BIA4 ICA 1 - Example projects / datasets

Below you will find a list of datasets and example tasks that you can use for ICA 1. These are just examples, you can decide to work on something completely different if you prefer!

1. *Drosophila melanogaster* wings

Description

This dataset contains images of Drosophila wings with various genotypes.

Reference

Sonnenschein et al, 2015 - An Image Database of Drosophila melanogaster Wings for Phenomic and Biometric analysis

Available from

http://gigadb.org/dataset/view/id/100141

Data type and size

Several Gb of TIFF files taken at 20x and 40x with different microscopes.

Example tasks

- Automatically detect the wing parts
- Automate measurements of fly wings
- Identify the genotype of the fly from the image

---

2. Retina scans

Description

This dataset contains images of retina scans from glaucoma positive and negative patients.

Available from

https://www.kaggle.com/sshikamaru/glaucoma-detection

Data type and size

650 JPEG files (~200 Mb).

Example tasks

- classify images to detect whether glaucoma is present
- segment blood vessels and determine their properties (length, tortuosity, etc.)

---

3. Calcium imaging in neurons

Description

These datasets contain calcium imaging data from neurons, as part of the "CodeNeuron" challenge.

Available from

http://neurofinder.codeneuro.org/

Example tasks

- Extract regions of interest for each neuron
- Determine the activity profile of each cell
- Determine which cells are active at the same time
- Determine the spatial distribution of active cells

---

4. Breast Cancer Cell Segmentation

Description

The BreaKHis dataset contains images from H&E stained biopsies from benign and malignant breast tumors.

Reference

Spanhol et al, 2016 - "A Dataset for Breast Cancer Histopathological Image Classification"

Data type and size

850 Mb of PNG files.

Available from

https://www.kaggle.com/forderation/breakhis-400x

Example tasks

- Counting cells
- Segmentation of nuclei
- Measurement of cell properties
- Classification of tumour type

---

5. Nematodes

Description

A dataset of images of different nematodes.

Reference

https://arxiv.org/abs/2103.08335

Data type and size

JPG images ~ 600 Mb.

Available from

https://github.com/xuequanlu/I-Nema

Example tasks

- Segment the nematodes
- Identify nematode species

---

6. Cell tracking

Description

The cell tracking challenge aims to improve algorithms in cell tracking in 2D and 3D.

Reference

Ulman et al, 2017 - An objective comparison of cell-tracking algorithms

Data type and size

13 different datasets of several Gb of 2D + time and 3D + time videos of moving cells.

Available from

http://celltrackingchallenge.net/

Example tasks

- Cell segmentation in 2D and 3D
- Cell tracking in 2D and 3D
- Measurement of cell properties over time

7. Tuberculosis Chest X-ray Database

Description

A large dataset of control and tuberculosis positive chest X-ray images.

Reference

Rahman et al. - Reliable Tuberculosis Detection using Chest X-ray with Deep Learning, Segmentation and Visualization - 2020

Data type and size

700 Mb of PNG files.

Available from

https://www.kaggle.com/tawsifurrahman/tuberculosis-tb-chest-xray-dataset

Example tasks

- Segmentation of chest X-ray images
- Classification of tuberculosis positive or control

8. Annotating images to create ground truth image sets

Create a tool allowing an user to open an arbitrary 2D/3D image or a video (e.g. those from example 6) and manually (or semi-automatically) label it for further processing.

9. Malaria species

Description

A dataset containing images from cells parasitised by different species of Plasmodium.

Available from

https://www.kaggle.com/saife245/malaria-parasite-image-malaria-species

Example tasks

- Segmentation of cells and parasites
- Detect the species of the parasite
- Count the number of parasites per cell

10. Mouse behaviour

Description

A dataset of videos of different behaviours from mice.

Reference

Jhuang et al, 2010 - Automated Home-Cage Behavioral Phenotyping of Mice

Data type and size

4200 short MPG videos of mice in different behaviours (~1Gb). A full dataset is also available, containing over 10.6 hours of continously labeled video (8 day videos and 4 night videos) for the eight behaviors of interest: drink, eat, groom, hang, micromovemnt, rear, rest, walk.

Available from

https://cbmm.mit.edu/mouse-dataset

Example tasks

- Identify and track the mouse
- Classify the mouse behaviour
- Measure the mouse speed or trajectory

11. Zebrafish tracking

Description

A large dataset of videos of zebrafish swimming in a tank, viewed from two different angles.

Reference

Pedersen et al., 2020 - 3D-ZeF: A 3D Zebrafish Tracking Benchmark Dataset

Data type and size

14 Gb of high-resolution videos + ground truth annotations

Available from

https://motchallenge.net/data/3D-ZeF20/

Example tasks

- Tracking multiple zebrafish in a tank
- Calculate the speed and 3D trajectory of each zebrafish
- Create heatmaps of the zebrafish location over time