EHT2018-IDFE-pipeline

Release 0.1.0

Iniyan Natarajan

Feb 14, 2023

CONTENTS:

CHAPTER

ONE

INTRODUCTION

This pipeline is used for performing image domain analysis of EHT 2018 data, focussing on the M87 science case.

The first part of the pipeline performs image domain feature extraction (IDFE) using the Ring Extractor module (REx, Chael 2019) from eht-imaging and Variational Image Domain Analysis (VIDA, Tiede et al. 2022) from VIDA.jl.

The second part of the pipeline applies the metronization (Christian et al. 2022) algorithm to perform Topographical Data Analysis for identifying ring-like structures in the input images, and consolidates the results of all the previous steps.

This documentation explains the steps necessary for installing and running the pipeline.

CHAPTER

TWO

DATA DESCRIPTION

In the 2018 observing campaign of the EHT, M87 was observed over multiple days and frequency bands. Two calibration pipelines, HOPS and CASA, were used to calibrate these data. The output of each calibration pipeline was processed by 5 imaging tools. To validate the performance of the imaging tools, 15 synthetic datasets (of which 5 are loosely classified as "validation" datasets), from 15 different source models, were generated for each day+band combination. More details on the outputs and conventions followed by the imaging tools are given below. Customarily, the results of the various imaging tools are copied to eht-cloud.

Days : 3644 (April 21) and 3647 (April 25).

Bands: b1, b2, b3, b4; some imaging tools also process combined band data i.e., b1+2, b3+4, b1+2+3+4

2.1 Comrade

Comrade/netcal - hosts all the synthetic and M87 datasets.

All the synthetic datasets are named **<model>_<day>_<band>**. All the M87 datasets are named **<casa/hops>_<day>_<band>**.

Comrade does not process band-combined datasets.

2.2 SMILI

smili - hosts all the synthetic and M87 datasets.

All the synthetic datasets are named **<model>_<day>_<band>**. All the M87 datasets are named **<casa/hops>_<day>_<band>**.

SMILI also includes band-combined datasets labelled b1+2, b3+4, and b1+2+3+4.

2.3 difmap

difmap - hosts all the synthetic and M87 datasets. Also hosts an entirely parallel set of datasets suffixed **_geofit**.

All the synthetic datasets are named **<model>_<day>_<band>**. All the M87 datasets are named **<casa/hops>_<day>_<band>**.

difmap does not process band-combined datasets.

2.4 THEMIS

THEMIS/synthetic_unblurred/netcal - hosts all synthetic datasets; on Cannon cluster, the path is slightly changed to **THEMIS/netcal**. **THEMIS/M87real** - hosts all M87 datasets, subdivided further under **<casa/hops>_raster+LSG_unblurred>**.

All the synthetic datasets are named **<model>_<day>_<band>**. All the M87 datasets are named **<casa/hops>_<day>_<band>**.

THEMIS also includes band-combined datasets labelled b1b2, b3b4, and b1b2b3b4, except for **hops_3644** datasets, which are labelled b12, b34, and b1234.

2.5 ehtim

CHAPTER

THREE

INSTALLATION

We recommend installing the pipeline inside a virtual environment running Python 3.10 or above. On a new Debian-based system, one may need to install, among other things, *python3-venv* via *apt-get*. As always, upgrade the following packages in a new virtual environment:

```
pip install --upgrade pip setuptools wheel
```

3.1 Prerequisites

If installing on a Debian-based system, ensure that *cmake* and *libboost-python-dev* are installed.

Clone the dev branch of *eht-imaging* and install using *pip*.

```
git clone -b dev https://github.com/achael/eht-imaging.git
cd eht-imaging
pip install .
```

Ensure that numpy version 1.23.x is installed (uninstall any version that is higher) for eht-imaging to work properly. For example:

```
conda install -c conda-forge numpy=1.23.5=py310h53a5b5f_0
```

Install other necessary packages:

```
conda install -c conda-forge seaborn joypy termcolor
```

Clone *ehtplot* and install using *pip*.

```
git clone https://github.com/liamedeiros/ehtplot.git
cd ehtplot
pip install .
```

Install *dionysus* from conda-forge **before** installing metronization.

```
conda install -c conda-forge dionysus
```

Install *metronization* from GitHub using *pip*.

```
git clone https://github.com/focisrc/metronization.git
cd metronization
pip install -e .
```

3.2 Installing the pipeline

Clone the IDFE pipeline and install using *pip*.

```
git clone https://github.com/iniyannatarajan/eht2018-idfe-pipeline.git
cd eht2018-idfe-pipeline
pip install -e .
```

CHAPTER

FOUR

RUNNING

The driver scripts *eht2018_idfe_driver_realdata.py* and *eht2018_idfe_driver_syndata.py* set values for variables necessary to locate topsets/posteror samples and run REx and VIDA on them. The output of the driver scripts are a set of REx and VIDA output files in HDF5 and CSV formats respectively.

These scripts use no comand-line arguments and all the values are set within the script to keep a record of what was run. The user must check these values at the beginning of every run.

Copy the required IDFE driver script to the output directory, modify the input variables, and run with python 3.6 or above. To see the available command-line arguments run with the -h option.

```
python <idfe-driver-script> -h
```

CHAPTER

FIVE

CONTRIBUTORS

- Iniyan Natarajan
- Yutaro Kofuji
- Paul Tiede

CHAPTER

SIX

INDICES AND TABLES

- genindex
- modindex
- search