PUSHING THE BOUNDARIES OF NEUTRINO PHYSICS WITH DEEP LEARNING

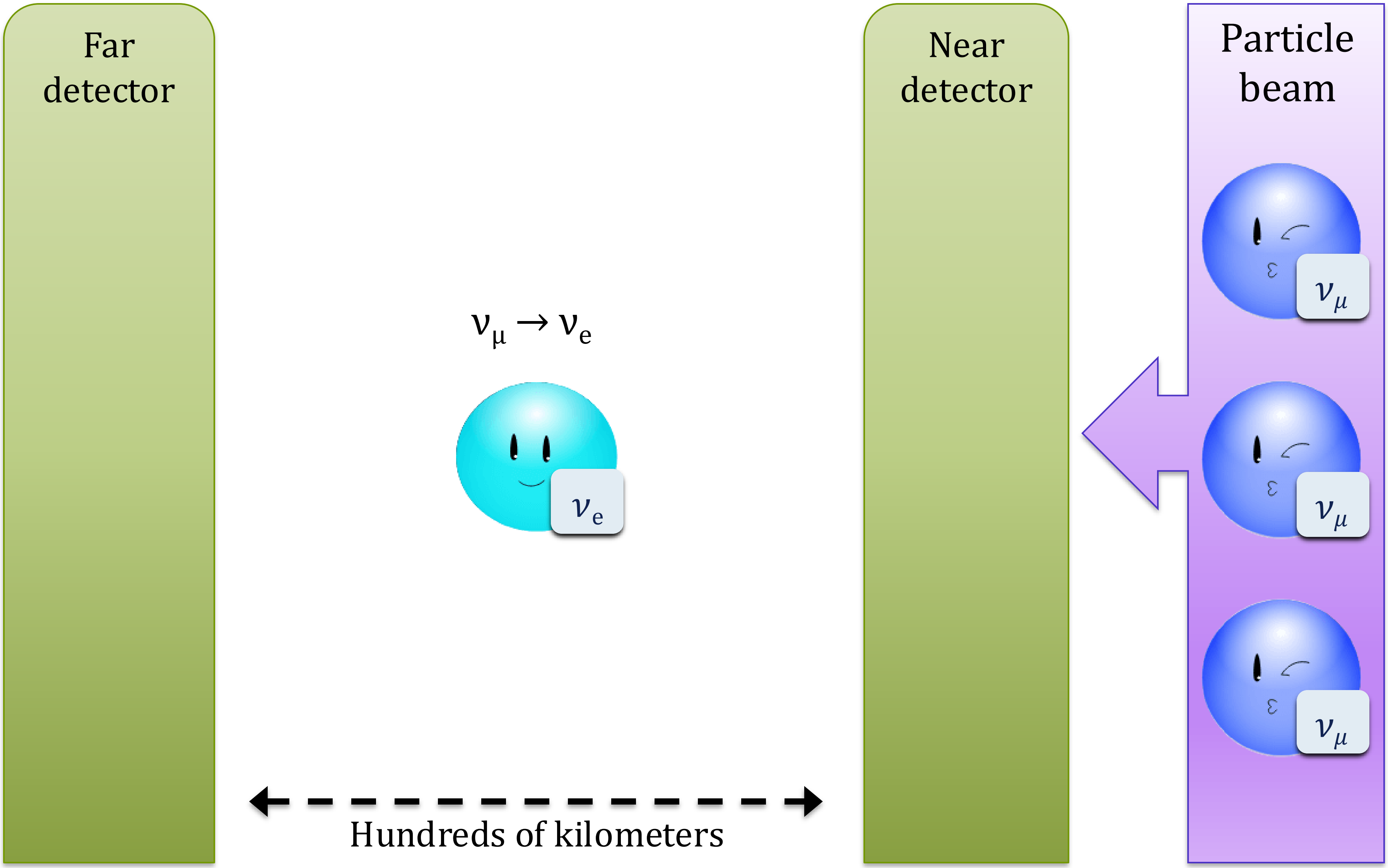
Dr. Saúl Alonso-Monsalve

ETH Zurich

Experimental Seminar, SLAC

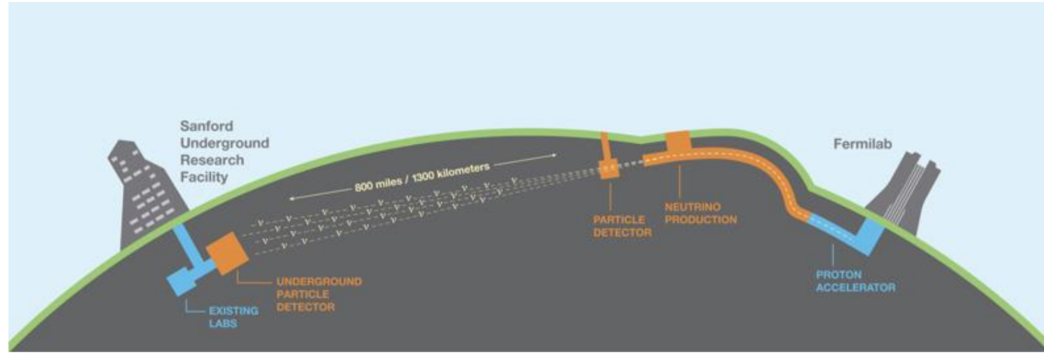
8 May 2025

Neutrino oscillation experiments



DUNE DEEP UNDERGROUND NEUTRINO EXPERIMENT

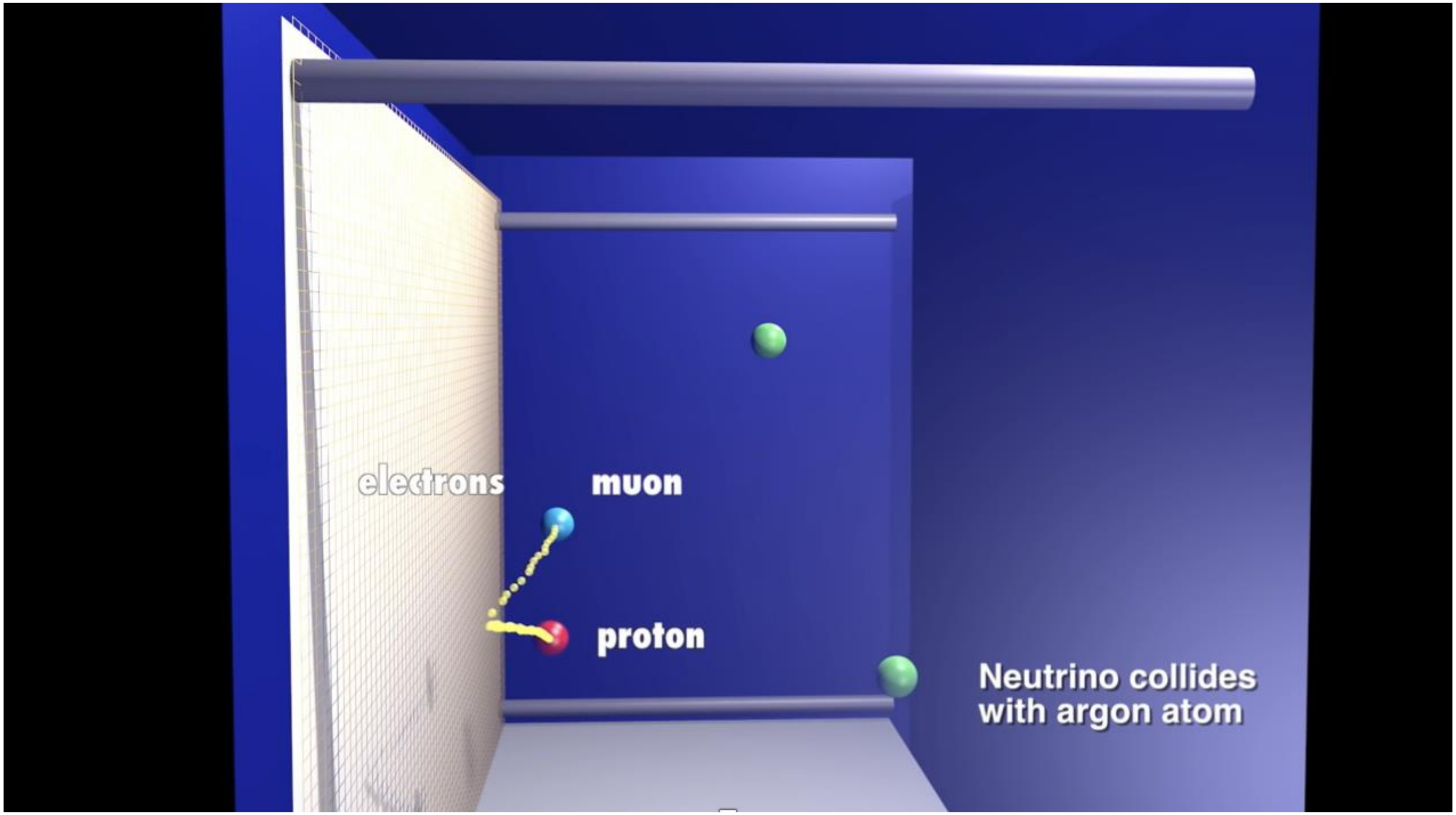
- International neutrino oscillation experiment in the USA.
- Far detectors located 1300 kilometres away from the source.
- Goal: measuring CP violation.



www.dunescience.org/

DUNE far detectors

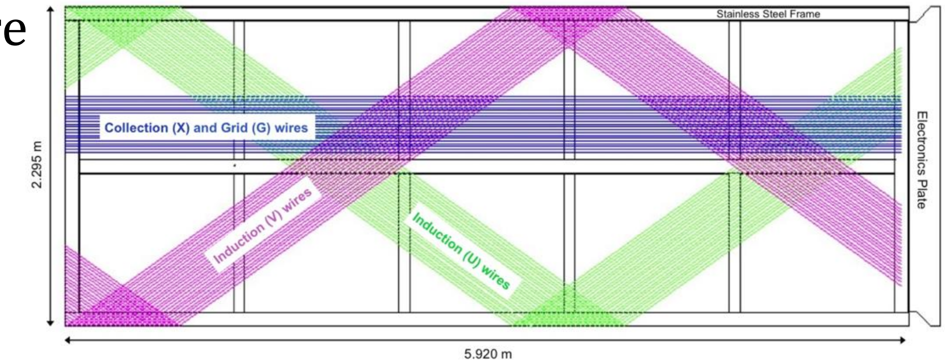
- Technology: Liquid Argon Time Projection Chamber (LArTPC).
  - Provides images of each neutrino interaction.



youtube.com/c/fermilab

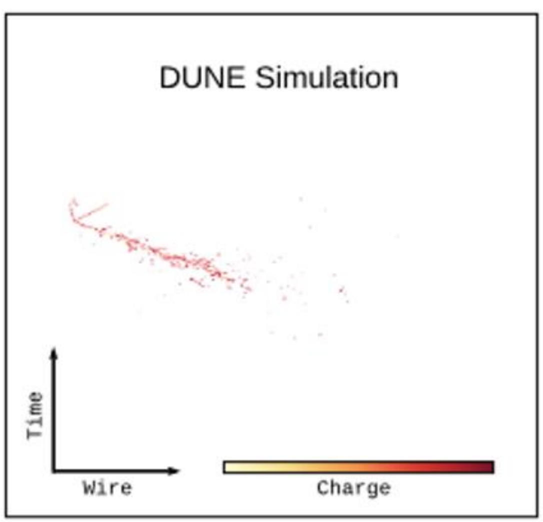
DUNE far detector data

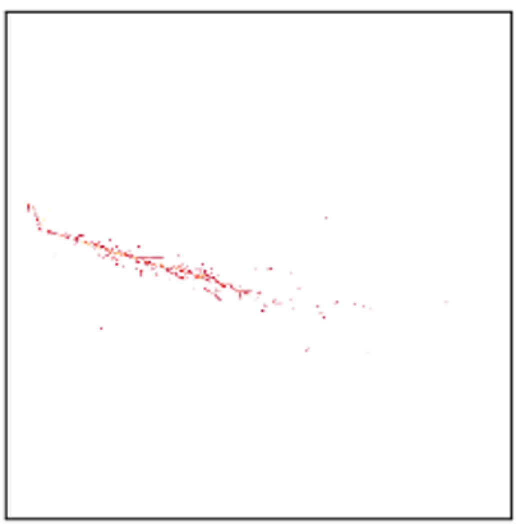
- The Far Detectors contain three wire readout planes.
  - This provides three “images” of each neutrino interaction (500x500 pixels each).

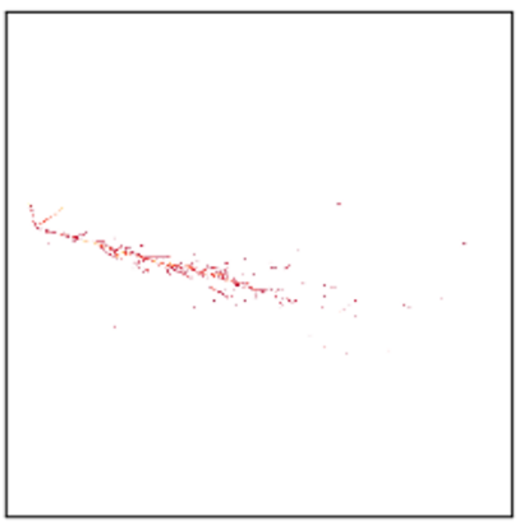


- Official simulated electron neutrino interaction:

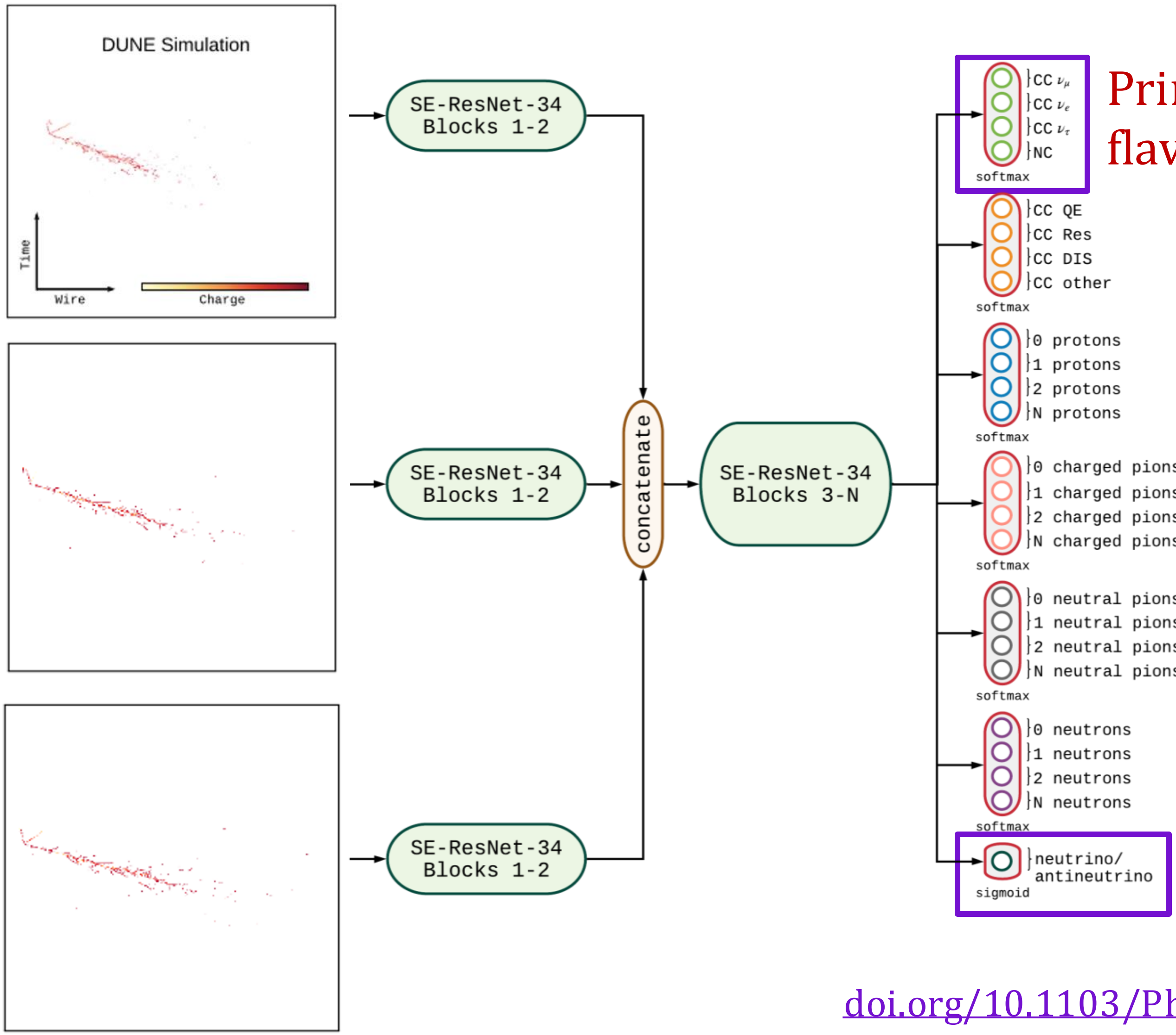
doi.org/10.1016/j.nima.2022.167217







Convolutional neural network in DUNE



Primary output: flavour identification

Secondary outputs: particle counting for exclusive final states

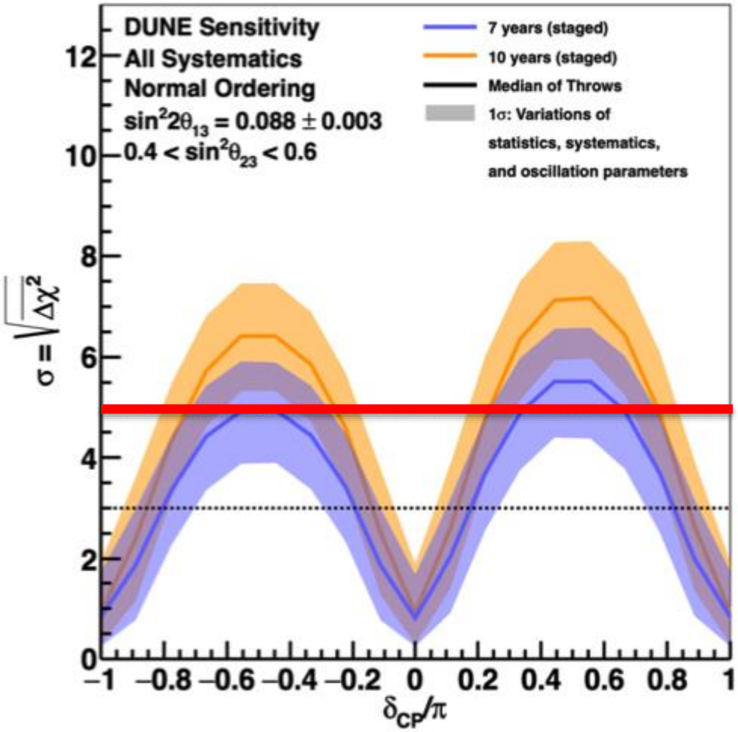
Neutrino / antineutrino

doi.org/10.1103/PhysRevD.102.092003

Results

- The flavour identification results are used in the official analysis of DUNE:
  - arXiv:2002.03005.
  - https://doi.org/EPJC/S10052-020-08456-Z.

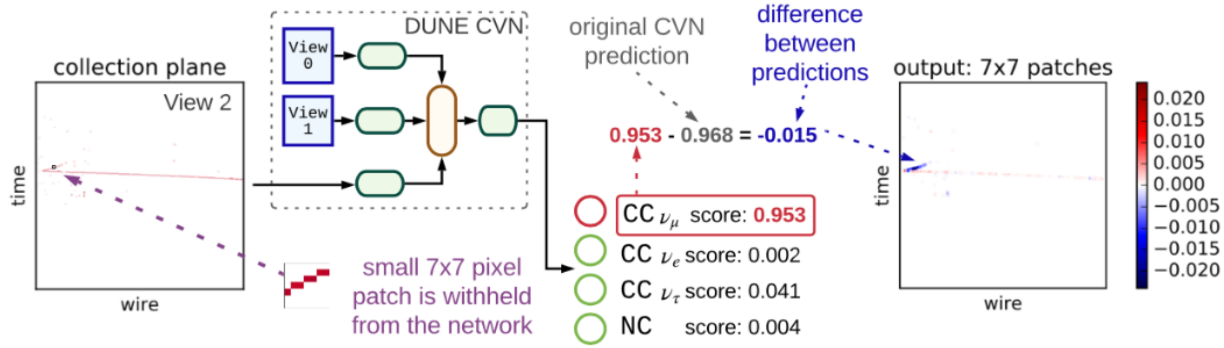
- Sensitivity to CP violation:



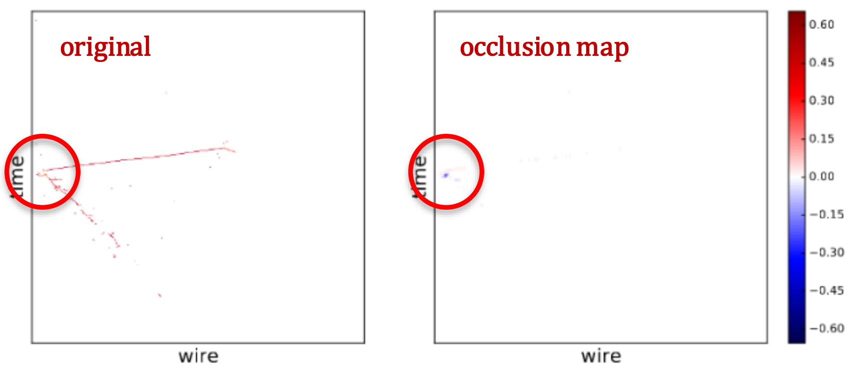
- 5σ after seven years of data collection.
- A milestone for the experiment!
- Made possible thanks to the convolutional neural network.

Understanding the CNN

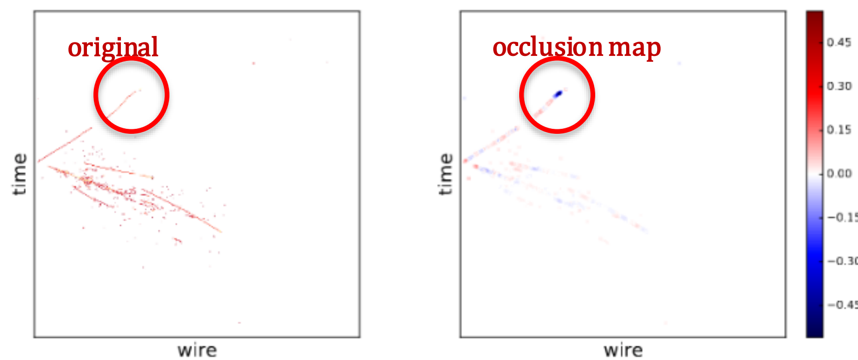
- Occlusion tests:
  - Hide parts of the images and check how the CNN reacts to the changes.



electron neutrino (ν_e)



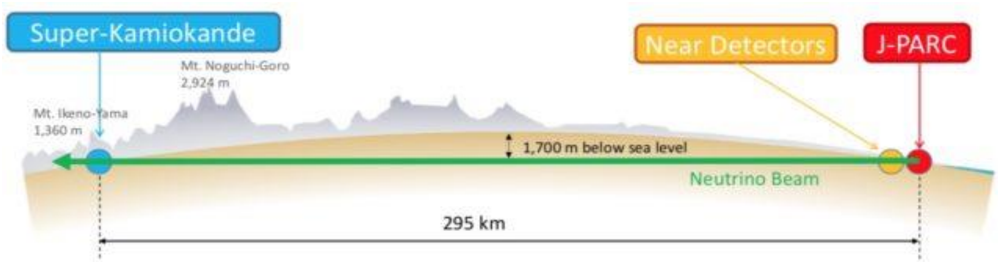
muon neutrino (ν_μ)



Removing the start of the electron shower reduces the ν_e score, as expected

The CNN finds the vertex a bit ambiguous, but it is using the end point of the muon to gain a handle on the event type.

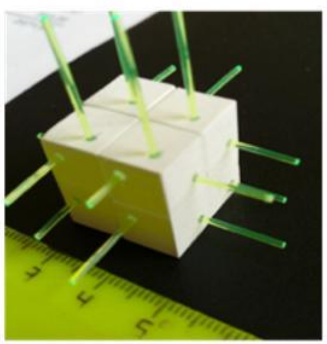
- T2K (Tokai to Kamioka) is an international neutrino oscillation experiment in Japan.



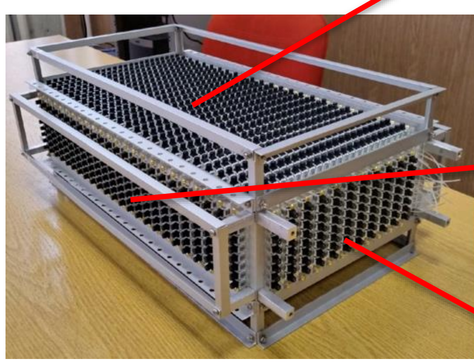
More than 500 collaborators from 80 institutions in 12 countries.

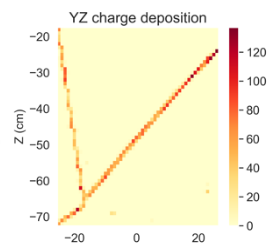
- SuperFGD: main upgrade to the near detector.

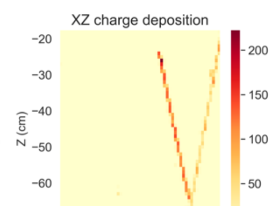
2D projections

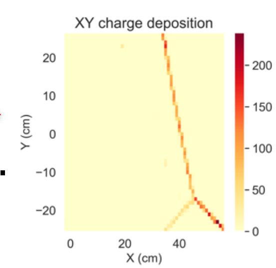


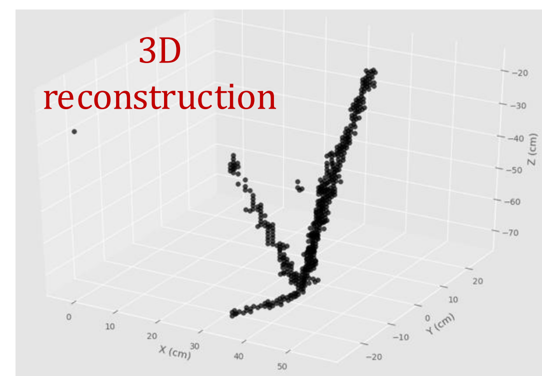
JINST 13 (2018) 02, P02006
NIM A936 (2019) 136-138





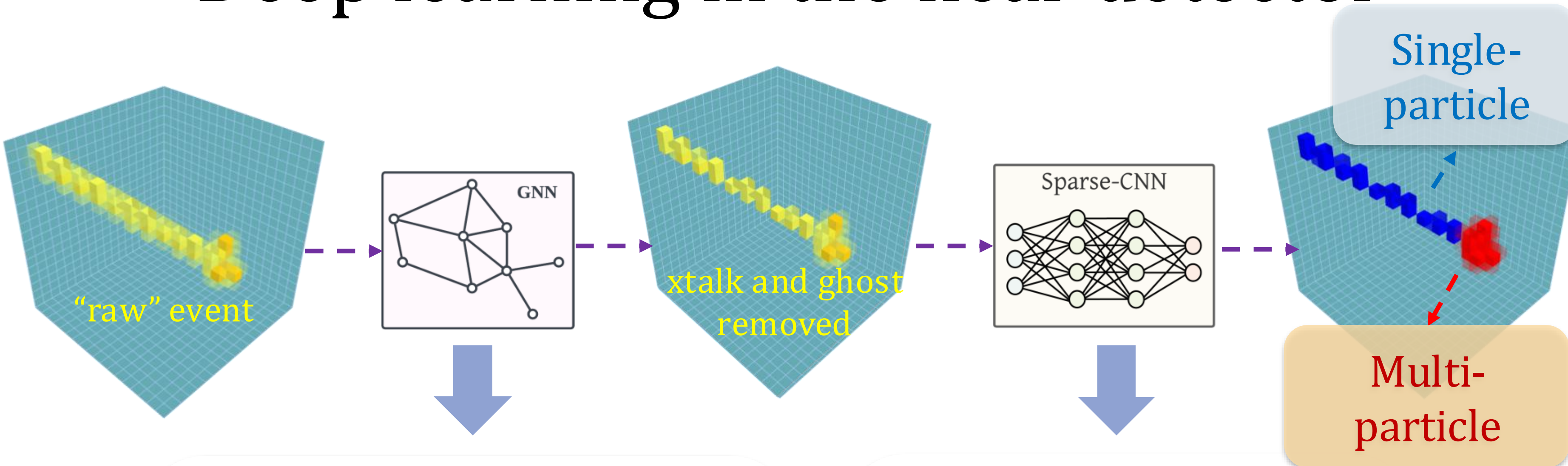


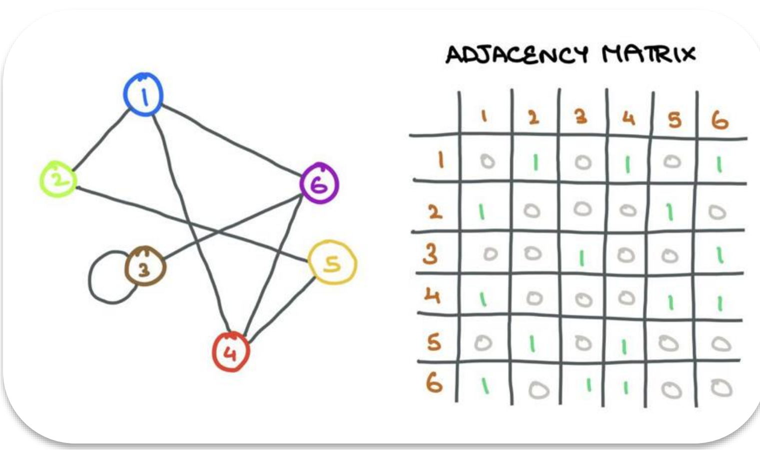


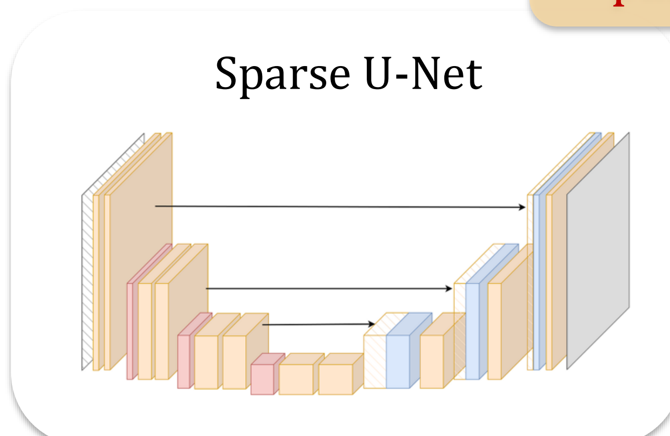


- 2M optically independent cubes, 1 cm^3 per cube.
- Spatial localisation of scintillation light.

Deep learning in the near detector





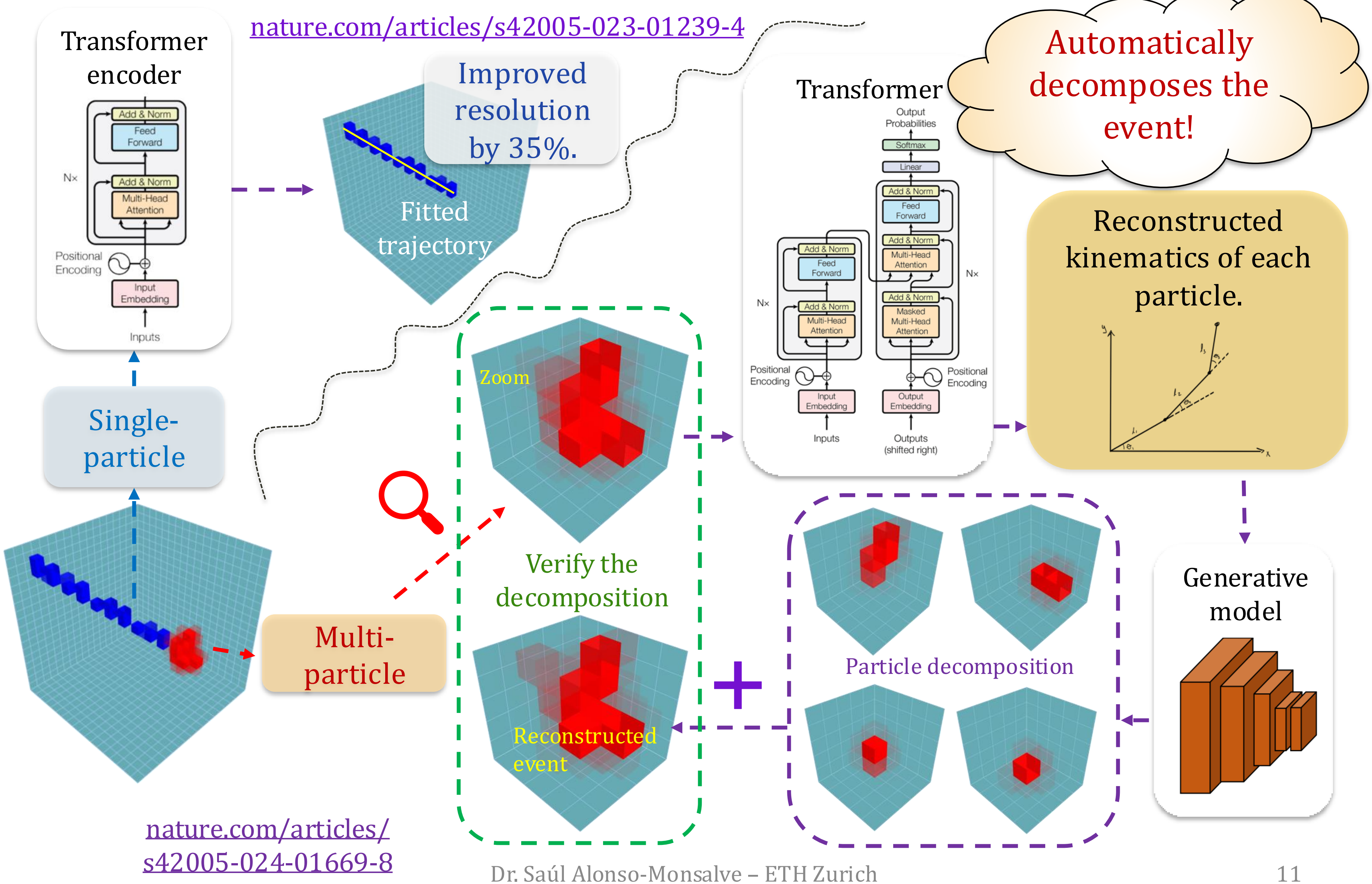


90% of the noise is eliminated without harming the signal.

doi/10.1103/PhysRevD.103.032005

~95% accuracy!

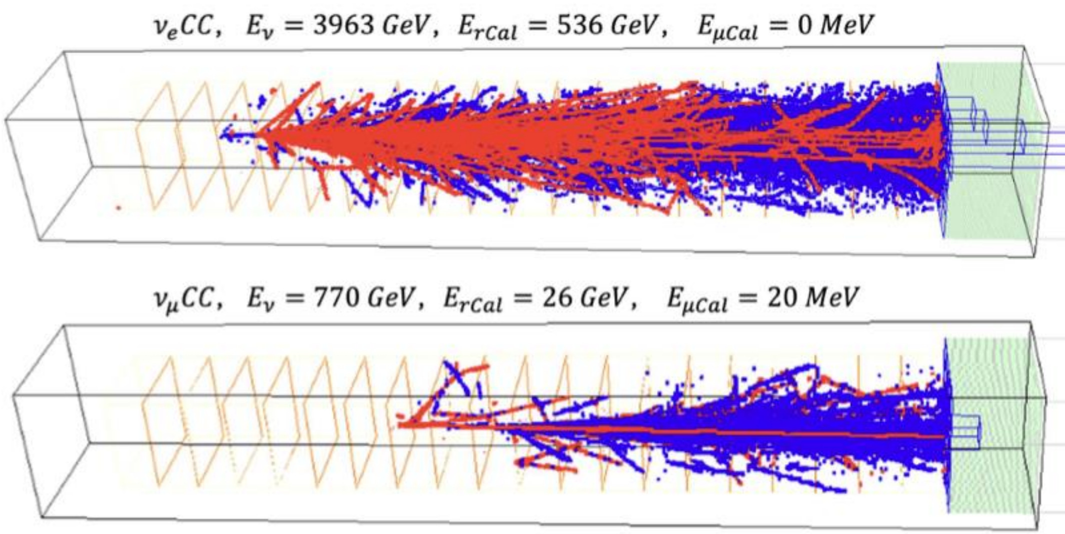
Deep learning in the near detector



- Experiment at CERN's Large Hadron Collider designed to search for light, weakly interacting particles and study high-energy neutrinos produced in proton-proton collisions.

- FASERCal is a proposed upgrade for FASER.
  - 3D precision calorimeter for high-energy neutrinos.
  - Inspired by the SuperFGD detector at T2K using plastic scintillation cubes.
  - Goals: detecting CC ν_e and CC ν_μ interactions, measuring differential cross-sections and identifying charm and tau events.
  - ETH Zürich is leading its design.

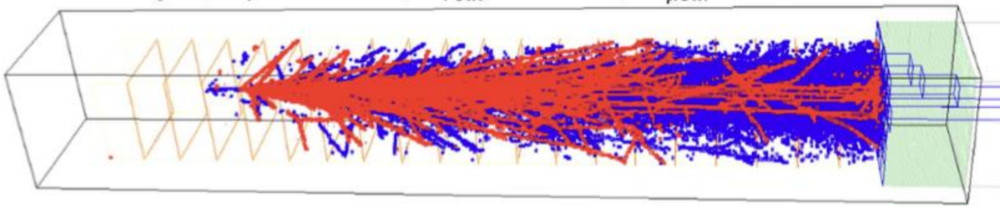
Simulated neutrino interactions in FASERCal



github.com/rubbiaa/FASER

Deep learning in FASERCal

- Neutrinos with energies up to the TeV scale!
  - Around 100K simulated events.
  - Single events with ~60,000 hits!



- Goals:
  - Stage 1: categorising each hit as (primary lepton / electromagnetic / hadronic / ghost).
  - Stage 2: flavour identification + regression tasks (energy, missing transverse momentum, jet momentum, primary lepton momentum).

- Both architectures use ConvNextV2 blocks (arxiv:2301.00808).
  - Implementation using MinkowskiEngine.
  - Depth-wise convolutions.
  - The model from stage 2 uses pretrained weights from stage 1.

Results (I)

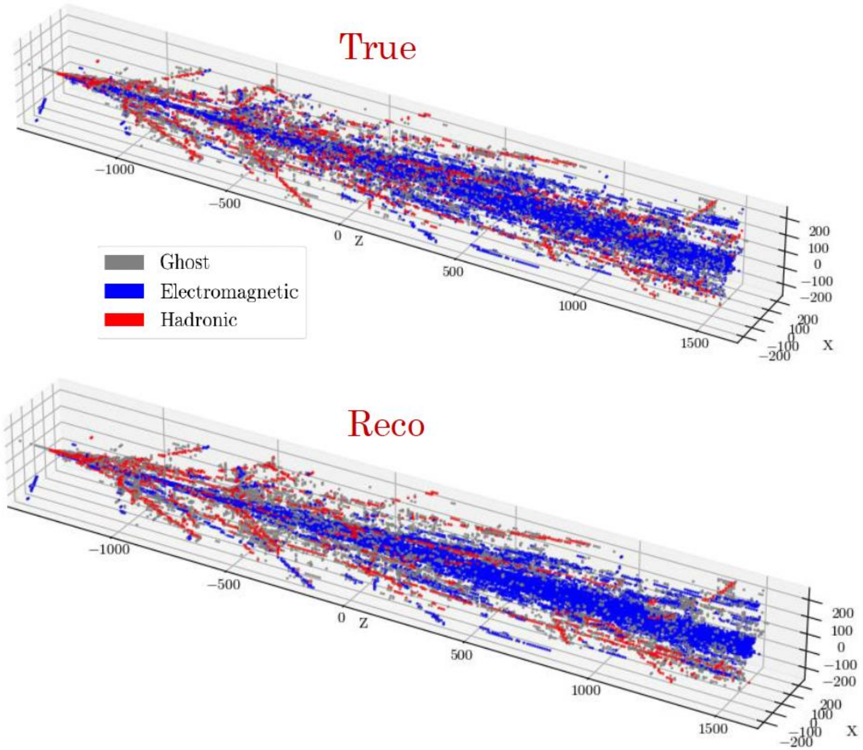
Voxel Level Results - I

Ghost/electromagnetic/hadronic:

	True ghost	True electromag.	True hadronic
Pred. ghost	**94,546,869**	23,683,123	9,399,787
Pred. electromagn.	17,920,388	**97,211,544**	12,555,489
Pred. hadronic	4,093,591	5,796,918	**29,519,436**

Purity:
- Ghost: 74%.
- Electromagnetic: 76%.
- Hadronic: 75%.

Efficiency:
- Ghost: 81%.
- Electromagnetic: 77%.
- Hadronic: 57%.



Results (II)

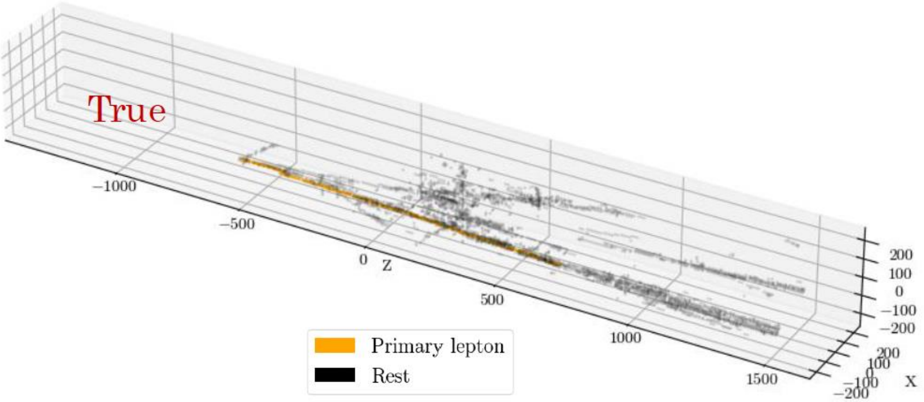
Voxel Level Results - II

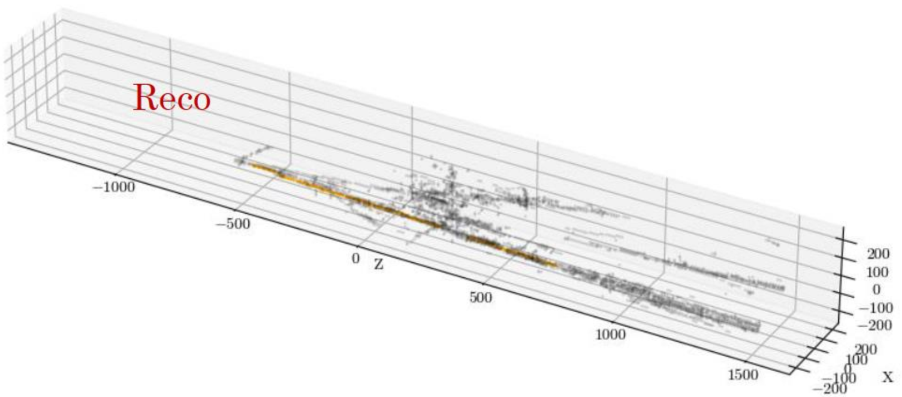
Primary lepton ID:

Overall accuracy: 99%	True primary lepton voxel	True rest
Pred. primary lepton voxel	**2,802,111**	435,835
Pred. rest	1,061,313	**290,427,886**

Purity:
- Prim. lepton: 87%.
- Rest: 100%.

Efficiency:
- Prim. lepton: 73%.
- Rest: 100%.

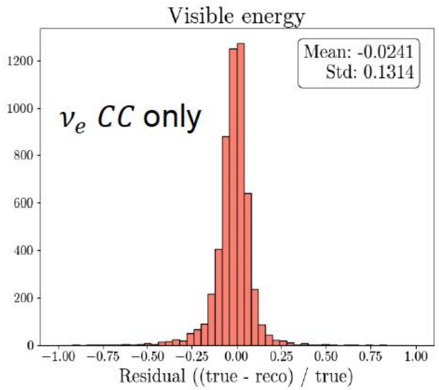




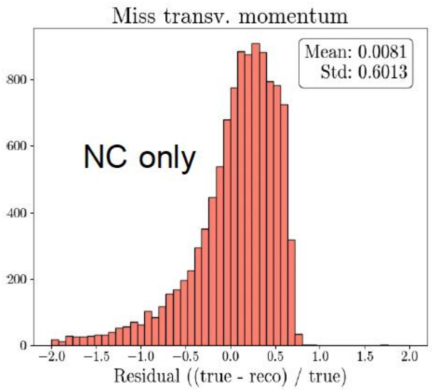
Results (III)

Event Level Results

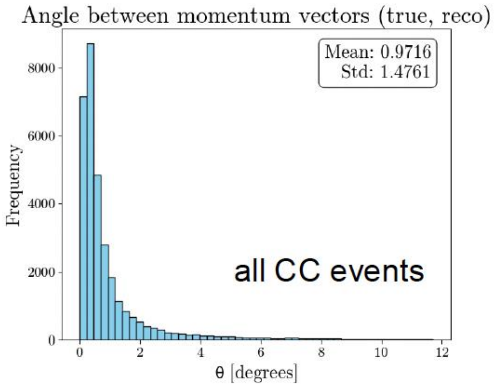
Neutrino Flavor identifications		
	Purity	Efficiency
$\nu_e\ CC$	88%	83%
$\nu_\mu\ CC$	87%	94%
NC	84%	71%
Overall accuracy 87%		



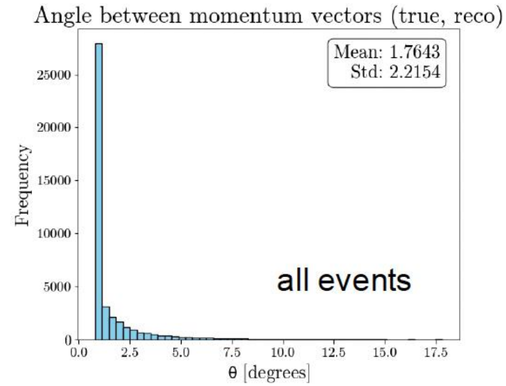
Visible Energy



Missing transverse momentum



Primary lepton momentum (direction)

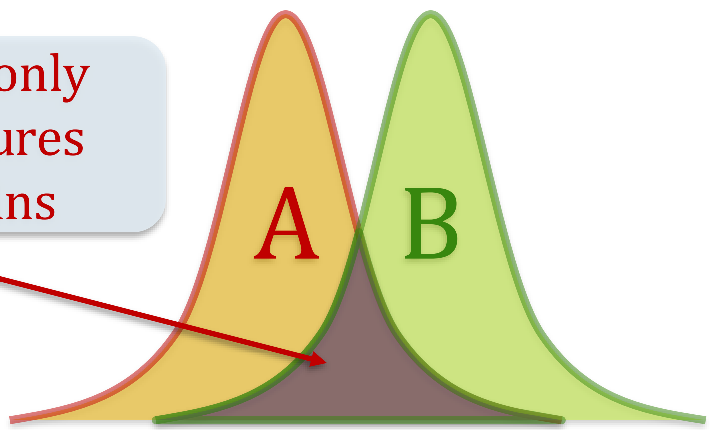


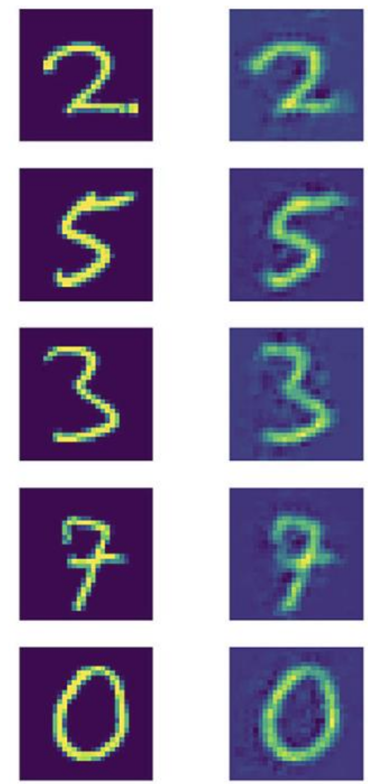
Jet momentum (direction)

Domain adaptation

- Machine learning models in particle physics are often trained with highly precise simulated data.
- The training data is "unlimited". It allows us to perform supervised learning.
- However, there's a risk that real data may not be exactly the same as simulated data.
- One solution: domain adaptation models.

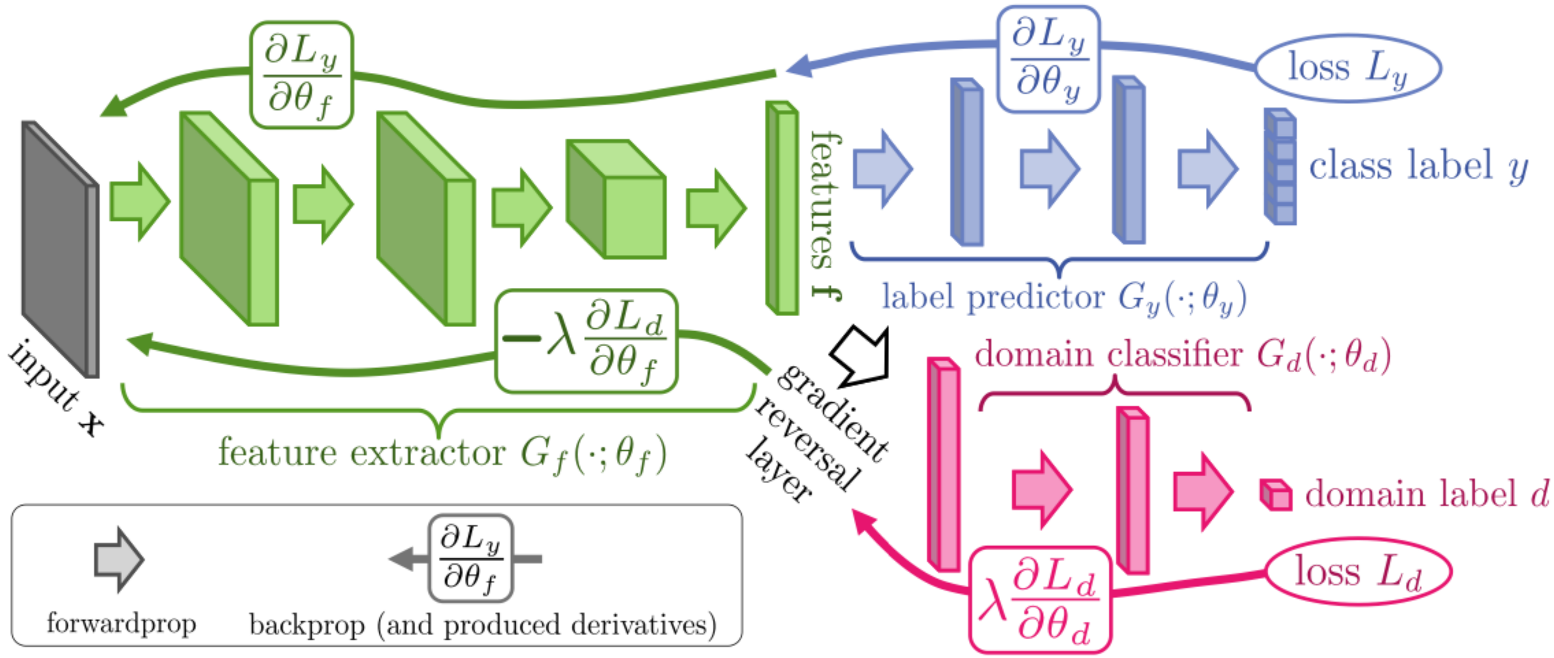
Force the model to only learn common features across both domains



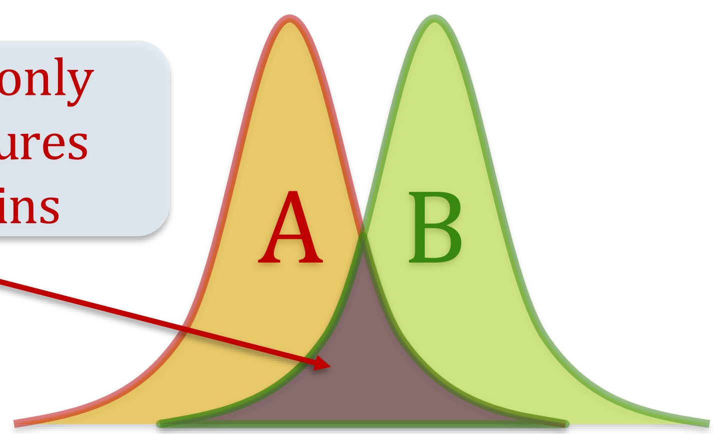


Example: Domain adversarial neural networks (arXiv:1505.07818)

Domain adversarial neural networks



Force the model to only learn common features across both domains



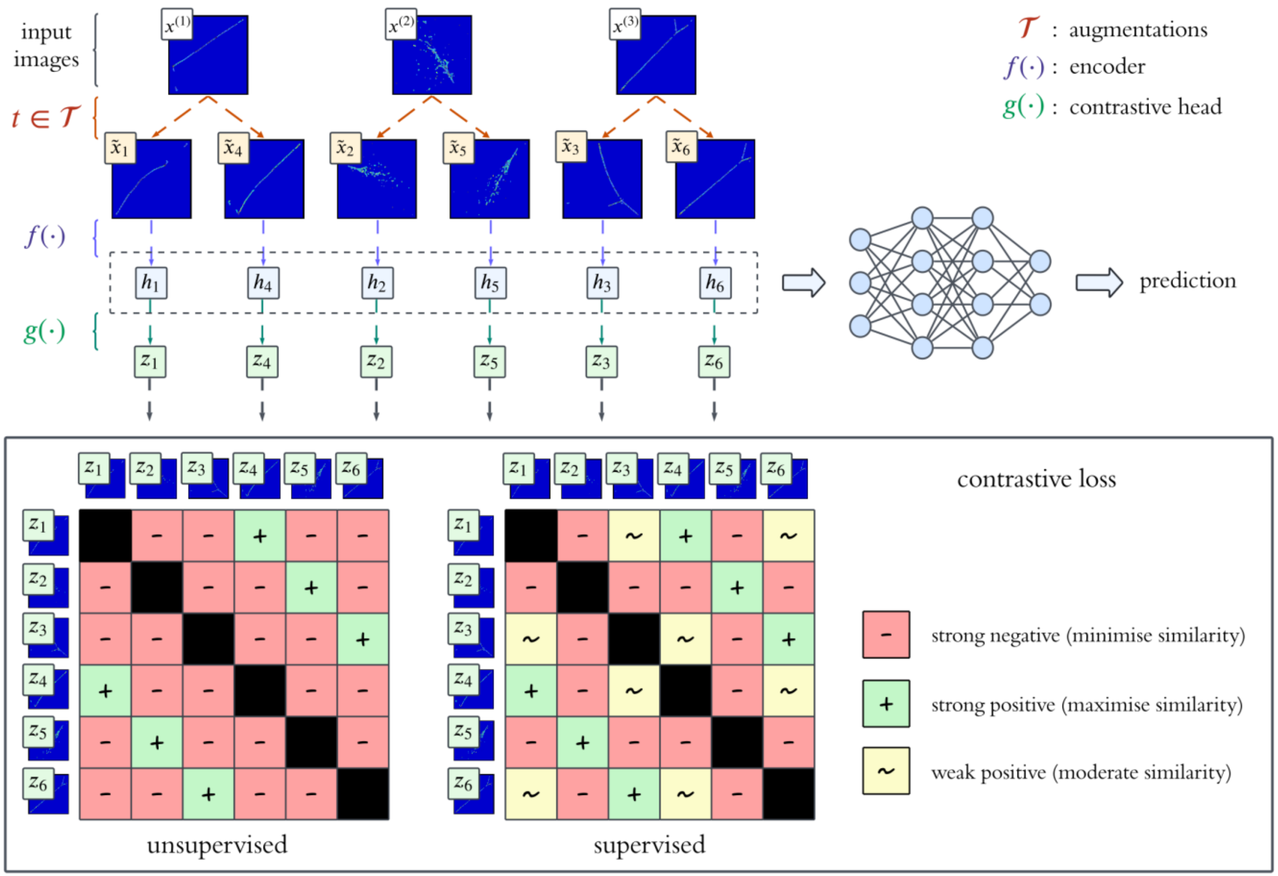
Example: Domain adversarial neural networks (arXiv:1505.07818)

Example on MINERvA (2018): doi.org/10.1088/1748-0221/13/11/P11020

Example on ICARUS (2022): doi/10.1103/PhysRevD.105.112009

Contrastive learning

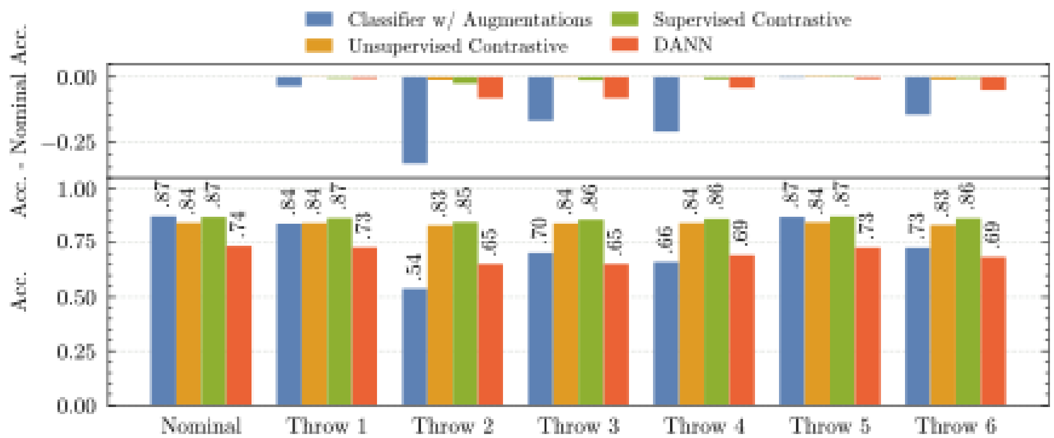
- Contrastive-learning recipe:
  - Pre-train a self-supervised model to produce similar embeddings for augmented views of the same example, and dissimilar embeddings for different examples.
  - Fine-tune a classification head on top of the pre-trained model.
  - Leverage the contrastive model's robust representations to boost performance.



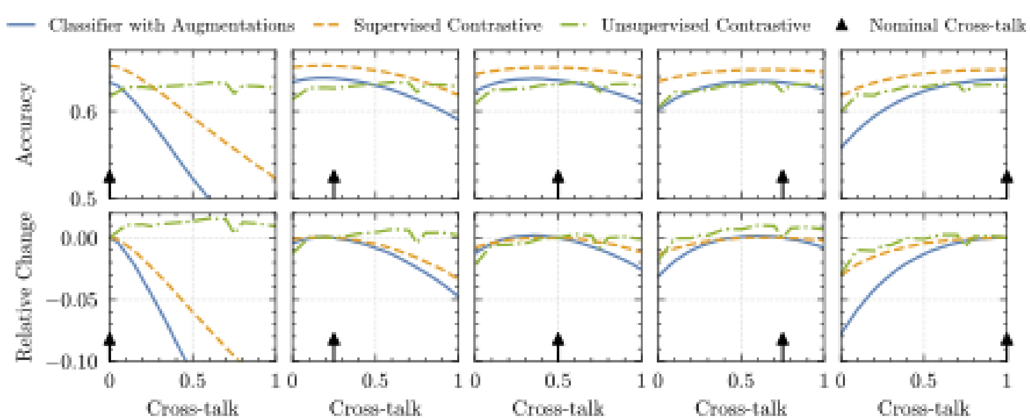
Augmentations include: rotations, voxel dropping, voxel-wise energy shifts, and translations.

Contrastive learning

- Tested on two datasets.
  1. Pixelated LArTPC detector.
     - Using larnd-sim.
     - 4 models compared.
     - Evaluated on various data throws (random shift in detector simulation parameters).



  2. Segmented plastic-scintillator detector.
     - Publicly available dataset: doi.org/10.5281/zenodo.10998285
     - Evaluated on different crosstalk levels.



Contrastive models consistently showing superior performance!

Recently accepted for publication (arXiv:2502.07724, link)

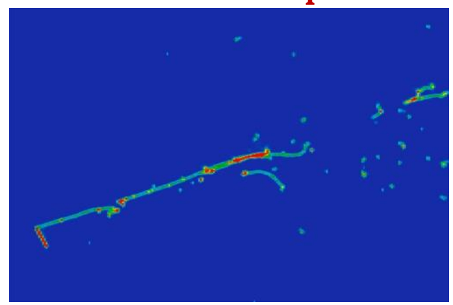
Bringing neutrino knowledge to other domains

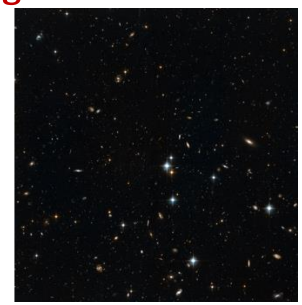
- In particle physics and astrophysics, data is often sparse, due to the nature of the objects being studied or the particles detected.

"Dense" image

- All pixels might be helpful for the target task.
- Ideal for standard CNNs.

"Sparse" images



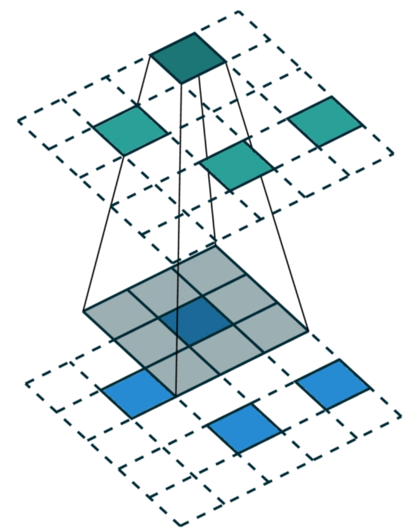


- Most pixels are background.
- A standard CNN would perform loads of useless computations.

[Source: britannica.com]

- Challenge for computer vision: standard CNNs are designed to work with dense data.

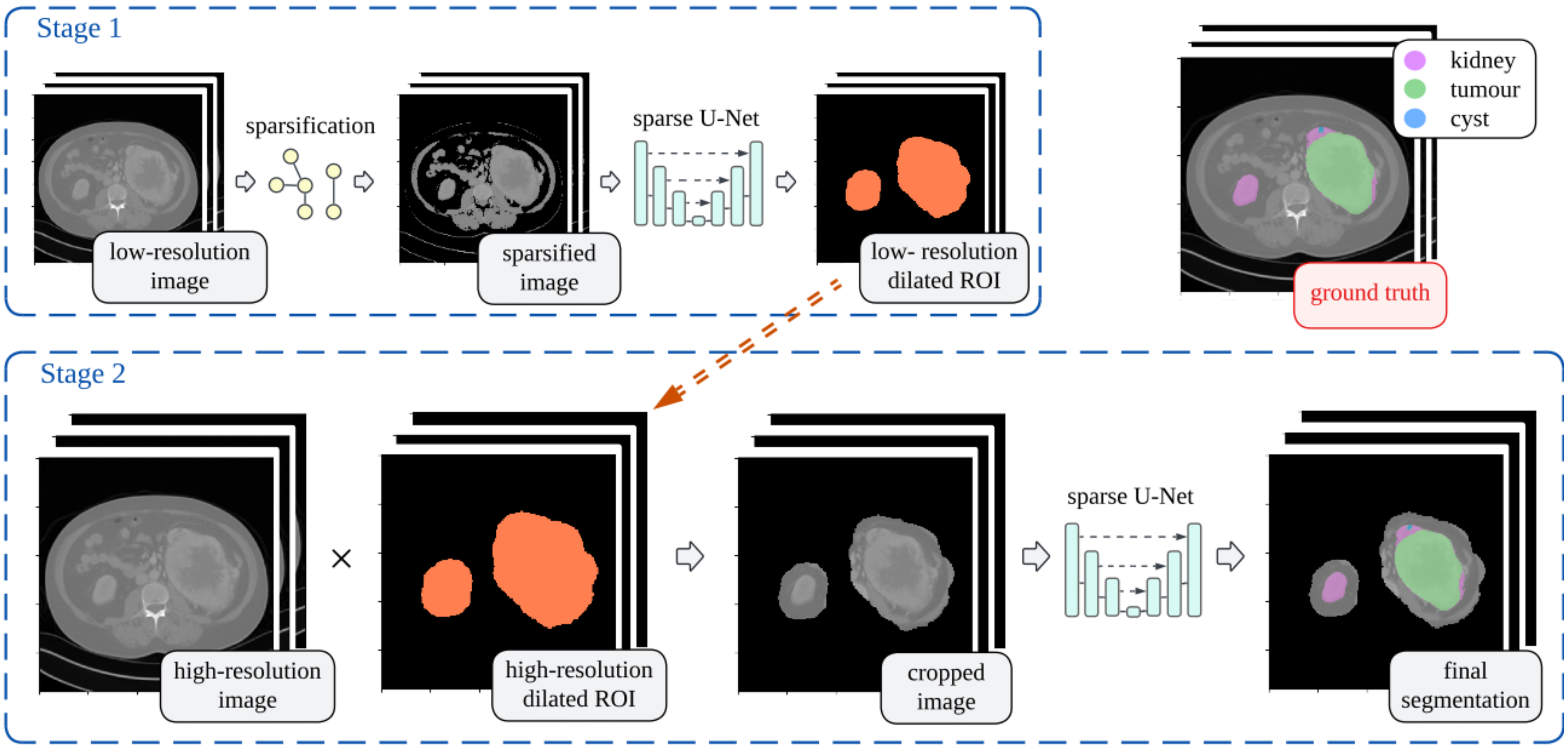
- Submanifold Sparse Convolutional Networks (SSCNs) perform the convolution **operation only on the non-zero elements of the data**, resulting in an efficient and accurate representation of the data.
  - ~16 times faster than standard CNNs on neutrino data! (arxiv.org/abs/2303.08812).



- Can we use them somewhere else?

Application to Computed Tomography

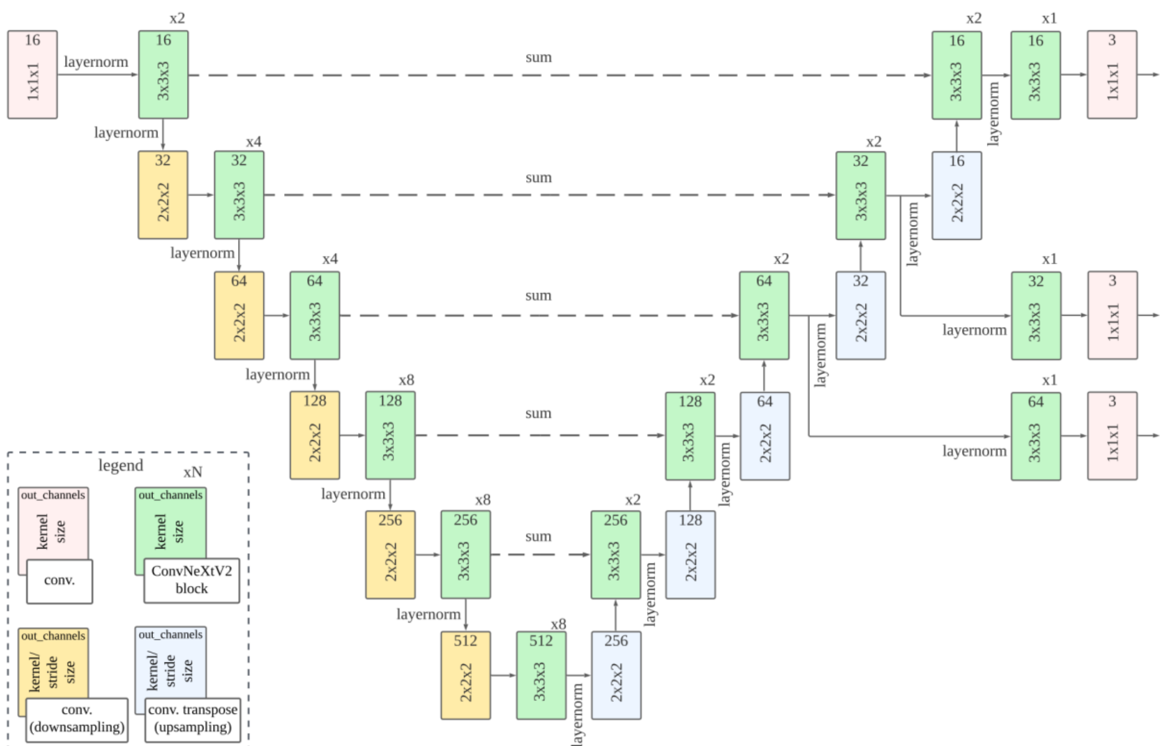
- Even though CT scans are not naturally sparse, we found that in segmentation tasks (organ and tumour finding) we could **“sparsify” the 3D images** by removing voxels outside the segmentation range, removing ~80% of background voxels while keeping 99% of the signal.
- Sparse submanifold networks on the sparse images.
- Kits23 (kits-challenge.org/kits23/): public challenge, only 489 scans, ~80M voxels per image.



Training

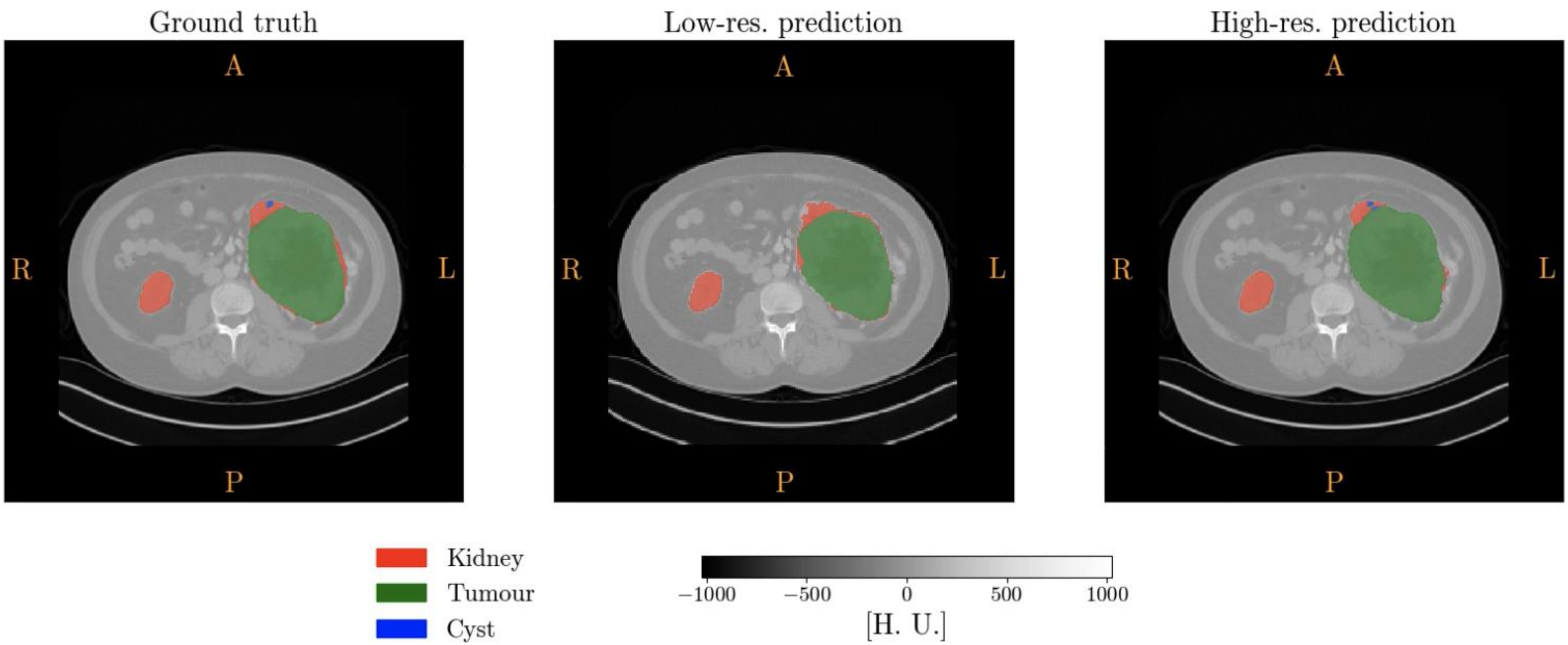
- Sparse-submanifold U-Net.
  - 5-fold cross validation.
  - Dice similarity coefficient as loss.
  - Deep supervision.
  - 500 epochs, warmup + cosine annealing.
  - Augmentations: affine transformations, flipping, intensity variations, noise and smoothing.

$$\mathrm{Dice\_loss}(\mathrm{pred}, \mathrm{target}) = 1 - 2\frac{pred \cap target}{pred + target + \epsilon}$$



Results

- Dice similarity coefficients of 95.8% for kidneys + masses, 85.7% for tumours + cysts, and 80.3% for tumours alone.
  - — **Comparable to the winners of the KiTS23 challenge!**



- Crucially, our method also offers **significant computational improvements**:
  - — Up to a 60% reduction in inference time and up to a 75% reduction in VRAM usage!

Article under review in Computer Methods and Programs in Biomedicine.

Summary

- Deep learning is revolutionising neutrino physics, enabling precise event reconstruction and physics analyses.
- Transformers, graph, and sparse convolutional networks have been deployed across multiple neutrino experiments (DUNE, T2K, FASERCal).
- Interpretability techniques (e.g., occlusion tests) help build trust in ML-driven analyses.
- Novel detector concepts (e.g., plenoptic systems) are pushing the boundaries of 3D reconstruction.
- Cross-domain applications demonstrate the broader impact of ML methods developed for neutrino physics.
- Domain adaptation and contrastive learning are key tools to bridge simulation–reality gaps.

PUSHING THE BOUNDARIES OF NEUTRINO PHYSICS WITH DEEP LEARNING

Dr. Saúl Alonso-Monsalve

ETH Zurich

Experimental Seminar, SLAC

8 May 2025