Machine Learning for Calorimetry Physics: Classification and Anomaly Detection

Bjartur í Túni Mortensen, Jens Peter Andersen and Julie Bojesen Koefoed

Copenhagen University

1. Introduction

FoCal-H is a hadronic calorimeter currently under development and is intended as a part of an upgrade to the ALICE experiment at the LHC at CERN.

As part of the development, regular data taking events called testbeams are performed where you fire single particles at the detector. Particle type, energy and angle of incident are known.

As part of an Applied Machine Learning course the goal was to explore machine learning approaches on these data, for instance classification and anomaly detection.

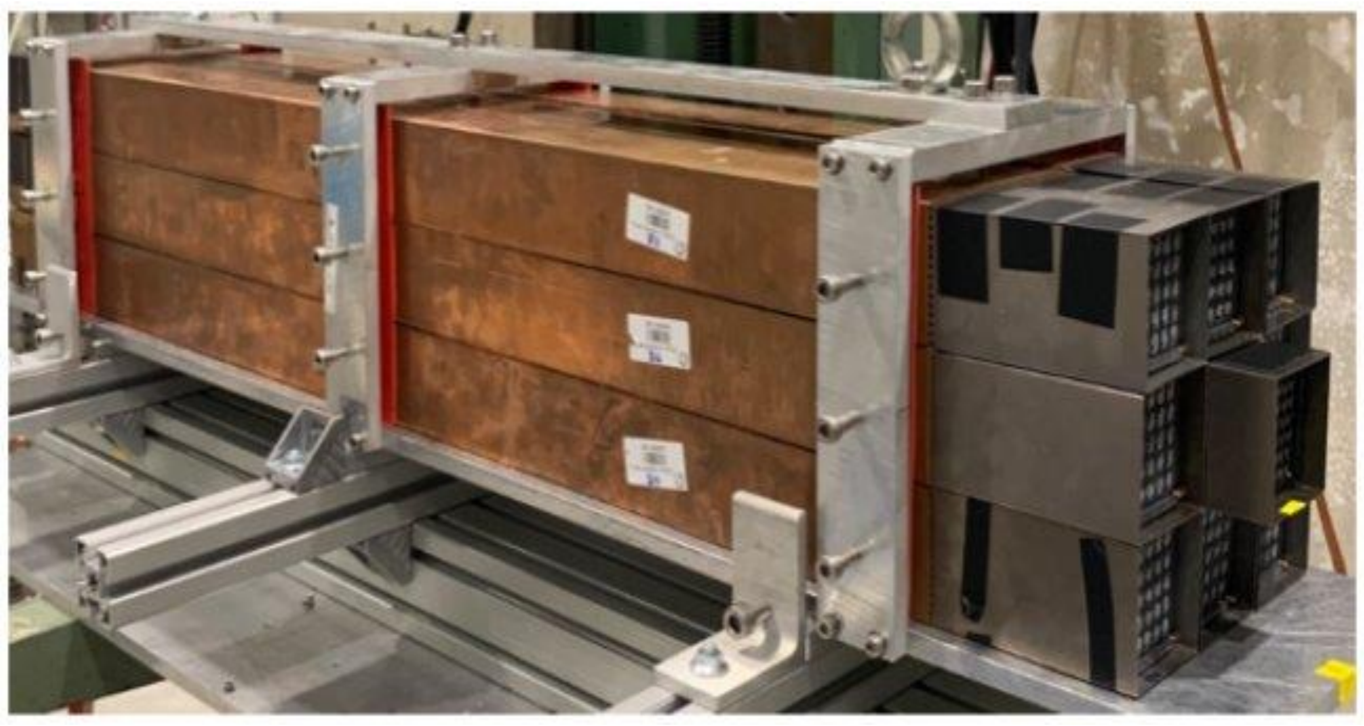

Fig. 1. FoCal-H Prototype 2. Image from a testbeam event September 2023.

2. Methods

Classification of particle type

Feature Engineering

- Data consists of 441 pixel image (upscaled from 7x7 + 8x5x5 sensors).
  - See fig. 2
- We calculate **Hu Moments** for each event.
  - **Hu Moments** are invariant under translation, rotation, scaling and reflection (except for one of the moments).
- Use the 7 **Hu Moments** for each event as features instead of original data.

Classification and Regression

- Train models like Gradient Boosting (**LightGBM**) to do classification and regression on labels (particle type, energy and angle).

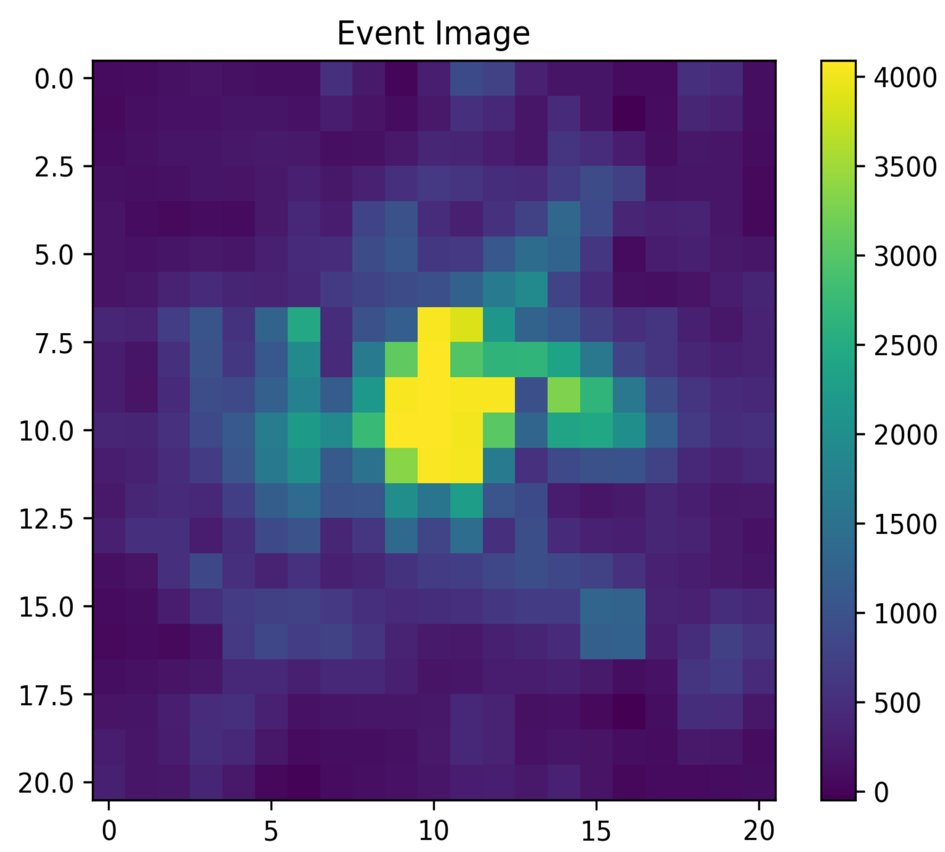

Fig. 2. Regular event. 21x21 pixel heatmap.

Anomaly Detection

Data Augmentation

- Create new events by picking out good events and translating, rotating and mirroring them.
- From 15000 events, filtered out 158 bad ones, and transformed good ones into a new 15 million **augmented** dataset.

Anomaly Detection

- Train a Convolution Neural Network to do anomaly detection on augmented dataset of good events.

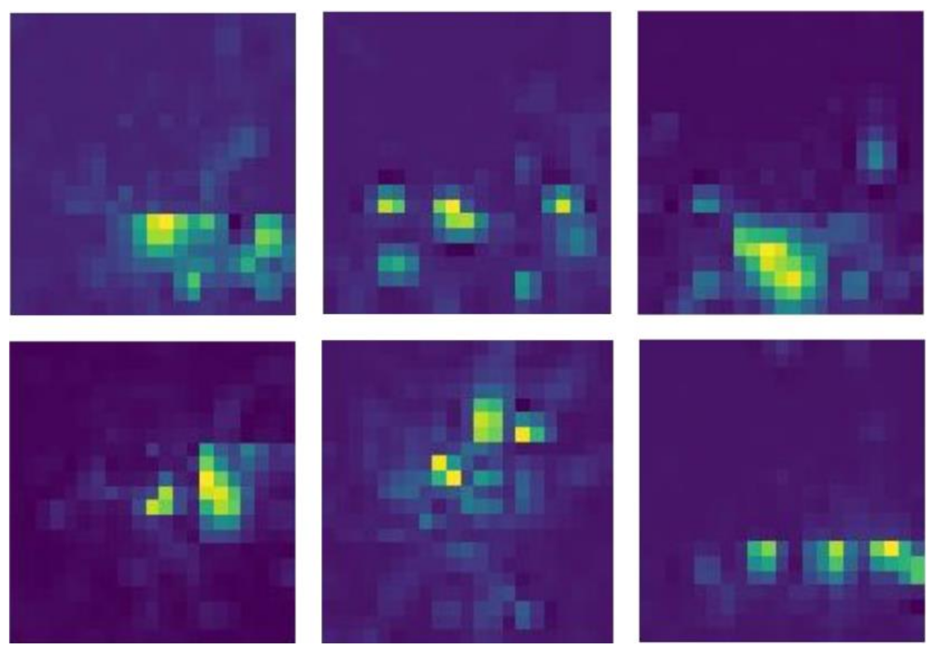

Fig. 3. Images found with anomaly detection with mixed results. Third image looks like a good event but 1,2 and 6 could be anomalies.

3. Results and Discussion

Classification and Regression

- Particle type classification with an accuracy of **97.8%**.
  - Caveat: We only trained on real data and suspect there could be bias in the data. The solution would be to train on simulated data.
- Regression on energy and angle were less successful.

Anomaly Detection

- Looked like a promising approach (fig. 3), but still finds good events as anomalies.
- Only a preliminary approach, successive iterations of this method could converge on better result.

4. Summary

- **Feature Engineering** by computing your own features from raw data is a good strategy to keep in mind.
- Classification worked well for particle type, but simulation data must be included to understand if there is bias.
- Given that good and bad events are part of the same dataset, some manual work seems necessary to do anomaly detection. **Data Augmentation** may be a powerful tool to aid in this.
- No quantitative evaluation of the anomaly detection but looked promising to us.

5. Acknowledgements

Special thanks to

- FoCal-H group and in particular Ian Pascal Møller.
- Applied Machine Learning course.

Bjartur í Túni Mortensen

Email: tcm955@alumni.ku.dk