



*cv*s to *numpy*/*hdf5* reading in **DarkFlow**

- DarkFlow is an effort by a few people to investigate Flow techniques on the DarkMachines unsupervised dataset
 - project started with Pratik Jawahar receiving a IRIS-HEP fellowship for this Summer
 - Pratik is working on developing a VAE-based anomaly detection solution and to power it with (normalizing, autoregressive, etc...) flows. More to come next week
 - We would like to also investigate different VAE architectures and compare them
 - We would contribute with one model (our best?) or document a few of them
 - We look forward to others joining our effort

- In this context, we put together a package that comes with a csv->hdf5 conversion tool, which gives access to data as numpy arrays

mpp-hep / DarkFlow

Unwatch 1 Star 0 Fork 0

Code Issues 0 Pull requests 0 Actions Projects 0 Wiki Security 0 Insights Settings

No description, website, or topics provided. Edit

Manage topics

5 commits 1 branch 0 packages 0 releases 1 contributor

Branch: master New pull request Create new file Upload files Find file Clone or download

Commit	Message	Time
✖ pierinim	added script to merge h5 files	Latest commit bea9000 21 days ago
data_preparation	added script to merge h5 files	21 days ago
models	added AE implementation	24 days ago
training	added AE implementation	24 days ago
README.md	Update README.md	27 days ago

README.md

DarkFlow

<https://github.com/mpp-hep/DarkFlow>

- store 4 momenta of highest-pT objects
 - first 10 jets
 - first 2 muons of each charge (total 4)
 - first 2 electrons of each charge (total 4)
 - first 4 photons
 - MET
- events meta data (process ID etc)

All of this could be changed, as long as we are OK with 0-padded fixed-representation format

```
[>>> import h5py
[>>> f = h5py.File("/eos/project/d/dshep/DARKMACHINES/sm_10fb.h5", "r")
[>>> f.keys()
[u'Bjets', u'EleMinus', u'ElePlus', u'EventContent', u'EventFeatures', u'Gamma', u'Jets', u'MuMinus', u'MuPlus', u'ParticleContent', u'ProcessID']
[>>> a = f.get("EventContent")
[>>> print(a[:])
['evtID' 'weight' 'METx' 'METy']
[>>> a = f.get("ParticleContent")
[>>> print(a[:])
['pX' 'pY' 'Eta' 'M']
[>>>
```

- No selection applied yet
- We would implement any and produce a reduced dataset if needed