**Discussion session** 

# Big Data Science in Astroparticle Research: sharing and exploiting data & knowledge

Martin Erdmann & Andreas Haungs 19-Feb-2020

# This workshop: AGAIN PROGRESS in 2019/2020

- Golden concept: variants of convolution algorithms
- Graph networks: sparse data without cartesian symmetry
- Include symmetries of data in network concepts
- Include operational reality of large-scale experiment in training data
- Employment perspectives: excellent for you in science, economy, society



# **Evaluation** Participants

I participate in this workshop as

- Student
- PhD student
- Postdoctoral researcher
- Senior physicist (>5 years PhD)
- Other

#### 19-Feb-2020, 31 answers of 64 participants

#### I participate in this workshop as



Answered: 30



Student (8)	
PhD student (17)	
Postdoctoral researcher (8)	
<ul> <li>Senior physicist (&gt;5 years PhD) (2)</li> </ul>	
Other (1)	
No answer (0)	
Not displayed (0)	

# University: courses, seminars, research







- Deep learning networks (6)
- Neural networks (shallow, 1-2 layers) (10)
- Boosted decision trees (16)
- None (16)
- = Don't know (0)
- Other (0)

#### Flipped 2017/18 -> 19/20 many of you active in deep learning

ungs

- Deep learning networks Neural networks (shallow)
- Graph Neural Networks **Reinforcement Learning**

# Physicists take advantage of new opportunities arising with deep networks

My research interests concerning deep learning networks are



Image pattern recognition (e.g. event reconstruction)

- Classification (e.g. particle identification, physical processes)
- Assignments (e.g. solving ambiguities with >2 identical particles)
- Regression (calculation of variables, e.g. energy corrections)
- Adversarial training
- Curiosity
- Applications not obvious
- Other applications









- Image pattern recognition (e.g. event reconstruction) (23)
- Classification (e.g. particle identification, physical processes) (22)
- Assignments (e.g. solving ambiguities with >2 identical particles) (6)
- Regression (calculation of variables, e.g. energy corrections) (23)
- Curiosity (currently no special research idea) (19)
   Applications are not
- Applications are no obvious (1)
   Other (0)
- Other (0)

### **GPU** resources

I have access to GPU resources for network training at my research institute



# Experiment data

For performing my research I need access to



#### Every form of data used, resource intensive

- Raw data (17)
- Calibrated data (16)
- Reconstructed data (15)
- High level data (direction, energy,...) (14)
- Simulated data (28)
- Other (1)



# Experiment analysis support from



Support is an obvious prerequisite for successful exploitation of data

Martin Erdmann, Andreas Haungs

My research group/institute Colleagues of my collaboration Information on websites

### Experiment open data

I have accessed open (*public*) data for my research purposes before

Never

- **Sometimes**
- Frequently

![](_page_8_Figure_5.jpeg)

Requires scientific experience to judge what to extract from which data

Martin Erdmann, Andreas Haungs

![](_page_8_Figure_8.jpeg)

Answered: 30

#### 

### Improve exploitation of new and old data

The following **measures** are important for achieving progress in my research

![](_page_9_Figure_2.jpeg)

- а. b. С. d. 2020 e. 6 8 10 12 14 16 18
  - User support deep learning
  - Common platform to collect and re-use networks
  - National GPU resources for basic research
  - Common national structure for deep learning applications
  - others

![](_page_9_Figure_9.jpeg)

**E. Other**: 6 (10.91%)

A. User support for deep learning: 18 (32.73%)

- User support deep learning (18)
- Common platform to collect and re-use networks (13)
- National GPU resources for basic research (18)
- Common national structure for deep learning applications in (astro-)particle physics (18)
- Other (1)

Need sustainable structures for accelerating research & development

### Workshops perspectives

For the future, I consider continuation workshops important on topic(s)

![](_page_10_Figure_2.jpeg)

Strong wish to benefit from each other's progress, still deep learning

- **Deep learning** a.
- Open data b.
- Open analysis software С.
- Analysis preservation d.
- Not important e.
- f. Other

#### Workshop advanced tutorial: Graph Networks

![](_page_11_Figure_1.jpeg)

1=strongly disagree, 2=disagree, 3=undecided, 4=agree, 5=strongly agree

Martin Erdmann, Andreas Haungs

# The example applications deepened my

### Workshop *beginners tutorial*

- Introduction to machine learning with deep neural networks •
  - The deep learning introductory tutorial was helpful

![](_page_12_Figure_3.jpeg)

mean=4.12

# understanding of deep networks

![](_page_12_Figure_6.jpeg)

1=strongly disagree, 2=disagree, 3=undecided, 4=agree, 5=strongly agree

![](_page_12_Figure_8.jpeg)

![](_page_12_Picture_9.jpeg)

mean=3.71

![](_page_12_Figure_11.jpeg)

2018

2017

1 (0) 2 (6) 3 (3) 4(14) 5 (13) No answer (0) Not displayed (0)

#### VISPA platform

1=insufficient, 2=sufficient, 3=ok, 4=good, 5=very good

![](_page_13_Figure_2.jpeg)

# what you liked about this workshop

- everything
- That DL has a widespread possibilities in physics and I've learned new things from which I didn't knew you could use DL in this (e.g. **FPGA** programming)
- The **beginner hands on**
- Very practical, and **hands-on**. Great examples!
- Good, well prepared talks, good tutorial
- A lot of information that I didn't know before, hands on workshop
- The **Tutorial on Graph** networks
- The **tutorials** and the talks about the **application** in astronomy
- The graph neural network tutorial was great some talks were really good
- Good **tutorial and talks** with a well defined topic
- Finally some **basics** presented
- The idea to have more like a discussion about topics than to only have polished presentations.
- I liked to see the variety of possibilities for ML/DL applications that might inspire future work also on  $\bullet$ other fields.
- I liked the graph network tutorial, which I found very useful. Moreover, I liked the interdisciplinarity of the workshop, which allowed me to come in contact with topics that are not directly related to my research.

# what you did not like about this workshop

#### nothing

- It was more like a conference with all the talks than a workshop. The **tutorial was too short** with just one afternoon. And maybe even for the **beginners** it could be more interesting to also use **data** from physics experiments.
- That it's **only one day**. I'd prefer to have a little more time to digest new concepts.  $\bullet$
- The talks about **FPGA**S were not my area of interest
- some talks were a bit **boring**  $\bullet$
- More **overview for beginners** please
- I would have liked to hear some more about struggles people have using NN and how to solve them

# your future suggestions for this workshop

- please continue lacksquare
- More hands on tutorial with real astroparticle data to get more into the topic and maybe, if the participants have to contribute something, maybe not just only in talks, but also with **posters**, so that there would be more time for deepening the **Tutorials** a bit in this three days.
- Even in the beginner hands on: explain how to optimize networks. Explain how to evaluate the networks I'd argue for a **school** every number of years. I'll be happy to send several students to such a school. (Very  ${\bullet}$ likely I'll send them next year for this workshop anyway.)
- make workshop a bit more interactive, use Jupyter notebooks to go through the steps so we can  ${\color{black}\bullet}$ understand what's actually happening
- More tutorials  $\bullet$
- A **middle level tutorials**. There is a big gap between the beginners and advanced specialist, which is left uncovered so far
- More tutorials, also over **longer days**. lacksquare
- announce the titles of the talks earlier have more tutorials maybe provide abstracts to the talks
- I would like more overview of different methods and what is best suited for what purpose
- Maybe the advanced tutorials should cover less topics but have more in depth explanations. Also I think it's lacksquarebetter to first go through the code (hands on writing) before exercises

### How to continue

Deep Learning in Astroparticle Research 2017, 2018, 2019, 2020

![](_page_17_Picture_2.jpeg)

In 4 years community has grown to apply deep learning in all large-scale experiments.

#### Big Data Science in Research on Universe & Matter (ErUM-Data)

Astroparticle, Astronomy, Particle, Hadron & Nuclei,

![](_page_17_Figure_6.jpeg)

#### Enlarge exchange community beyond astroparticle physicists.

?

#### Accelerators, Neutrons, Sychrotron Radiation, Nuclei Probes

#### From the Discussion

- Change format from a pure HAP workshop (HAP project financing is gone) to an *ErUM-Data workshop* where  $\bullet$ funding is supposed to arrive in a year from now.
- Astroparticle community has advanced very fast, every large collaboration has a deep learning paper. Survey 2017/18 only few people with deep learning experience on this workshop, 2019/20, quite a lot of people bring to exchange deep learning experiences. With this maturity we could open up to neighbor communities and learn from each other. Not all of them but well selected community.
- Observation that workshop is smaller by 1/3, only last minute we filled the program. We should react to this for the next workshop or whatever format. Suggestion: Organizers can send more email reminders.
- The Big Data Science in Astroparticle Research is a flagship of our community. No other ErUM community has a comparable constant effort on Big Data Science with focus on deep learning over several years.
- It was good to say at the beginning, tell us the physics context and tell us why you wanted to apply this method. It helped the workshop.

### From the Discussion

- Make a hackathon for medium scale between beginners and far-advanced people so that one gets into contact and has hands-on on a real problem [Like on the ML4Jets workshop (New York Jan 2020), we had a data set including a new resonance signal which was to be found. Or a Kaggle challenge].
- Participants could bring challenges to the workshop: we'd be very happy if participants bring tutorials or challenges, it is a lot of work.
- Super experts are coming here together, give a small round of the workshop a task to be solved together within the workshop. Or help people not yet so far advanced with their problems.
- Offer a whole week but stage it according to the interests, 1<sup>st</sup> day newcomer tutorials, 2<sup>nd</sup> day advanced  $\bullet$ tutorials, 3<sup>rd</sup> day talks on deep learning, 4<sup>th</sup> day talks on big data,...
- Make a system or workshop for discussing failures.
- Prefer the hands-on, the tutorial was good, the talks were not so important. Rather make it a school type  $\bullet$ workshop, make it longer, a whole week.
- There are time constraints for the organizers, 3 days you can affort, 9 days (Erlangen school) is impossible. 5  $\bullet$ days would be already a challenge.
- The exact form is to be thought about, do we want a school, a workshop, a hackathon, a mixture, ...

#### From the Discussion

- This workshop has different topics, one is to produce the big data, another one is to analyze the big data.
- Keep the astroparticle physics community central in this workshop and invite people from other field to ۲ contribute and exchange. Don't let us be swamped by the particle physics community, we cannot be a diminishing part on this workshop, otherwise we lose our identity here.
- When inviting neighbor communities, plan it with a criterion, e.g. in the way, that the technology method is the overlap. Otherwise everybody brings something and it is a potpourri. For example announce a workshop on convolutional deep learning methods in astroparticle and astronomy
- If new forms of the workshop are to be found, make a smooth transition, evolution, don't make a revolution.  $\bullet$
- Level of talks should be designed for the audience, extract at a more abstract level the most important  $\bullet$ aspects relevant to the people in the audience. Yes, but: we also invite young people to give their first ever workshop presentation on the deep learning topic they were working on. For them it is a unique chance, but they cannot give a talk at the level of an experienced lecturer.

### Thanks very much!

Workshop administration: Sabine Bucher, Melanie Strothotte

Local RWTH workshop team: Teresa Bister, Peter Fackeldey, Jonas Glombitza

VISPA tutorial team: Benjamin Fischer, Niklas Langner, Dennis Noll, and many more contributing!