

SFT Group Meeting

Status Update #1

Konrad Helms

25th May 2023

15 May - 25 May

- got familiar with Khrystyna's code
- wrote data loading function
- implemented the benchmark algorithm
- AutoML

AutoML

-  GitHub
 - developed by Uni Freiburg, Uni Hannover and Uni Tübingen
1. auto-sklearn
 - automated machine learning toolkit
 - drop-in replacement for a scikit-learn estimator
 2. auto-PyTorch
 - Automatic architecture search
 - hyperparameter optimisation for PyTorch
- all ML techniques were trained/fitted on the same amounts of data as Khrystyna's CNN
 - today (after lunch): discovered a horrible bug ⚡

AutoML - Sklearn Results

No Normalisation of Data

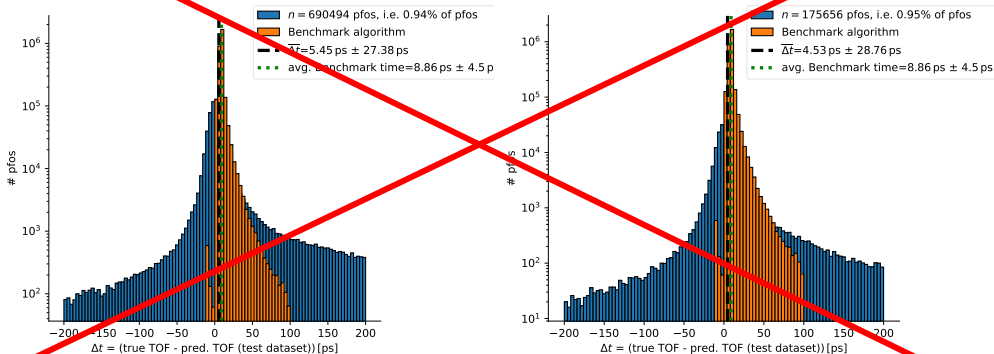


Figure: Auto-sklearn: ensemble of BDTs, using 3 (left) or 10 (right) hits

AutoML - Sklearn Results Including Normalisation of Data

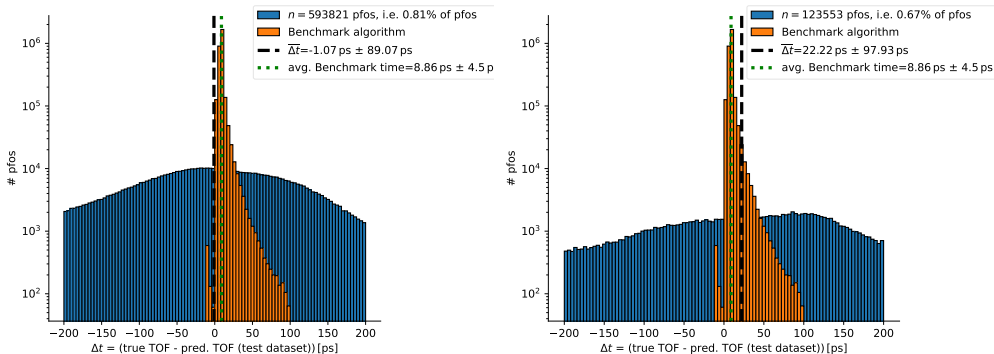
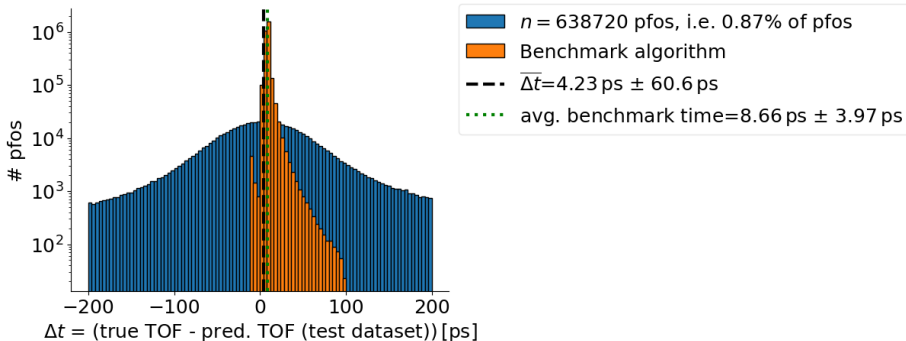


Figure: Auto-sklearn: ensemble of BDTs, using 3 (left) or 10 (right) hits

AutoML - PyTorch Results



Multilayer perceptron (MLP), using 10 hits

Outlook

- benchmark algorithm correct?!
- similar performance when using 3 and 10 hits → more hits not a solution, clever data pre-processing?
- no automl model export: manually implement the BDTs and AutoPyTorch models → optimise hyperparameters
- analyse the models → what do they pay attention to (... physics)
- $\vdots$
- *far future*: look at PointNet(++), point cloud data, ...