GEOMAPLEARN 1.2

This folder contains python codes that implement the two fold-detection methods described in "GEOMAPLEARN 1.2: Detecting structures from geological maps with machine learning: The case of geological folds" by Oakley, Loiselet, Coowar, Labbe, and Callot.

Folders

Make_Figures: This folder contains python scripts to create and save Figures 2-14.

For figures that show the synthetic models, you will need to first run the scripts in Synthetic_Maps to create the models.

For figures that show results of the unsupervised learning method, you will need to run main.py in Unsupervised_Clustering_Method for each of the cases shown in the figure. See directions under the heading for that folder.

Neural_Network_Method: This folder contains the code and the trained network weights for the U-NET based supervised learning method of fold detection.

The trained model that was used for the paper is saved in the folder "DetectAreasAndAxisModel" with its history in the file "DetectAreasAndAxisModel_history.npy". If you wish to recreate the training, first run the script "MakeRandomModels.py" in the "RandomModels" folder and then run the script "FitNeuralNetwork.py." However, since the trained model is provided, it is not necessary to recreate the training in order to run the rest of the scripts.

The folder "HyperParameterTest" contains the training history for each of the hyperparameter variations shown in Figure 4. If you wish to rerun the training of these models, change the relevant options in FitNeuralNetwork.py and run it.

Real_Maps: This folder contains the code to analyze the real-world Lavelanet and Esternay maps and produce the results shown in Figures 11, 12, 13, and 14.

Data

The data for each map are contained in the folders Esternay and Lavelanet, which contain the following files, in which <name> is replaced with either Esternay or Lavelanet:

- <name>_map_clipped_merged.shp: This is a geologic map using data from the BD Charm-50 dataset, downloaded from http://infoterre.brgm.fr/formulaire/telechargement-cartes-geologiques-departementales-150-000-bd-charm-50 (last accessed July 3, 2024). The BD Charm-50 datasets within each map area have been merged and clipped to the necessary area, and the Lavelanet map has been edited slightly to be more similar to the Souquet et al. (1984) map, but elevation and relative age information have not yet been added.

- <name>_map.shp: This is the map shapefile with the addition of elevation and relative age information, which is used by the unsupervised clustering-based method.
- <name>_raster.tif: This is a raster version of <name>_map.shp, which was created using the "Feature to Raster" option in ArcGIS Pro.
- <name>_DEM.tif: This is a digital elevation model (DEM) covering the area of each map. These DEMs are portions of the BD ALTI 25M dataset, downloaded from https://geoservices.ign.fr/bdalti (last accessed July 4, 2024), which have been merged and converted to tif format using ArcGIS Pro's "Mosaic To New Raster" function.
- <name>.npz: This file contains numpy arrays storing the geology and elevation rasters to be used by the U-NET based supervised learning method.
- UnitRelAges_<name>.csv: This is a table showing how the "NOTATION" field in the map shapefiles is converted into relative ages.
- PrepareMap_<name>.py: This is a python script to prepare the data needed for the unsupervised clustering-based method, by adding the elevation and relative age information to the map shapefile. It takes <name>_map_clipped_merged.shp, <name>_DEM.tif, and UnitRelAges_<name>.csv as input and produces <name>_map.shp.
- PrepareUNET_<name>.py: This is a python script to prepare the data needed for the UNET-based supervised learning method. It takes <name>_raster.tif and <name>_DEM.tif as inputs and produces <name>.npz.

Data Licensing Information

The BD Charm-50 geologic map data and the The BD ALTI 25M elevation data are covered by the etalab 2.0 license. A copy of this license has been included in the Real_Maps folder.

Synthetic_Maps: This folder contains the code to create the synthetic maps discussed in the paper as well as the resulting shapefile and raster maps.

To create these models run the scripts "MakeSyntheticModels_Vector.py" and "MakeSyntheticModels_Raster.py".

Unsupervised_Clustering_Method: This folder contains the code for the unsupervised learning method of fold detection that uses hdbscan clustering.

To use this, run the script "main.py". Change options in the top section as desired.

Under "CHOOSE your case:" in main.py, there are a series of blocks of commented text. Each of these contains the settings for one of the models shown in the paper. Uncomment the one you want to run and comment out all the rest. Do this alternately for each of the models in order to run all the analyses. This will need to be done before running the scripts to plot figures showing these results.

Python Packages

Due to package incompatibilities, two different python environments were used: one for scripts that require tensorflow, but not geopandas, and one for those that require geopandas, but not tensorflow. The lists below give package versions in the two environments, with which all the scripts ran successfully. They may or may not run with other versions, but if you encounter errors, we suggest trying with these versions.

In both environments, the spyder IDE was used to run the scripts.

Environment 1 (Without Tensorflow, with Geopandas)

- geopandas 0.12.2
- hdbscan 0.8.33
- matplotlib 3.7.4
- numba 0.58.1
- numpy 1.26.2
- pandas 2.1.4
- pyogrio 0.7.2
- python 3.9.18
- regex 2023.10.3
- scikit-learn 1.3.2
- scipy 1.11.4
- shapely 2.0.2
- spyder 5.4.3

Environment 2 (With Tensorflow, without Geopandas):

- cudatoolkit 11.2.2
- cudnn 8.1.0.77
- imageio 2.31.4
- keras 2.10.0
- matplotlib 3.8.0
- numba 0.59.0
- python 3.9.18
- spyder 5.4.3
- tensorflow 2.10.1

Change Log

Version 1.0: Original version:

Version 1.1: This version contains data and code for the real-world Lavelanet and Esternay maps.

Version 1.2: This version contains scripts and files necessary for making the final versions of figures, and it removes the old scripts for making figures.