CiML 2018 @NIPS : TrackML, a Particle Physics Tracking Machine Learning Challenge

Jean-Roch Vlimant (Caltech), Vincenzo Innocente, Andreas Salzburger (CERN), Isabelle Guyon (ChaLearn), Sabrina Amrouche, Tobias Golling, Moritz Kiehn (Geneva University),David Rousseau*, Yetkin Yilmaz (LAL-Orsay), Paolo Calafiura, Steven Farrell, Heather Gray (LBNL), Vladimir Vava Gligorov (LPNHE-Paris), Cécile Germain, Victor Estrade(LRI-Orsay),Edward Moyse (University of Massachussets), Mikhail Hushchyn, Andrey Ustyuzhanin (Yandex, HSE)

October 10, 2018

The Large Hadron Collider (LHC) located at the European Center for Nuclear Research (CERN, Switzerland) is a unique particle accelerator complex providing proton beams at unprecedented energies. It allowed the Higgs boson discovery in 2012 (acknowledged by the 2013 Nobel prize in physics). Experiments at the Lhc are collecting events (images, see Figure 1) of the proton collisions of increasing complexity at an increasing rate. The CPU time to reconstruct the trajectories (helices) from the measurement (3D points) is increasing year after year faster than the hardware resources. Exploration of completely new approaches to pattern recognition is needed.

To reach out to Computer Science specialists, a Tracking Machine Learning challenge (TrackML[1]) has been run in 2018 by a team of particle physicists tracking experts and Computer Scientists, building on the experience of the successful Higgs Machine Learning challenge[2] in 2014 [1]. A dataset consisting of an accurate simulation[3] of a LHC experiment has been created, listing for each event the measured 3D points, and the list of 3D points associated to each true track. The orders of magnitude are : ten thousand events, 100 million tracks, 1 billion points, 100 GB, 10^500 combinations per event. No limit on the training resources has been imposed. The participants to the challenge should find the tracks in an additional test dataset, meaning building the list of 3D points belonging to each track.

The challenge has been run in two phases:

1. During the first "Accuracy" phase (which has run August to September on Kaggle[4]), a metric reflecting the accuracy of the model at finding the proper point association

*Corresponding author : rousseau@lal.in2p3.fr

[1]https://sites.google.com/site/trackmlparticle/

[2]https://www.kaggle.com/c/higgs-boson

[3]https://cern.ch/acts

[4]https://www.kaggle.com/c/trackml-particle-identification

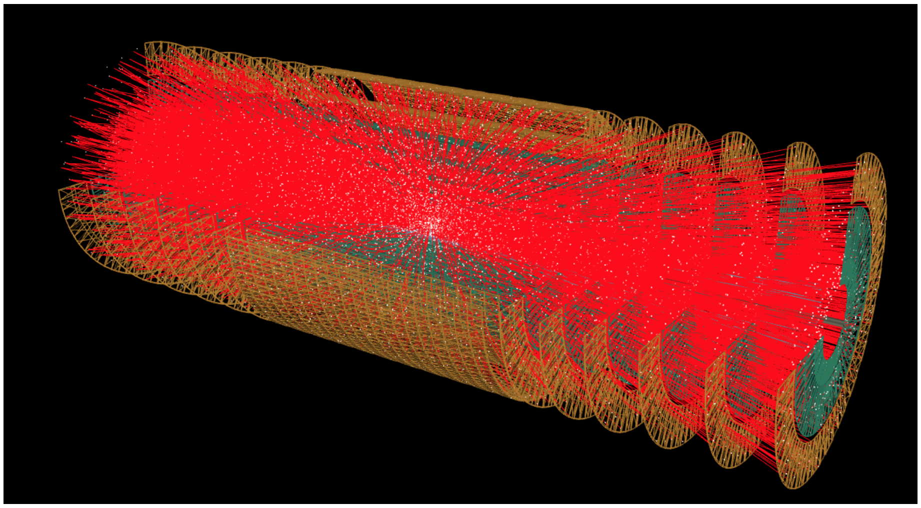

Figure 1: Etched-out collision image : measurement points are white, trajectories are red. The tracking volume is a cylinder of 1m radius and 5m length. The measurement points have precision of 10 μm to 3 mm. The goal of TrackML is to cluster points into trajectories.

that matters to most physics analysis has allowed to elect the programs. The metric is based on the overall fraction of points associated to a good track.

2. The second "Throughput" phase (running at the time of writing until 5 November on Codalab[5]) is focussed on optimising the inference speed on two CPU core (a number of software from top performers of the first phase have been released as possible starting point). The training speed remains unconstrained. The ranking score is combining the accuracy score with the speed. This is an official NIPS 2018 competition.

The Accuracy phase has had 656 participants. It has revealed a variety of approaches, from deep learning to more traditional combinatorial approaches including mixed approaches. The Throughput phase is on-going at the time of writing.

There has been very limited use of Machine Learning algorithms in particle physics tracking so far, however there is a strong potential for application of machine Learning techniques. The problem can be related to representation learning [2] as in [3], to combinatorial optimization as in [4], to clustering, and even to time series prediction [5]. An essential challenge is to efficiently exploit the a priori knowledge about geometrical constraints [6], with recent work in the generative, e.g. [7] and [8] as well as discriminative approaches e.g. [9] for combining structural priors and nonlinear state estimation with neural network.

[5]https://competitions.codalab.org/competitions/20112

References

[1] C. Adam-Bourdarios, G. Cowan, C. Germain, I. Guyon, B. Kégl, and D. Rousseau, "The Higgs boson machine learning challenge." in *HEPML@ NIPS*, 2014, pp. 19–55. [Online]. Available: http://www.jmlr.org/proceedings/papers/v42/cowa14.pdf

[2] Y. Bengio, A. Courville, and P. Vincent, "Representation learning: A review and new perspectives," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 8, pp. 1798–1828, 2013.

[3] N. Wang and D.-Y. Yeung, "Learning a deep compact image representation for visual tracking," in *26th NIPS*, 2013, pp. 809–817.

[4] I. Bello, H. Pham, Q. V. Le, M. Norouzi, and S. Bengio, "Neural combinatorial optimization with reinforcement learning," 2016. [Online]. Available: https://arxiv.org/abs/1611.09940

[5] P. Ondruska and I. Posner, "Deep tracking: Seeing beyond seeing using recurrent neural networks," in *13th AAAI Conference on Artificial Intelligence*, 2016, pp. 3361–3368.

[6] R. Urtasun, "Incorporating structure in deep learning," in *4th ICLR*, 2016, keynote. [Online]. Available: http://videolectures.net/iclr2016_urtasun_incoporating_structure/

[7] R. G. Krishnan, U. Shalit, and D. Sontag, "Deep kalman filters," *CoRR*, 2015. [Online]. Available: http://arxiv.org/abs/1511.05121

[8] M. Johnson *et al.*, "Composing graphical models with neural networks for structured representations and fast inference," in *29th NIPS*, 2016.

[9] T. Haarnoja *et al.*, "Backprop kf: Learning discriminative deterministic state estimators," in *29th NIPS*, 2016, pp. 4376–4384.