Eclip

Eclip is a package that trains a CNN model to recognise phased (good) and unphased (bad) electron density maps. This package reads in the electron density maps (in .map format) and slices them into two dimensional images. These images are then used to train a modle that will be able to give a label to each image, indicating how likely it is that the protein is phased. These labels are then averaged to give an overall score for the map.

The package is split into two sections. The first is used to train the model, the second is used to implement this model on new data. It also comes with some pretrained models for use with certain data types.

The github repositary for this package can be found here.

Getting Started

These instructions will get Eclip running on your local machine.

How to use

To use Eclip, it is simplest to run first the training, and then implementing programs. However, each individual section can also be run independantly. This allows for more felxibility in the arguments used.

The training program is run by calling RunTrain from the command line. The implementing program is similarly run by calling RunPred from the command line.

You can learn more about how to do this, and the arguments required, here:

Contents of Eclip

Details on individual functions and their arguments can be found here, as well as details on the pretrained models available:

Pretrained Models

There are two pretrained models in this package. Their performance and data type is found in the table below.

Model	Data	Maps	Images	Training	Predicting	Image Accuracy	Map Accuracy
1	Feature Enhanced	786	35370	5h49min	34sec	97%	93%
2	Heavy Atoms						

The total number of maps and images used in the creation of these models (the sum of those used for training and predicting) are in the columns 'Maps' and 'Images' respectively. The runtimes are under 'Training' and 'Predicting'. The training runtime was calculated for the program running on a 4 Nvidia Tesla GPU cluster, and the Predicting runtime as calculated for a single MGA-G200e GPU.

Details on the data types and how they were produced can be found here:

To use the pretrained models to predict new scores, make sure the input data has been preprocessed in the same way as the data used to train the model (see above). Then call RunPred with the model location (see example below). The $CHOICE_OF_DATA_TYPES should be replaced with either 'FeatureEnhanced' or 'HeavyAtoms'.

```
RunPred --model=/$YOUR_LOCATION_FOR_ECLIP/eclip/eclip/pretrainedmodels/$CHOICE_OF_DATA_TYPE/
```

Utils Subpackage

The utils subpackage is a package containing all the functions that could be needed more than once.

Its contents can be found here:

Modules

eclip.RunTrain module

eclip.RunPred module

eclip.ConvMAP module

eclip.EP_success module

eclip.learn module

eclip.predictest module

eclip.EP_add_new module

eclip.predic module

Acknowledgements

- Supervisor: Melaine Vollmar
- co-supervisor: James Parkhurst
- Principle Investigator: Gwyndaf Evans