Editor: Nir Shalmon

User's Guide

Functionality

This application is aimed to allow the users to:

1. Predict the location of alpha-helixes given a cryo-EM density maps of proteins, using a pre trained model.
2. Train a model to predict the location of alpha helixes given a labeled database of cryo-EM density maps.

Requirements

The application requires the installation of pytorch, numpy, mrcfile, and should be run on a computer with a CUDA GPU.

User Interface

The user interface is a command line interface.

The interface is as follows: python helix_predictor_UI *command*

Supporting the following commands:

1. --help: prints the interface of the supported commands.

2. -train output_net_path dataset_path [evaluation_results_output]
   output_net_path: The path in which the trained network will be saved.
   dataset_path: A path to a directory that contains both the protein density maps and their labeled helix maps. The expected format is specified under the "In-depth – Training" section.
   evaluation_results_output: An optional argument, a path for the results of the evaluation.

3. -predict net_path protein_mrc_path output_mrc_path [binary_output_mrc_path]
   protein_mrc_path: a path of the protein density map.
   output_mrc_path: a path to where the prediction will be saved at.

binary_output_mrc_path: an optional parameter. If it is supplied, a binary version of the prediction results (where each voxel will be marked as a helix or not a helix) will be saved as an mrc file at that path.

In depth

Training
The purpose of the training stage is to train a Convolutional Neural Network to predict the location of alpha helixes given a density map of a protein.

The training process requires a labeled database that consists of pairs of a protein density map and its labeled helix mask. The dataset is a directory, the pairs' format should be "protein.mrc" for the protein density map and "protein_helix.mrc" for its corresponding helix map.
A protein density map and its helix mask should be of the same dimensions.

To achieve the best possible performance, the model should be trained on similar proteins to the protein whose helices the user wants to predict.

Due to randomness in the training process, it is recommended that the network be trained multiple times. Some training results will be better than the rest. Use the evaluation reports to select the best network for your use.

Prediction
The application predicts the location of alpha helices given a protein density map and a trained neural network.
The input is protein density map in mrc file format.
The output is an mrc file of the same dimensions as the input protein. Each voxel will contain a real number – the closer that number is to 1, the more likely it is that that voxel is part of a helix in the input protein.

Pre-trained models
The application provides 6 pre-trained models:

1. Super family 4Å
2. Super family 6Å
3. All alpha class 4Å
4. All alpha class 6Å
5. Alpha and Beta (a/b) 4Å
6. Alpha and Beta (a/b) 6Å

All models were trained with grid size of 1Å