

ZeroCostDL4Mic User Manual

Henriques and Jacquemet labs
v1.0 – April 2020

Table of Contents

A.	What is ZeroCostDL4Mic?	3
A.1.	Overview of the workflow	3
A.2.	Networks currently implemented	4
A.3.	Online resources	5
B.	Preparing the data for ZeroCostDL4Mic	6
B.1.	Obtaining the dataset	6
B.2.	Uploading the dataset to Google Drive.....	6
C.	Getting started with the notebook and initialising the Colab session	7
C.1.	Opening the notebook.....	7
C.2.	Initialising the Colab session.....	8
C.3.	Installing network components and related dependencies	9
C.4.	Setting training parameters.....	9
D.	Training the model and performing quality control.....	12
D.1.	Training.....	12
D.2.	Quality control on the trained model.....	13
E.	Using the trained model to obtain predictions on new data.....	16
F.	Final notes.....	18
G.	References	19

A. What is 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. Training data can be uploaded to the Google Drive from where it can be used to train models using the provided Colab notebooks in a web-browser. Predictions on unseen data can then be performed within the same notebook, therefore not requiring any local hardware or software set-up.

This guide describes the workflow of a typical notebook run-through in a step-by-step fashion. You can also refer to this video for a complete walkthrough.

<https://www.youtube.com/watch?v=GzD2gamVNHI&feature=youtu.be>

A.1. Overview of the workflow

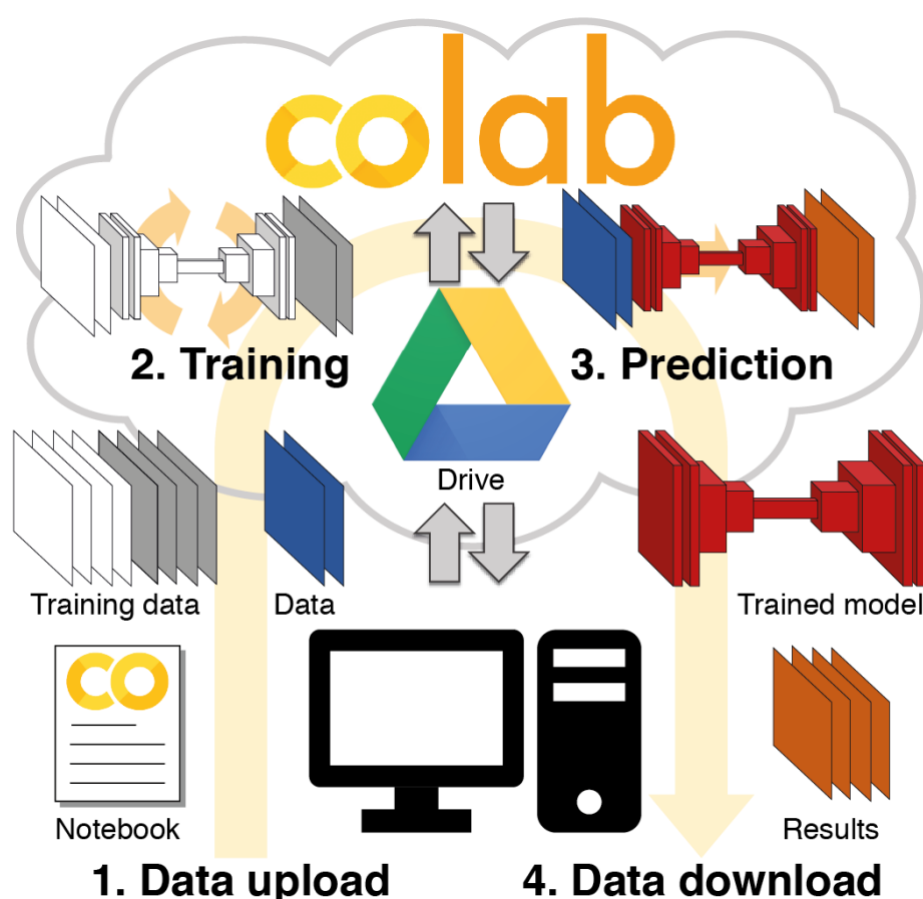


Figure 1: Workflow of ZeroCostDL4Mic. From data upload, execution of training and prediction on the cloud, to results and trained model download.

ZeroCostDL4Mic exploits online resources made available through Google Colab. **The general workflow comprises the following steps and is showcased in following sections A-E:**

- i) Uploading data to Google Drive
- ii) Opening Colab Notebook online in Google Colab
- iii) Connecting the Google Drive to the current Colab session
- iv) Installation of network components dependencies
- v) Set training parameters and run the training
- vi) Run Quality control on the trained model to ensure the validity of the trained model
- vii) Run predictions on unseen dataset
- viii) Download the trained network and the predictions on the unseen datasets

A.2. Networks currently implemented

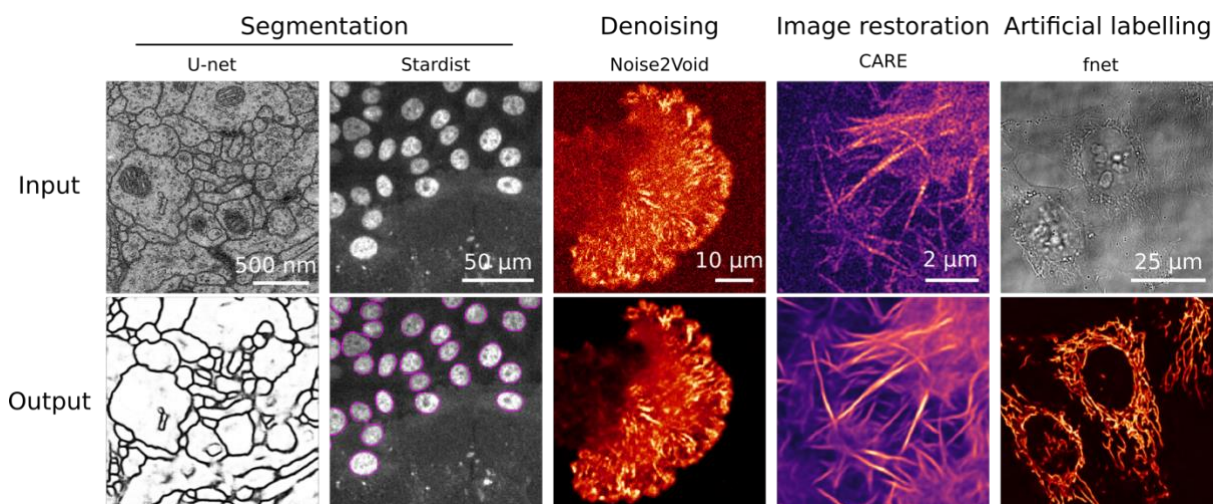


Figure 2: Networks and tasks currently implemented in ZeroCostDL4Mic. Segmentation, denoising, image restoration and artificial labelling are currently available.

- **U-net:** for segmentation of bioimages, typically from electron microscopy data or bright-field imaging^{1,2}.
- **Noise2Void (2D and 3D):** for denoising directly from the raw data, unsupervised so no need for paired training dataset³.
- **Stardist (2D and 3D):** for segmentation of cell nuclei in fluorescence images, specifically designed to detect round/oval objects^{4,5}.
- **CARE (2D and 3D):** provides image restoration (denoising and resolution enhancement, among other features) using fully supervised training⁶.
- **Label-free prediction (fnet):** predicts pseudo-fluorescence images from bright-field or EM data. The code is originally based on the code found here⁷.

A.3. Online resources

Code development is a never-ending endeavour. We keep the framework alive by constantly developing new capabilities and adding new tasks and networks. Therefore, we make available many useful pieces of information via [our GitHub page](#).

Below are few links to specific online resources:

Google Colab

<https://colab.research.google.com/notebooks/intro.ipynb>

Link to our bioRxiv paper

<https://www.biorxiv.org/content/10.1101/2020.03.20.000133v1>

Wiki page on our GitHub repository

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

Reporting bugs or issues

<https://github.com/HenriquesLab/ZeroCostDL4Mic/issues>

Page to the latest releases of the notebooks

<https://github.com/HenriquesLab/ZeroCostDL4Mic/releases>

Tips, tricks and FAQs

<https://github.com/HenriquesLab/ZeroCostDL4Mic/wiki/Tips,-tricks-and-FAQs>

Zenodo (dataset repository, use *ZeroCostDL4Mic* in the search bar)

<https://zenodo.org>

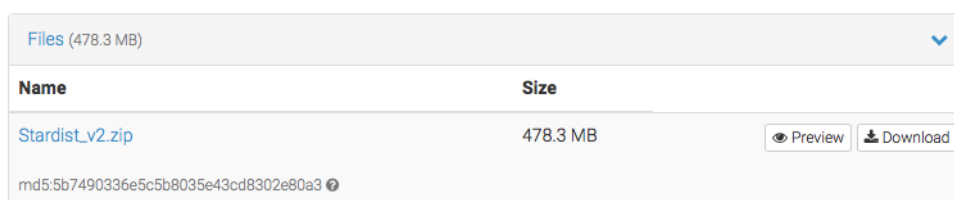
B. Preparing the data for ZeroCostDL4Mic

First, you should decide what task they want to perform using ZeroCostDL4Mic: we currently provide published networks that can perform image segmentation, denoising, restoration and artificial labelling. Please refer to our main Wiki page for info on the currently implemented notebooks. Users should also read our bioRxiv preprint⁸ on the general framework to understand whether that is the right tool for the application that the user has in mind.

B.1. Obtaining the dataset

Now, you need some training datasets. You can either test the networks on the example training and test dataset that we provide or generate your own training dataset to be used on your own data. The different pages of our Wiki describe how to acquire the data you need in order to train the different networks.


If you decide to simply test the networks on our example training and test dataset, you can follow the different **Zenodo links that we provide** on our main Wiki page. You can download the whole dataset by clicking on the **Download** button on the Zenodo page. For U-net, we do not provide dataset ourselves but point towards an available dataset originally generated for the ISBI 2012 Segmentation challenge, see the U-net page for details.



Files (478.3 MB)		
Name	Size	
Stardist_v2.zip	478.3 MB	Preview Download
md5:5b7490336e5c5b8035e43cd8302e80a3		

Figure 3: Obtaining our example training and test dataset: Zenodo download link, from <https://zenodo.org/record/3715492#.Xo8bki2ZNQK>

B.2. Uploading the dataset to Google Drive

In order for Google Colab to have access to these datasets, they need to be uploaded to your [Google Drive](#). You can do that simply by logging into your Google Drive account and clicking on  and then use the *Folder upload* option. **Please ensure that ALL the data is uploaded properly before proceeding further.**

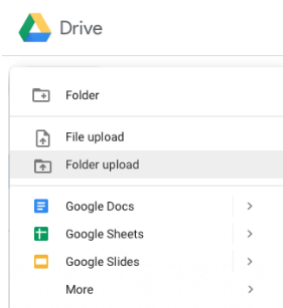



Figure 4: Uploading data to Google Drive.

C. Getting started with the notebook and initialising the Colab session

C.1. Opening the notebook

You can directly open the Colab notebook from our main Wiki page, by clicking on the corresponding **Open In Colab badge** . This will automatically open the notebook in Google Colab. Here, we have chosen to use the Dark mode of the notebook (white text on black background) for improved visibility.

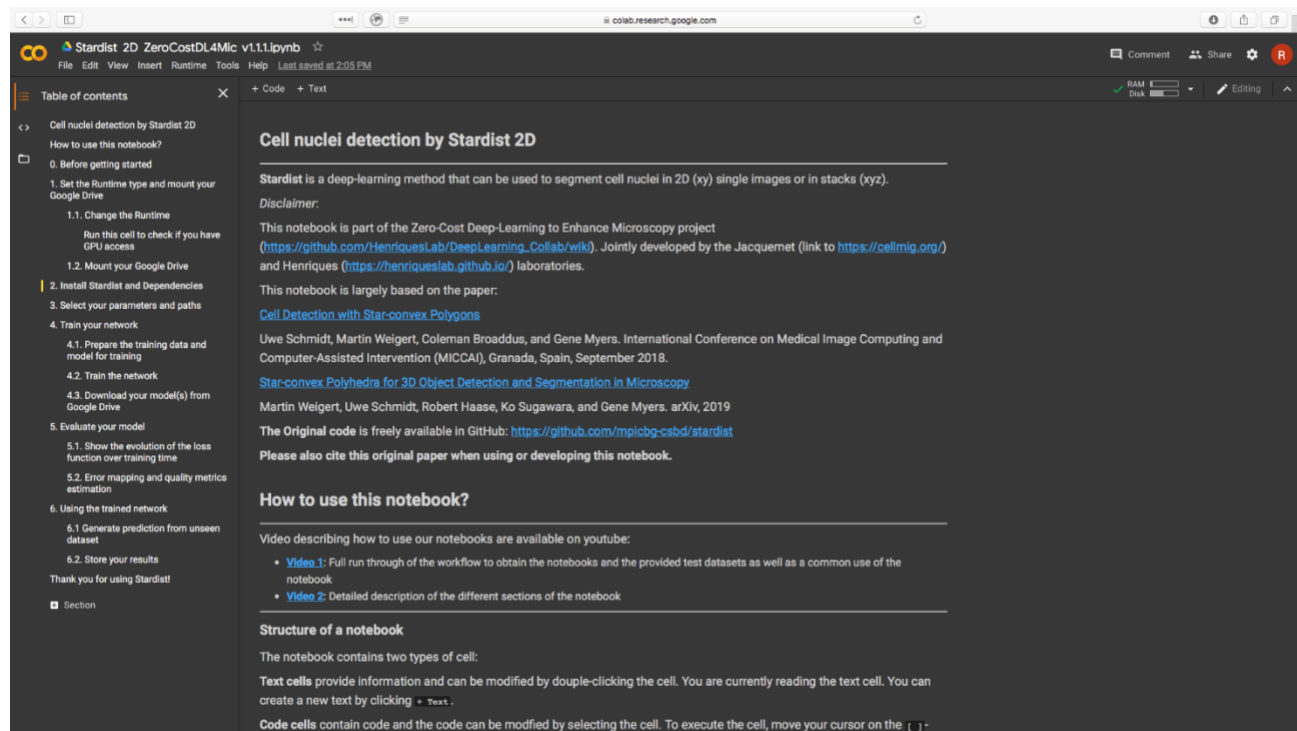


Figure 5: The ZeroCostDL4Mic notebook layout. The notebook shown here is that of Stardist (2D). The table of contents on the left highlights the different sections in the notebook.

On the left-hand side you will find a **table of contents** that can be navigated through by clicking on the different items. It is important that you read the **0. Before getting started** section.

At that stage, you will not be able to make changes to the notebook unless you **save a copy of the notebook**. This can be easily done by going into *File* and *Save a copy in Drive....* If you are signed into a Google Drive account, this will automatically save a copy of the notebook in your Google Drive in the Colab Notebooks folder.

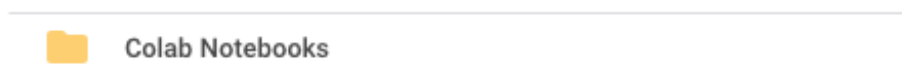


Figure 6: Colab Notebooks folder in Google Drive where all copies of existing notebooks appear by default. The notebooks can be subsequently moved to other folders.

C.2. Initialising the Colab session

Now you're in the notebook, the first step is to check whether Google Colab will currently allow GPU access and, if so, what type. You can do this by running the **1.1 Change the runtime** section. The GPU type is described on the last line of the Cell output. Here, we have been allocated a **Tesla P100-PCIE-16GB**.

```

1.1. Change the Runtime
Go to Runtime -> Change the Runtime type
Runtime type: Python 3 (Python 3 is programming language in which this program is written)
Accelerator: GPU (Graphics processing unit)

Run this cell to check if you have GPU access

TensorFlow 1.x selected.
You have GPU access
[{"name": "/device:CPU:0",
  device_type: "CPU",
  memory_limit: 268435456,
  locality {
  },
  incarnation: 4583164427646872838, name: "/device:XLA_CPU:0",
  device_type: "XLA_CPU",
  memory_limit: 17179869184,
  locality {
  },
  incarnation: 2273771063487264438,
  physical_device_desc: "device: XLA_CPU device", name: "/device:XLA_GPU:0",
  device_type: "XLA_GPU",
  memory_limit: 17179869184,
  locality {
  },
  incarnation: 12138839317506808780,
  physical_device_desc: "device: XLA_GPU device", name: "/device:GPU:0",
  device_type: "GPU",
  memory_limit: 15956161332,
  locality {
    bus_id: 1
    links {
    }
  },
  incarnation: 486436365595342333,
  physical_device_desc: "device: 0, name: Tesla P100-PCIE-16GB, pci bus id: 0000:00:04.0, compute capability: 6.0"]}

```

Figure 7: Checking GPU availability and type.

Then, you need to give Google Colab access to the data that you uploaded into your Google Drive. This step is called **Mounting the drive** and is executed in the **1.2 Mount your Google Drive** section. You will need to be logged in and provide an access code accessible via the provided link in the notebook.

```

1.2. Mount your Google Drive

To use this notebook on the data present in your Google Drive, you need to mount your Google Drive to this notebook.

Play the cell below to mount your Google Drive and follow the link. In the new browser window, select your drive and select 'Allow', copy the code, paste into the cell and press enter. This will give Colab access to the data on the drive.

Once this is done, your data are available in the Files tab on the top left of notebook.

Play the cell to connect your Google Drive to Colab

• Click on the URL.
• Sign in your Google Account.
• Copy the authorization code.
• Enter the authorization code.
• Click on "Files" site on the right. Refresh the site. Your Google Drive folder should now be available here as "drive".

... Go to this URL in a browser: https://accounts.google.com/o/oauth2/auth?client\_id=947318989803-6bn6qk8qdg4n4g3pfee6491bc0brc4i.apps.googleusercontent.com
Enter your authorization code:

```

Figure 8: Mounting Google Drive to the Colab session. The link provided in the cell enables the authentication of the Google Drive and will provide the corresponding authorization code.

Once mounted, the drive and all its content are accessible via the dedicated tab on the left of the page.

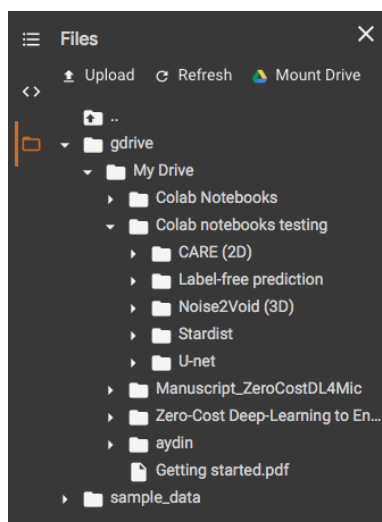


Figure 9: The File tab on the left of the notebook page gives access to the content of the mounted Google Drive.

C.3. Installing network components and related dependencies

Then, we are ready to install the network components and the required dependencies on the virtual machine made available to us by Google Colab. This is done by running the cell corresponding to the **section 2** of the notebook.

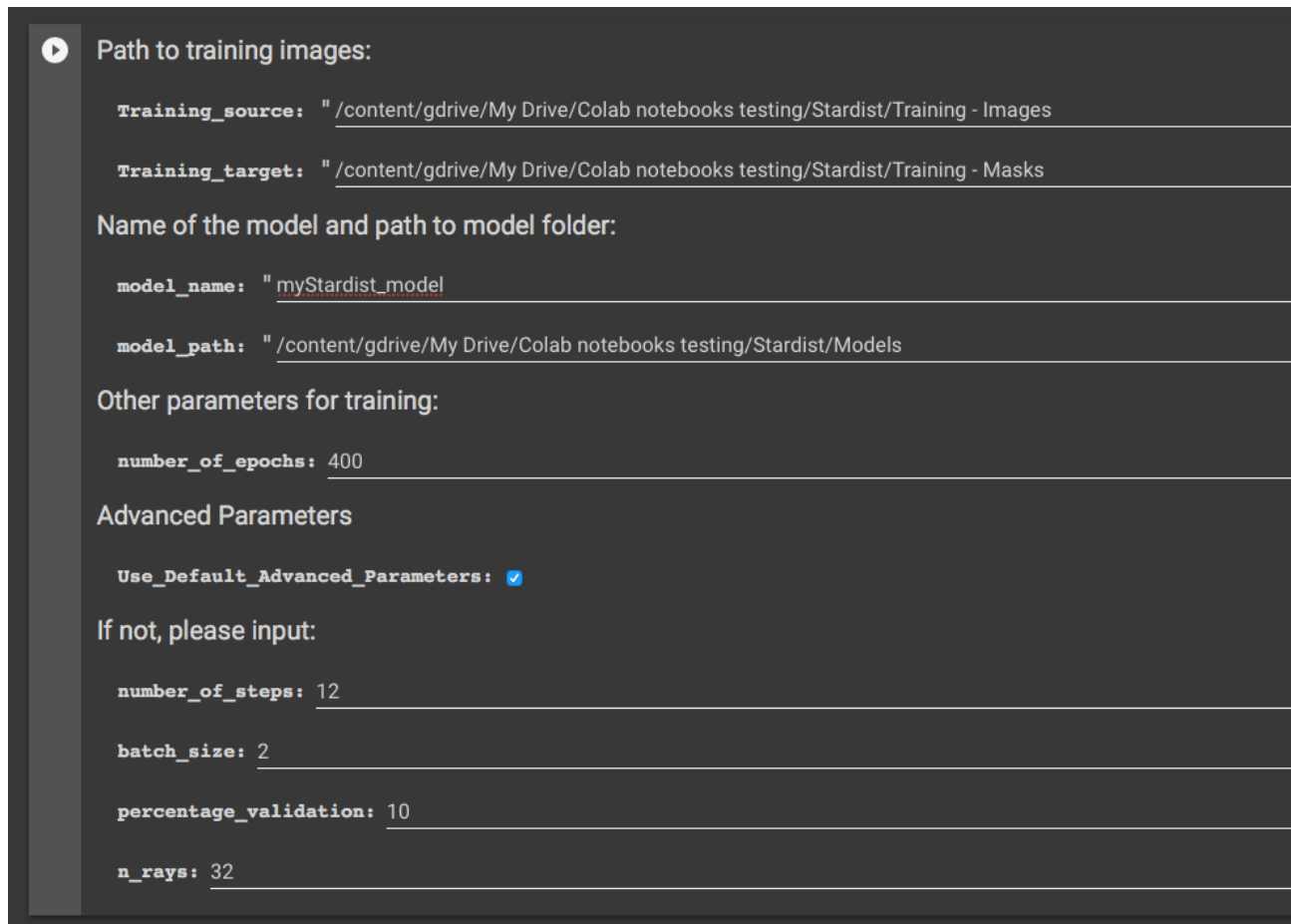
Important note: After running section 2, if you already have a trained model and wish to perform quality control on it, you can simply jump to **5. Evaluate your model** section. Similarly, if you only wish to run predictions on new unseen dataset, you can jump directly to **6. User the network** section.

C.4. Setting training parameters

Section 3 of the notebook then provides very important information and requires important user input to prepare for the network training. The first part of section 3 highlights and describes the different parameters that the user will need to provide to prepare training the model. This will include **providing the location (path, on the Google Drive) of the different parts of the training dataset** (source and target for fully supervised networks such as U-net, Stardist, CARE and Label-free prediction, or only the source training dataset for unsupervised networks like Noise2Void). The other parts of the user input will be the training parameters, **e.g. number of epochs, number of steps, batch size**. If you are unsure about the meaning of these parameters, you can also refer to our [Glossary page](#).

Some parameters are currently considered as advanced and can simply be left as their default values if the user wants to get started with a simple training session.

A typical section 3 user input will look like the figure below. **Do not forget to run this cell for the notebook to take your input into account.**



▶ Path to training images:

Training_source: "/content/gdrive/My Drive/Colab notebooks testing/Stardist/Training - Images"

Training_target: "/content/gdrive/My Drive/Colab notebooks testing/Stardist/Training - Masks"

Name of the model and path to model folder:

model_name: "myStardist_model"

model_path: "/content/gdrive/My Drive/Colab notebooks testing/Stardist/Models"

Other parameters for training:

number_of_epochs: 400

Advanced Parameters

Use_Default_Advanced_Parameters: ☒

If not, please input:

number_of_steps: 12

batch_size: 2

percentage_validation: 10

n_rays: 32

Figure 10: Setting the training parameters. Should be provided: paths to training dataset, name for the new model, path to where the new model will be created and saved, network-specific training parameters. Advanced parameters are also accessible.

For some networks, the notebook will highlight a randomly selected pair of source and target images from the training dataset for inspection that the data was uploaded and mounted properly.

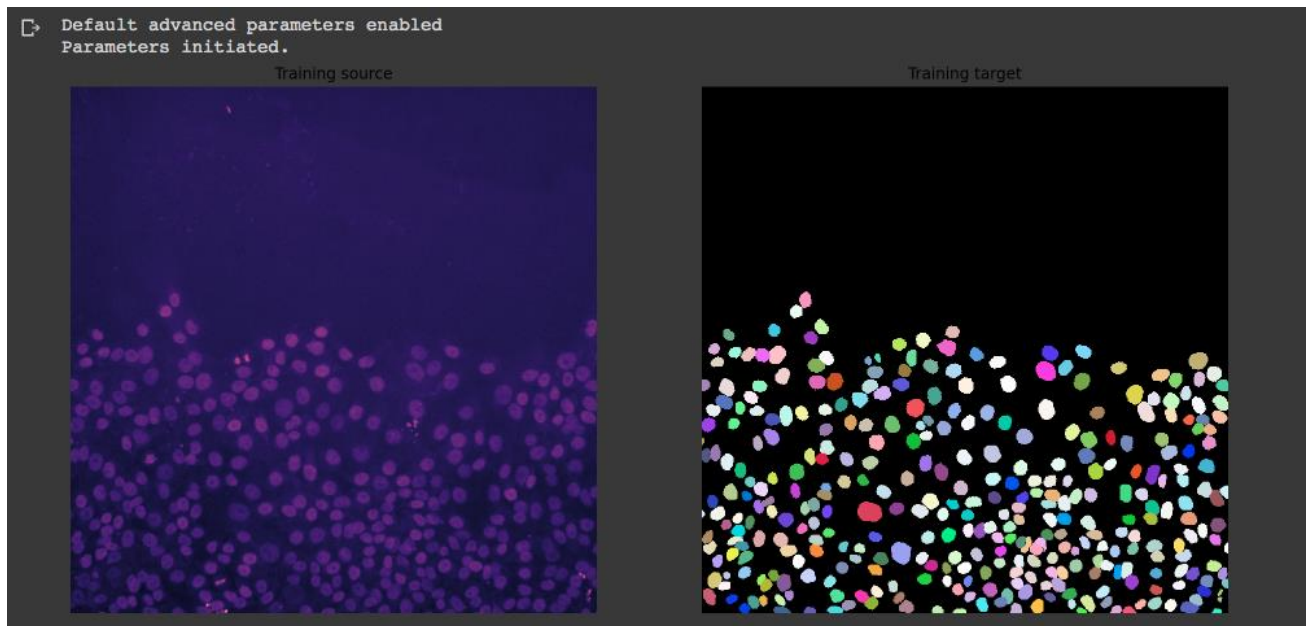


Figure 11: Display of an example of training dataset pair for Stardist 2D. This dataset was taken from the example training dataset that we provide.

D. Training the model and performing quality control

D.1. Training

Here we go! Now, you can start the training by running the **4.2 Train the network** section. In some cases, the network may throw some minor warnings about TensorFlow versions, *this is not an issue and can be ignored*. We have enforced specific versions of TensorFlow in order to improve reliability of the notebooks.

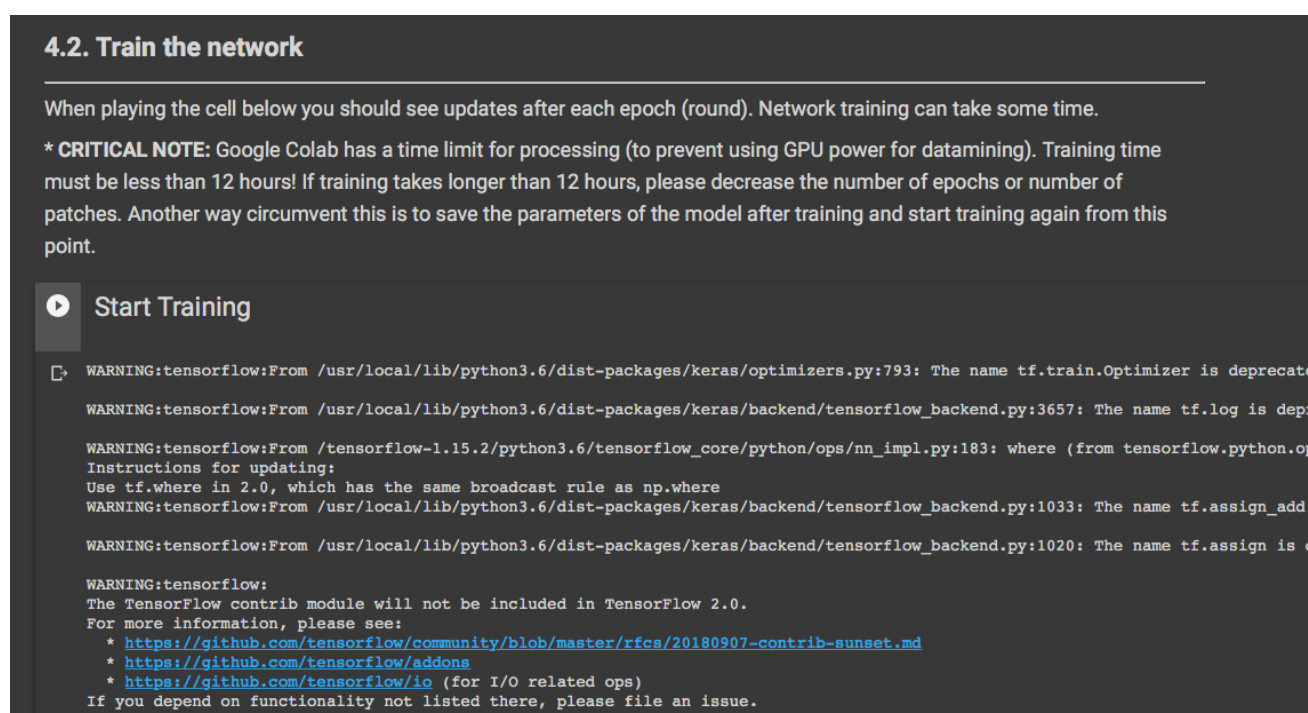


Figure 12: Training the model with the user-set parameters and training dataset.

This step can take a few minutes to a few hours depending on the network, training parameters and training dataset size. So be patient! Just ensure that the training time does not exceed 12 h, as this represents the current time limit of a Google Colab session. The time taken by each epoch is displayed for some networks as an output of the network training. This can be used to estimate how long the training will take (here shown as ~2s / epochs and 400 epochs, it should take about ~15min, so yes, you have time for a coffee). The total time taken for the training will be indicated once it is complete.

```

Epoch 73/400
12/12 [=====] - 2s 133ms/step - loss: 0.6066 - prob_loss: 0.1177 - dist_loss: 2.4445 - prob_kld: 0.0249
  
```

Figure 13: Network training output for each epoch, including the time taken for the epoch. This can help estimating the total time that the training will take.

After training is complete, **the trained model is automatically saved** into the **model_path** folder previously selected. This allows you to set the training running and ignore the risk of running into time-out issues after the training is over. The following sections of the notebooks can be run as a new Colab session if sections 1 and 2 of the notebook are run again.

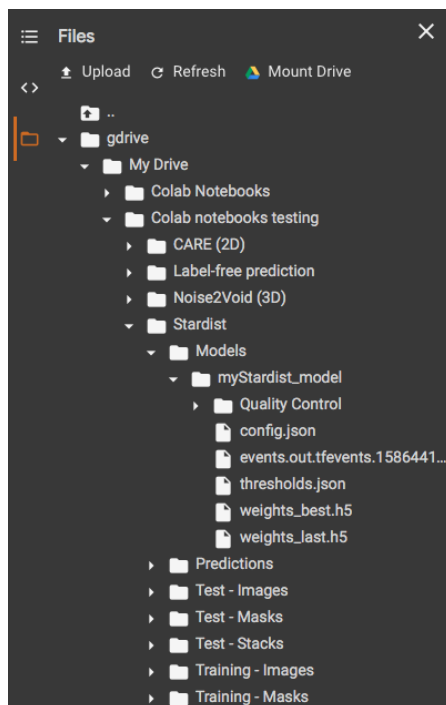


Figure 14: The newly trained model is automatically saved in the selected model folder.

You can download the trained model from your Google Drive for safekeeping or for sharing with another user (provided the limitations of using pre-trained model highlighted in our bioRxiv paper⁸).

D.2. Quality control on the trained model

Section 5 of the notebook is very important! It allows you to perform some important quality control on the trained model. This can be evaluated using two complementary approaches:

- (1) Observe the evolution of the loss function as a function of training time. This allows for the inspection of the presence of over- or under-fitting of the model to the training dataset.
- (2) Compute quantitative image metrics on Quality control dataset (also often called "Test dataset" in the machine learning field), where the known ground truth can be compared to the prediction from this trusted dataset.

The quality control datasets should **not** be included in the training dataset during training, otherwise the network performance will be over-estimated.

First, you will need to choose whether you are performing Quality control on the current trained model (as obtained in section 4), or whether you will be doing on a model that was

previously trained in a different session. For the latter, you will need to provide model path and name.

5. Evaluate your model

This section allows the user to perform important quality checks on the validity and generalisability of the trained model.

We highly recommend to perform quality control on all newly trained models.

▶ Do you want to assess the model you just trained ?

Use_the_current_trained_model: ☒

If not, please provide the name of the model and path to model folder:

During training, the model files are automatically saved inside a folder named after the parameter 'model_name' (see section 3). Provide the name of this parent folder in 'QC_model_path'.

QC_model_name: " Insert text here "

QC_model_path: " Insert text here "

Figure 15: Choice of trained model to perform quality control. The path and name of a previously trained model can be given.

Below are examples of loss function curves over training time, both from the training and validation dataset. If this doesn't make sense, please see our [Glossary page](#) and also this review by Moen *et al.*⁹, which explains how to interpret the curves very well.

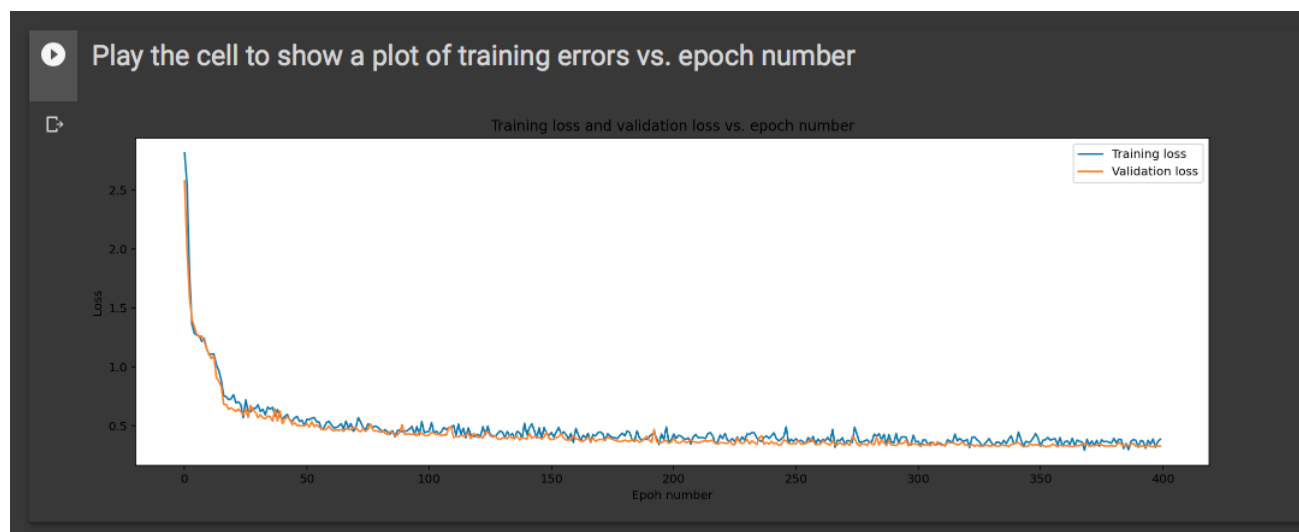


Figure 16: Evaluation of the model performance by inspection of the training and validation loss curves. A significant deviation over time between the two curves highlight a risk of over-fitting of the model to the training dataset.

The second stage of quality control is to see whether the network is able to generalise to unseen dataset, using a quality control dataset. This is performed here by the user providing a dataset with the source images and the equivalent known ground truth targets. Depending on the type of network, we use **Root Squared Error** (RSE) and **Structural SIMilarity** (SSIM)

or **Intersection over Union** (IoU) as metrics in order to visually and qualitatively assess whether the model can provide accurate output from unseen data.

Here, in the case of Stardist, because the model predicts a segmentation image, the metric used is Intersection over union. We describe the meaning of these metrics in details in the Supplementary information of our paper.



Figure 17: Quantitative assessment of a Stardist model performance using masks overlay and Intersection over union metric. The image on the left is the source file to be segmented. The green mask represents the manually segmented ground-truth image, the magenta mask is the network prediction, the overlay shows the agreement and disagreement between the two. The Intersection over union metric is indicated over the overlay image.

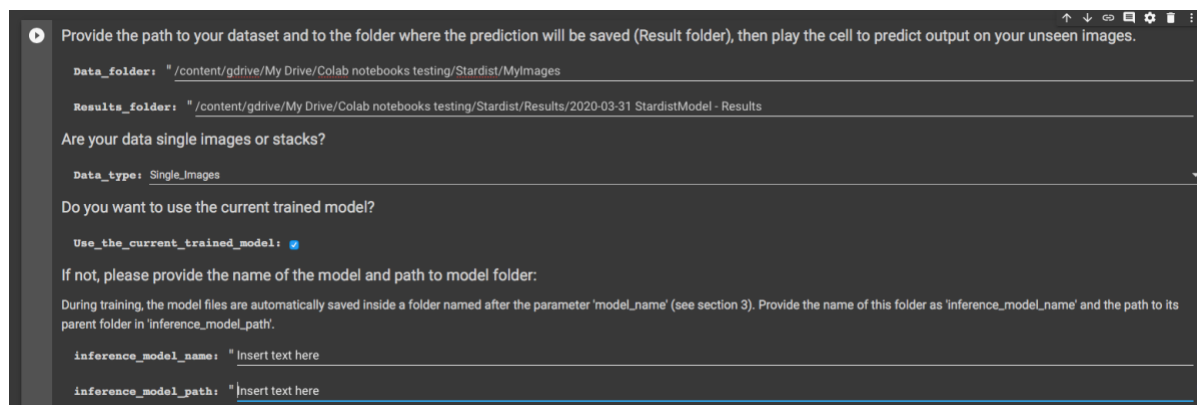
Once the metrics have been calculated the predictions and the metrics maps and indices are all automatically saved in the **model_path** folder in *Quality Control* folder. You are strongly encouraged to take a look and analyse the data obtained on the test dataset in order to build confidence on the validity of the trained model to analyse the data in the Quality control dataset.

Important note: When wondering if a pre-trained model may be suitable to analyse new type of data. This section of Quality control can also be used to quantify and estimate if the model will perform well with this new type of data.

E. Using the trained model to obtain predictions on new data

The last step, available in **Section 6**, is what you have been waiting for the whole time: It is the generation of predictions from unseen data using the trained model. This can be performed by giving the path to the Google Drive directory containing the new unseen data that you wish to run the prediction on.

A different model from the currently loaded model can also be used by providing the name and path to the model to be used.



Provide the path to your dataset and to the folder where the prediction will be saved (Result folder), then play the cell to predict output on your unseen images.

Data_folder: `"/content/gdrive/My Drive/Colab notebooks testing/Stardist/Myimages"`

Results_folder: `"/content/gdrive/My Drive/Colab notebooks testing/Stardist/Results/2020-03-31 StardistModel - Results"`

Are your data single images or stacks?

Data_type: `Single_Images`

Do you want to use the current trained model?

Use_the_current_trained_model: ☒

If not, please provide the name of the model and path to model folder:

During training, the model files are automatically saved inside a folder named after the parameter 'model_name' (see section 3). Provide the name of this folder as 'inference_model_name' and the path to its parent folder in 'inference_model_path'.

inference_model_name: `"Insert text here"`

inference_model_path: `"Insert text here"`

Figure 18: User input to run predictions on unseen data. In the particular case of Stardist, it is possible to run it directly on stacks (or on single images), and therefore there is an option to select this.

The notebook will show you an example of prediction chosen at random from the set of data provided. In this example we trained a Stardist model and the notebook shows an overlay of the original input image with the mask image obtained from the prediction.

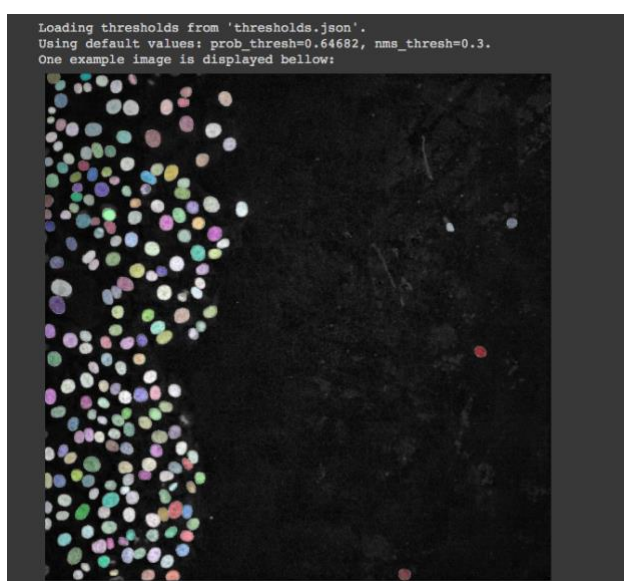


Figure 19: Prediction output display. The notebook chooses a random dataset to display for the visual assessment of the prediction.

The predictions obtained from the unseen data is now available in the Results folder as chosen earlier and can therefore be downloaded from your Google Drive for further analysis.

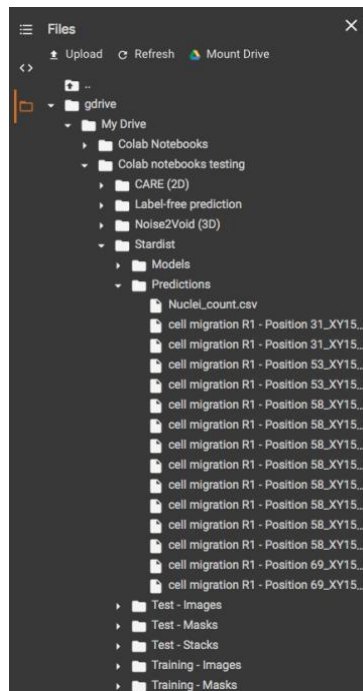


Figure 20: The predictions are automatically saved in the selected Results folder.

F. Final notes

Many thanks for trying out **ZeroCostDL4Mic**! Whether you find it useful, intuitive or difficult and buggy, we want to hear from you and always welcome constructive feedbacks. Feel free to report issues in the GitHub issue page or simply Tweet your results using #ZeroCostDL4Mic.

G. References

1. Ronneberger, O., Fischer, P. & Brox, T. U-Net: Convolutional Networks for Biomedical Image Segmentation. in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* 234–241 (2015). doi:10.1007/978-3-319-24574-4_28.
2. Falk, T. *et al.* U-Net: deep learning for cell counting, detection, and morphometry. *Nat. Methods* **16**, 67–70 (2019).
3. Krull, A., Buchholz, T.-O. & Jug, F. Noise2Void - Learning Denoising from Single Noisy Images. *ArXiv181110980 Cs* (2019).
4. Schmidt, U., Weigert, M., Broaddus, C. & Myers, G. Cell Detection with Star-Convex Polygons. in *Medical Image Computing and Computer Assisted Intervention – MICCAI 2018* (eds. Frangi, A. F., Schnabel, J. A., Davatzikos, C., Alberola-López, C. & Fichtinger, G.) vol. 11071 265–273 (Springer International Publishing, 2018).
5. Weigert, M., Schmidt, U., Haase, R., Sugawara, K. & Myers, G. Star-convex Polyhedra for 3D Object Detection and Segmentation in Microscopy. 8 (2020).
6. Weigert, M. *et al.* Content-aware image restoration: pushing the limits of fluorescence microscopy. *Nat. Methods* **15**, 1090–1097 (2018).
7. Ounkomol, C., Seshamani, S., Maleckar, M. M., Collman, F. & Johnson, G. R. Label-free prediction of three-dimensional fluorescence images from transmitted-light microscopy. *Nat. Methods* **15**, 917–920 (2018).
8. von Chamier, L. *et al.* ZeroCostDL4Mic: an open platform to simplify access and use of Deep-Learning in Microscopy. <http://biorxiv.org/lookup/doi/10.1101/2020.03.20.000133> (2020) doi:10.1101/2020.03.20.000133.

9. Moen, E. *et al.* Deep learning for cellular image analysis. *Nat. Methods* (2019)
doi:10.1038/s41592-019-0403-1.