



ZEROCOSTDL₄MIC

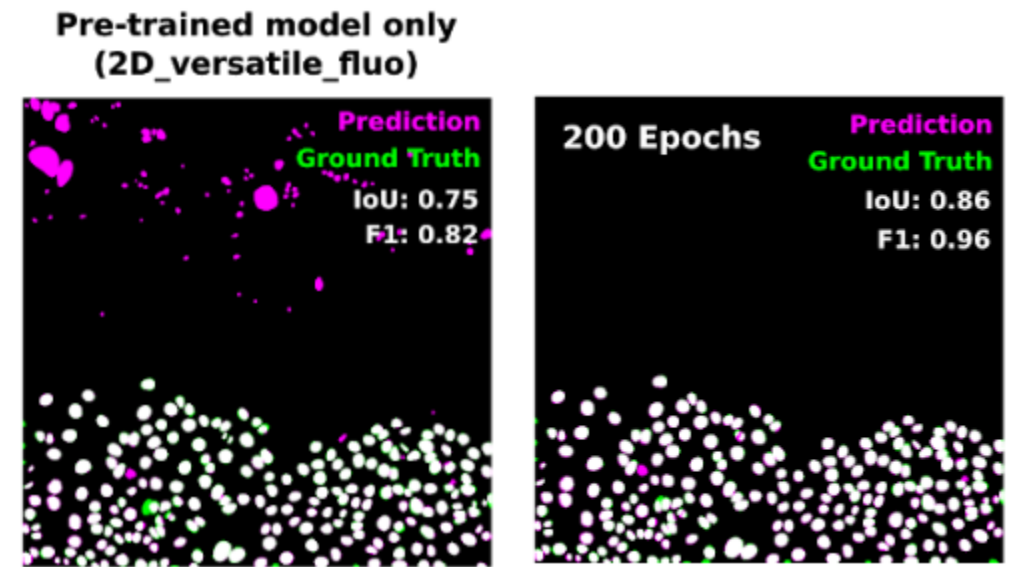
Quality assesment of Deep Learning models - Part 2

QUALITY CONTROL

DL segmentation

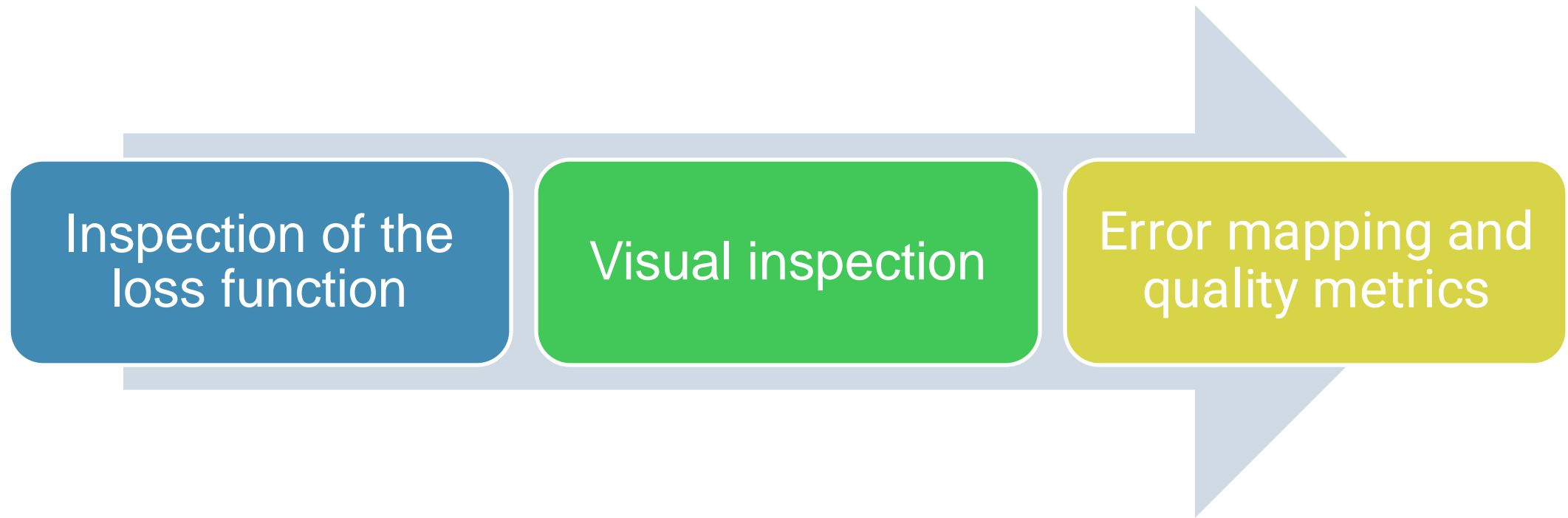
Why do we do Quality Control

- Ensures the model correctly identifies and classifies images, reducing errors.
- Helps detect and mitigate biases in data or model outputs that could lead to incorrect predictions.
- Verifies that the model performs well not just on training data but also on new, unseen images (test data).
- Ensures the model is resistant to noise or variations in input images, ensuring consistent performance.



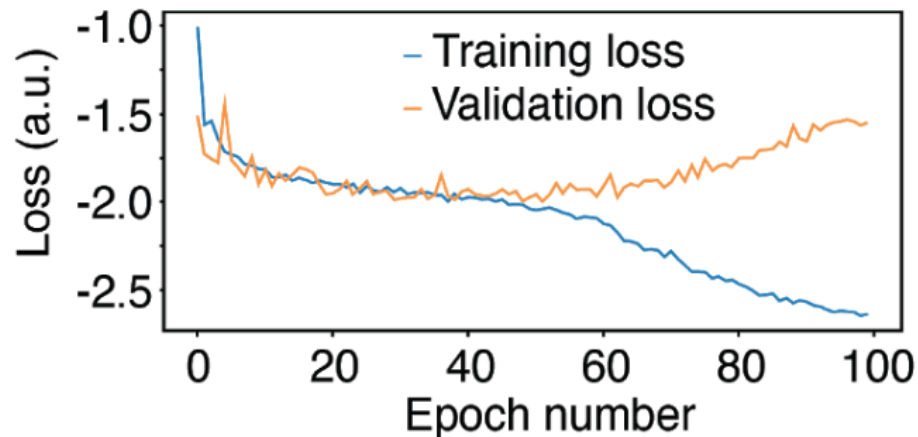
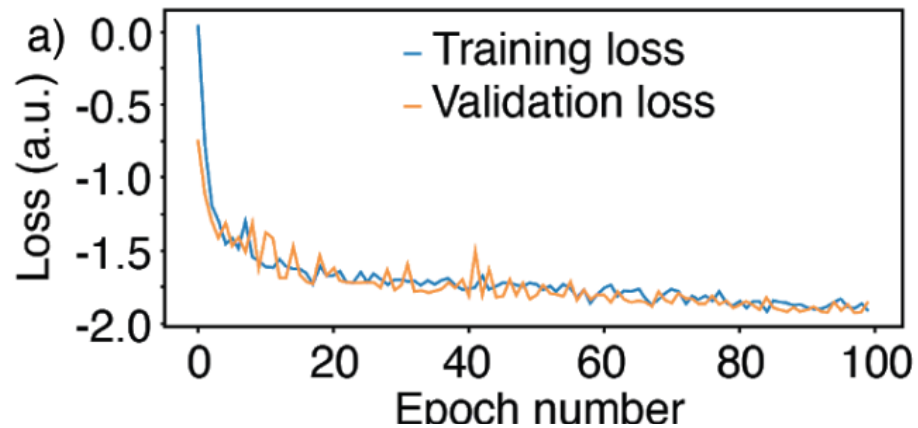
[Chamier., *et al.* Nature Communications (2021)]

Steps in Quality control



Step #1.4: Repeat

- Train model for hundred of EPOCH



Quality Control #1: Inspection of the loss function

Always evaluate the training progress, compare the training loss with the validation loss.

Training loss describes an error value after each epoch for the difference between the model's prediction and its ground-truth target.

Validation loss shows how well the network performs on the validation data.

Decreasing Training Loss and Validation Loss:

- **Action:** Continue training for more epochs, train for more epochs until the validation loss plateaus.

Curves Are Flattening Out:

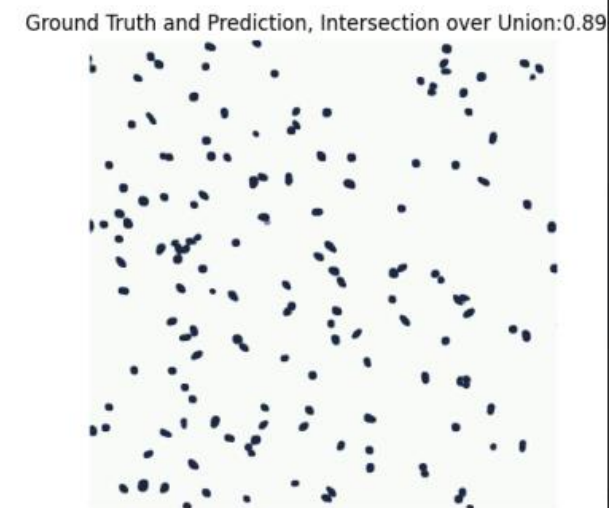
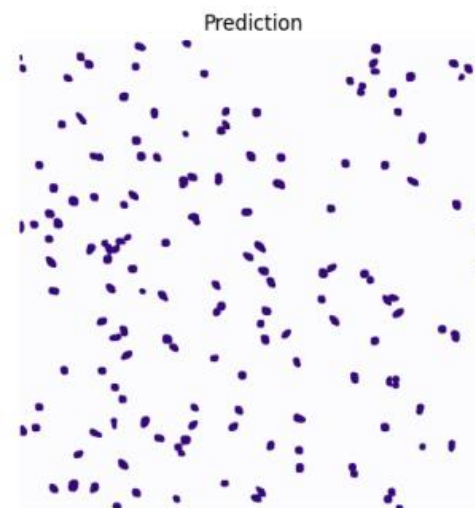
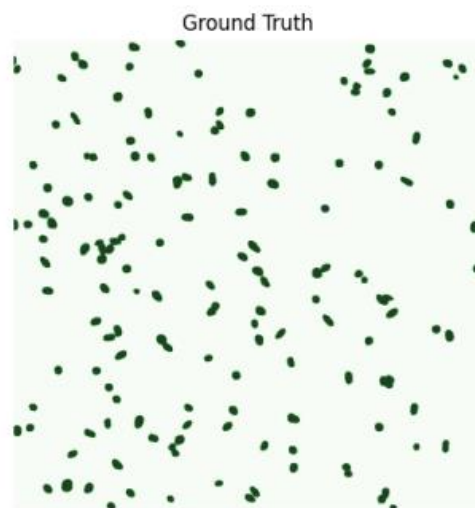
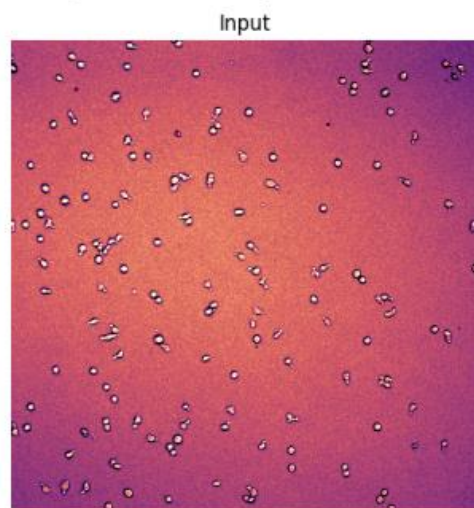
- **Action:** No more training may be needed.

Validation Loss Increases While Training Loss Decreases:

- **Diagnosis:** The network is overfitting.
 - Overfitting occurs when the model is memorizing the training data rather than generalizing well to unseen data.
- **Solutions to Overfitting:**
 1. Increase the size of the training dataset
 2. Augment your data

Quality Control Step #2: Visual inspection

Example Quality Control Visualisation

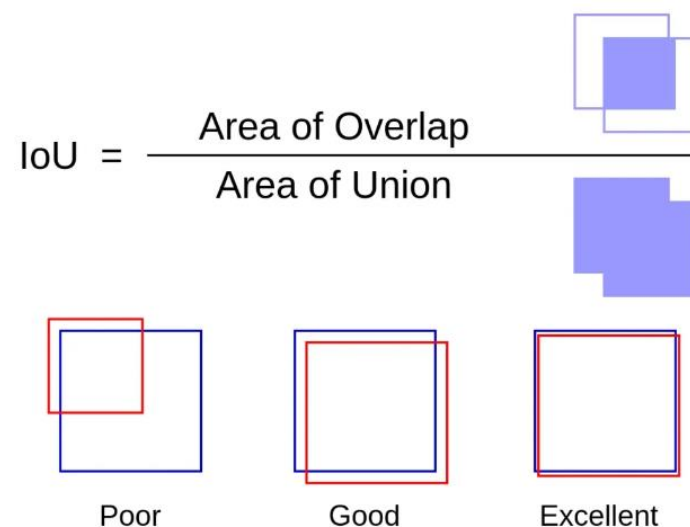


Quality Control3:

Error mapping and quality metrics estimation

| image name | Prediction v. GT Intersection over Union | false positive | true positive | false negative | precision | recall | accuracy | f1 score | n_true | n_pred | mean_true_ score | mean_matche d_score | panoptic _quality |
|----------------------------------|------------------------------------------------|-------------------|------------------|-------------------|-----------|---------|----------|----------|--------|--------|---------------------|------------------------|----------------------|
| Training_source _ICAM 1-1.tif | 0.89414007 | 2 | 153 | 2 | 0.987096 | 0.98709 | 0.974522 | 0.987096 | 155 | 155 | 0.873075 | 0.884487 | 0.873075 |

- Percent overlap between the target mask and your prediction output.
- **The closer to 1, the better the performance.**
- This metric can be used to assess the quality of your model to accurately predict nuclei. (whole image)



True and false positives

| image name | Prediction v. GT Intersection over Union | false positive | true positive | false negative | precision | recall | accuracy | f1 score | n_true | n_pred | mean_true_ score | mean_matche d_score | panoptic _quality |
|----------------------------------|------------------------------------------------|-------------------|------------------|-------------------|-----------|---------|----------|----------|--------|--------|---------------------|------------------------|----------------------|
| Training_source _ICAM 1-1.tif | 0.89414007 | 2 | 153 | 2 | 0.987096 | 0.98709 | 0.974522 | 0.987096 | 155 | 155 | 0.873075 | 0.884487 | 0.873075 |

“true positive”

- When a segmented object has an IoU above 0.5 (compared to the corresponding ground truth)

“false positive”

- “n_pred” - “true positive”

“false negative”

- “n_true” - “true positive”

n_true

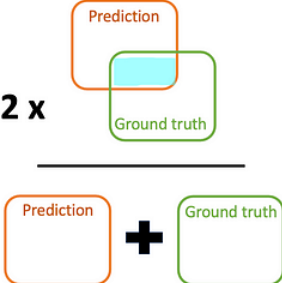
- number of objects in ground truth

n_pred

- number of objects in prediction

f1 score(Dice Coefficient) and mean matched score

| image name | Prediction v. GT Intersection over Union | false positive | true positive | false negative | precision | recall | accuracy | f1 score | n_true | n_pred | mean_true_ score | mean_matched_ d_score | panoptic _quality |
|----------------------------------|------------------------------------------------|-------------------|------------------|-------------------|-----------|---------|----------|----------|--------|--------|---------------------|--------------------------|----------------------|
| Training_source _ICAM 1-1.tif | 0.89414007 | 2 | 153 | 2 | 0.987096 | 0.98709 | 0.974522 | 0.987096 | 155 | 155 | 0.873075 | 0.884487 | 0.873075 |

$$\text{Dice} = \frac{2 \times \text{Area of overlap}}{\text{Total area}} = \frac{2 \times \text{Prediction} \cap \text{Ground truth}}{\text{Prediction} \cup \text{Ground truth}}$$


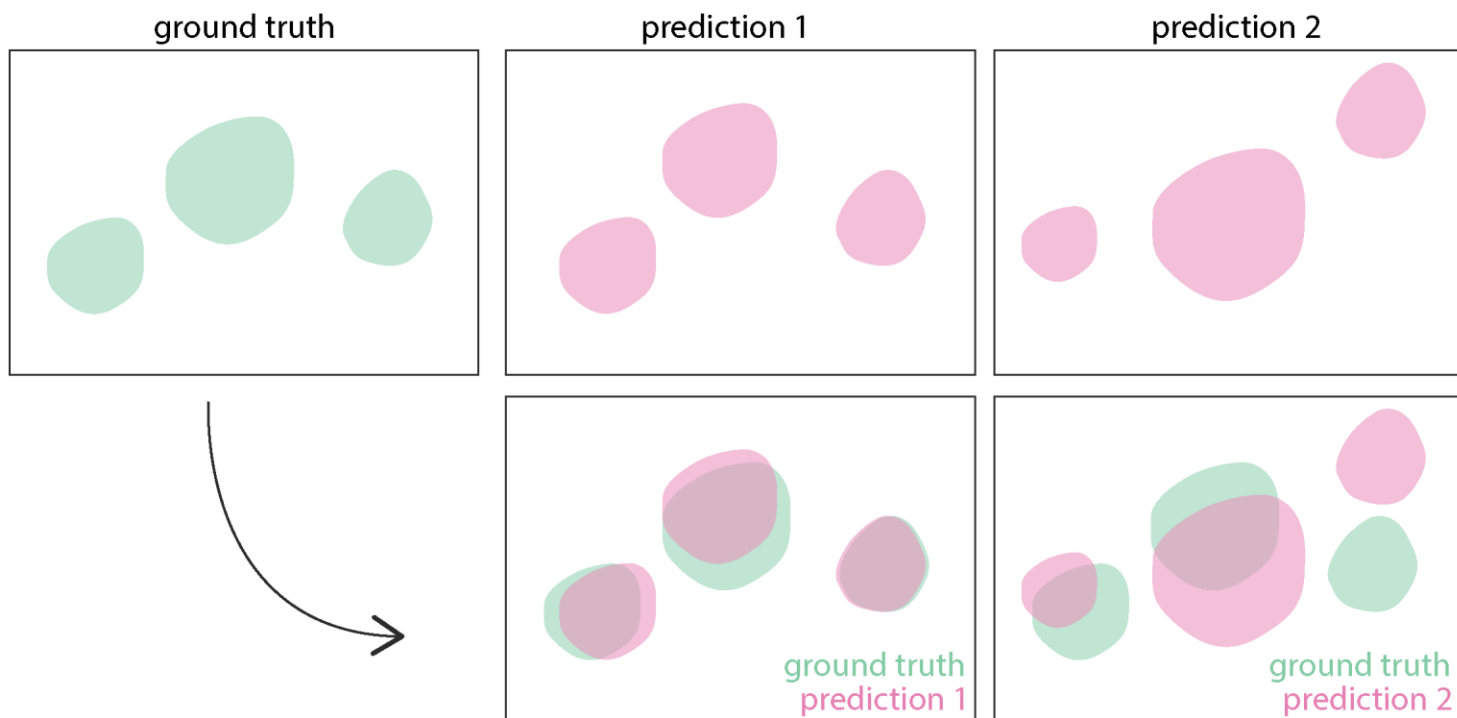
f1 score / dice score

- 2 * the Area of Overlap divided by the total number of pixels in both images.

The mean_matched_score

- is the mean IoUs of matched true positives.

Quality control overview



High IoU and F1 scores
true positive = 3
false positive = 0
false negative = 0

Low IoU and F1 scores
true positive = 2
false positive = 1
false negative = 1

▼ 5. Evaluate your model

This section allows you 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 path to the model folder:

`QC_model_folder:` "/content/gdrive/Shareddrives/Pasteur_image_analysis_course_2023/3-StarDist model/Pasteur_model_80epocs"

[Show code](#)

- ▶ Play the cell to show a plot of training errors vs. epoch number

[Show code](#)

- ▶ Choose the folders that contain your Quality Control dataset

`Source_QC_folder:` "/content/gdrive/Shareddrives/Pasteur_image_analysis_course_2023/4-Quality_source"

`Target_QC_folder:` "/content/gdrive/Shareddrives/Pasteur_image_analysis_course_2023/5-Quality_target"

Quality report

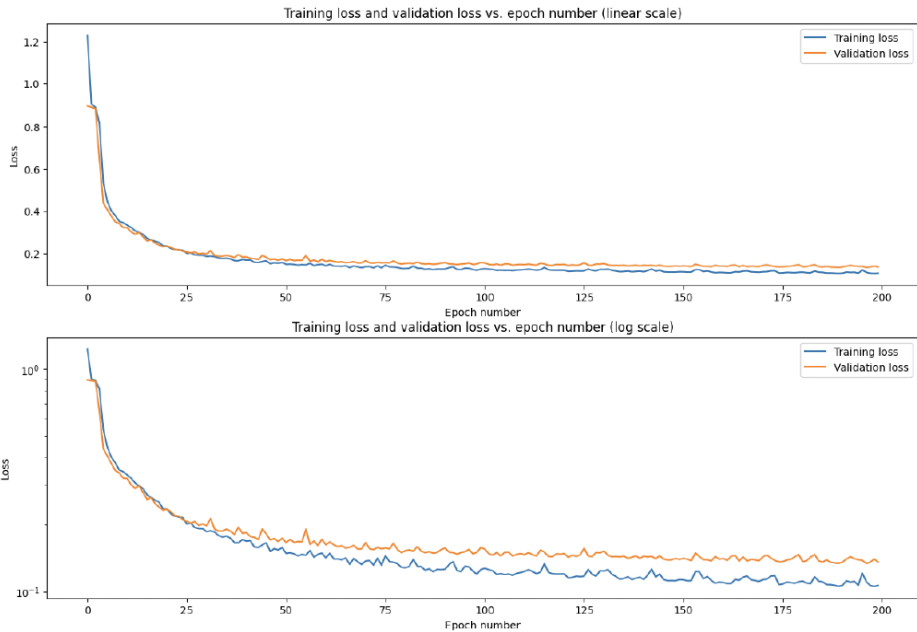
Loss inspection

Visual assessment

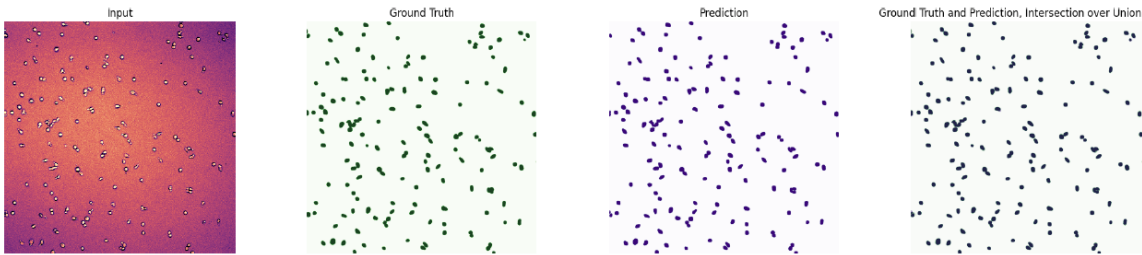
Quality metrics

Quality Control report for Stardist 2D model (turku2024_model_200epocs)
Date: 2024-09-13

Development of Training Losses



Example Quality Control Visualisation

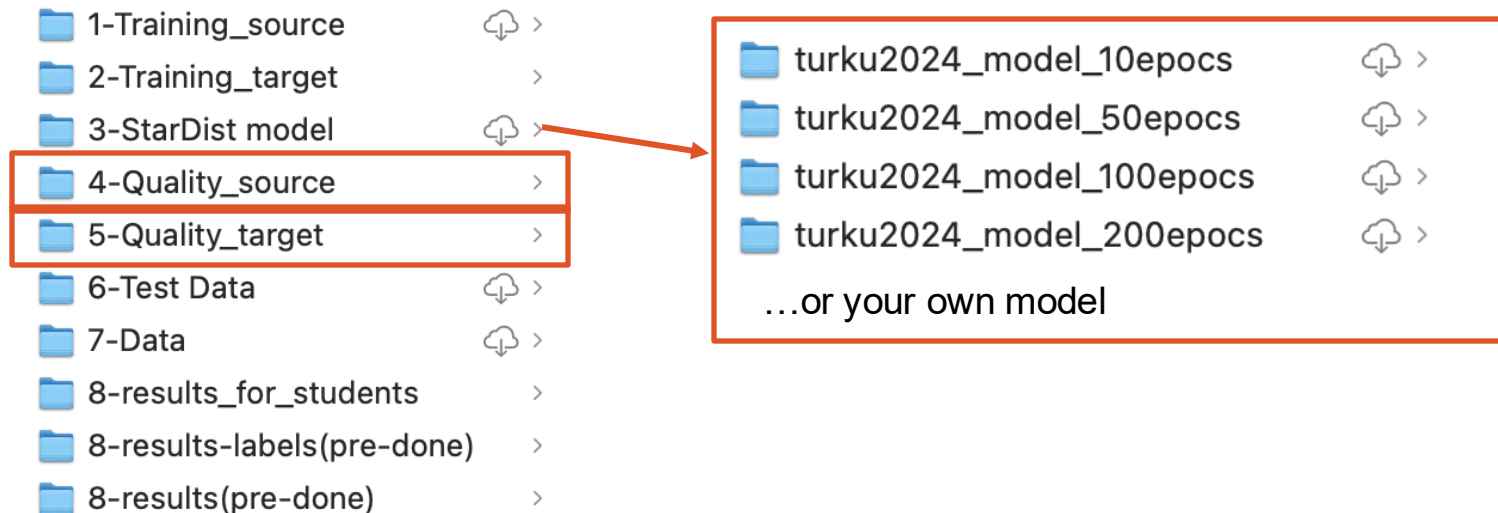


Quality Control Metrics

| image # | Prediction v. GT IoU | false pos. | true pos. | false neg. | precision | recall | accuracy | f1 score | n_true | n_predicted | mean_true_score | mean_matched_score | panoptic_quality |
|---------|----------------------|------------|-----------|------------|-----------|--------|----------|----------|--------|-------------|-----------------|--------------------|------------------|
| 1 | 0.891 | 1 | 153 | 2 | 0.994 | 0.987 | 0.981 | 0.99 | 155 | 154 | 0.872 | 0.884 | 0.875 |

Let's try!

- Go to your google drive and open your StarDist notebook
- Run Steps 1-2 again
- Run steps under 5. Evaluate your model



AND WE ARE DONE 😊