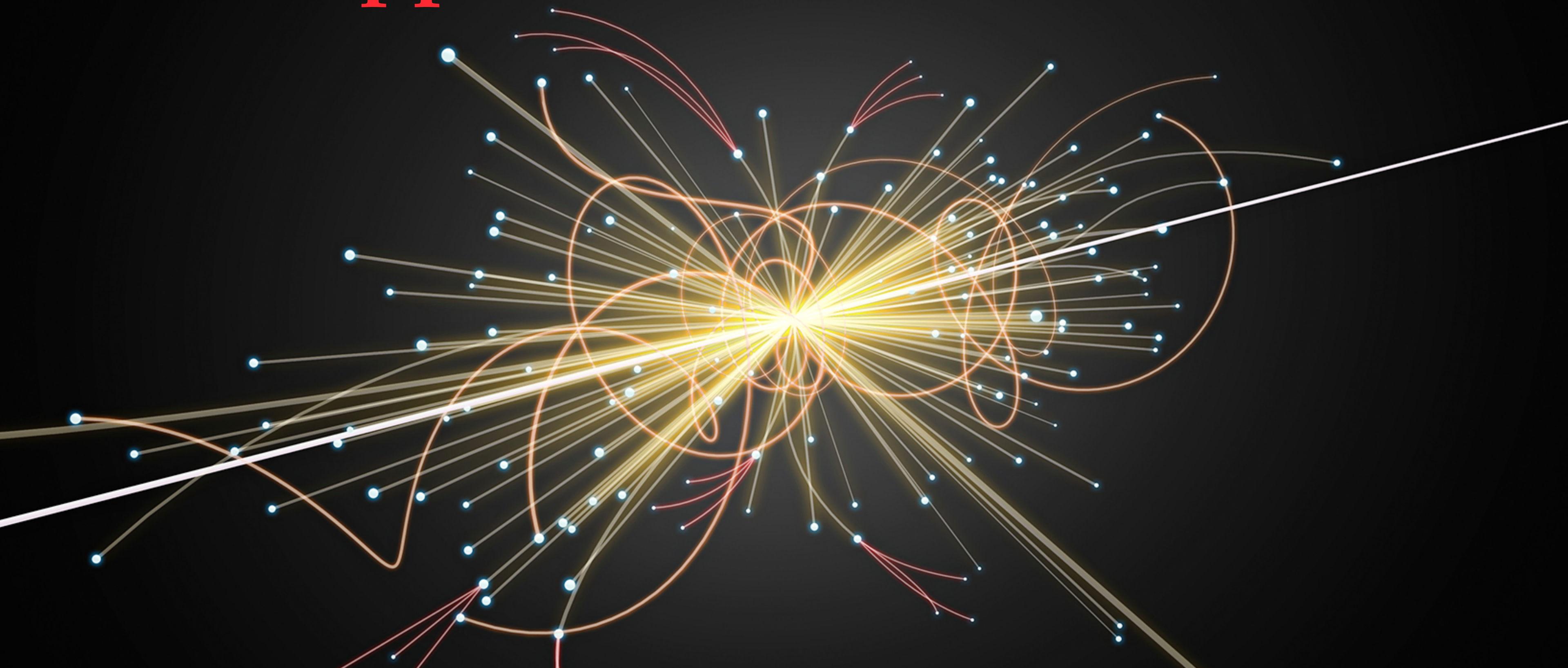
Deep NN & other ML techniques applied to science cases



Maher Sahyoun, Haixing Fang, Juan Carmona, Oswin Krause,
Toby Perring, and Troels C. Petersen (chair)

Workshop on Perspectives and Applications of Deep Learning for Accelerated Scientific Discovery in Physics, 14-15 May 2020

Outline

The outset was the following questions:

1. In what ways can we use Deep Learning and other ML techniques to:
a) Get new insights, b) Enhance physics research, c) Spur new ideas?
Examples: Discovering new useful features and relations, ML-based de-correlation, Predicting uncertainties, etc.

2. How to minimise biases from training on simulation?
Examples: Changes in data taking / content / quality, Planning of future projects, Augmenting simulation to resemble data, etc.

3. How to find other smart/fast/fitting ML-related methods for specific problems.
Examples: Optimum in very high dimensional spaces, Finding simpler methods, that will work with fewer parameters under many conditions.

We decided to simply through ourselves at challenges at random, and then went to fit the answers into the above categories.

The panel was also joined by Oswin Krause, who is a computer scientist, and who came with many great technical insights.

ML techniques

1. In what ways can we use Deep Learning and other ML techniques to:

a) Get new insights, b) Enhance physics research, c) Spur new ideas?

A.

While ranking input features and providing uncertainties can provide some insights, it seemed that this was a hard point to get to with ML. However, a case (see next page) was discussed.

B.

The enhancements were “obvious” in ML simply providing better classifications and estimates. However, the problem might be to get researchers to embrace the new methods. Also, image analysis was pointed out as a great tool, and the example with “teaching patterns by drawing” [Anders Dahl] was great.

C.

For new ideas, making data public was probably the best idea. This gives inherent competition for showing state-of-the-art, and can provide “a kick” [Toby] to a field of research.

Training/simulating data

2. How to minimise biases from training on simulation?

<u>Lively discussion circled around the following idea:</u>
We often simulate situations from first principles (i.e. understanding), and use this for training ML algorithms. However, simulation does not match real data perfectly, and now the problems start…
Would it be possible for an ML algorithm to do a "small" transformation of simulated data to make it match real data in known situations?

The answer is "yes" [Oswin] and the method is a Wasserstein GAN, but the problem is specifying what "small" is?
And a better output is probably a re-weighing of the simulated data, rather than a transformation.

And finally, given a good transformation / re-weighing, do we trust it to span other situations? Can we learn from it? **What is wrong/inaccurate in the simulation?**

Simpler methods?

3. How to find other smart/fast/fitting ML-related methods for specific problems.

Often we run into the trouble of requiring:

- Too much training data
- Too much training time

Can simple(r) methods do the trick and if so, how do we find them?

It was pointed out, that linear methods pieced together by an ML algorithm [Juan] was often a very good and fast solution. This leads towards “ensemble methods”.

Returning to the simulation-to-data transformation (last page), one can probably extract some “distribution or directions” (PCA or FFT like) of the transformation, from which one can apply simpler (physics motivated?) transformations to the simulation.

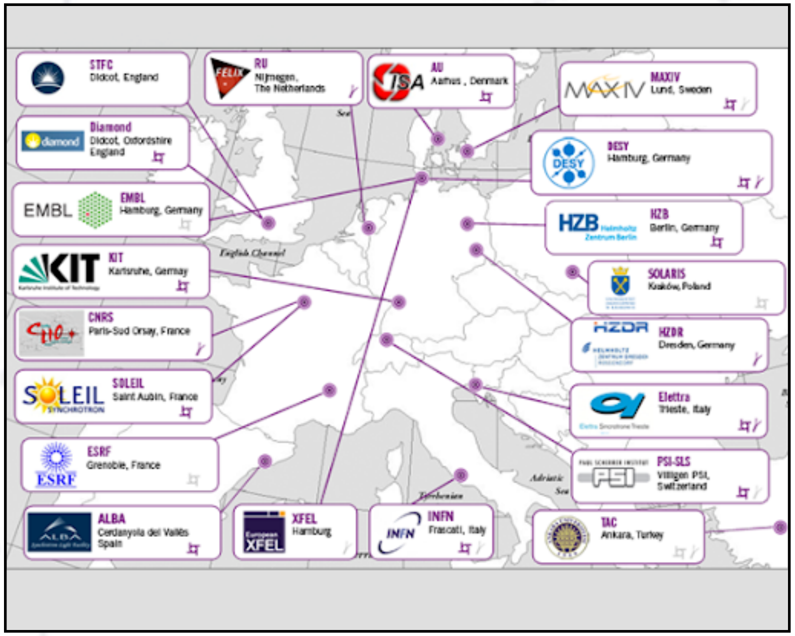
We should probably teach ourselves to “also like simple methods”, for speed and simplicity.

The unexpected suggestion

It has been mentioned often, that open data with fitting descriptions of the problem could accelerate a field.

However, it was also pointed out, that different facilities have different setups, resolutions, detectors, problems, etc.

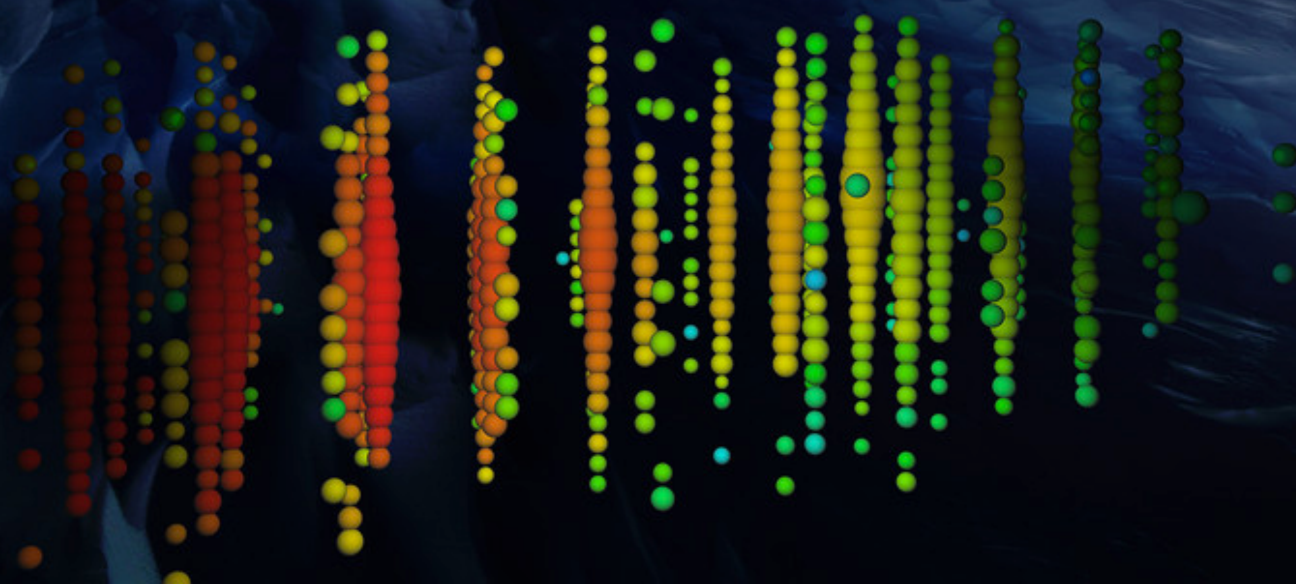
A simple solution would be to take a few “standard” samples and take them to all facilities [Maher]. That would be a very interesting sample to tackle, and would probably teach the community a few things.



Ideas for the future

Part of what ML can do, is to speed up the process of prediction / reconstruction, and do so in a less humanpower intensive way.

A quote from a student two days ago: “I always give new data to a clustering algorithm, and see what it finds… then I have a quick, fast, and unbiased first view of the data!”

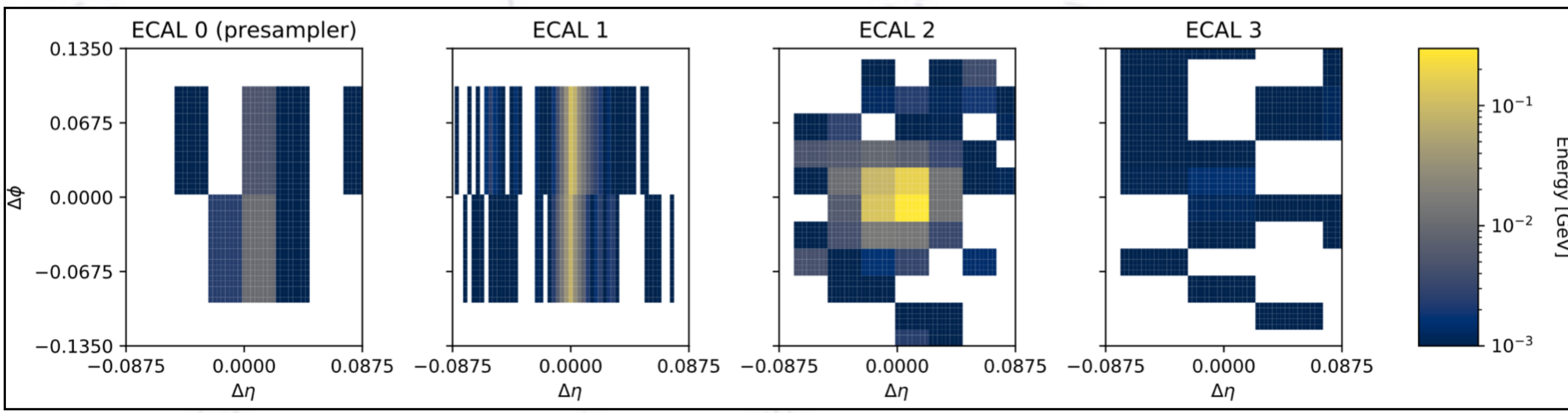


Conclusions

We agreed that there were infinitely many great perspectives for ML in science:

- Sharpening of almost every classification/quantity extracted from data.
- New insights from taking .
- However, in some cases, it makes all the difference.

Maybe we should remember to start to see our (research) world with ML eyes!
An example of that is given below…



The "image" of a simulated electron in the four layers of the ATLAS calorimeter - great CNN case!

Bonus slides