EXPLORING THE ROLE OF DEEP LEARNING FOR PARTICLE TRACKING IN HIGH ENERGY PHYSICS

Mayur Mudigonda,* Steve Farrell, Jesse Livezey,Prabhat, Paolo Calafiura
Lawrence Berkeley National Laboratory, Redwood Center for Theoretical Neuroscience,
Berkeley, California

Dustin Anderson, Jean-Roch Vilmant, Josh Bendavid, Maria Spiropoulou, Stephan Zheng
California Institute of Technology, Pasadena, California

Aristeidis Tsaris, Giuseppe Cerati, Jim Kowalkowski, Lindsey Gray, Panagiotis Spentzouris
Fermi National Accelerator Laboratory, Batavia, Illinois

ABSTRACT

Tracking particles in a collider is a challenging problem due to collisions, imperfections in sensors and the nonlinear trajectories of particles in a magnetic field. Presently, the algorithms employed to track particles are best suited to capture linear dynamics. We believe that incremental optimization of current LHC (Large Halidron collider) tracking algorithms has reached the point of diminishing returns. These algorithms will not be able to cope with the 10-100x increase in HL-LHC (high luminosity) data rates anticipated to exceed O(100) GB/s by 2025, without large investments in computing hardware and software development or without severely curtailing the physics reach of HL-LHC experiments. An optimized particle tracking algorithm that scales linearly with LHC luminosity (or events detected), rather than quadratically or worse, may lead by itself to an order of magnitude improvement in the track processing throughput without affecting the track identification performance, hence maintaining the physics performance intact. Here, we present preliminary results comparing traditional Kalman filtering based methods for tracking versus an LSTM approach. We find that an LSTM based solution does not outperform a Kalman fiter based solution, arguing for exploring ways to encode apriori information.

1 INTRODUCTION

Deep Learning has played a phenomenal role in making advances in many fields such as computer vision Goodfellow et al. (2014), speech recognition Hinton et al. (2012) and robotics Levine et al. (2016) amongst other fields. While there has been some work on applying deep learning techniques to searching for particles in high energy physics (HEP) Baldi et al. (2014) there has not yet been any concerted effort in applying it to problems of tracking.

In this work, we explore the role of deep learning for problems of tracking in high energy physics experiments. First, we present the complexities of the problem in detecting particles. Next, we present preliminary results on the applications of LSTMs to tracking particles in a detector array. We hope, with this work, to reach out to the broader machine learning community to both present our findings and seek out methods for solving challenging problems in high energy physics .

2 THE PATTERN RECOGNITION PROBLEMS IN HIGH ENERGY PHYSICS DETECTORS

In a typical HEP experiment, building-sized underground detectors collect TBs of data per second coming from high-energy collisions of two particle beams. The detectors are composed of concentric cylindrical layers of electronic sensors surrounding the collision region. Each collision event

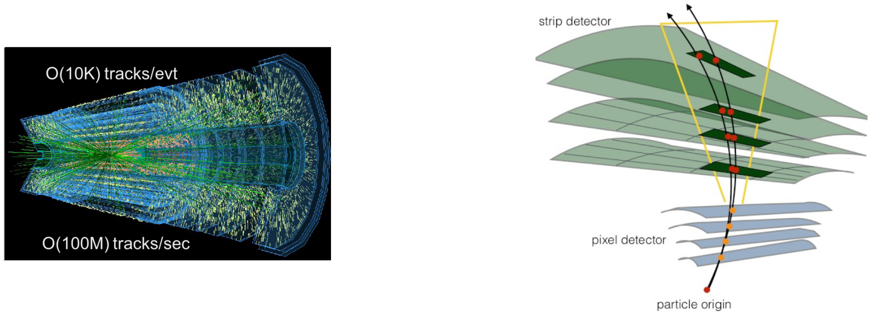

Figure 1: **left** Visualization of hits from trajectories from the ATLAS general-purpose LHC. The figure shows a slice view of the detector ands hits on the various layers of the detectors. **right** a schematic describing how a single particle generates hits that are used as inputs. The particle travels from its origin and passes through pixel detectors which help form the seed for track fitting. The particle then continues to travel through the various detector layers sometimes resulting in multiple or missed hits in layers due to various physical interactions

consists of $O(10^3)$ particles that traverse the detectors in various directions and different charge, energy, and momentum as seen in Figure 1. The topologies of the events offer insight into the nature of the collisions, allowing to probe the properties of elementary particles and the fundamental laws of nature on a statistical basis.

The most demanding pattern recognition task in HEP is to reconstruct the trajectories (”tracks”) of millions of charged particles per second as they propagate through a tracking system of a detector. Given a 3D image $I(x, y, z)$ with triplets of inputs where each pixel has a binary value, with 1 signifying a hit on the detector layer, the pattern recognition task is to group together all hits generated by each particle as seen in Figure 1. This task is made complicated by detector effects (such as noise in the sensors and an imperfect magnetic field) as well by stochastic perturbations to the particle trajectory derived from particle interactions with detector material.

The similarities in problems between those explored in computer vision, robotics and the HEP field are obvious. The obvious differences lie in the fact that in the case of HEP-LHC typically we would need to estimate the parameters of millions of tracks in parallel. Further, the required reliability of a model is significantly higher. For example, the existing state of the art methods can detect tracks with an efficiency between 90-99% depending of the particle type and its momentum.

3 Modeling

For the tracking problem, one is provided with a *seed*. A seed is a n-tuple of three points in 3D space, where n is the minimum number of points required to fit a parametric curve to a set of points in 3D.

Seed generation is a pattern recognition problem in of itself. But given the seed, our approach has been to fit an LSTM to predict the location of the next hit. The loss function in this case is to minimize the predicted hits across an entire sequence of a trajectory.

Input vectors are fed into LSTM units (at least 5 in number), the output of the LSTM units are then fed into two fully connected layers which produce the prediction (or the next time-step). The weights are then learnt through gradient descent.

We compare our method against a Kalman filter whose transition matrices are not learnt but manually set with knowledge of the physics Frhwirth (1987). That is, we encode the transition matrices that describe the dynamics in the latent space and their projections back onto the observation space based on the approximated analytical forms that the particles are expected to take as they make their way out of the detectors. Of further importance, these Kalman filters have unique transition matrices for each detector layer to better capture the expressive nature of dynamics.

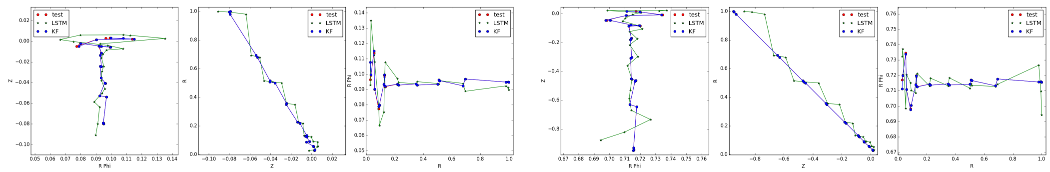

Figure 2: For a given track (red), we compare the Kalman Filter solutions (blue) and the LSTM solution (red). The three subplots show the comparision of $R\phi$ vs z, z vs R and R vs $R\phi$ **(left)** Shows a case where the KF and LSTM solution very closely match the data. **(middle)** Compares the LSTM and KF solution on a track where they differ the most.

Measurement	Euclidean Distance
Meas - KF	0.208
Meas - LSTM	1.834

Table 1: Comparing average Euclidean distance between measurements and predictions from both the Kalman Filter and LSTM

The limitations with these methods are that they inherently cannot capture non-linear dynamics and the physics is known only up to an approximation that is further exacerbated by noisy measurements. We wish to explore the role RNNs like LSTMs could play in modeling these dynamics.

4 Discussion

Here we present a summary of our preliminary findings. Using the ACTS simulation software we simulated around 50,000 charged particle tracks. From this, for convenience of analysis , we sampled trajectories of step length 22 resulting in 16,275 samples. We sampled 200 examples to form a test set. Each sample consists of three dimensions - $R\phi$, z and R. R is the distance of the detector from the origin determined by the geometry of the detector, ϕ is the angle swept across the detector by the particle and finally z is a shift along the slice swept out by ϕ.

We then fit an LSTM with 10 hidden units and two fully connected units of sizes 20 and 2 (since R is known apriori for all tracks) to produce the prediction for the next time step. We used Adam Optimizer Kingma & Ba (2014) to train the weights with an initial learning rate set to 0.001. We also experimented with changing the number of fully connected layers and the types of recurrent network units (for e.g. GRUs with varying number of hidden units), although we make no claim for an exhaustive search of these architectures.

We find that an LSTM based approach can filter states comparably in some cases to an ideal model based on the Kalman Filter as seen in Figure 2 and Table 1. Yet, there still remains a **large gap in performance**.

Our ideas moving forward is to look towards **combining prior knowledge** about the problem **with a learning based approach**. For example, we hope to train models that have access to information such as the magnetic field (say). Further we hope to explore models which can encode the geometry of the detector to better be able to make predictions between layers?

Track fitting is just one step of the puzzle in high energy physics. The goal of our HEP.TrkX project[1] is to prototype an end-to-end solution for the HL-LHC track pattern recognition challenge. Current solutions for this have a combinatorial approach that would make the latency larger when the data throughput is higher. The motivation for this submission is to seek advice and inputs from the larger representation learning community on models and methods.

[1]https://heptrkx.github.io/

REFERENCES

Pierre Baldi, Peter Sadowski, and Daniel Whiteson. Searching for exotic particles in high-energy physics with deep learning. *Nature communications*, 5, 2014.

R. Frhwirth. Application of kalman filtering to track and vertex fitting. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 262(2):444 – 450, 1987. ISSN 0168-9002. doi: http://dx.doi.org/10.1016/0168-9002(87)90887-4. URL `http://www.sciencedirect.com/science/article/pii/0168900287908874`.

Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. Generative adversarial nets. In *Advances in neural information processing systems*, pp. 2672–2680, 2014.

Geoffrey Hinton, Li Deng, Dong Yu, George E Dahl, Abdel-rahman Mohamed, Navdeep Jaitly, Andrew Senior, Vincent Vanhoucke, Patrick Nguyen, Tara N Sainath, et al. Deep neural networks for acoustic modeling in speech recognition: The shared views of four research groups. *IEEE Signal Processing Magazine*, 29(6):82–97, 2012.

Diederik Kingma and Jimmy Ba. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014.

Sergey Levine, Chelsea Finn, Trevor Darrell, and Pieter Abbeel. End-to-end training of deep visuomotor policies. *Journal of Machine Learning Research*, 17(39):1–40, 2016.