The TrackML Particle Tracking Challenge

David Rousseau, Sabrina Amrouche, Paolo Calafiura, Victor Estrade, Steven Farrell, Cecile Germain, Vladimir Gligorov, Tobias Golling, Heather Gray, Isabelle Guyon, et al.

The TrackML Particle Tracking Challenge

David Rousseau* , Sabrina Amrouche†, Paolo Calafiura‡, Victor Estrade §, Steven Farrell‡, Cécile Germain§, Vladimir Vava Gligorov¶, Tobias Golling†, Heather Gray‡, Isabelle Guyon§, Mikhail Hushchyn‖, Vincenzo Innocente**, Moritz Kiehn†, Andreas Salzburger**, Andrey Ustyuzhanin††, Jean-Roch Vlimant‡‡ and Yetkin Yilmaz*

* LAL, Orsay
† University of Geneva
‡ Lawrence Berkeley National Laboratory
§ UPSud, INRIA, University Paris-Saclay
¶ LPNHE-Paris
‖ Yandex, MIPT
** CERN
†† Yandex, HSE
‡‡ CalTech

***Abstract*—Can Machine Learning assist High Energy Physics in discovering and characterizing new particles? With event rates already reaching hundred of millions of collisions per second, physicists must sift through ten of PetaBytes of data per year. Ever better software is needed for processing and filtering the most promising events. This will allow the Large Hadron Collider (LHC) to fulfill its rich physics programme, understanding the private life of the Higgs boson, searching for the elusive dark matter, or elucidating the dominance of matter over anti-matter in the observable Universe. To mobilize the scientific community around this problem, we are organizing the TrackML challenge, which objective is to use machine learning to quickly reconstruct particle tracks from points left in the silicon detectors. Therefore this challenge offers an interesting new puzzle to the Computational Intelligence community, while addressing pressing needs of the Physics community.**

I. Introduction

Tracking is an important subfield of computer vision [1], [2]. The general goal is to predict future position of multiple moving objects based on their previous positions, with numerous applications such as video surveillance, vehicle navigation, and autonomous robot navigation. These applications focus on identifying a few target objects in complex environments.

In this paper we describe the design of the *TrackML challenge* that is part of the competition program of WCCI2018: a competition that focuses on tracking in an original context. The problem considered refers to recognizing trajectories in the 3D images of proton collisions at the Large Hadron Collider (LHC) at CERN. Think of this as the picture of a fireworks: the time information is lost, but all particle trajectories have roughly the same origin and therefore there is a correspondence between arc length and time ordering. Given the coordinates of the impact of particles on detectors (3D points), the problem is to "connect the dots" or rather the points, i.e. return all sets of points belonging to alleged particle trajectories.

The challenge shares the basic setting of the classical computer vision problem: the general goal of reconstructing a trajectory based on low-level data with no metadata. However, it departs from classical tracking on two majors features: the considerable multiplicity of objects to track, in the order of 10^4, while the objects as much simpler, in the order of a ten of points; and the fact that there is no hierarchy of objects: all, or at least most of, the image points must be associated with a track.

Current methods employed for tracking particles in the LHC experiments will be soon outdated. The augmented data throughput creates a major scaling bottleneck for the associated pattern recognition-tracking task. To mobilise the scientific community for devising radically new approaches, a collaboration between the CERN and Machine Learning scientists organizes this challenge. The challenge has been carefully devised in order to be inclusive and not to require any physics knowledge.

The rest of this paper is organized as follows. We first give an overview of the physics, and of the Machine Learning approaches that could be exploited. In the following, we describe the challenge features, including the newly generated data sets, evaluation metrics and baselines. We also provide a preliminary analysis of the results of the challenge, and discuss the first lessons learned[1]. Given the complexity and originality of the task, as well as its importance for the LHC experiments, the data sets, competition and this design paper will motivate further research.

II. Track reconstruction at the LHC

In the LHC, proton bunches circulates and collide at high energy. The challenge addresses the critical reconstruction step, where raw data are transformed into structured events. Raw data are energy traces (the *hits*) left in the silicon detectors, which are to be clustered into particle trajectories (fig. 1) afterwards called the *tracks*. Tracking is one of the most important tasks in a high-energy physics experiment, as it provides high-precision position and momentum information of charged particles. Such information is crucial for a diverse variety of physics studies - from Standard Model tests to new particle searches - which requires robust low-level optimization without information loss and will be further refined in the analysis for narrower and more specific physics contexts.

By 2025, there will be a major upgrade of the LHC to fulfill its rich physics program: understanding the characteristics of the Higgs boson, searching for the elusive dark matter, or elucidating the dominance of matter over anti-matter in the observable Universe. The upgrade requires to revisit the real-time preprocessing and filtering of the collisions data, to make them scale.

While the energy of the LHC will remain nearly constant, the number of proton collisions will be increased 10-fold progressively until 2025 so that the number of particles per proton bunch collision will also increase from about 1000 to 10,000. In addition, the ATLAS and CMS experiments plan a 10-fold increase of the readout rate. The explosion in combinatorial complexity for the tracking task is mainly due to the increase of the probability of confusion between tracks. It will have to be dealt with with a flat budget at best. The projection of CPU computing power gain with the already highly optimized production software leaves at least a 10-fold gap.

III. Related work

From the machine learning point of view, the problem can be treated as:

- **A latent variable problem:** A data generating process first drew at random particles with given characteristics (momentum, charge, mass), then drew points along a trajectory originating near a collision focus (with uncertainties including diffusion/scattering and imperfections of detectors). "Particle memberships" are the latent variables to be inferred. This is similar to a clustering problem.
- **A tracking problem:** Using the correspondence between arc length and time ordering, one can treat the trajectories as time series and use tracking techniques, including Kalman filters.
- **A pattern de-noising problem:** Considering the collision snapshot as a 3D image, through the data acquisition process, the original trajectory lines were degraded into dotted lines with just a dozen points per line (the human eye cannot see the lines); the problem can therefore be thought of as signal enhancement of an "in-painting" problem (filling in missing data).

In physics, the field of particle tracking is well developed with a specialized conference [3], [4]. While early methods included mathematical transformations such as the Hough transform, the methods offering the best speed/accuracy tradeoff have concentrated on variants of Kalman filters in recent years, combined with various local pattern recognition methods. For an in depth review of the pre-Machine Learning state of the art, see [5].

The preliminary attempts of applying Machine Learning to particle physics pattern recognition-tracking indicate a strong potential [6]. [7] analyzes a simplified and smaller 2D version of the problem. Several promising machine learning and neural network solutions have emerged, including LSTM (Long Short-Term Memory) [8], [9]. Optimization methods such as MCTS (Monte Carlo Tree Search) were also successfully used.

The problem relates to representation learning [10] as in [11], to combinatorial optimization as in [12], neural-network based clustering [13], and even to time series prediction [14] (even though the time information is lost, it can safely be assumed that particles were coming from the center of the detector and have successively crossed the nested layers of the detector).

A possible approach is to efficiently exploit the a priori knowledge about geometrical constraints [15]. Indeed, trajectories are close to segments of helices, as shown in figure 2. The generative approach [16], [17], in particular with introduction of supervision in variational autoencoders [18], [19], as well as the discriminative approaches [20] could be exploited for combining structural priors and nonlinear state estimation with deep neural networks.

[1] The analysis of results will be provided for the camera ready version, if the paper is accepted, as the competition will still have some time to go at this time

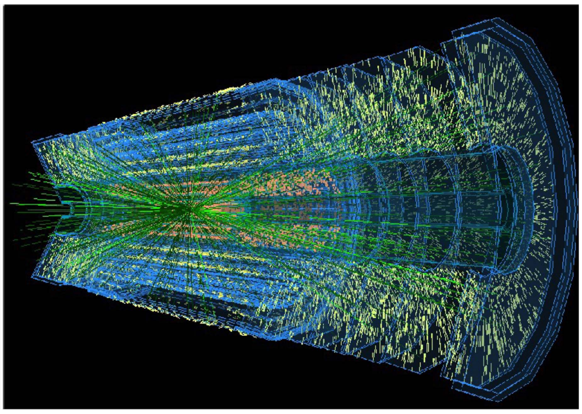

Fig. 1. An etched-out high-multiplicity collision image in the future detector, measurements in yellow, trajectories are green.

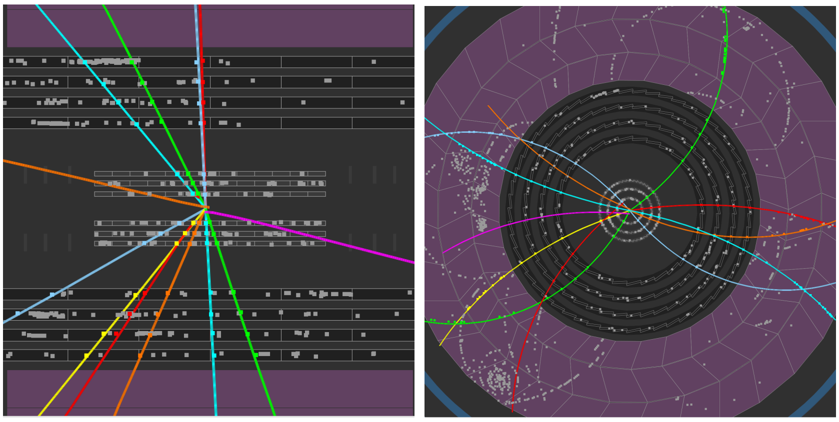

Fig. 2. Projection of the tracks in the longitudinal and transverse planes, for low multiplicity events in the current detector

IV. Challenge overview

We propose a challenge that aims at exercising the latest research advances from the pattern recognition, and more generally machine learning, community in devising fast and accurate particle tracking algorithms. The methods will be evaluated on a very large dataset simulated with a realistic simulator anticipating the new LHC architecture to be deployed by 2025. Thus the ground truth of particle trajectories will be known. The data created for the challenge are representative of the real HEP experimental scenarios.

A. The task

A particle traversing the detector is measured on a number of discrete points, in average 12, but as low as 1 and up to 20. Each point is 3D measurement (euclidean coordinates) with some non isotropic measurement error. The participants should associate 3D points together to form tracks. While the task can be formally stated as a clustering problem, the ratio between the number of clusters ($\sim$ 10K) and their size ($\sim$ 10 points), is highly unusual, and drastically limits the performance of off the shelf clustering algorithms. Typically, at least 90% of the true tracks should be recovered.

The tracks are slightly distorted arc of helices with axes parallel to a fixed direction, and pointing approximately to the interaction center. On figure 2, the arcs appear as lines on the longitudinal projection and circles on the transverse one. Robustness with respect to these distortions and approximate pointing are enforced by the metric and are a de facto requirement.

This task correspond to the first step of particle physics tracking, which is to attribute the points to the track they associated to the same true track, bent in a magnetic field. Further steps, like deriving the parameters of the track trajectory given the 3D points, is not part of the competition.

B. Schedule

The challenge is organized in two phases. The first one, the *Accuracy* phase, which is presented here, targets innovative algorithms reaching the highest accuracy, with no speed concern. It will run from March to May. Fact sheets describing the algorithms to be submitted by end of May. The results will be announced early June.

A second phase of the challenge called the *Throughput* phase will be run from mid-June 2017 to October 2017. This second phase will be focused on the evaluation speed of the algorithms exposed during the first phase (the training speed will not be constrained), while maintaining a similar accuracy. The speed will be evaluated by the challenge platform.

A final workshop will be organized at CERN in spring 2019, where winners of both phases of the challenge will be invited.

V. Dataset and metrics

A. Data

The dataset consisting of a simulation of a typical full Silicon LHC detector lists for each event the measured 3D points coordinates, and the list of 3D points associated to each true track (ground truth), an event corresponding to the tracks of one collision. The simulation engine uses the ACTS[2] [21] simulator both fast (1s per event) and accurate. Realistic collisions yielding 10.000 tracks per event have been simulated with a sufficient level of details to make the task almost as difficult as for real events: points are measured with a precision of approximately 50 microns, some tracks are grouped in dense "jets" (increasing the possibility of confusion), multiple scattering distorts the tracks, points are some times missing, some tracks stop early.

The data set is large in order to allow the training of data intensive Machine Learning methods. The orders of magnitude are : 10^4 events, with each 10^4 tracks, for a total dataset size of 100 GBytes. The events are independent. The public and private evaluation datasets need to be much smaller, about 100 events (1 GBytes), but are large enough to evaluate the metrics within a per mille of statistical uncertainty.

The dataset has been generated for the purpose of the challenge, and can be publicly released.

B. Evaluation

A perfect algorithm will uniquely and correctly associate each point to the track it belongs to. An imperfect algorithm will miss some tracks altogether, miss one or more points for an otherwise valid track, associate wrong points to an otherwise valid track, find tracks from random association of points, find multiple variants of the same track.

Because the data come from simulation, we know which particle created each hit (point), in other words the ground truth. For brevity, we note R-tracks the proposed solutions and T-tracks the ground truth. A point must belong to at most one R-track, but it is not required to list all points. The score is defined as follows.

- R-Tracks with 3 points or less have a zero score, as they do not allow to compute any meaningful physical quantity in further analysis.
- R-tracks and T-tracks are uniquely matched by the combination of the following rules.
  - For each R-Track, the matching T-track is the one to which the majority of the R-track points belong; if there is no such particle, the score for this track is zero.
  - The R-Track should have the majority of the points of the matching T-track, otherwise the score of this track is zero

  These two requirements guaranty a one to one relation M between all remaining R-tracks and T-tracks.
- The score of a R-track r is the weighted count of the points in the intersection $r \cap M(r)$.
- The score of the event is the sum of the scores of the R-tracks, normalized by the sum of the weights of all points. This actually normalizes the score to the $[0, 1]$ range, with 1 being the score of a perfect reconstruction.
- Finally, the overall score is the average over the 100 test events. We have evaluated the statistical uncertainty to be a few 10^{-4}.

The weights are incentives to get physically meaningful reconstructions, along two directions: the weight of a point is the product of two independent quantities *weight_order* and *weight_pt*.

- *weight_order* The points at the beginning of the track, close to the collision, and at the end, are more important than the ones in the middle. The weights reflect this hierarchy and are normalized to sum to 1.
- *weight_pt* The high energy particles (large transverse momentum p_T) are the most interesting ones. As the bulk of the tracks have low p_T, we have to explicitly favor high p_T. *weight_pt* is 0.2 if $p_T < 0.5$GeV and 1. for $p_T > 3$GeV, with a linear interpolation in between. Note that the lower the p_T, the larger the geometrical curvature; at large p_T tracks appear as straight lines.
- Particles which generates 3 hits (points) or less are considered spurious, the weights of the associated points are set to zero.

As the weights disclose important information, they are provided along with the points in the training data, but will be kept hidden for the test data.

An alternative metric would be the *adjusted Rand index* (with weights) clustering metric [22], [23]. But

[2]https://gitlab.cern.ch/acts/acts-core

the Rand_index is the sum of two quantities (up to a multiplicative constant): the number of true positives, i.e. the pairs of points that agree to be in the same cluster, and the number of true negatives, i.e. the pairs of points that agree to be in different clusters. (false positives would be the pair of points that are in the same cluster in the reconstructed track, but in different clusters in truth, etc.).

The setting is different between tracking and clustering in the sense that clustering has to assign each and every point to a cluster, thus the need to (indirectly) penalize false assignments. Also and more importantly, in our case, there is a one-to-one assignment of reconstructed clusters (R-Tracks) to true clusters (T-tracks) through the the double majority rule. Our score counts only the true positives, but penalizes the false positives and negatives in the matching procedure. This score is consistent with the various and more complicated metrics used in physics reconstruction.

VI. Protocol and evaluation

A. Evaluation

The Accuracy phase of the challenge, which is the subject of this paper, is run as a traditional Kaggle challenge where the participants do not upload software but *solution files*. As usual, the dataset is partitioned into training, public test and private test. The challenge platform will be Kaggle.

A solution is a list of associated points, each association being an assumed track, with an arbitrary unique numbering. The participant will develop their code (without any restriction on the language or libraries), and train their models on their own computing resources. They will apply their resulting evaluation algorithm to a test file (with hidden ground truth) by uploading the solution (list of points associated together) to the challenge platform. The challenge platform will compute a score value of the metric for the test sets and display the score on the public test set on a leaderboard. The final ranking will be based on the private score only in order to avoid overfitting the public test set. It will not be disclosed until the end of the challenge. The precise numbers of leaderboard submissions per day and final submissions will be finalized with the platform provider.

B. The tracks

The competition is organized in two tracks. Participants competing for a prize in any of the two tracks will be requested to release their software and to self-assess the CPU usage for both training and testing.

- The *Performance* track will be solely based on the previously described metric in section V-B.
- The *Algorithm* track will be based on a jury (with experts in particle physics tracking algorithms and machine learning) which will select the submission with the most promising balance between the score, speed and originality.

C. Helpers

A set of helpers have been designed in order to facilitate accessing the challenge. They will be made available on the challenge platform from the first day:

Two light events (with just 100 tracks instead of 10.000) will allow first steps and easier development of auxiliary tools.

The starting kit includes iPython notebooks implementing a complete workflow, various python procedures for accessing the data as well as creating and evaluating solutions and the baseline algorithms. A 3D web based visualization tool is under development.

Two baselines will be provided.

- DBScan: this clustering algorithm demonstrates non-trivial performance, although far from the requested ones. The main goal is to provide a simple method to demonstrate the workflow.
- Hough transform; where the 3D hit space is mapped onto a track parameter space, where maxima (corresponding to tracks) are found, and then moved back to the original 3D space to associate the points. This technique has a linear complexity; however it does not allow to reach the maximum efficiency.

VII. Conclusion

The HEP (High Energy Physics) experiments have embraced Machine Learning, originally for supervised classification as a routine tool in the final analysis stage, and in the past few years for exploring more diverse applications. The preliminary attempts of applying Machine Learning to particle physics pattern recognition-tracking indicate a strong potential. Considering the success of the Higgs Boson ML Challenge [24], the HEP-ML collaboration for this challenge can be expected to produce high impact results.

References

[1] A. Yilmaz, O. Javed, and M. Shah, "Object tracking: A survey," *ACM Comput. Surv.*, vol. 38, no. 4, Dec. 2006.

[2] X. Li, W. Hu, C. Shen, Z. Zhang, A. Dick, and A. V. D. Hengel, "A survey of appearance models in visual object tracking," *ACM Trans. Intell. Syst. Technol.*, vol. 4, no. 4, pp. 58:1–58:48, Oct. 2013.

[3] R. Frhwirth, E. Brondolin, B. Kolbinger, and W. Waltenberger, Eds., *Proceedings, Connecting The Dots 2016*, vol. 127, 2016.

[4] C. Germain, H. Grasland, A. Lounis, D. Rousseau, and D. Varouchas, Eds., *Proceedings, Connecting The Dots / Intelligent Tracker (CTD/WIT 2017)*, vol. 150, 2017.

[5] T. Cornelissen, M. Elsing, S. Fleischmann, W. Liebig, and E. Moyse, "Concepts, Design and Implementation of the ATLAS New Tracking (NEWT), ATL-SOFT-PUB-2007-007," 2007.

[6] S. Farrell *et al.*, "The HEP.TrkX Project: deep neural networks for HL-LHC online and offline tracking," *EPJ Web Conf.*, vol. 150, p. 00003, 2017.

[7] S. Amrouche *et al.*, "Track reconstruction at LHC as a collaborative data challenge use case with RAMP," *EPJ Web Conf.*, vol. 150, p. 00015, 2017.

[8] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Comput.*, vol. 9, no. 8, pp. 1735–1780, Nov. 1997.

[9] Z. C. Lipton, "A critical review of recurrent neural networks for sequence learning," *CoRR*, vol. abs/1506.00019, 2015.

[10] 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.

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

[12] 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

[13] G. Aad *et al.*, "A neural network clustering algorithm for the ATLAS silicon pixel detector," *Journal of Instrumentation*, vol. 9, no. 09, 2014.

[14] 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.

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

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

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

[18] D. P. Kingma, S. Mohamed, D. J. Rezende, and M. Welling, "Semi-supervised learning with deep generative models," in *Advances in Neural Information Processing Systems 27*, 2014, pp. 3581–3589.

[19] S. Narayanaswamy, B. Paige, J. van de Meent, A. Desmaison, N. D. Goodman, P. Kohli, F. D. Wood, and P. H. S. Torr, "Learning disentangled representations with semi-supervised deep generative models," in *Advances in Neural Information Processing Systems 30*, 2017, pp. 5927–5937.

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

[21] C. Gumpert, A. Salzburger, M. Kiehn, J. Hrdinka, and N. Calace, "ACTS: from ATLAS software towards a common track reconstruction software," *J. Phys. Conf. Ser.*, vol. 898, no. 4, p. 042011, 2017.

[22] W. M. Rand, "Objective criteria for the evaluation of clustering methods," *Journal of the American Statistical Association*, vol. 66, no. 336, pp. 846–850, 1971.

[23] M. Meilă, "Comparing clusterings—an information based distance," *J. Multivar. Anal.*, vol. 98, no. 5, pp. 873–895, 2007.

[24] 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