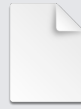


Forams Classification 2025

Semi-supervised classification of forams from volumetric images

header

ファイル · 28 KB



[Overview](#)
[Data](#)
[Code](#)
[Models](#)
[Discussion](#)
[Leaderboard](#)
[Rules](#)
[Team](#)
[Submissions](#)

Overview

Goal:

Welcome to the Forams Classification 2025 Challenge! Your task is to classify volumetric μ CT images of foraminifera tests—the calcium carbonate shells of planktonic foraminifera. The goal is to develop an efficient method for species classification with minimal time spent on annotation. To support this, we have provided a dataset of 18,426 volumes, of which only 210 are labeled. This challenge can be addressed using a semi-supervised classification model.

Why does it matter?

Planktonic foraminifera (aka forams) are tiny organisms that inhabit the sea waters. Forams produce a unique calcium carbonate shell, called a test, which can consist of multiple chambers and exhibit elaborate structures. Foraminifera tests are preserved in sea sediments for millions of years. Since certain types of foraminifera only live in specific conditions, identifying the composition of the foraminifera sediments allows us to gain insights into past environmental conditions.

With μ CT scanning, we can scan thousands of forams in a single scan. By determining the species composition, we can efficiently understand how the environment has evolved.

Start

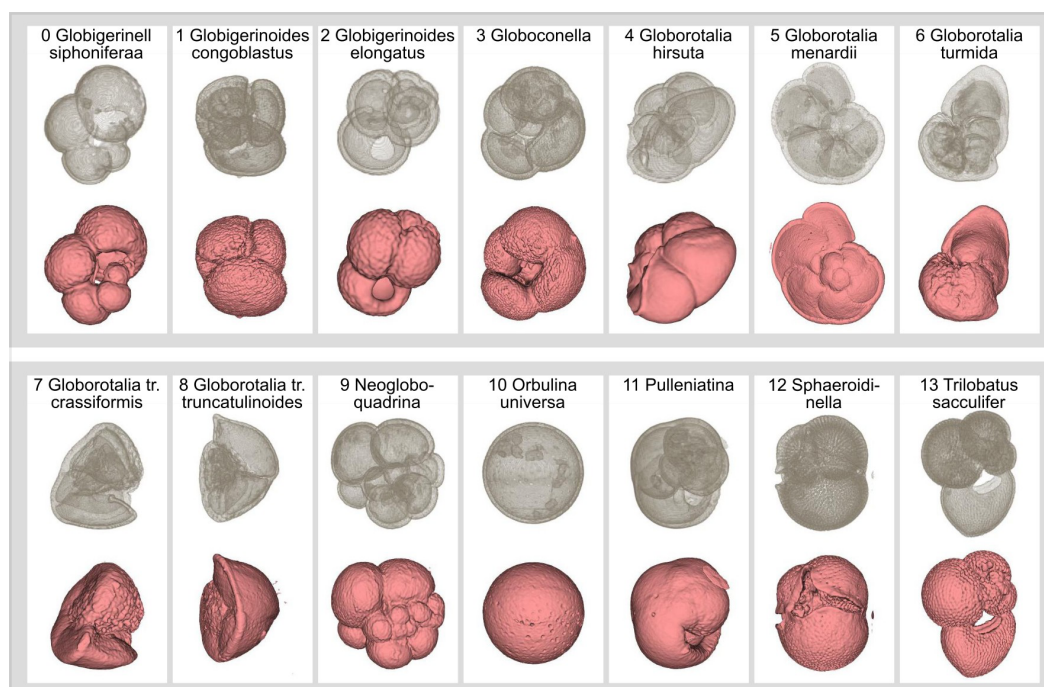
19 days ago

Close
a month to go

Description

[linkkeyboard_arrow_up](#)

The 14 types of forams are shown below. The task is to classify 18,216 volumetric scans of forams into these 14 types, plus a 15th 'unknown' class, which includes scans of foram fragments, sand, or other similar objects. To learn these types, you are provided with only 15 labeled volumes per type. Each image is 128-by-128-by-128 voxels.



Evaluation

[linkkeyboard_arrow_up](#)

See "Data" section for overview over submission format and evaluation metric.

Citation

[linkkeyboard_arrow_up](#)

We will make a paper describing the dataset, but for now, please acknowledge the QIM Center (qim.dk). Forams Classification 2025. <https://kaggle.com/competitions/forams-classification-2025>, 2025. Kaggle.

Dataset Description

Foram Dataset and Submission Guidelines

The Forams Dataset contains 18,426 volumes of 128-by-128-by-128 voxels. This is a semi-supervised classification challenge, so only a few labeled samples are given. The data contains 210 labeled volumes and 18,216 unlabeled volumes. The task is to classify the unlabeled volumes into one of the labeled species or an unknown class. The unknown class should be used for broken fragments of forams and other unclassifiable objects, which may occur in unlabeled volumes. However, no labeled examples of the unknown class are provided.

Folders:

Volumes - contains all the volumes split into two folders called 'labelled' and 'unlabelled'. Each volume file name contains the scaling factor used when scaling the files, with underscore used in the file name instead of the decimal point.

Filenames for the unlabeled volumes:

foram_[5 digit id]_sc_[1 digit integers scale factor]_[3 digits decimal scale factor].tif

Filenames for the labeled volumes:

labelled_foram_[5 digit id]_sc_[1 digit integer scale factor]_[3 digits decimal scale factor].tif

For example, a file name foram_09707_sc_0_673.tif contains:

- id: 9707
- scaling factor: 0.673. This means that the largest side of the original volume was $128/0.673 = 190$ voxels. In the original volume, voxels are 3.7 micro-meter, so the physical side length of the scaled volume is approximately 0.7 mm (190 voxels times 3.7 micron/voxel).

Visualizations - contains one RGB image for each volume, showing a volume rendering and a surface rendering of the same volume.

'Visualizations' is structured similarly as 'Volumes', it contains two folders, 'labelled' and 'unlabelled'.

Filenames for images:

foram_[5 digit id].jpg for the files in the 'unlabelled' folder and labelled_foram_[5 digit id].jpg in the 'labelled' folder.

For example, a file name labelled_foram_00017.jpg contains:

- file id: 17

- rendering of the labelled_foram_00017_sc_0_680.tif file

labelled.csv - contains two columns. The first is the id of the files, and the second is the label of each of the 210 volumes included in the labelled set. The order is the same as in the 'labelled' folders in the Volumes and Visualization folders. The labels 0 to 13 come in sorted order, such that the first 15 rows are for files with label 0, the next 15 are for label 1, etc.

unlabelled.csv - contains two columns. The first is the id of the files, and the second is empty and should be filled out for submission. The labels must be 0 for the first class, 1 for the second, etc. The label for unknown must be 14.

sample_submission.csv - example of a submission.

Evaluation

Predictions are evaluated by matching the predicted label to the true labels and calculating the F1-score. The evaluation is based on the f1_score from scikit-learn and is used to calculate the F1 score between the ground truth labels and the submission.