

Journée Projets au LPNHE



Développements IA au LPNHE

Biennale LPNHE



Anja, Vava, Nabil

Plan

Activités LHCb / RTA

Activités ATLAS

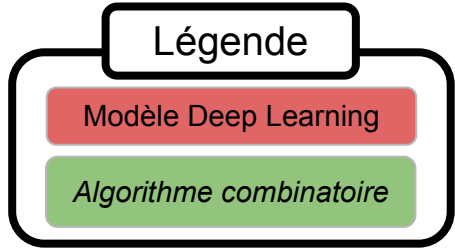
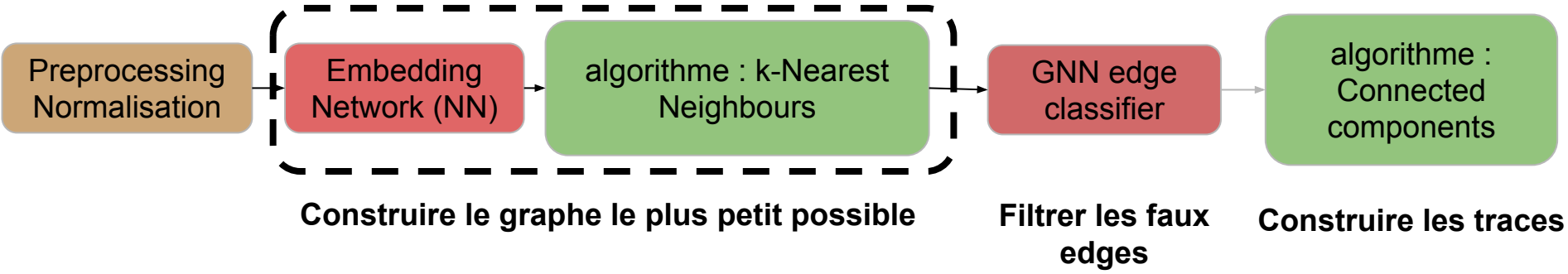
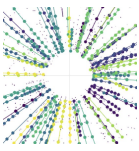
Activités des autres groupes LPNHE basées sur l'IA et le Machine Learning

Interactions possibles :

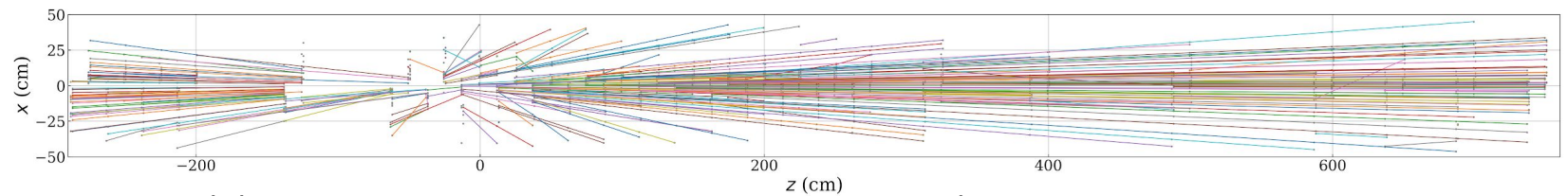
Mise en commun des efforts : déploiement du framework Allen vers d'autres types de reconstruction

ETX4VELO: Pipeline Python

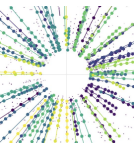
Vava, Anthony, Fotis, Nabil,
Bertrand Granado (LIP6)



Pour l'inférence :
Les modèles de deep learning doivent être exporté vers Allen.
La pipeline associe le deep learning et la combinatoire (intrinsèque à la construction du graphe)

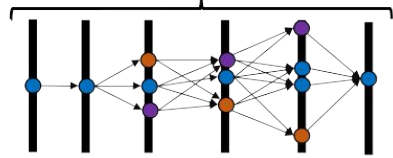


ETX4VELO: Inference Pipeline (Python) : Graphe-Edges

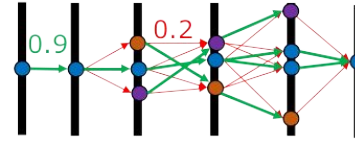


Au début : Graphe basé sur les edges (connexions entre hits)

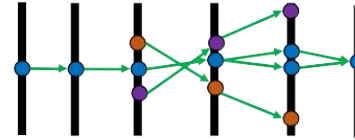
Output of Embedding + kNN



GNN edge classifier
score for every edge



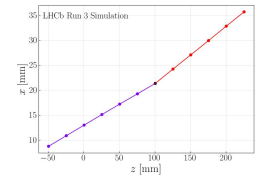
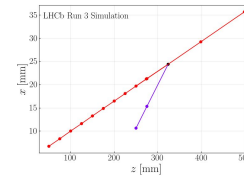
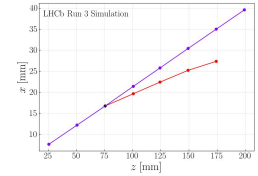
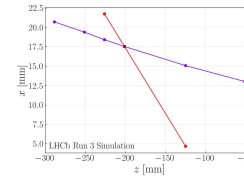
Edge score cut



Metric	Allen	etx4velo
Efficiency	98.17%	46.23%
Clone rate	3.07%	0.47%
Hit efficiency	95.35%	98.89%
Hit purity	99.67%	93.89%

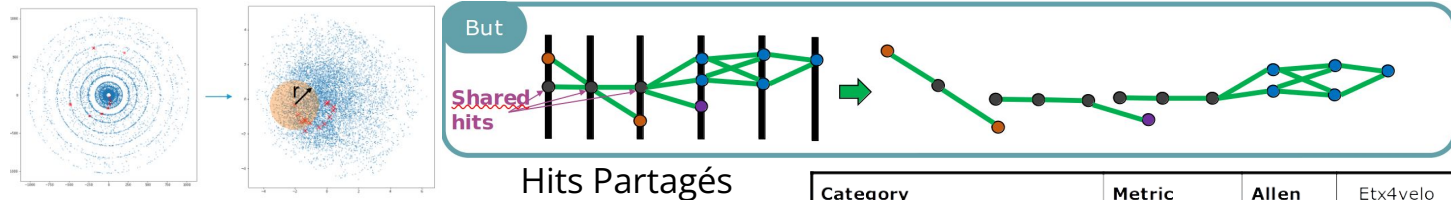
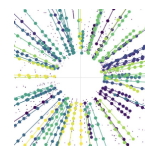
Évalué sur 5000 événements

La reconstruction pour les électrons est mauvaise : les hits en commun sont difficiles à classer



~ 55 % des électrons partagent des hits avec un autre électron

ETX4VELO: Pipeline python GNN-triplets



Maintenant : GNN basé sur les triplets :
 efficacité supérieure à la combinatoire

Entraînement : 700 000 events (~4 jours)

Évalué sur 5000 événements

MLP : 31 kilo paramètres

GNN : 7,1 million paramètres

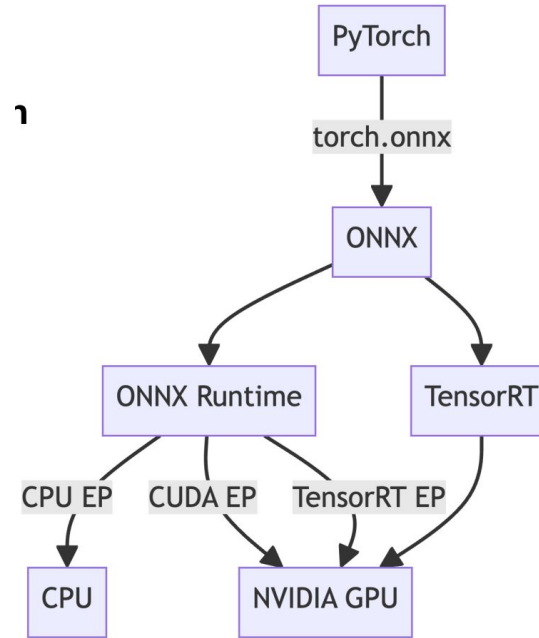
Proceedings of the CTD 2023
 PROC-CTD2023-XX October 26, 2023
 Graph Neural Network-based pipeline for track finding
 in the Velo at LHCb
 Talks : CDT2023 / JRJC 2023/ WP2 LHCb 2023

Category	Metric	Allen	Etx4velo
Long, no electrons ✓ In acceptance ✓ Reconstructible in the velo ✓ Reconstructible in the SciFi ✓ Not an electron	Efficiency	99.26%	99.28%
	Clone rate	2.54%	0.96%
	Hit efficiency	96.46%	98.73%
	Hit Purity	99.78%	99.94%
Long electrons ✓ In acceptance ✓ Reconstructible in the velo ✓ Reconstructible in the SciFi ✓ Electron	Efficiency	97.11%	98.80%
	Clone rate	4,25%	7.42%
	Hit efficiency	95.24%	96.54%
	Hit purity	97.11%	98.46%
Long, from strange ✓ In acceptance ✓ Reconstructible in the velo ✓ Decays from a strange <i>Good proxy for displaced tracks</i>	Efficiency	97.69%	97.50%
	Clone rate	2.50%	0.92%
	Hit efficiency	97.69%	98.22%
	Hit purity	99.34%	99.68%
X	Ghost rate	2.18%	0.76%

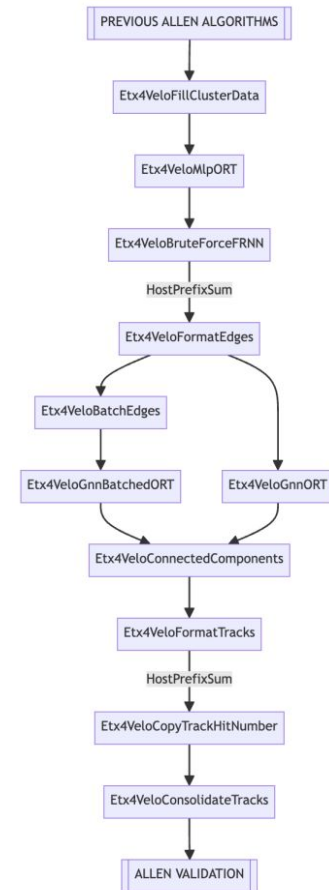
ETX4VELO: Pipeline C++/ CUDA

Problématique :

1. Intégrer la pipeline python dans Allen en C++/CUDA
2. Optimiser la taille des hyperparamètres pour accélérer l'inférence
3. Proposer des solutions open sources mais aussi optimisées pour les GPU nvidia : ONNX Runtime et TensorRT



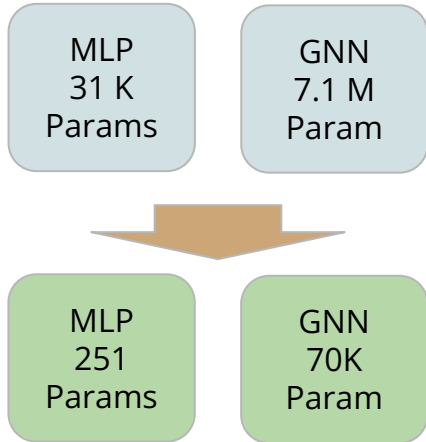
Graphe d'intégration
ONNX Runtime / TensorRT



Séquence
d'algorithmes
dans Allen 6

ETX4VELO: Optimisation de la pipeline Cuda

Réduction du nombre de paramètres pour augmenter la vitesse d'inférence sans perdre en efficacité



NVIDIA GeForce RTX 2080Ti. node n4050101. 500 events

Pipeline	Step	Throughput		
		ONNX Runtime (FP32)	TensorRT (FP32)	TensorRT (INT8)
ETX4VELO	MLP	50K events/s	260K events/s	520K events/s
	GNN	0.31K events/s	0.86K events/s	-
	Velo tracks	-	0.83K events/s	-
ALLEN	Velo tracks	510K events/s		

Encore des optimisations possibles:
streams, float 16, pruning, quantization

https://gitlab.cern.ch/gdl4hep/etx4velo_cuda

Présenté prochainement en séminaire Machine Learning LPNHE

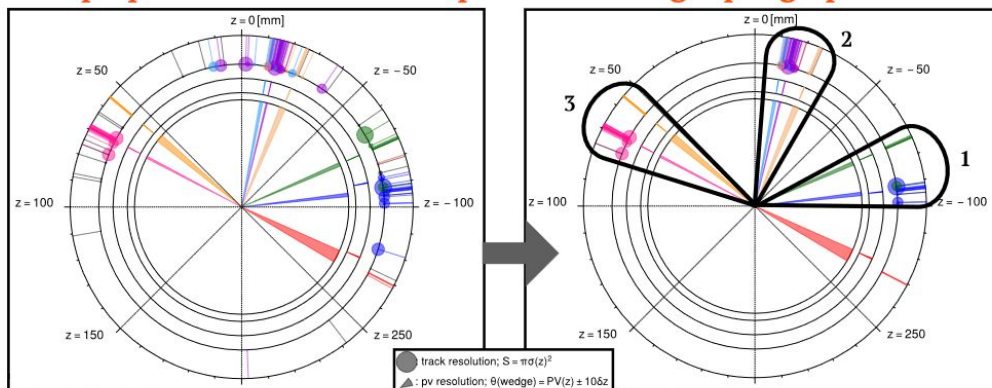
Please Visit Fotis' Poster at the Biennale

Talk : WP2
LHCb 2024

Adaptation de la Pipeline pour la reconstruction des PVs

Graph building for GNN implementation

Data preparation is a crucial step when building input graphs



1. **Ordering tracks** in z values along the beamline.
2. **Filtering tracks** with simple isolation criteria ($|\Delta z_{i-1}|$ or $|\Delta z_{i+1}|$) < 0.5 mm
3. **Forming groups** based on gaps, $d_{ij} > 20$ mm, between consecutive tracks
4. **Build graphs** from groups with **edges connecting nodes (i.e. tracks)** if $d_{ij} < 5$ mm

Exemple d'adaptation de la pipeline par Simon Akar, pour un autre problème, la reconstruction des PVs :

https://indico.nikhef.nl/event/4875/contributions/20317/attachments/8200/11681/AkarS_EuCAIFCon24.pdf

https://indico.nikhef.nl/event/4875/contributions/20317/attachments/8200/11960/PVFinder_EuCAIFCon24.pdf

Summary:

- ▶ **GNN models appear quite versatile** where similar models achieve good performances for different tasks (tracking vs PV finding)
- ▶ **GNN and hybrid models achieve similar intrinsic physics performances...**
- ▶ ...but only partial overlap meaning both models did not learn exactly the same relations from identical input data!

ATLAS

4 Axes principaux:

- Génération d'événements et simulations de détecteurs
- Calibration des objets ou de paires d'objets
- Reconstruction du temps au niveau d'un cluster
- Analyse des données (discrimination signal/background)

Jet calibration (Laura's QT)

Improvement of **electron identification** and photon triggers with ATLAS Egamma Trigger (Edmar, Bertrand)

Development of ML techniques for time reconstruction (XTalk mitigation) in LAr calorimeter (Marton, Mateus, Bertrand)

Identification of prompt **boosted photon** pairs for low mass diphoton resonance search (Romain, José)

Identification of **late diphoton pairs** for Long Lived Axion Like Particles search (Artur, Bertrand)

Identification of W bosons based on Lund Jet Plane (Reina)

Outside ATLAS

From Uncertainties to measurements : integration of heteroscedastic uncertainties in ML algorithms (training, inference)

Development of new methods for **unfolding** (proof of concept) (Anja, Bogdan) -> to be applied in ATLAS

Generative models for event generation (Anja)

Generative models for detector simulation (Bertrand, Edmar, Bogdan, Anja, project with école polytechnique students, Rio and LPTHE)

Chercheur.euses

- Anja Butter
- Reina Coromoto Camacho Toro
- Bertrand Laforge
- Bogdan Malaescu
- Edmar Egidio Purcino de Souza
- José Ocariz

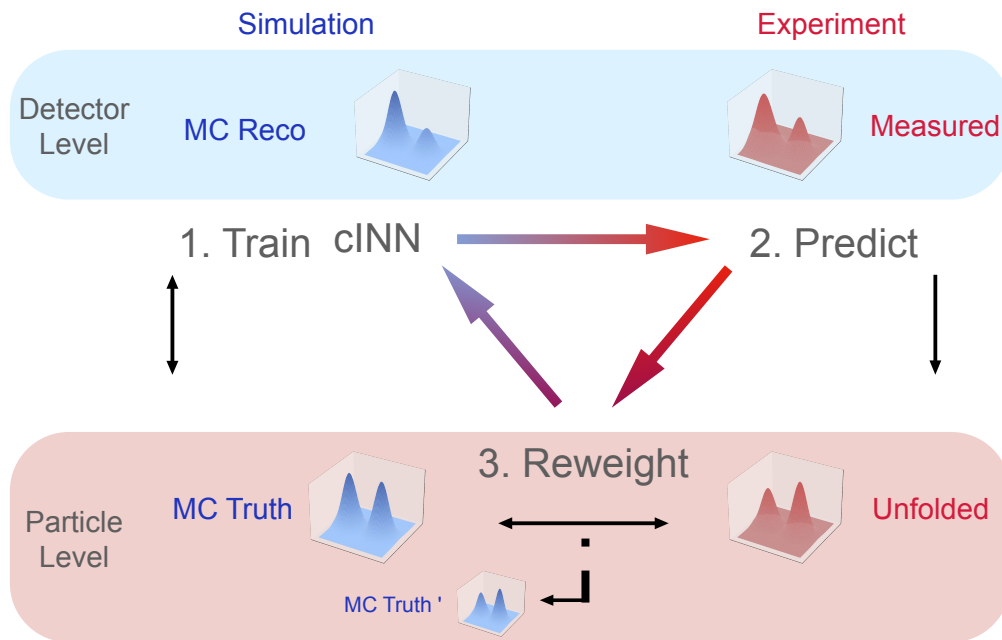
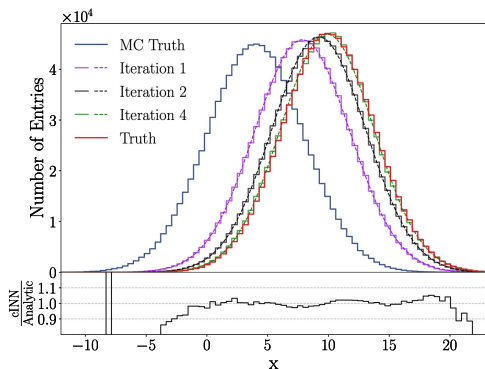
Étudiant.es

- Artur Cordeiro Oudot Choi
- Romain Van der Brouck
- Laura Boggia
- Mateus, Marton (Brazil)

ATLAS, Anja, Bogdan : Conditional iterative unfolding

Travail en cours

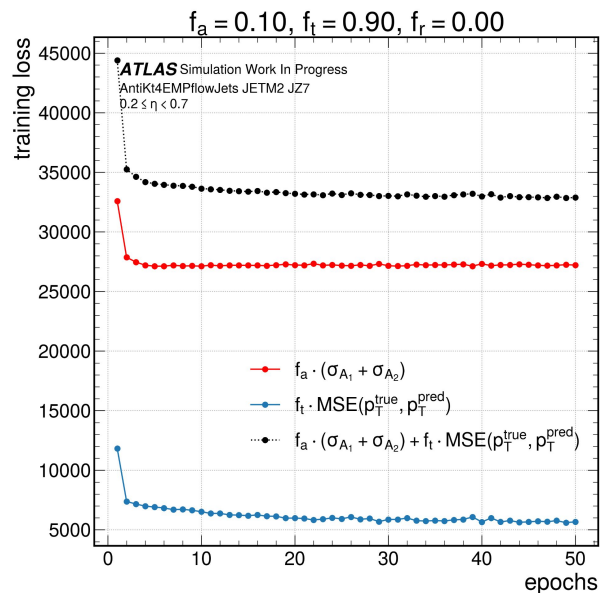
Unfolding basé sur le deep learning avec un cINN (conditional invertible NN)



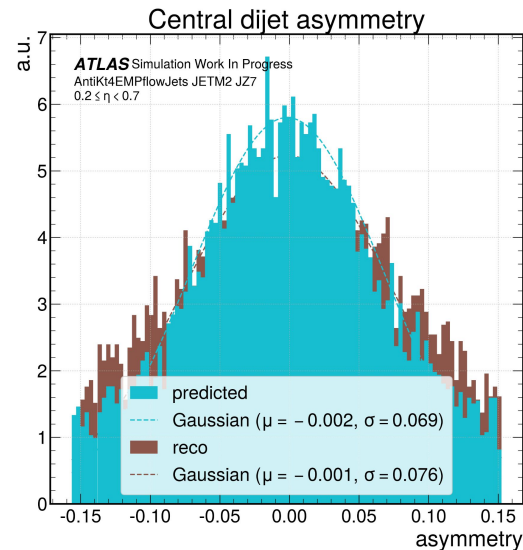
Cf : LPNHE : Machine Learning Group seminar, dec 2023

ATLAS, Anja, Bogdan, Laura : Jet Energy Calibration

Quelques résultats récents sur la calibration des jets



fonction loss sur 50 epochs

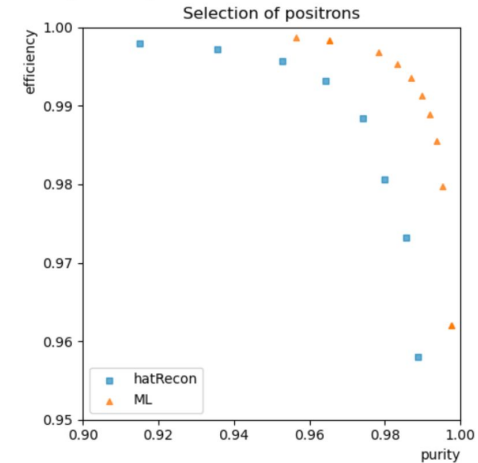
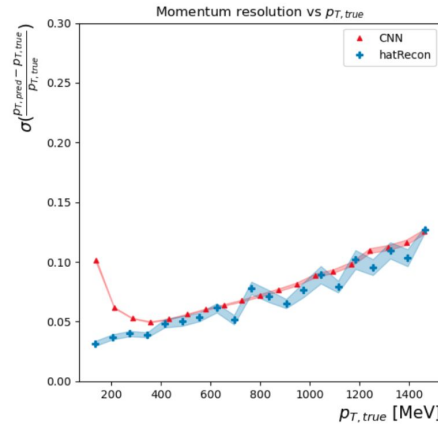
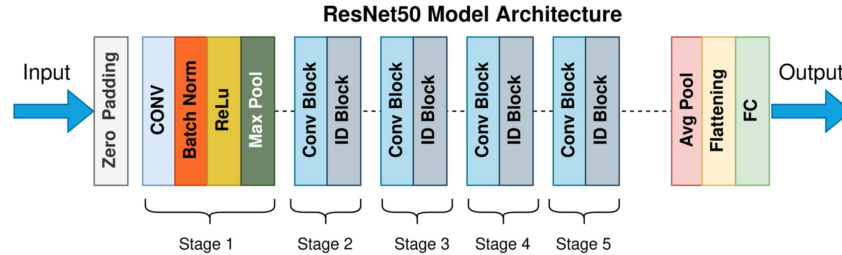


Asymétrie prédite par NN est
moins large que la reconstruction

Groupe Neutrinos : Reconstruction des traces dans les nouvelles TPC grand-angle (Anaëlle, Claudio, Mathieu, Marco...)

Utilisation d'un CNN (ResNet50) pour extraire la position, la pente et la courbure des traces

Les performances sont similaires pour l'impulsion mais le CNN semble mieux marcher pour le PID



Autres activités, à suivre

GRAND : Trigger basé sur le deep-learning, inférence sur FPGA avec TensorFlow lite. Cf le travail de Sandra Le Coz / Jean-Marc Colley

Xenon : Travail de Jean-Philippe Zopounidis sur le CNN pour la reconstruction des signaux dans les TPCs.

Conclusion

Activités variées et dynamiques: Il faut favoriser la **transversalité dans les activités IA** au **LPNHE** : Appliquer la technique des **GNN pour d'autres architectures**, il faut interagir avec les autres groupes du LPNHE aussi bien au niveau scientifique que technique : Nous avons de la matière à partager.

IN2P3 : AllenCore : Un des rares frameworks parallèles (jusqu'à 1000 événements) donnant la possibilité d'associer les algorithmes combinatoires et deep learning sur GPU.

Approfondir l'inférence sur d'autres architectures : **FPGA, ARM** (transversalité avec l'embarqué), **RISC V** (Vava)

Pérenniser et partager le savoir faire sur le développement des algorithmes de training.

Merci

Support

Séminaires IA à venir

Prochains séminaires Groupe Machine Learning:

Anthony Correia : Reconstruction de traces par GNN dans LHCb : training et inférence C++/Cuda : Done

Nabil Garroum : Qualité logicielle dans le Machine Learning : Tutoriel MLOps

A discuter : Utilisation du deep Learning dans le groupe Neutrinos

Atlas : publications depuis Octobre 2022

2022

[1] "Two Invertible Networks for the Matrix Element Method"
A. Butter, T. Heimel, T. Martini, S. Peitzsch, T. Plehn, Published in: SciPost Phys. 15 (2023) 094, [2210.00019]

[3] "MadNIS -- Neural Multi-Channel Importance Sampling"
T. Heimel, R. Winterhalder, A. Butter, J. Isaacson, C. Krause, Published in: SciPost Phys. 15 (2023) 141, [2212.06172]

[4] "An unfolding method based on conditional Invertible Neural Networks (cINN) using iterative training"
M. Backes, A. Butter, M. Dunford, B. Malaescu, [2212.08674], SciPost Phys. Core 7, 007 (2024)

[5] "Performance versus Resilience in Modern Quark-Gluon Tagging"
A. Butter, B. M. Dillon, T. Plehn, L. Vogel, [2212.10493], under review

2023

[6] "Searching for dark matter subhalos in the Fermi-LAT catalog with Bayesian neural networks"
A. Butter, M. Krämer, S. Manconi, K. Nippel, JCAP 07 (2023) 033, [2304.00032]

[7] "Jet Diffusion versus JetGPT -- Modern Networks for the LHC"
A. Butter, N. Huetsch, S. Palacios Schweitzer, T. Plehn, P. Sorrenson, J. Spinner, [2305.10475], under review

[9] "Precision-Machine Learning for the Matrix Element Method"
T. Heimel, N. Huetsch, R. Winterhalder, T. Plehn, A. Butter, [2310.07752], under review

[10] "Event-by-event Comparison between Machine-Learning- and Transfer-Matrix-based Unfolding Methods"
M. Backes, A. Butter, M. Dunford, B. Malaescu, [2310.17037], under review

Others

[2] "TF07 Snowmass Report: Theory of Collider Phenomena"
F. Maltoni, et al., Contribution to: Snowmass 2021, [2210.02591]

[8] "Artificial Intelligence for the Electron Ion Collider (AI4EIC)"
C. Allaire(IJCLab, Orsay), et al., Contribution to: AI4EIC, [2307.08593]

ML workshops and conferences

ML4Jets conference: Together with colleagues from LPNHE, LPTHE and IJCLab, 5 day conference (around 400 registered participants, 150 on-site) in November

Workshop on ML unfolding financed via France-Berkeley Fund, taking place in June

Contrib. to ECSQARU 2023
Z. Waysenson, M. Saita, B. Laforge, Propagation of uncertainties in deep neural networks, Sept. 2023