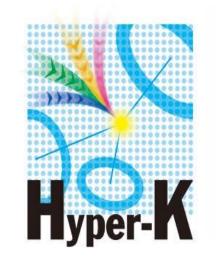
GRANT & WatChMaL: Towards a unify framework for Machine Learning based reconstruction

@HK 2024 Collaboration Meeting

ILANCE – LLR | Erwan Le Blévec



L

Λ

N

C = C

International Laboratory for Astrophysics, Neutrino and Cosmology Experiments



Let's begin: Context

• I. The situation

• II. The problem(s)

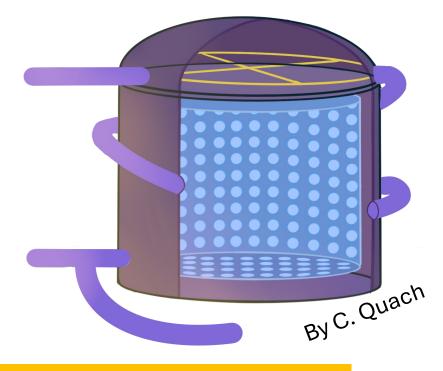
• III. My task

Context

1. The situation

Context 1. The situation

FitQun's n°1 drawbacks: Time





Can take from 1 minute up to > 10 to process one event

With the huge MC production mandatory for err_stat + err_sys < 1%: it is too long for us

Deadline is 2027

Context 1. The situation

- Conclusion I: Time limits is not negligeable for us
- Conclusion II: Opening to alternatives
- Another approach : Neural Networks

Context 1. The situation

Neural Networks effort - Software

WatChMaL:



- ➤ Started 4 years ago.
- > Events as images

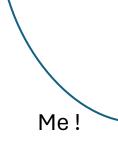
Grant:

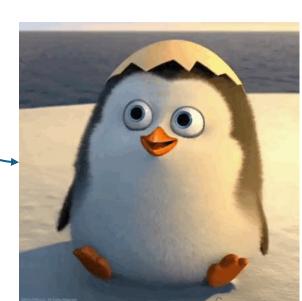


- ➤ Started 3 years ago
- > Events as graphics

Realized Independently

(Until Dec. 2023)





Context

2. The problem

Context 2. The problem

Neural Networks effort - State

- Lot of redundancy in the programs
- **Difficulties to exchange good features**. *Programs are incompatible*
- **Difficulties to compare**The metrics are differents, and the programs are metrics-dependant)

Not the best situtation when efficiency in time is required

Context

3. My task

Context 3. My task

To merge softwares so all these problems vanish

- Implement the necessary features (for both)

- **Simplify** the use of the software

- Keep the interesting features

- Cross check results

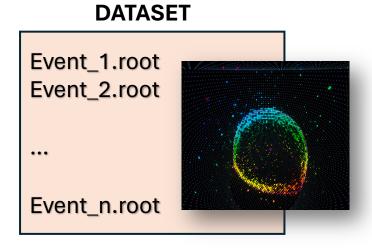
Started: December 2023

Let's go back to basics:

What is « Training a neural network »?

In 7 steps

The world of NN Step 1. Datasets



Neural networks use examples to be tuned

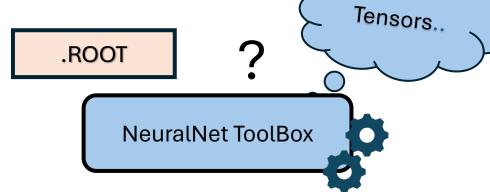
We need datasets

Step 1 : Create the datasets

Step 1. Datasets

	WatChMaL	GRANT
Detector supported	WCTE / IWCD	WCTE / HK
Format (raw data)	.root	.root
MC Data production	WCSIM (v. ?), Root 6	WCSim Nuprism, (Root 5)
Logs of the data	Yes	No
Storage	Local (disk)	Local (disk)
Apptainer image of the configuration (to simulate more)	Yes	No

Step 2: Conversion



15

Constraints from the software



we use:



Step 2: Convert our data to these Tensors

Step 2: Conversion

	Watchmal	GRANT
Root -> Numpy	pyroot	pyroot
Software to perform the conversion	DataTools	None
Convert to flat root?	No	Yes
Depends on WCSIM's version ?	Yes	No
ML library	Pytorch	Pytorch
Storage format	.h5 (easy to convert to .pt)	.pt (PytorchTensor)
Storage 04/06/2024	Local (disk)	Local (disk)

The world of NN Step 3: Loading

- Dataset is on the **disk**
- Fast computation is made on the memory

We want fast so we create subdataset that fits memory

→ Step 3 : Loading

Step 3: Loading

	Watchmal	GRANT
Data as	Images	Graphes only
Disk to Memory	Lazy Loading	Full loading
GPU support	Single and multi GPUs	Single GPU
GPU VRAM issue	No	Yes
Dynamic data transformation at loading	Yes (but dataset- dependant)	No

Step 4: The model

Let's create the model!

Step 4: The model

Step 4: The model

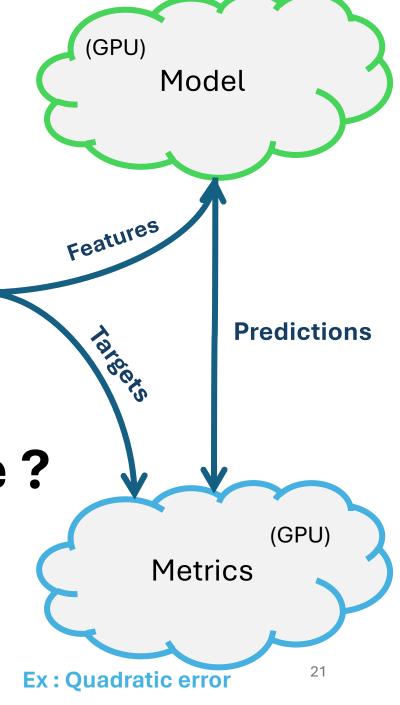
	Watchmal	GRANT
Arbitrary choice of model (no modification of the source code required)	Yes	No (designed for 1 model)
GPU support	Single and Multi	Single
Kind	Convolutional Neural Network (CNN)	Graph Neural Networks (GNN)

Step 5: Metrics

Dataset (Tensors) on the disk

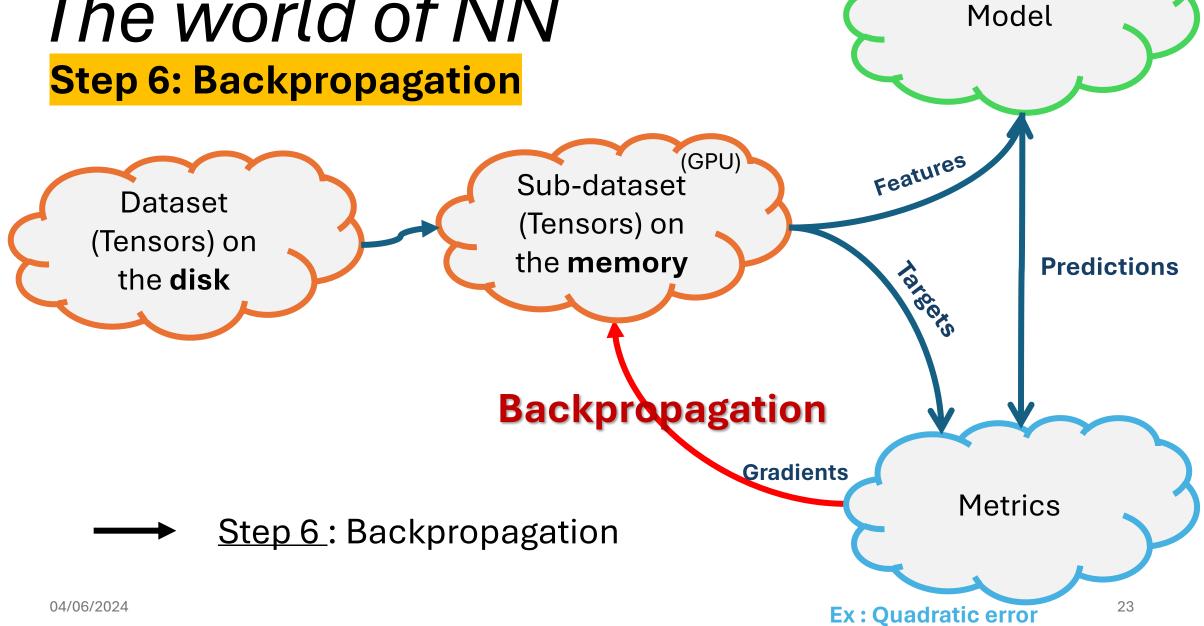
How to compare ?

→ Step 5 : Metrics



Step 5: Metrics

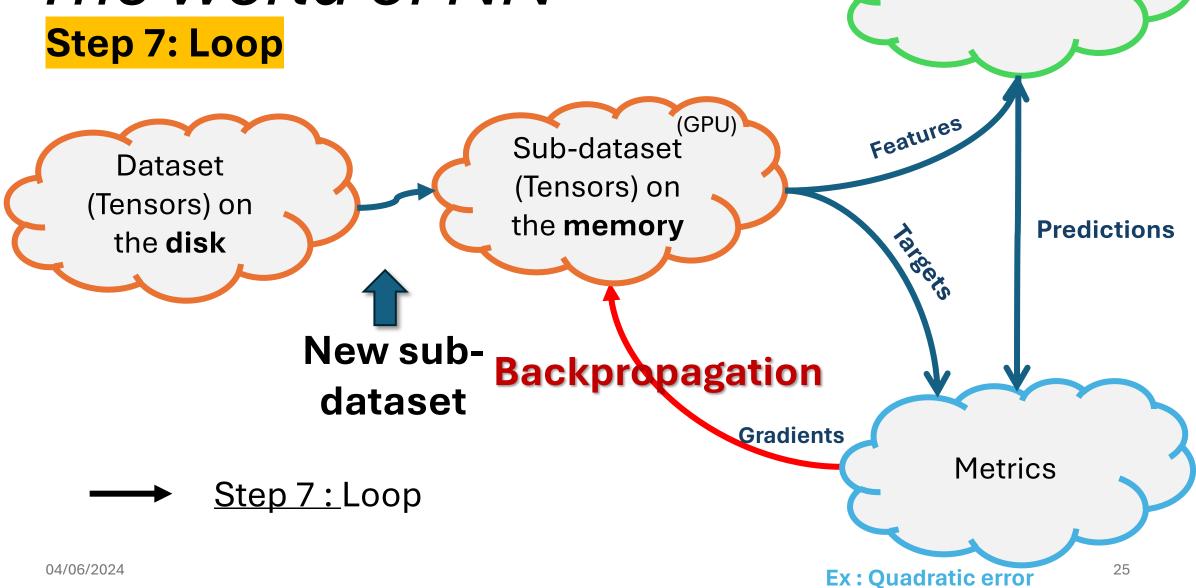
	Watchmal	GRANT
Arbitrary choice of metrics (no modification of the source code required)	Yes	No (3 possibles)
GPU support	Single and Multi	Single



(GPU)

Step 6: Backpropagation

	Watchmal	GRANT
Arbirary choice of optimizer	Yes	No
Scheduler support	Yes	No
Callbacks support	No	Yes
Model checkpointing	Yes	Yes



(GPU)

Model

Additional features

wandb: Weight and Biaises

	Watchmal	GRANT
Run manager	Yes (100% Hydra)	Yes (100% wandb)
Run outputs logging	Local (Disk)	Local (Disk) + Remote (Cloud)
Automatic hyperparameters search	None	Wandb Sweep

Crucial for Neural Network optimization

04/06/2024 26

Results

Status of the merge

		WatChMaL	GRANT	New
_	Detector	WCTE / IWCD	WCTE / HK	WCTE / IWCD / SK / HK
	MC data production	WCSIM v. ?, (Root 6)	WCSim Nuprism, (Root 5)	WCSim 12.2 – 12.12 (Root 6.24)
	MC data logs	Local	None	Local (disk) + Remote

Step 1

			Watchmal	GRANT	New
	_	Detector	WCTE / IWCD	WCTE / HK	WCTE / IWCD / SK / HK
Step 1		MC data production	WCSIM v. ?, (Root 6)	WCSim Nuprism, (Root 5)	WCSim 12.2 – 12.12 (Root 6.24)
Step 2 Ste		MC data logs	Local	None	Local (disk) + Remote
		Root -> Npy	Pyroot	Pyroot	Uproot
Ste		Storage format	.h5	.pt	Both supported

		Watchmal	GRANT	New
١	Detector	WCTE / IWCD	WCTE / HK	WCTE / IWCD / SK / HK
Step 1	MC data production	WCSIM v. ?, (Root 6)	WCSim Nuprism, (Root 5)	WCSim 12.2 – 12.12 (Root 6.24)
Ste	MC data logs	Local	None	Local (disk) + Remote
Step 2	Root -> Npy	Pyroot	Pyroot	Uproot
Į	Storage format	.h5	.pt	Both supported
Step 3-7	GPU support	Multi	Single	Multi
Ste	Model kind	CNN	GNN	CNN + GNN

			WatChMaL	GRANT	New
	Γ	Detector	WCTE / IWCD	WCTE / HK	WCTE / IWCD / SK / HK
Step 1		MC data production	WCSIM v. ?, (Root 6)	WCSim Nuprism, (Root 5)	WCSim 12.2 – 12.12 (Root 6.24)
Š		MC data logs	Local	None	Local (disk) + Remote
Step 2		Root -> Npy	Pyroot	Pyroot	Uproot
Ste		Storage format	.h5	.pt	Both supported
Step 3-7		GPU support	Multi	Single	Multi
Ste		Model kind	CNN	GNN	CNN + GNN
		Run manager	Hydra	Wandb	Hydra + wandb
lagers 		Run logs	Local	Local + Remote	20% Local – 80 % Remote (both enhanced)
Mana	0	Hyperparamet 4er2Search	None	Sweep	Sweep - Multi GPU support 31

Next Steps Future work

- New features to be added
 Interesting features not implemented in any software
- **Result checks**Finishing to reproduce both CNNs and GNNs results on all reconstruction
- Complete documentation

Next HK CM: How to use BootCamp

Backup

Hey!

- Engineer in Data Science

- Research Engineer at ILANCE - LLR

- PhD position beginning (ILANCE – LLR)

Educational background:

- Applied mathematics
- Computer science

Graduated in September 2023

October 2023 – October 2024

October 2024



Contexte

Photo de SK, IceCude?

1. The situation

How do we do that?

Use experience: adapt already existing tools

ApFit, FitQunn..

Push limits: invest in innovative-driven tools

Contexte

1. The situation

Figure avec FitQun écrit dessus

First choice has been invested for a while.

FitQun HKFD tuning: B. Quilain - LLR

Very good global performances

PID classification, energy or vertex reconstruction

→ Still presents limits

Contexte

1. The situation

FitQun's n°1 drawbacks: Time

Can take 1 to > 10 minutes
to process one event

This is too long for us.

Deadline is 2027

