

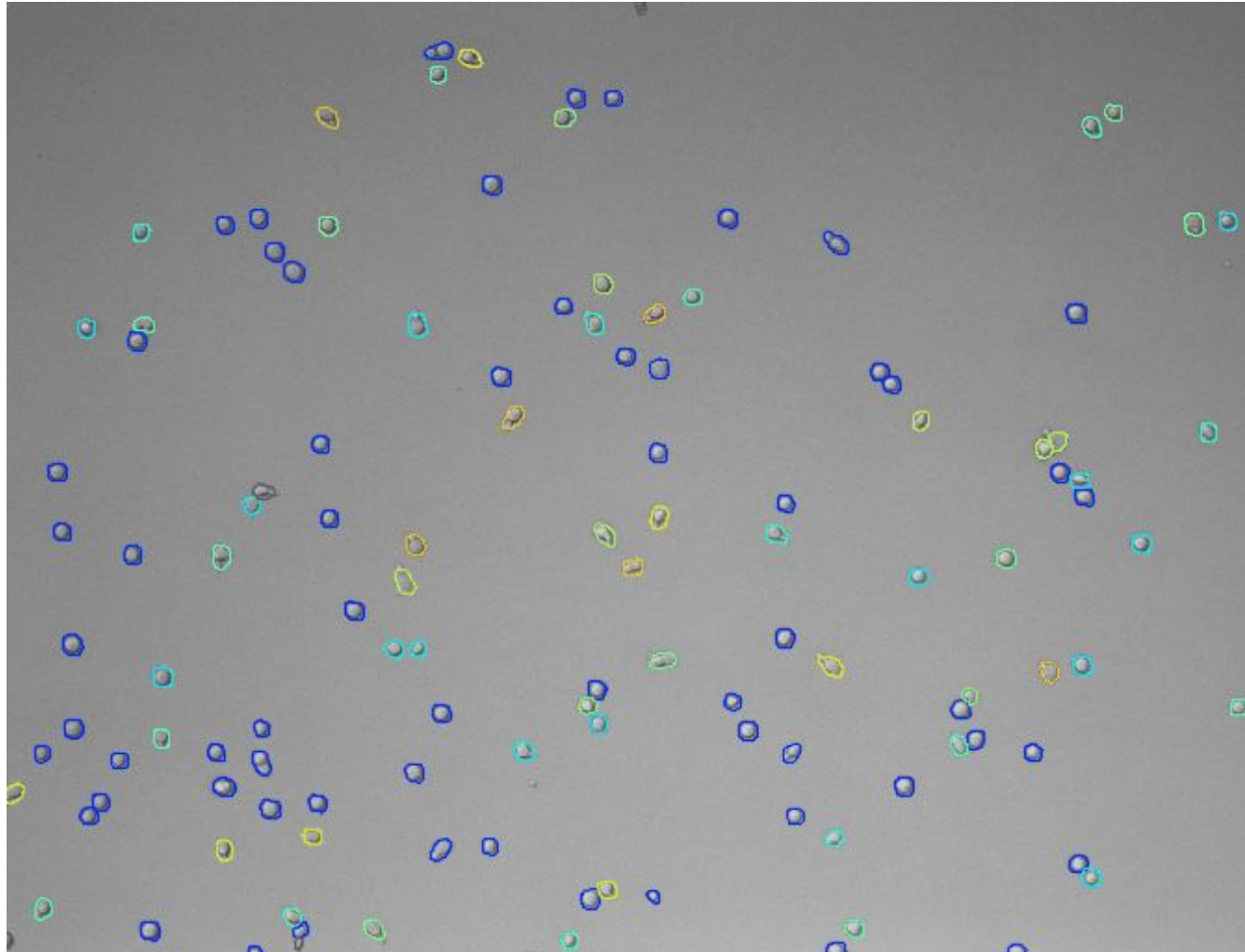


ZEROCOSTDL4MIC

Hands-on training - Part 1
Preparing data for Deep learning – Annotation
Training DL model using ZeroCostDL4Mic

Sujan Ghimire

T-cell Dataset



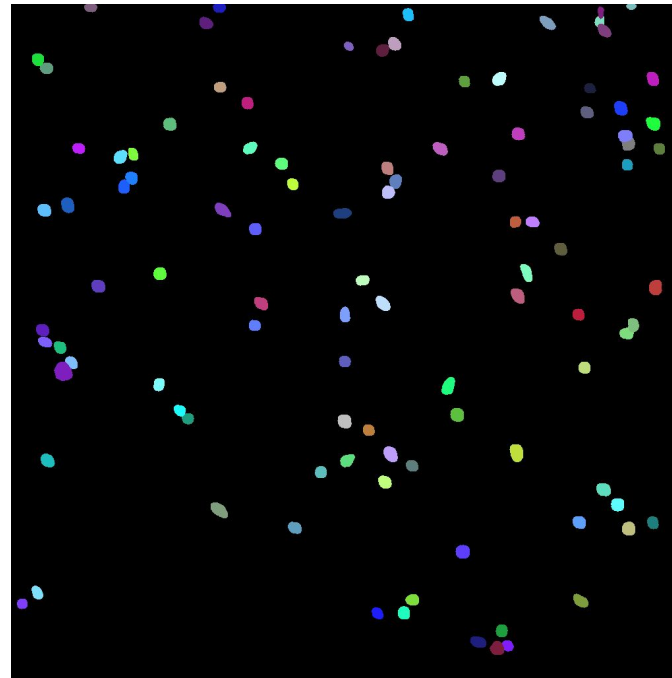
We want to:

1. Segment the cells
2. Track their movements

Nathan H. Roy, & Guillaume Jacquemet. (2020). Combining StarDist and TrackMate example 2 - T cell dataset (Version 1) [Data set]. Zenodo.
<https://doi.org/10.5281/zenodo.4034929>

Deep learning training

- We will train a StarDist model to segment cells
- Models need to be trained using paired images



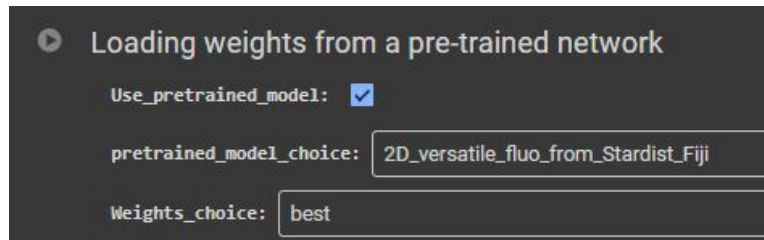
Types of training

Regular training

- Training new models from scratch using freshly annotated datasets
- Requires longer training time (more epochs)
- Requires larger training datasets
- Quality and generalization of the model depends on the datasets

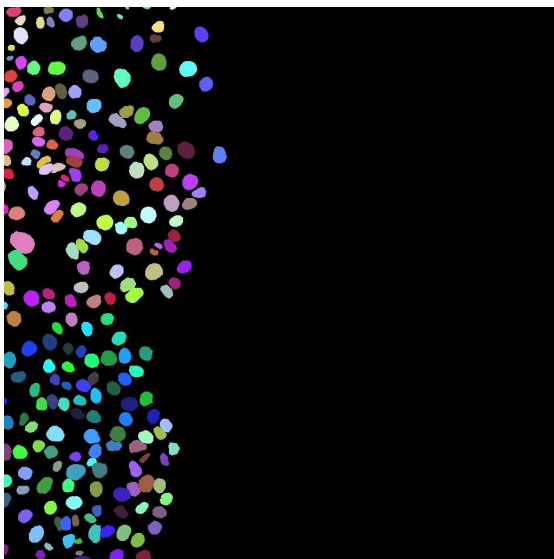
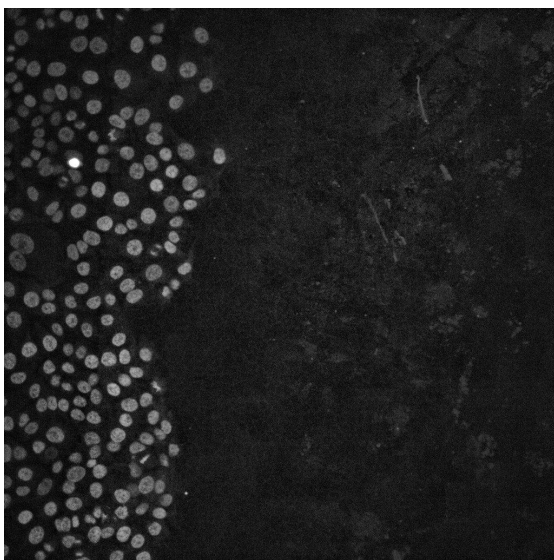
Transfer Learning

- Training new model using previously trained model such as versatile nuclei
- Shorter training and fewer datasets maybe enough

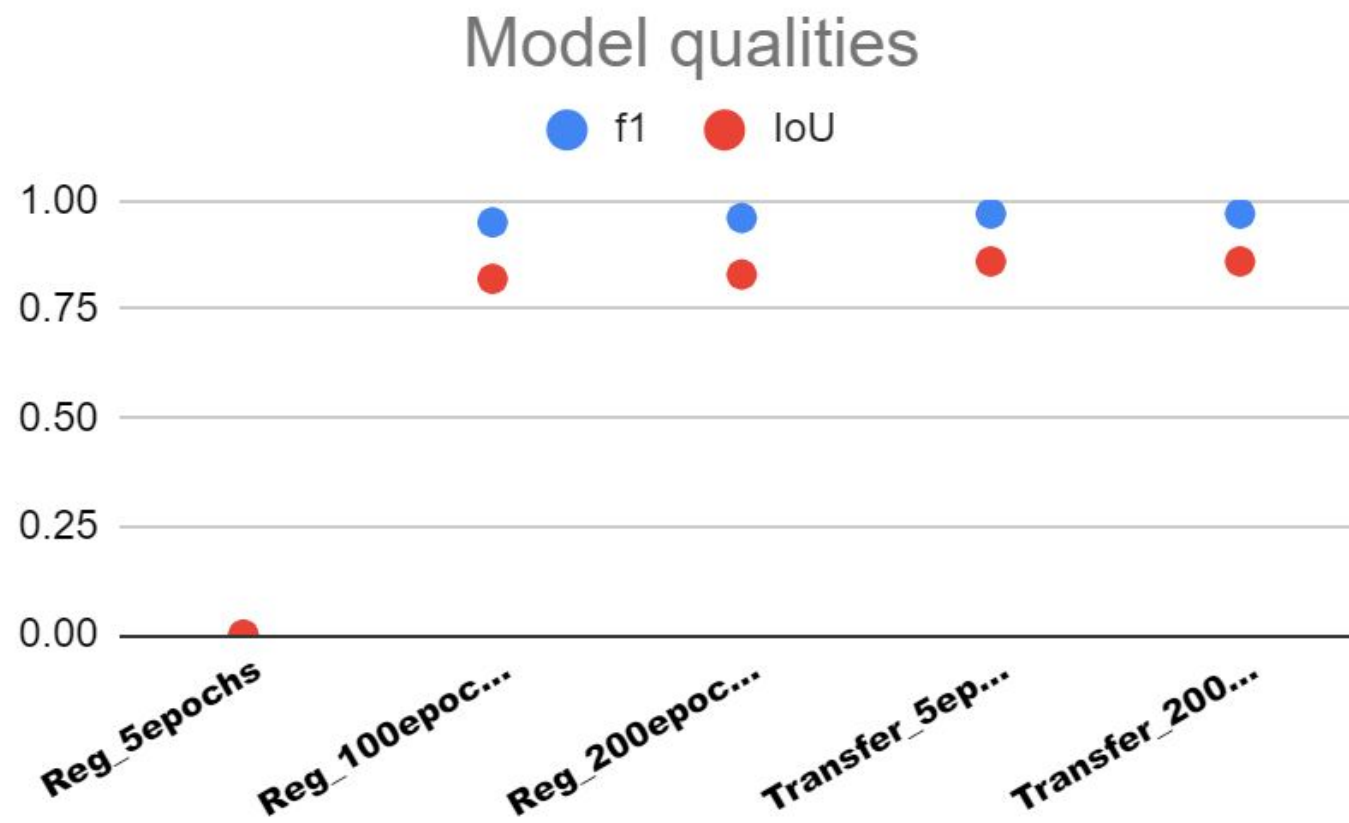


A screenshot of a software interface with a dark background. At the top, there is a header "Loading weights from a pre-trained network" with a play button icon to its left. Below the header, there are three settings: "Use_pretrained_model:" followed by a checked checkbox, "pretrained_model_choice:" followed by a text box containing "2D_versatile_fluo_from_Stardist_Fiji", and "Weights_choice:" followed by a text box containing "best".

```
Loading weights from a pre-trained network  
Use_pretrained_model: ☒  
pretrained_model_choice: 2D_versatile_fluo_from_Stardist_Fiji  
Weights_choice: best
```



Regular vs Transfer learning



Importance of high-quality training data

- Model is solely based on the training datasets
- Training data needs to be clean, diverse and representative
 - For example: datasets from different experiments
- Needs to be balanced without bias
 - No conflicting annotations within the training datasets
- Diverse data helps model to learn broader and prevents overfitting
- Clean data ensures effective and efficient training
 - Training focuses in learning relevant patterns

Steps for model training in ZeroCost

Create
training
dataset

Upload
data to
G-drive

Install
dependenci
es + check
GPU access

Define
paths to
training
data



Train

Asses the
quality of
your model

Run
predictions
on unseen
data

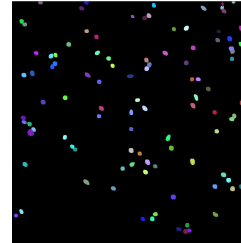


How to create a training dataset for StarDist?

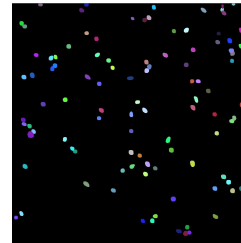
- Open Fiji – activate LOCI update site
- Restart Fiji
- Open your image to be annotated *Image1-T-cells*
- Select the Oval  or Freehand selection tool 
- Start drawing selections around each (yes each!) cell
- After every nucleus press “t” -> the selection will be stored to the ROI manager
- Repeat until all cells are in the ROI manager
- When done *Plugins > LOCI > ROI Map*
- Save the generated ROI map with the same title as the original image in a separate folder
- At the end you have one folder with the original images, another with the ROI maps

Suggested folder structure for StarDist

1-Training_source
2-Training_target
3-StarDist model
4-Quality_source
5-Quality_target
6-Test Data
7-results



209 image pairs, 10 % for validation



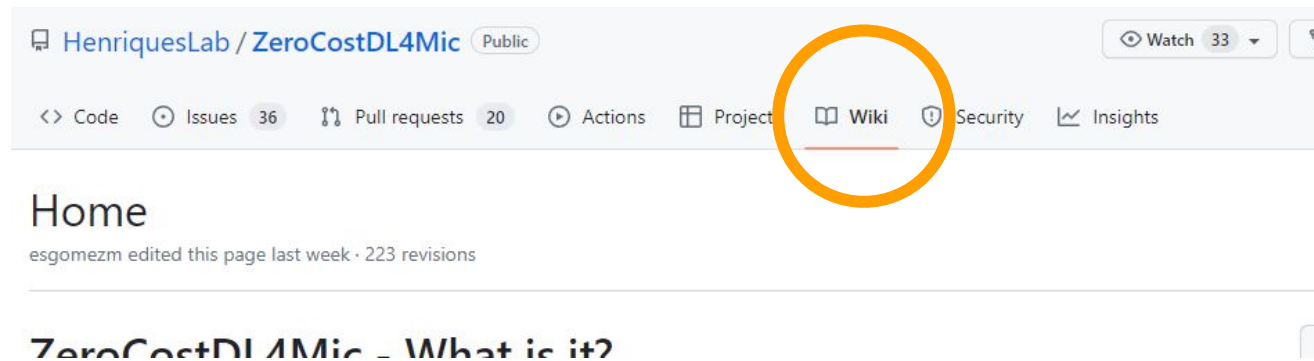
1 QC image pair



Data

Do you all have Google drives?

- Log in to your google account
- Using the same browser go to <https://github.com/HenriquesLab/ZeroCostDL4Mic/wiki>
- Google it: Github ZeroCostDL4Mic
- Make sure you are at the wiki sites

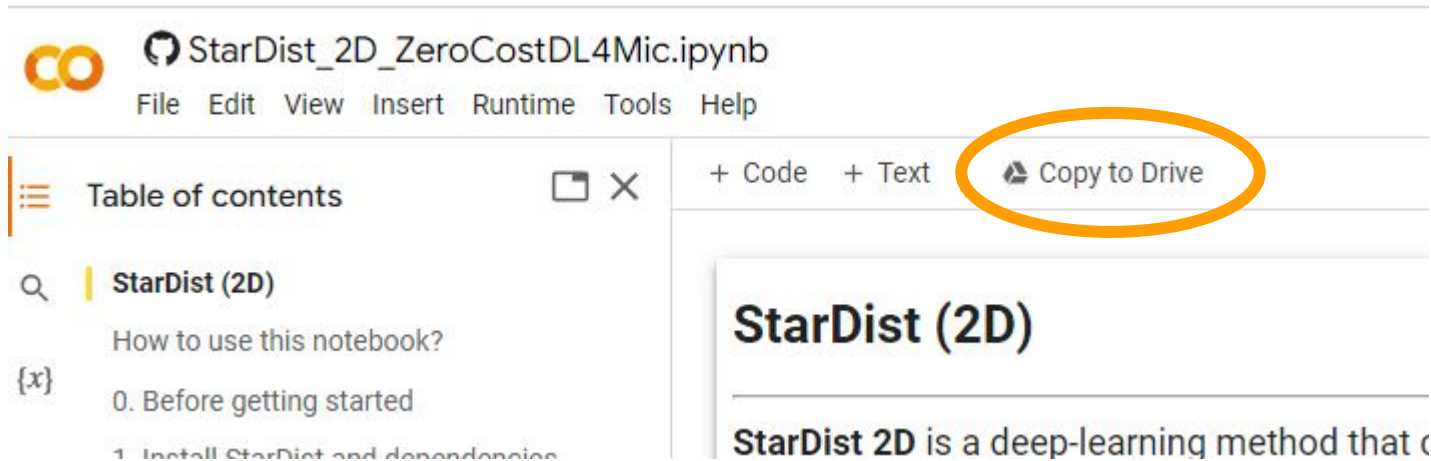


Open the notebook

- Look for the notebook, called **StarDist (2D)** and open it

StarDist (2D)	here and here	Instance segmentation	Fully supported	19/05/23 ✓ working (EGM)	here	 Open in Colab
---------------	---	-----------------------	-----------------	-----------------------------	----------------------	---

- Copy it to your drive and give it a name – close the original notebook



The screenshot shows the Google Colab interface for a notebook titled "StarDist_2D_ZeroCostDL4Mic.ipynb". The menu bar includes "File", "Edit", "View", "Insert", "Runtime", "Tools", and "Help". On the left, there is a "Table of contents" sidebar with a search icon and a list of sections: "StarDist (2D)", "How to use this notebook?", "0. Before getting started", and "1. Install StarDist and dependencies". In the top right of the notebook area, there are buttons for "+ Code", "+ Text", and "Copy to Drive", with the "Copy to Drive" button circled in orange. The main content area displays the title "StarDist (2D)" and the beginning of the text: "StarDist 2D is a deep-learning method that c".

- Will be stored in your google drive

 Colab Notebooks

Train your own model

▶ Path to training images:

Training_source: "/content/gdrive/Shareddrives/Image analysis course 2024/Day_2/2-Workshop/T cell dataset/1-Training_source"

Training_target: "/content/gdrive/Shareddrives/Image analysis course 2024/Day_2/2-Workshop/T cell dataset/2-Training_target"

Name of the model and path to model folder:

model_name: "Turku_IAC2024_xxepochs_YOURNAME"

model_path: "/content/gdrive/Shareddrives/Image analysis course 2024/Day_2/2-Workshop/T cell dataset/3-StarDist model"

Other parameters for training:

number_of_epochs: 10

Advanced Parameters

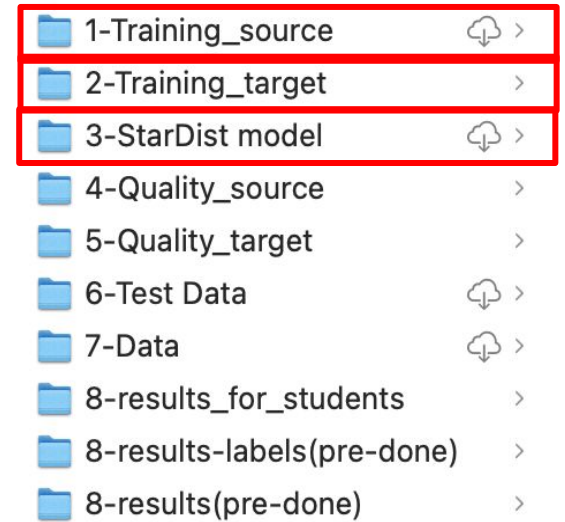
Use_Default_Advanced_Parameters: ☒

▶ Use_Data_augmentation: ☐

▶ Loading weights from a pre-trained network

Use_pretrained_model: ☐

pretrained_model_choice: 2D-versatile_fluo_from_Stardist_Fiji



More epochs = training slower

Choose between 5-15 (should take 5-10 min)

Leave both unticked