APPLICATION OF MACHINE LEARNING TOWARDS PARTICLE COUNTING AND IDENTIFICATION

University of Essex

Samuel Engel, Rahul Singh, Plamen Boutachkov
University of Essex, Colchester, England, GSI Helmholtzzentrum für Schwerionenforschung, Darmstadt, Germany

GSI

An exploration into the application of three machine learning (ML) approaches to identify and separate events in the detectors used for particle counting at the GSI Helmholtz Centre for Heavy Ion Research. A convolutional neural network (CNN), a shape-based template matching algorithm (STMF) and Peak Property-based Counting Algorithm (PPCA) were developed to accurately count the number of particles without domain-specific knowledge required to run the currently used algorithm. The three domain-agnostic ML algorithms are based on data from scintillation counters commonly used in beam instrumentation and represent proof-of-work for an automated particle counting system. The algorithms were trained on a labelled set of over 150 000 experimental particle data.
The results of the three classification approaches were compared to find a solution that best mitigates the effects of particle pile-ups. The two best-achieving algorithms were the CNN and PPCA, achieving an accuracy of over 99%.

Data collection and transformation

- High-resolution experiment data → **150 677 peaks**
- Low-resolution experiment data → artificial **downsampling**
- Bootstrapped high-resolution experiment data
- Synthetic laser data → 416 **validation peaks**

Sample rate: $2.5 \cdot 10^9$ samples per second
Extraction rate: **260 000** particles per second
Distance between each data point: **0.4 nanoseconds**

Peak Property-based Counting Algorithm

- Peaks have mathematical properties that define them
- Peak width, height, prominence and peak distances are used as weights
- The optimal combination of weights most representative of all the peaks are found iteratively
- The algorithm maintains accuracy at a 10% resolution equivalent to $2.5 \cdot 10^8$ samples per second

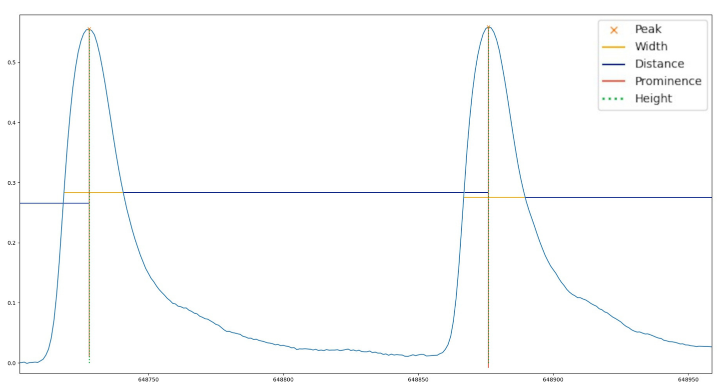

Figure 1: A visualization of the mathematical peak properties captured

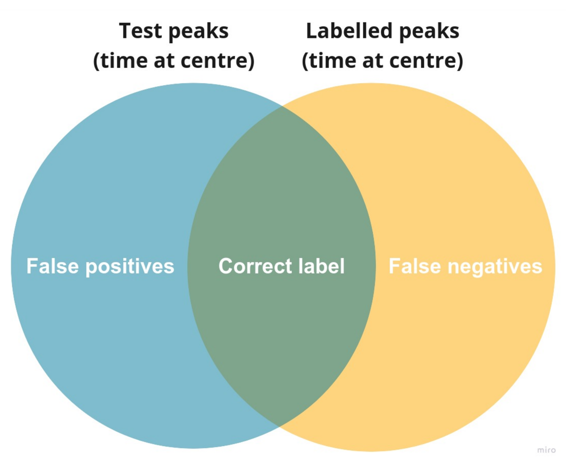

Figure 2: A diagram of the loss for an iteration as the sum of false negatives and positives

Shape-based Template Matching Framework

- A template is constructed for each peak group using an averaging scheme
- Time series segments are compared using dynamic time warping (DTW)
- The algorithm uses cubic-spline dynamic time warping averaging function for template creation [1]

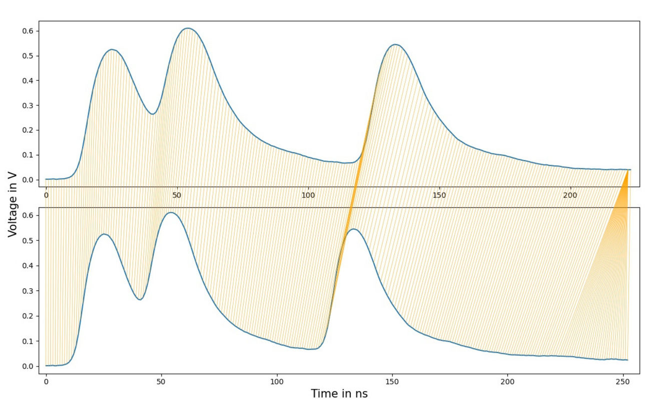

Figure 3: A DTW mapping between two 3-peak time series segments.

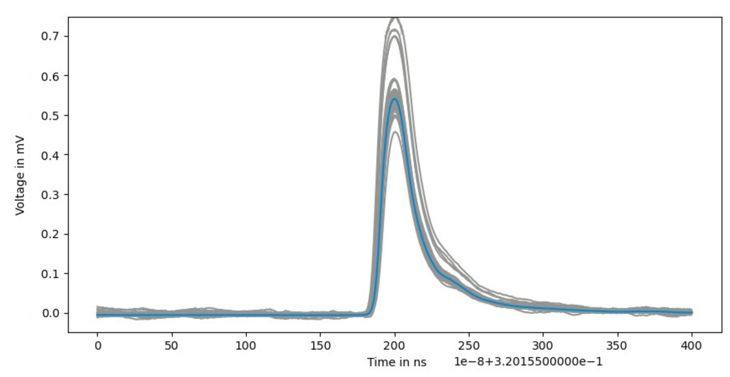

Figure 4: A template (blue) representative of training peaks (grey).

Aim

Develop a machine learning algorithm that **accurately counts** the number of particles, **regardless of particle pile-ups**.

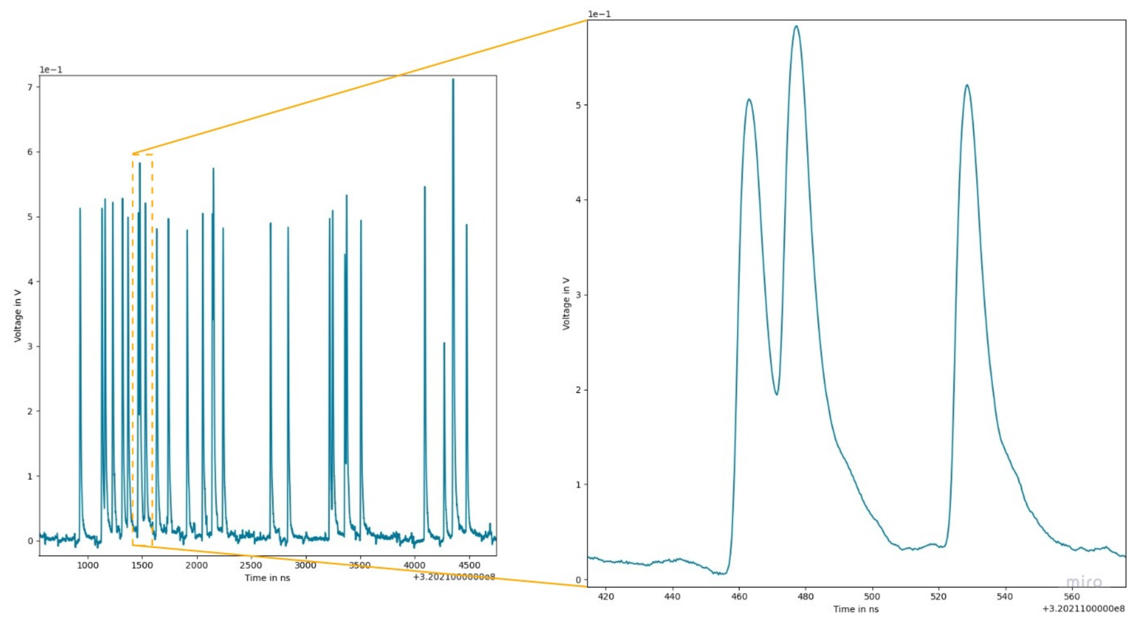

Figure 5: A segment of a time series with a close-up of a peak pileup

Convolutional neural network

- One-dimensional convolutional neural network using discriminative supervised learning
- Intrinsic features are learned from raw data rather than engineered features
- Even-sized window input allows for real-time processing
- A TensorFlow model allows for an implementation to a microcontroller

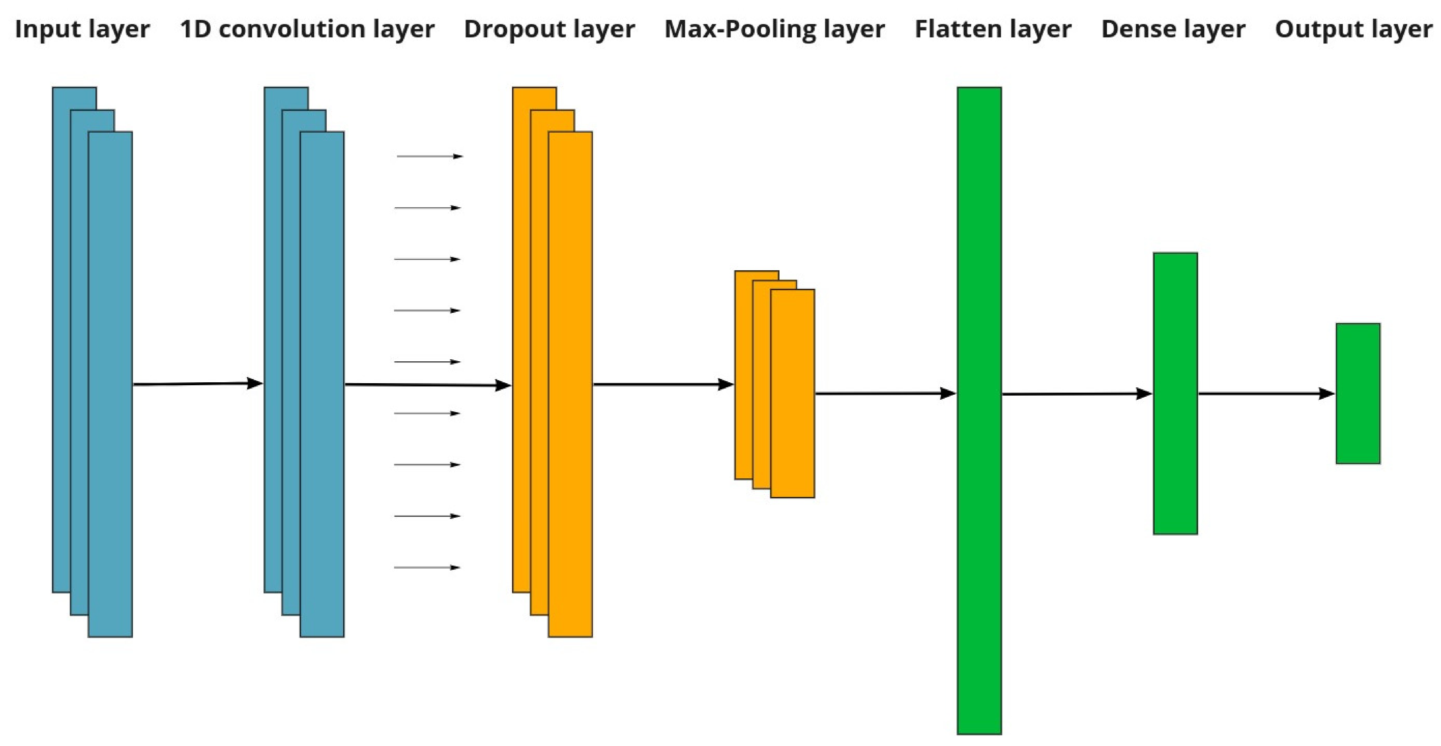

Figure 6: Diagram illustrating the capture mathematical peak properties.

Results

- CNN has fewer constraints on adaptability
- PPCA offers more transparency
- STMF fails with increased complexity

	Accuracy (%)	Storage requirements of the model
PPCA	99.97	<1kB
STMF	96.71	<20kB
CNN	99.84	<1MB

Table 1: Accuracy and space performance of the 3 algorithms

References

[1] V. Niennattrakul, D. Srisai, and C. A. Ratanamahatana, “Shape-based template matching for time series data,” Knowledge-Based Systems, vol. 26, pp. 1–8, Feb. 1, 2012, ISSN: 0950-7051. DOI: 10.1016/j.knosys.2011.04.015.