worth learning (adapting) a highly specialised tool that is a perfect fit
- Learn to master a few tools that cover different problems
- Go for polyglot tools when possible
    - E.g. ImageJ/Fiji, or better yet Python, specifically ZeroCostDL4Mic

Tools that we commonly use (the big 5)

- ImageJ/Fiji
  - Does ~everything and is known by ~all biologists
  - Thousands of plugins, now also for pretrained NNs
- CellProfiler + CellProfiler Analyst
  - No-code workflows, for cell segmentation and classification
- Ilastik
  - No-code pixel classification and more
- QuPath
  - Digital pathology/histology/FL, superb UI and documentation, DL interface
- Imaris
  - Commercial, no-code IA for big 3D data, now with interface to open source ML

When the big-5 isn’t enough

- More specialized stand-alone tools
  - see http://biii.eu/ for annotated list of 1’000+ tools (BioImage Informatics Index)
- Python
  - Data-wrangling code (big part of most projects)
  - Implement open source tools, e.g. **StarDist** or **CellPose** or **CSBDeep**
  - **Scikit-image**
  - **ZeroCostDL4Mic**

StarDist

From 2018

For segmentation

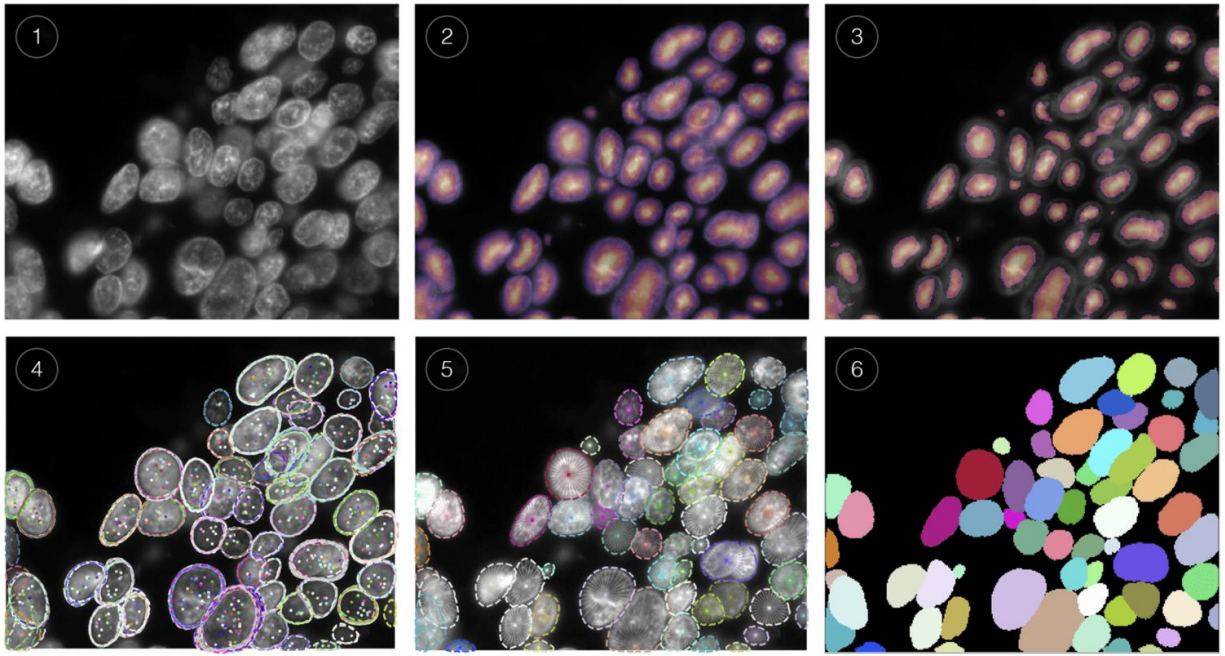
Very robust and accurate

Allows overlap

Only star-convex objects

Pre-trained Fiji plugin

ZeroCost notebooks



https://github.com/stardist/stardist Uwe Schmidt and Martin Weigert and Coleman Broaddus and Gene Myers

CellPose

From 2020

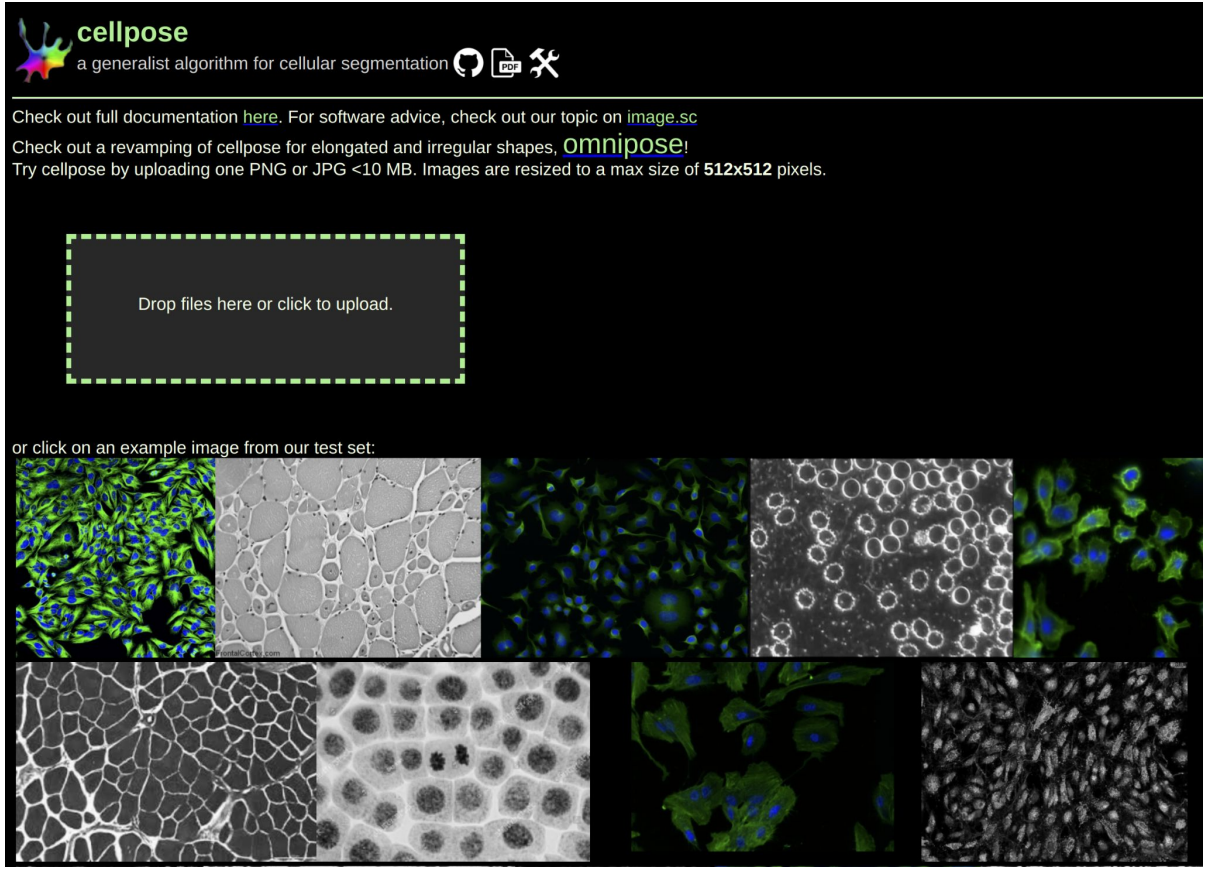
Versatile segmentation

Also non-covex objects

Web interface

GitHub

QuPath



http://www.cellpose.org/ Carsen Stringer, Michalis Michaelos, Marius Pachitariu

CSBDeep - a DL toolbox for microscopists

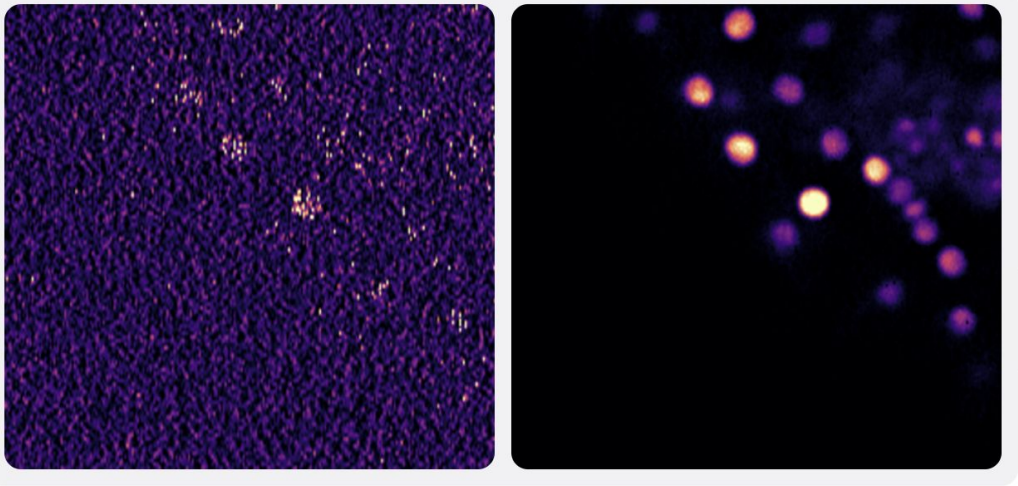
From ~2018

Florian Jug et al

Spearheaded the application of DL for microscopy image analysis

Initially just restoration

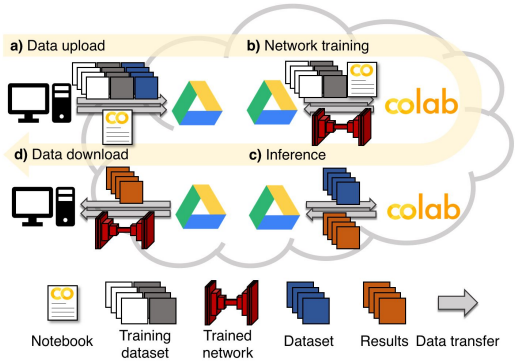
Denoising in 3D (Tribolium nuclei)



http://csbdeep.bioimagecomputing.com/

ZeroCostDL4Mic

- “ZeroCostDL4Mic is a toolbox for the training and implementation of common Deep Learning approaches to microscopy imaging. It exploits the ease-of-use and access to GPU provided by **Google Colab**.”
- Research grade DL for busy people (image analysts and biologists)
- **Python**: polyglot and *de facto* standard in DL
- **Google Colab**: GPU and environments taken care of
- Limitations: Power and storage (money can solve this)



https://github.com/HenriquesLab/ZeroCostDL4Mic/wiki

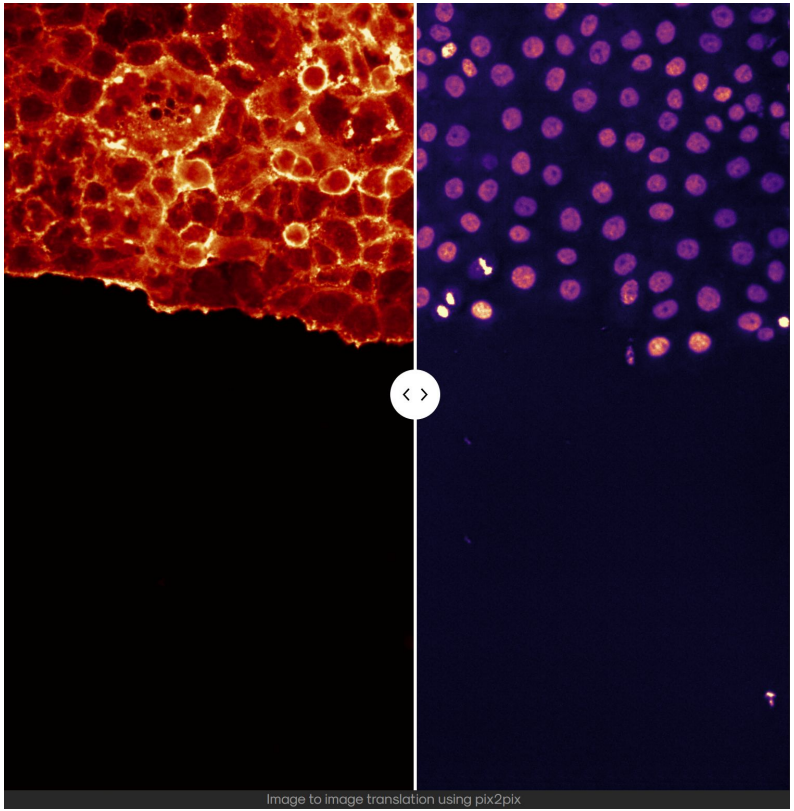
ZeroCostDL4Mic

Image to image translation (pix2pix)

Left side is input
Right side is the prediction

Similar to “Synthetic Fluorescence”

Video link:
https://videopress.com/v/1zQCIjlF

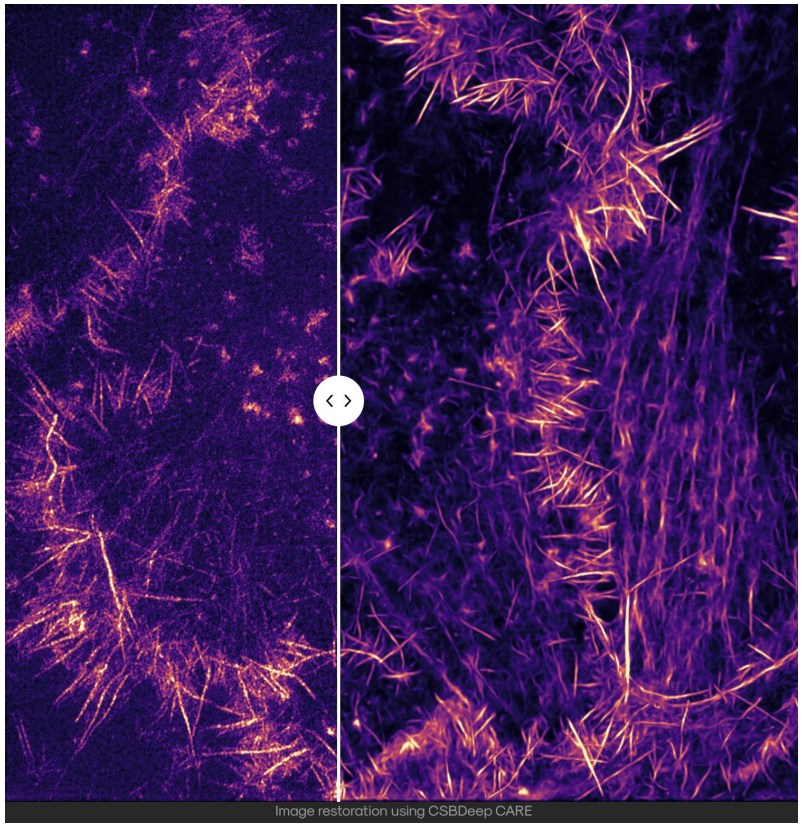


https://cellmig.org/image-analysis/

ZeroCostDL4Mic

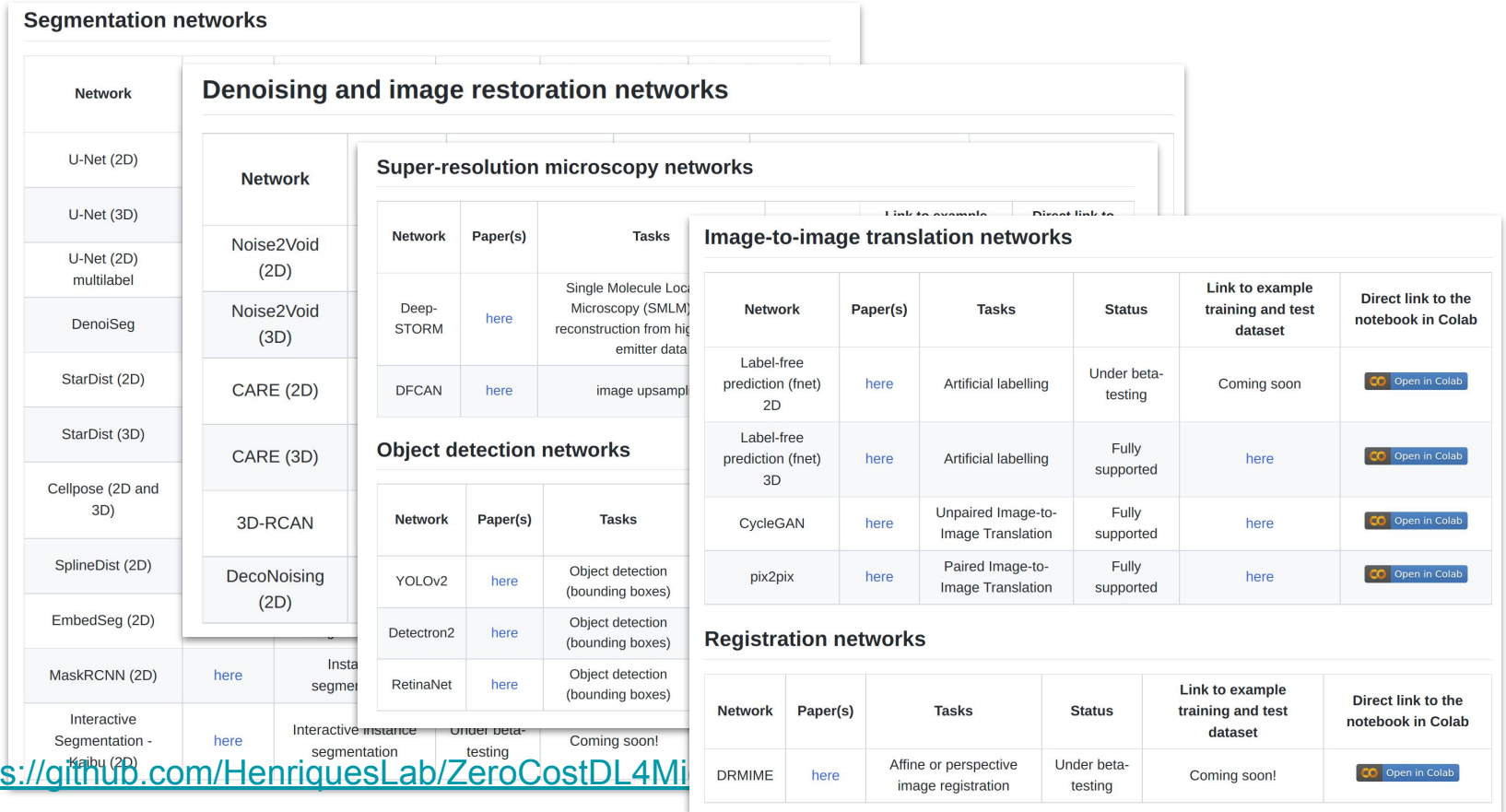
Image restoration using CARE

Unet trained on pairs of images: low and high quality



https://cellmig.org/image-analysis/

ZeroCostDL4Mic



Segmentation networks

Network	Paper(s)	Tasks	Status	Link to example training and test dataset	Direct link to the notebook in Colab
U-Net (2D)	here	Segmentation (binary)	Fully supported	here	Open in Colab
U-Net (3D)	here	Segmentation (binary)	Fully supported	here	Open in Colab
U-Net (2D) multilabel	here	Segmentation (multilabel)	Fully supported	here	Open in Colab
DenoiSeg	here	Joint denoising and segmentation	Fully supported	here	Open in Colab
StarDist (2D)	here	Segmentation (star-convex objects)	Fully supported	here	Open in Colab
StarDist (3D)	here	Segmentation (star-convex objects)	Fully supported	here	Open in Colab
Cellpose (2D and 3D)	here	Segmentation (arbitrary shapes)	Fully supported	here	Open in Colab
SplineDist (2D)	here	Segmentation (star-convex objects)	Fully supported	here	Open in Colab
EmbedSeg (2D)	here	Instance segmentation	Fully supported	here	Open in Colab
MaskRCNN (2D)	here	Instance segmentation	Fully supported	here	Open in Colab
Interactive Segmentation - Kaibu (2D)	here	Interactive instance segmentation	Under beta-testing	Coming soon!	

Denoising and image restoration networks

Network	Paper(s)	Tasks	Status	Link to example training and test dataset	Direct link to the notebook in Colab
Noise2Void (2D)	here	Denoising	Fully supported	here	Open in Colab
Noise2Void (3D)	here	Denoising	Fully supported	here	Open in Colab
CARE (2D)	here	Denoising	Fully supported	here	Open in Colab
CARE (3D)	here	Denoising	Fully supported	here	Open in Colab
3D-RCAN	here	Super-resolution	Fully supported	here	Open in Colab
DecoNoising (2D)	here	Denoising	Fully supported	here	Open in Colab

Super-resolution microscopy networks

Network	Paper(s)	Tasks	Status	Link to example training and test dataset	Direct link to the notebook in Colab
Deep-STORM	here	Single Molecule Localization Microscopy (SMLM) reconstruction from high emitter data	Fully supported	here	Open in Colab
DFCAN	here	image upsampl	Fully supported	here	Open in Colab

Object detection networks

Network	Paper(s)	Tasks	Status	Link to example training and test dataset	Direct link to the notebook in Colab
YOLOv2	here	Object detection (bounding boxes)	Fully supported	here	Open in Colab
Detectron2	here	Object detection (bounding boxes)	Fully supported	here	Open in Colab
RetinaNet	here	Object detection (bounding boxes)	Fully supported	here	Open in Colab

Image-to-image translation networks

Network	Paper(s)	Tasks	Status	Link to example training and test dataset	Direct link to the notebook in Colab
Label-free prediction (fnet) 2D	here	Artificial labelling	Under beta-testing	Coming soon	Open in Colab
Label-free prediction (fnet) 3D	here	Artificial labelling	Fully supported	here	Open in Colab
CycleGAN	here	Unpaired Image-to-Image Translation	Fully supported	here	Open in Colab
pix2pix	here	Paired Image-to-Image Translation	Fully supported	here	Open in Colab

Registration networks

Network	Paper(s)	Tasks	Status	Link to example training and test dataset	Direct link to the notebook in Colab
DRMIME	here	Affine or perspective image registration	Under beta-testing	Coming soon!	Open in Colab

https://github.com/HenriquesLab/ZeroCostDL4Mi

ZeroCostDL4Mic

Founding

- Lucas von Chamier
- Johanna Jukkala
- Christoph Spahn
- Martina Lerche
- Sara Hernández-Pérez
- Pieta K. Mattila
- Eleni Karinou
- Seamus Holden
- Ahmet Can Solak
- Alexander Krull
- Tim-Oliver Buchholz
- Florian Jug
- Loïc A Royer
- Mike Heilemann
- Romain F. Laine
- Guillaume Jacquemet
- Ricardo Henriques

Newcomers

- Elias Nehme
- Lucien Weiss
- Yoav Shechtman
- Christophe Leterrier
- Daniel Krentzel
- Martin Jones
- Wei Ouyang
- Estibaliz Gómez de Mariscal
- Erlantz Calvo
- Ignacio Arganda-Carreras
- Amin Rezaei
- Ainhoa Serrano
- Seyed M. Hosseini

Researchers providing guidance and recommendations

- Martin Weigert
- Uwe Schmidt

https://github.com/HenriquesLab/ZeroCostDL4Mic/wiki

Research

- We don’t really have time for our own research projects
- Except when we can hire a postdoc
- Example: Nelly Hajizadeh
  - Worked on denoising of cryo-electron microscopy images
  - Implemented new type of loss function with better frequency response
  - “Image quality measurements and denoising using Fourier Ring Correlations
  - https://arxiv.org/abs/2201.03992
- Research requires funding, which requires research, which …
- Only grant chances are blind reviews and similar
  - Novartis Freenovation
  - SNF Spark

Challenges

- How to deal with “being forgotten” in publications
  - Forgotten due to physical absence (big campus, different buildings)
  - Forgotten due to time-lag between service and publishing
  - Approach: Bring this up at beginning of project, get written agreement, follow up
- How to deal with changing charging model and targets
  - ETH sets price: Since 2013, 0 → 75 → 80 → 40 → 80
  - ScopeM sets targets. This year: 1’000 hours, i.e. 80’000 - 140’000 CHF should be billed
  - Approach: Some flexibility in fraction of hours actually charged - write grants with users
- How do deal with users asking for cost estimate
  - Give estimate if possible, otherwise explain this is research and therefore hard to estimate
  - Suggest open tab with reporting every 20-40 hours spent on project
  - Say “40h” then absorb additional time (good for user, bad for ScopeM?)
- Generic issue to avoid: Telling people both **what** to do and **how** to do it
  - Example: Sell X hours, but not more than Y per group and charge Z per hour and ...
  - Tried under communism: 5-year plans don’t really work for anything but killing an economy

Benefits of image analysis facilities

- For analysts
  - Exposure to wide variety of scientific problems
  - Chance to teach (good if you like it)
  - No publish-or-perish, just publish-or-languish
  - Get to play with many new developments
- For the user
  - Access to image analysis expertise with a human interface
  - Can be hard for a user to search web if they don’t know the image analysis jargon
  - They don’t have to stay up to date on technological developments and implementations
- For host institution
  - Increases quality of data analysis done
  - Allows smaller groups to do research that requires many areas of expert knowledge
  - Teaching of what is relevant, in small and flexible courses

Last slide. Thank you. Here some links

- SwissBIAS: Swiss BioImage Analysts' Society
  - https://swissbias.ch
- Forum: *The* place online for bioimage analysis knowledge
  - https://forum.image.sc/
- ZIDAS 2022: SwitZerland's Image and Data Analysis School
  - https://www.zidas.org/

Insights

- Be clear about your role and what you want
  - Insist on co-authorship or be forgotten
  - Charging is *independent* of intellectual contribution (postdocs are paid and authors)
- Passive versus active project acquirement
  - Passive (users come to us) tends to be smaller, less interesting project
    - Part of job, always keep resources for these
  - Active (we come to users) more likely to lead to longer collaborations
    - Can be hard to find the time, but is rewarding
- New technologies
  - We are the experts and have to identify new areas, then go there
  - Example: Moved into deep learning ~five years ago, considerable effort but paying off now
  - Sometimes enough to stay informed, without becoming expert–time is limited

Who to hire for your Image Analysis Facility?

- Questions to answer first
  - How many will you be (secured funding for the first years, later)?
  - What will your tasks be (EM, LM, HCS, medical, biology, materials science)?
  - What do you want your facility to be (service, research, development, teaching)?
- “Horses for courses”, example from ETH
  - Simon: Likes order, teaching, analysing problems
    - Role: support the others, teach, management
  - Szymon: Enjoys multitasking, teaching, coding, helping users, creating start-ups
    - Role: Many small-medium size projects, teaching
  - Andrzej: Enjoys single-tasking, data analysis, building IT infrastructure
    - Role: Few large projects, grant writing, IT
  - Post-docs: For deep work, research projects, teaching and support as desired
    - Role: Research postdoc, learn and do

What is the ideal image analysis facility?

- User perspective
  - Easy access: Just show up with project
  - Fast solutions: Work starts now and is finished in days/weeks (yesterday is best)
  - Free of charge
- Staff perspective
  - Interesting projects, not too many at the same time
  - Acknowledgement of contribution, i.e. treated as equal
  - Minimum administrative overhead (charging hours, logging every hour of the day)
  - Mixture of permanent and temporary staff: keep knowledge in house and gain fresh input

History of IDA

- Formed in December 2012, in response to experience from earlier model:
  - Then: Part of light microscopy (LM) and screening facility
  - Then: Only serving one institute in one department
  - Then: Two units - image analysis (one person) and data analysis (one person)
  - Issues: No clear rules for engagement leading to:
    - Analysts chose projects they found interesting
    - A few very happy users and some unhappy users
  - Response: Form one unit and start charging for projects
- Since 2014
  - Light and electron microscopy (EM) facilities merged into ScopeM
  - ScopeM serving all of university
  - IDA still strongly attached to LM side for historical reasons

Publishing

- Why do we want to publish?
  - Fairness: Intellectual contributions should be acknowledged
  - Need papers for winning grants
  - Career: Not all stay as staff, need proof of productivity
- What do we do?
  - Contribute methods descriptions, figures, tables, statistics
  - Write entire papers (ongoing/future)
- Factoids
  - Zero publications first four years (2013-2016)
  - A few each year after that
  - Main change: We started insisting on co-authorship