ML based PFlow + Super-resolution for calorimetry

Francesco Di Bello[2] Sanmay Ganguly[3,1]
Jonathan Shlomi[1] Eilam Gross[1] Marumi Kado[2] Lorenzo Santi[2]
Weizmann Institute Of Science[1] Roma Sapienza[2]
ICEPP-UTokyo[3]

Eur. Phys. J. C (2021) 81:107

ML4Jets2021 - ITP Heidelberg

07/07/2021

Motivation

1703.10485

Particle-flow algorithm is a generic event reconstruction technique.

Energyflow algorithm :

Cluster the cells into calorimeter topoclusters and subtract the (statistically) expected energy deposit by charged tracks matched to the topocluster.

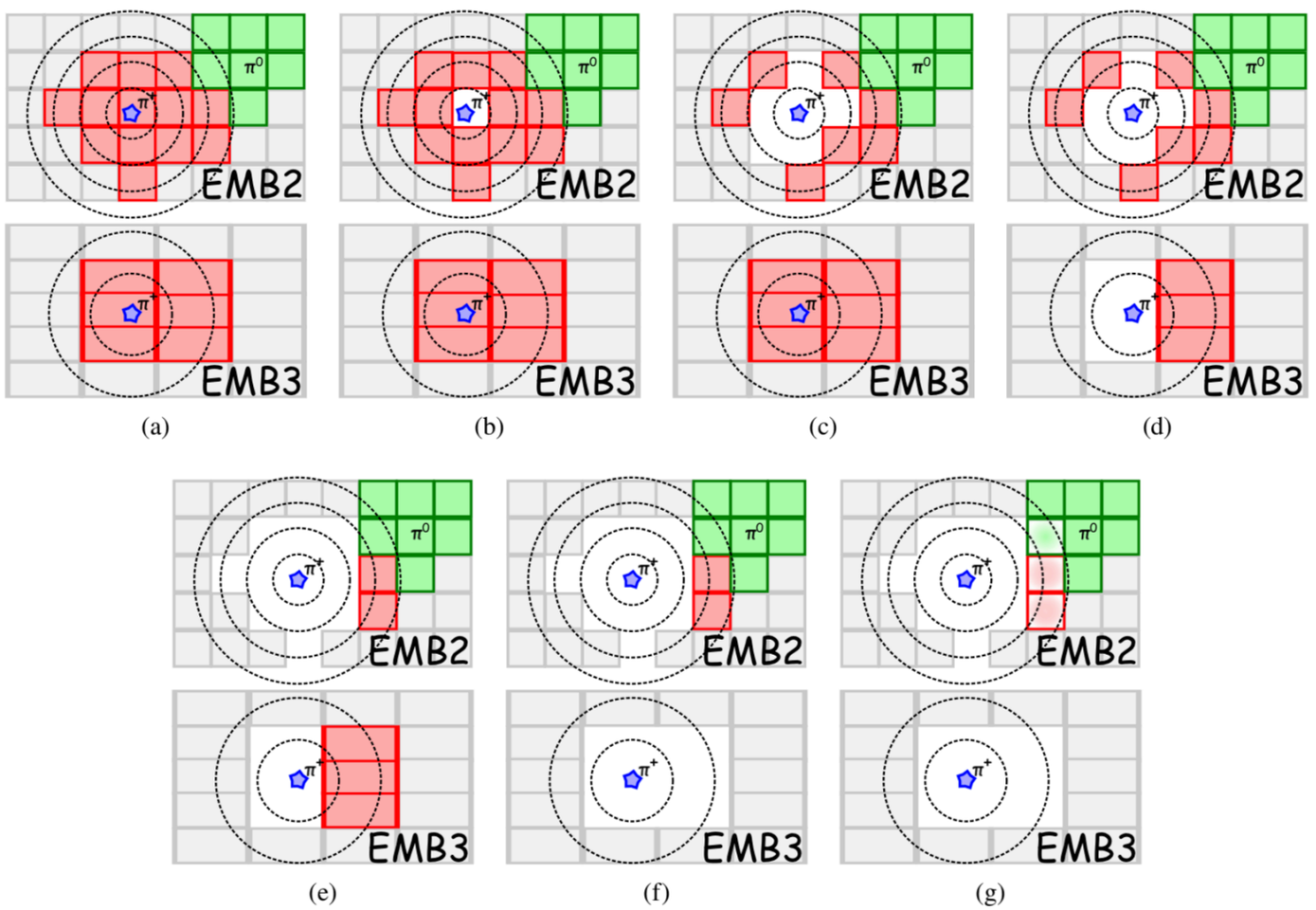
The task : predict fraction of neutral energy per cell of the topocluster

The Challenge : Different calorimeter layers have different resolutions.

Our proposal :

Implement a deep learning based method to extract the fraction of neutral energy for each cell in each layer of the ECAL and HCAL calorimeter layers.

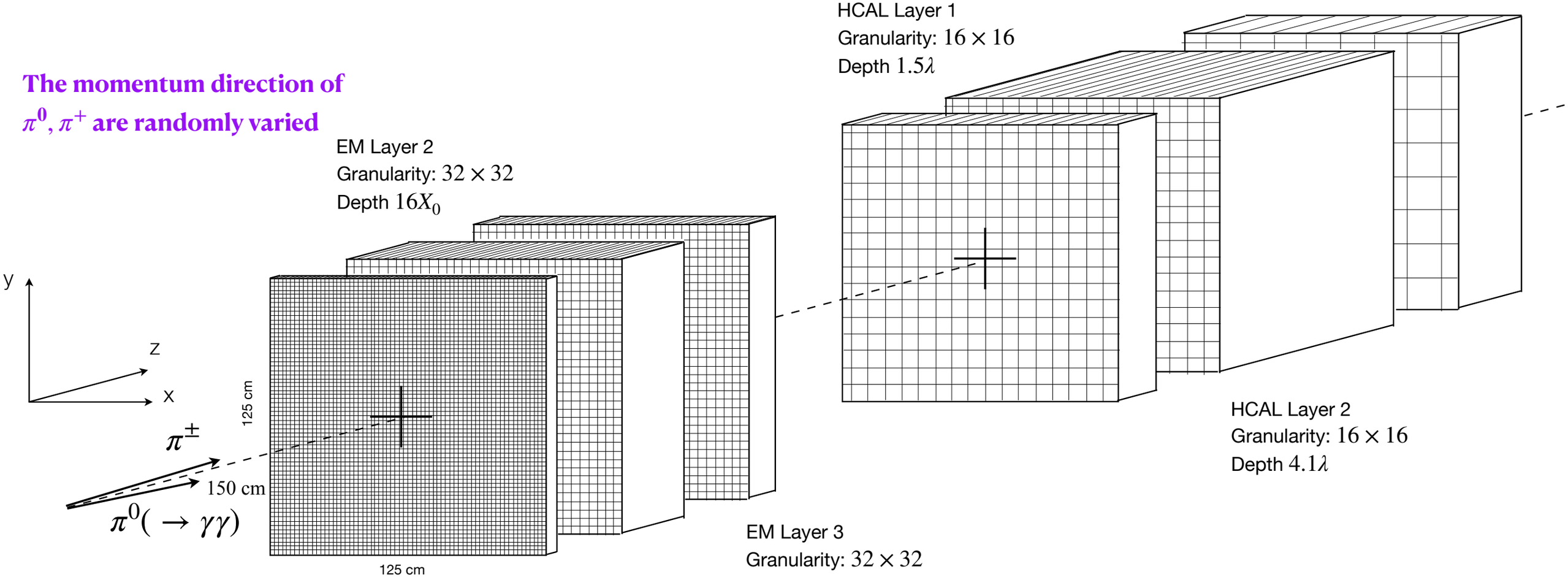
The existing algorithm as of now :



Calorimeter with GEANT

The momentum direction of π^0, π^+ are randomly varied

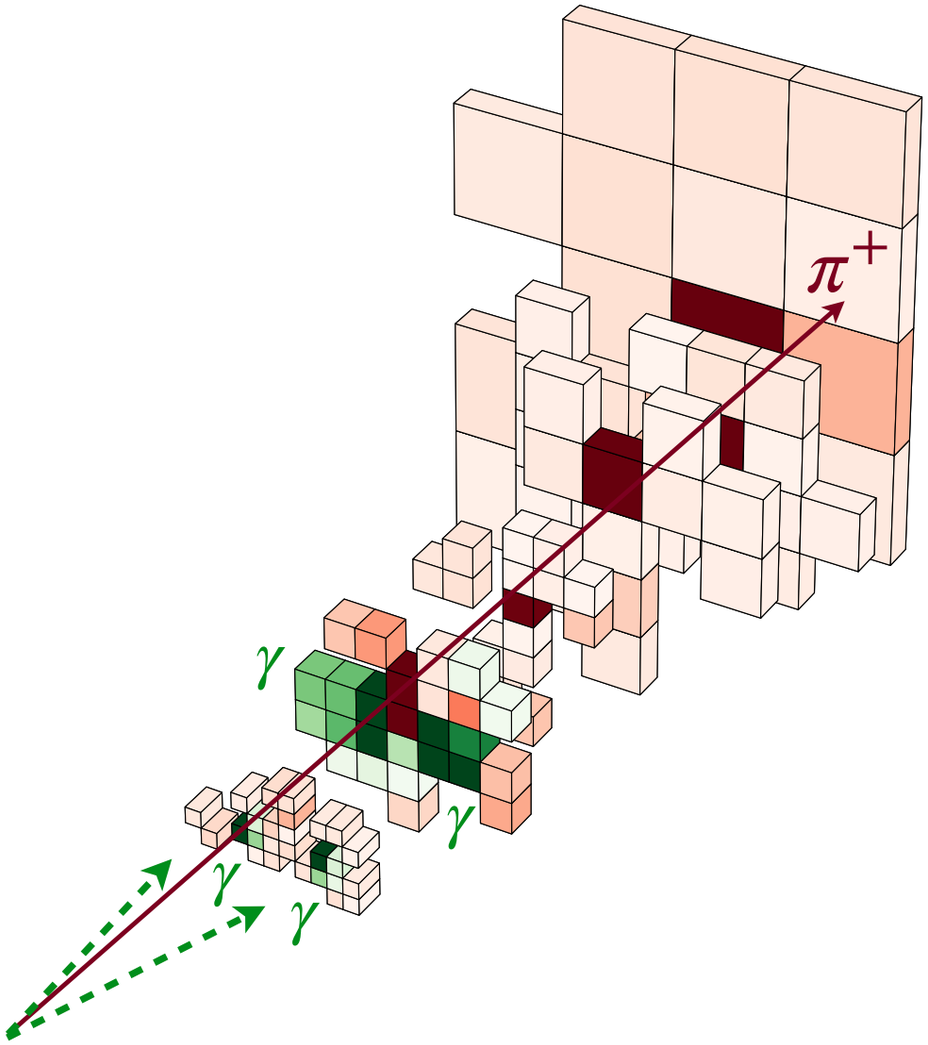
The impact of the PFlow highly depends on the granularity of the designed calorimeter.



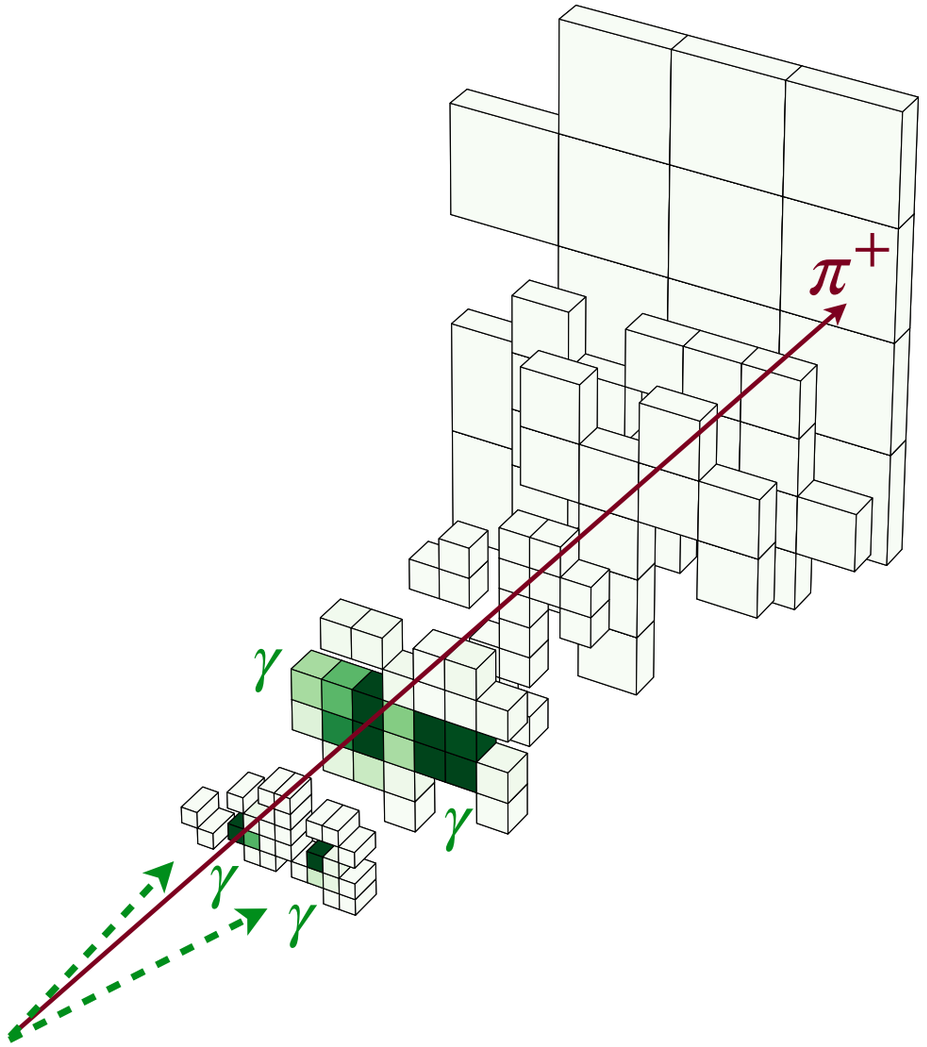
Particles are thrown randomly from a corner of 20 cm X 20 cm square.

We make sure that there is a significant energy overlap per cell.

A 3-D view for topoclusters only



$\pi^0 \rightarrow \gamma\gamma$



$\pi^0 \rightarrow \gamma\gamma$

Machine learning task

We want to regress the neutral energy per cell

Input :

6 channel image

(Signal + Noise)

Layer1 –> 64 X 64
Layer2 –> 32 X 32
Layer3 –> 32 X 32
Layer4 –> 16 X 16
Layer5 –> 16 X 16
Layer6 –> 8 X 8

+

Track Layer

Output :

6 channel image

(neutral energy frac)

Layer1 –> 64 X 64
Layer2 –> 32 X 32
Layer3 –> 32 X 32
Layer4 –> 16 X 16
Layer5 –> 16 X 16
Layer6 –> 8 X 8

$$L_{event} = \frac{1}{E_{tot}} \sum_{c} E_c (f_t^c - f_d^c)^2$$

A simple L2 loss function doesn't serve the purpose, we need to put extra weights on highest seed cells inside a topocluster

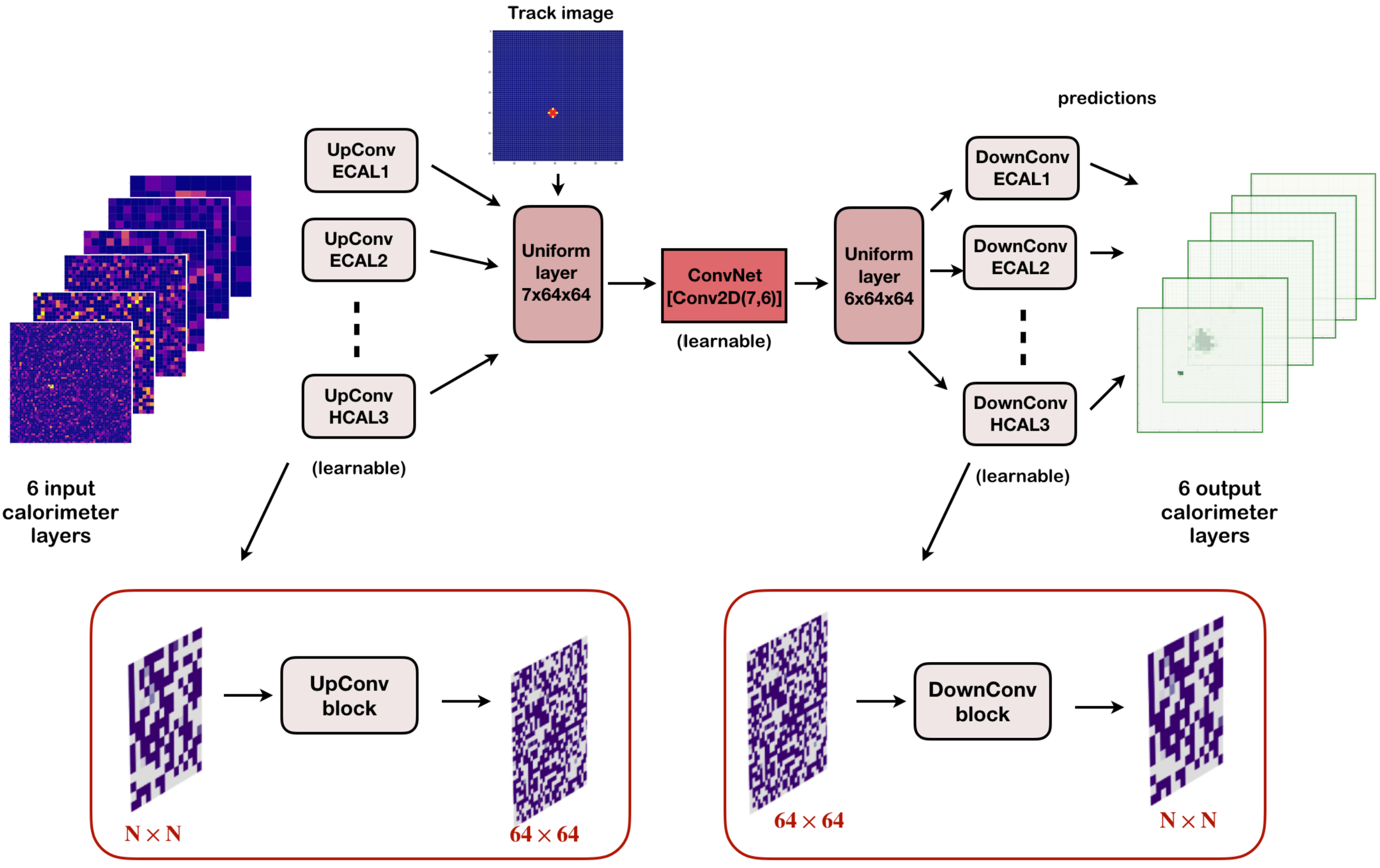
A weight factor of $\sqrt{E}$ won't work either .

E_c : Energy of a cell, $E_{tot} = \sum E_c$

f_t^c : target neutral energy fraction

f_c^d : predicted neutral energy fraction

The neural network architecture (cPFlow)



The graph network

https://arxiv.org/pdf/1801.07829.pdf

$$(x')_i^{l+1} = max_{j \in \mathcal{N}(i)}\ \Theta_x(x_j^l - x_i^l) + \Phi_x(x_i^l)$$
$$(e')_i^{l+1} = mean_{j \in \mathcal{N}(i)}\ \Theta_e(e_j^l - e_i^l) + \Phi_e(e_i^l)$$

$(\overrightarrow{x_1}, E_1)$
$(\overrightarrow{x_2}, E_2)$
$(\overrightarrow{x_3}, E_3)$
.
.
.
$(\overrightarrow{x_N}, E_N)$

→ $\mathbf{EC_{L1}(d_{in} = 3,\ d_{out} = M)}$ →

(z_1^M, E_1^{L1})
(z_2^M, E_2^{L1})
(z_3^M, E_3^{L1})
.
.
.
(z_N^M, E_N^{L1})

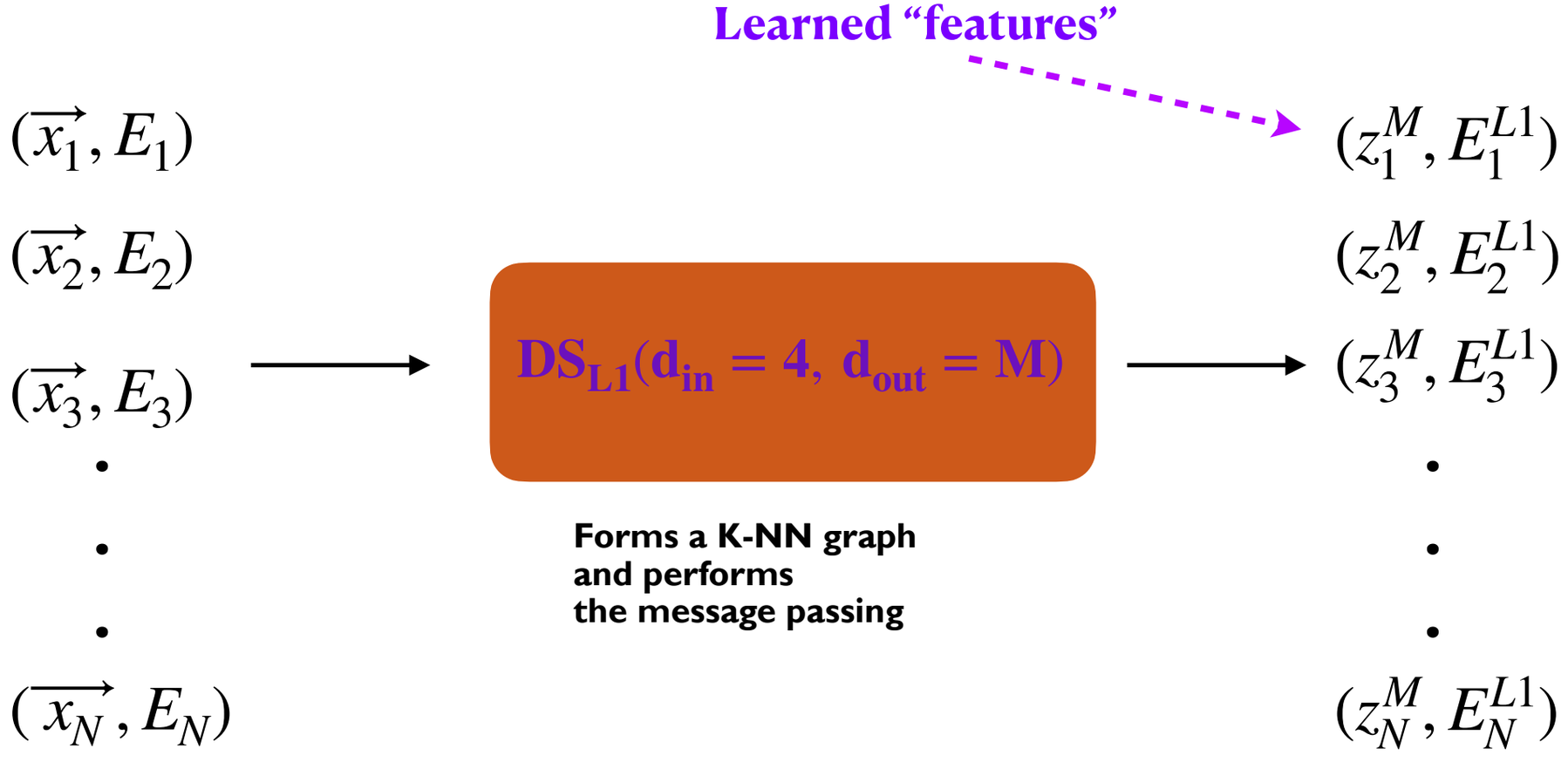
Forms a K-NN graph and performs the message passing

After p message passing layers , the q-th node has following energy representation :

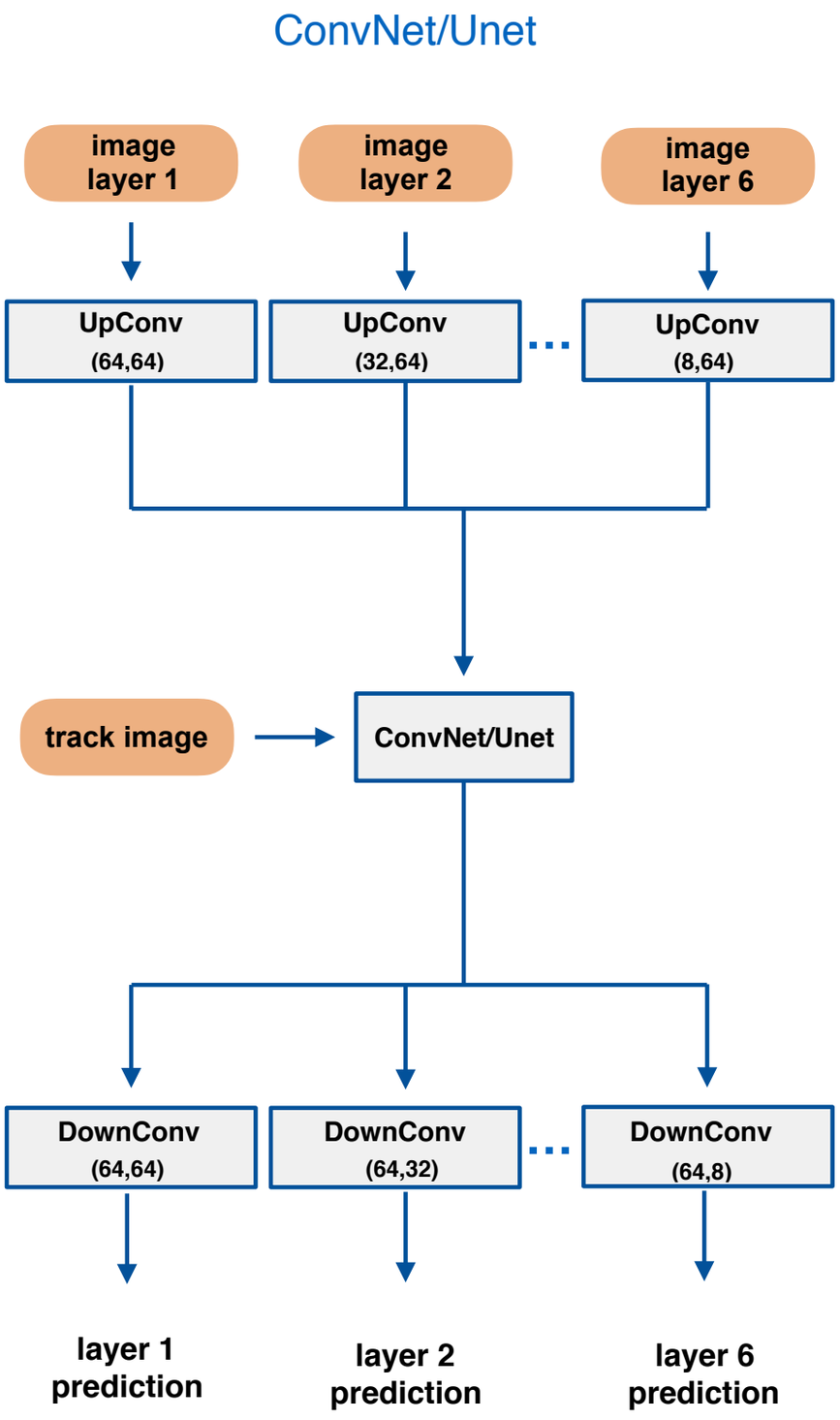
$\left(\mathbf{E_q}, \mathbf{E_q^{L1}}, \mathbf{E_q^{L2}}, \ldots, \mathbf{E_q^{Lp}} \right)$ → **MLP** → f_q^{nu}

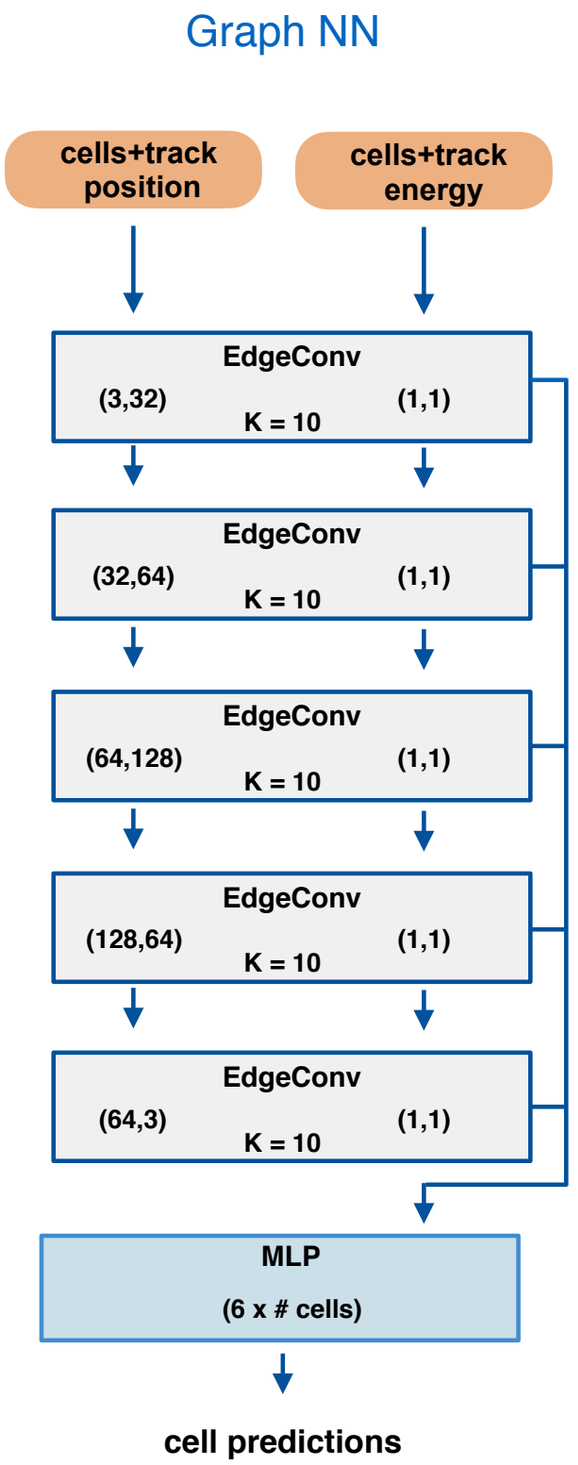
The deepset network

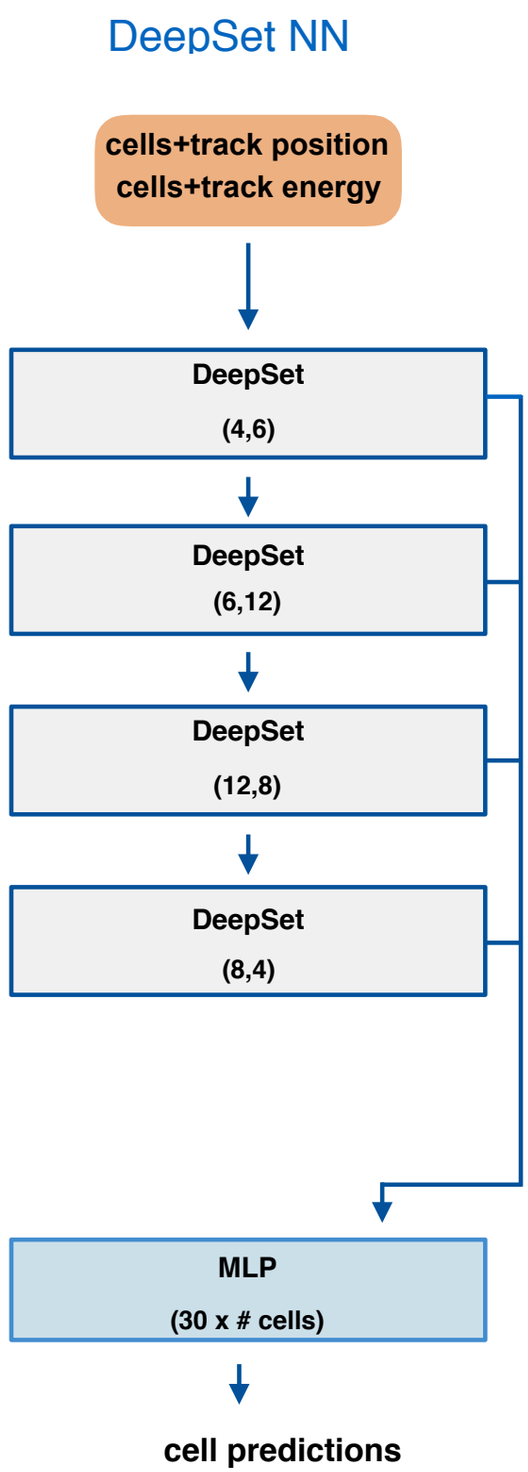
$$f(\mathbf{x}) = \sigma\Big(\gamma\ \mathrm{I}\ \mathbf{x} - \lambda\ mean(\mathbf{x})\ \mathrm{I}\ \Big),\ x = (\overrightarrow{x}, E)$$



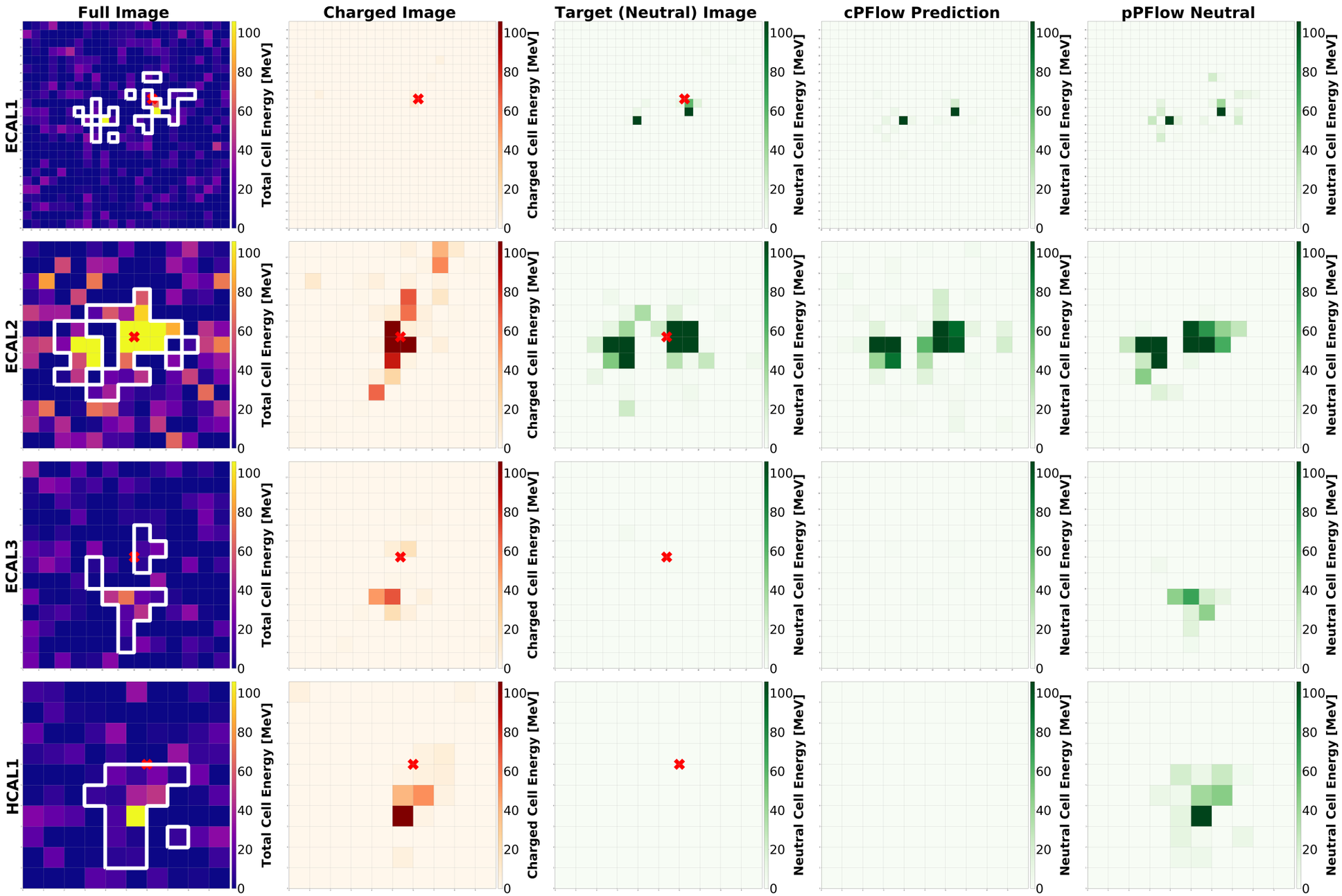
The network layout comparison







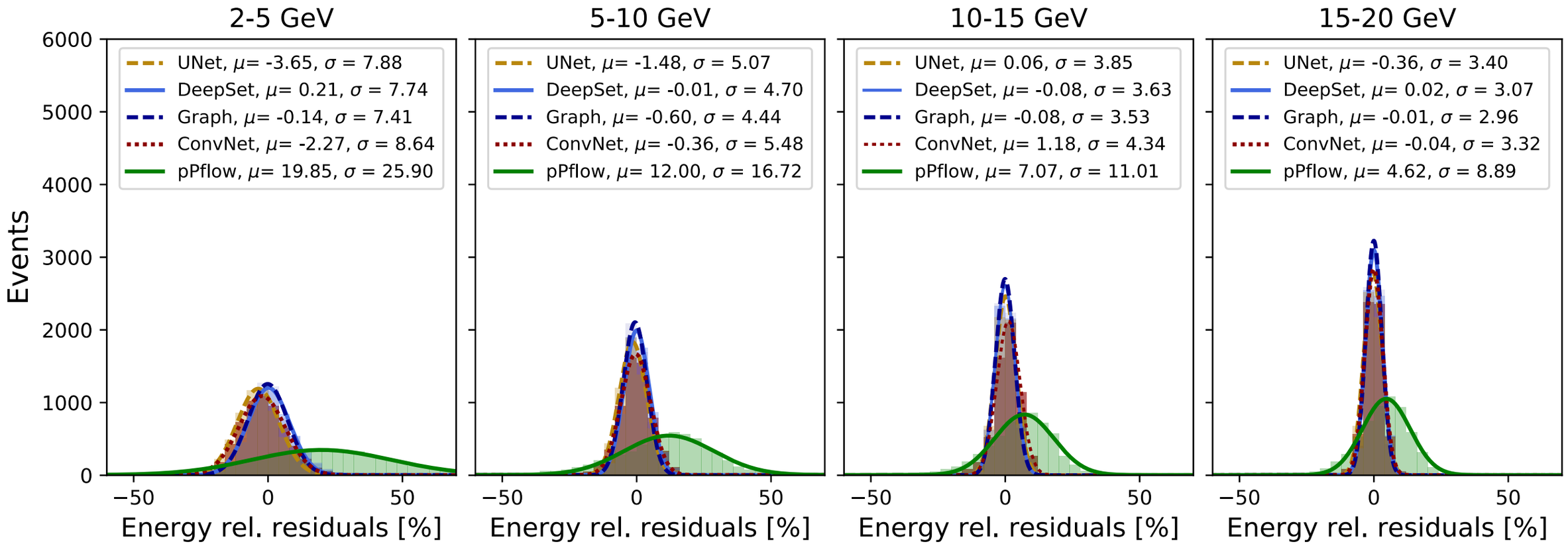
An event display (2-5 GeV)



The trained network efficiently suppress the noise contribution

Energy response comparison

With the current model, we observe a small bias + a Non-Gaussian tail
For image based methods.



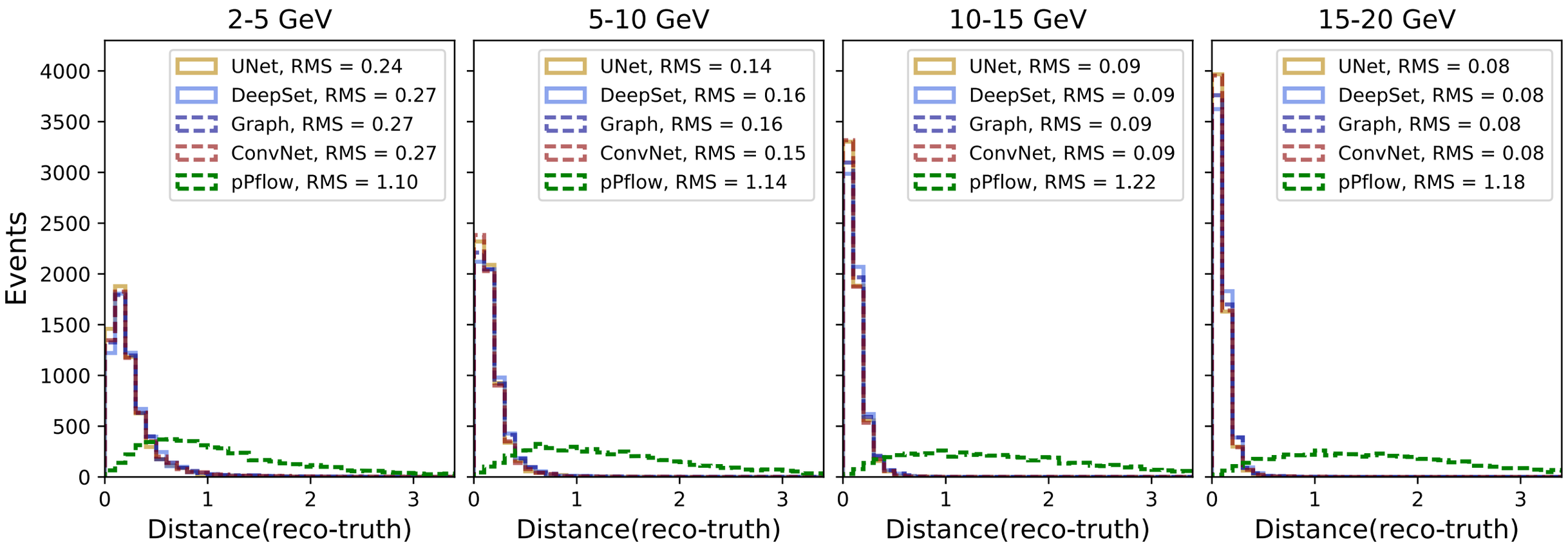
- The current NN trained on full images, the performance is evaluated only within topocluster.
  Hence an underestimate of the mean residual within topoclusters.

- The soft cells are treated as per noise & hence the NN predicts zero energy for those cells.
  Compensates by over-predicting the hard cells, leading to the Non-Gaussian tail.

- The bias in the pPFlow originates from systematically lower energy & size of the topocluster of single π^+ samples.

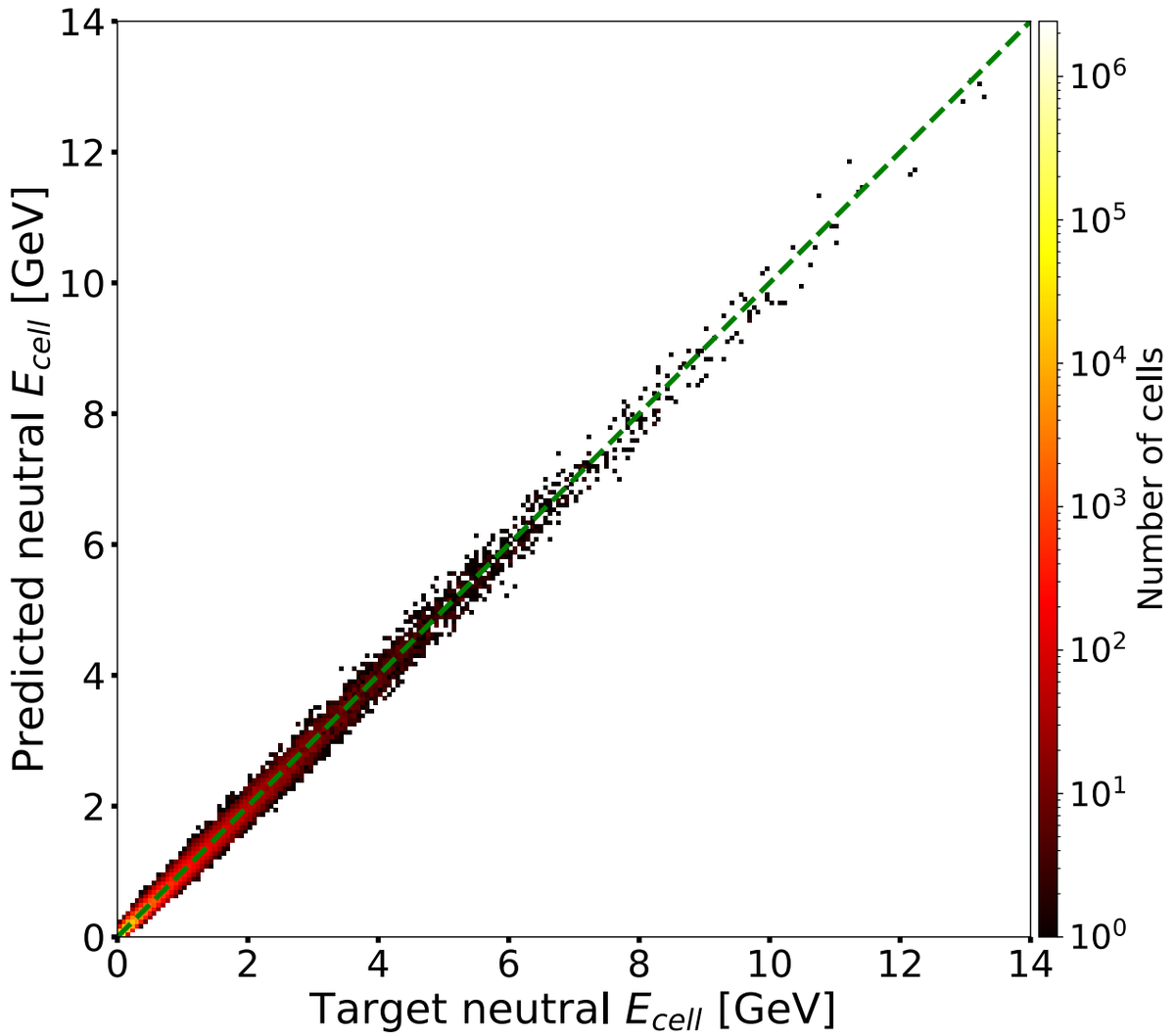
Direction response comparison

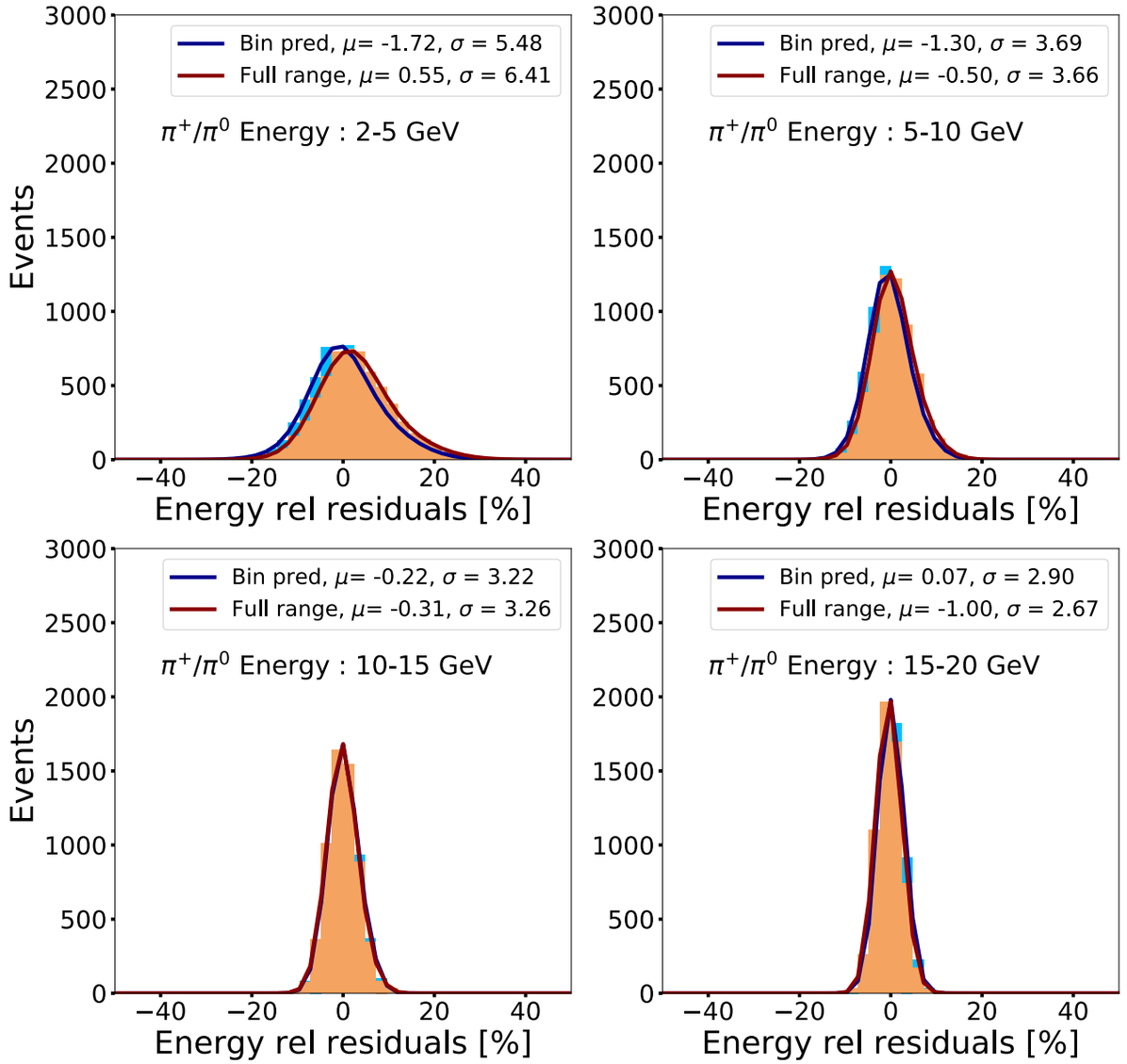
The distance computed in number of cells between the barycenter of the predicted and truth neutral energy in the ECAL2 layer.



The cPFlow algorithm has much better (upto 6 X) spatial resolution than traditional PFlow

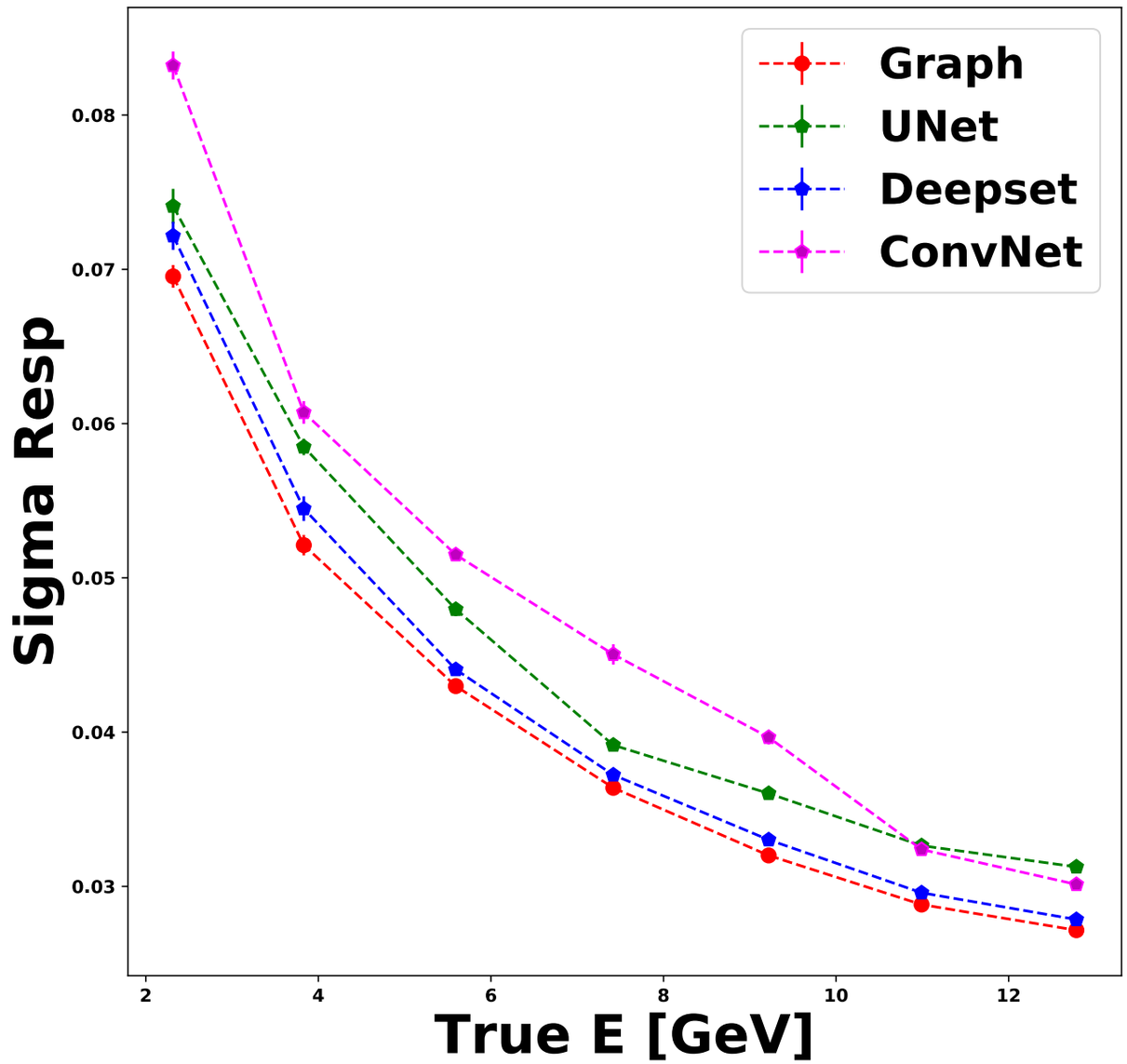
Per cell level performance





The networks in general have good noise removal abilities.

Stability of different networks

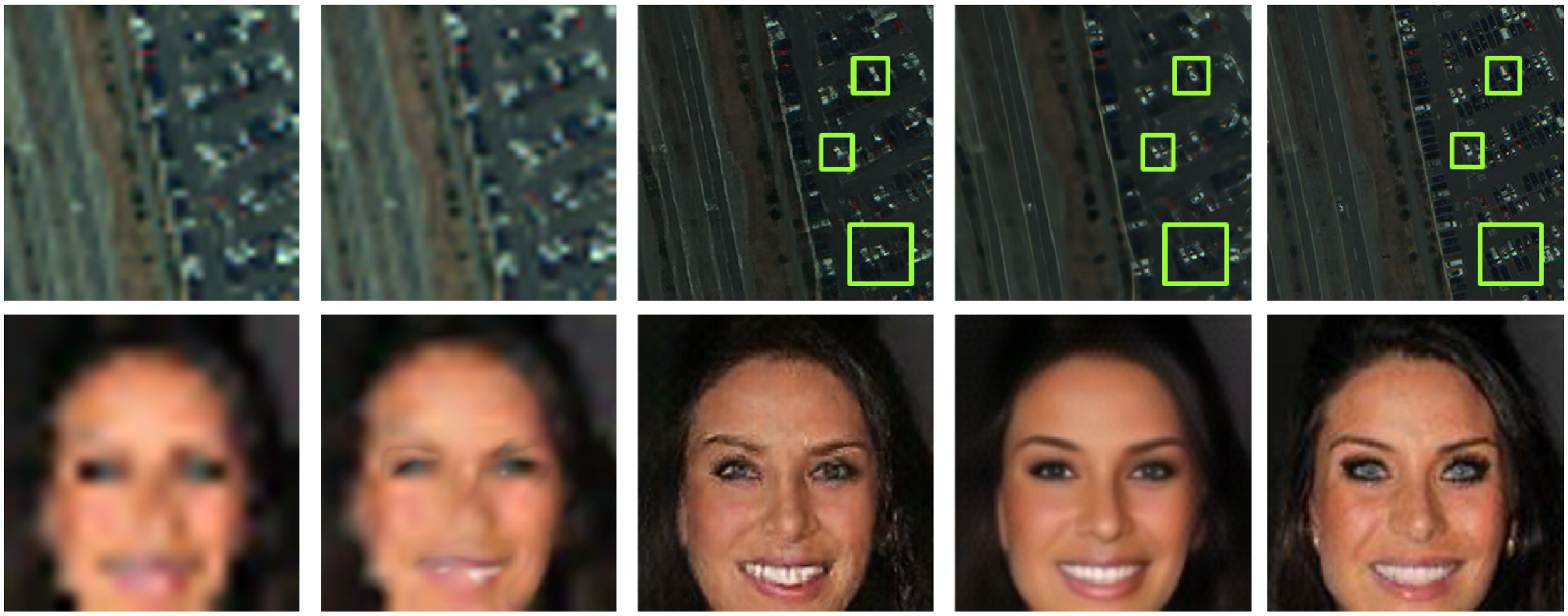


Graph networks have best stability and resolution followed by deepsets.

The transition to Super-resolution

What is super-resolution ?

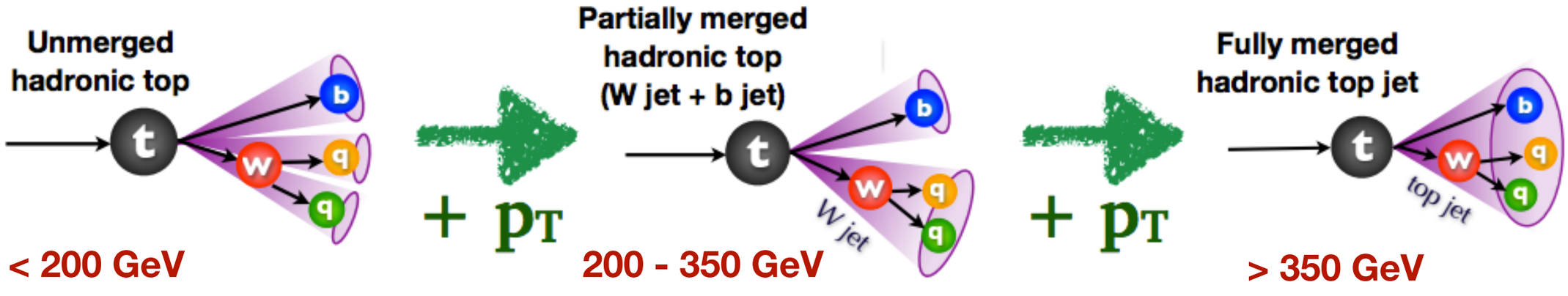
In the context of images : a systematic algorithm to sharpen the images



Low Resolution | MSSRNet | SRGAN | Our Method | High Resolution

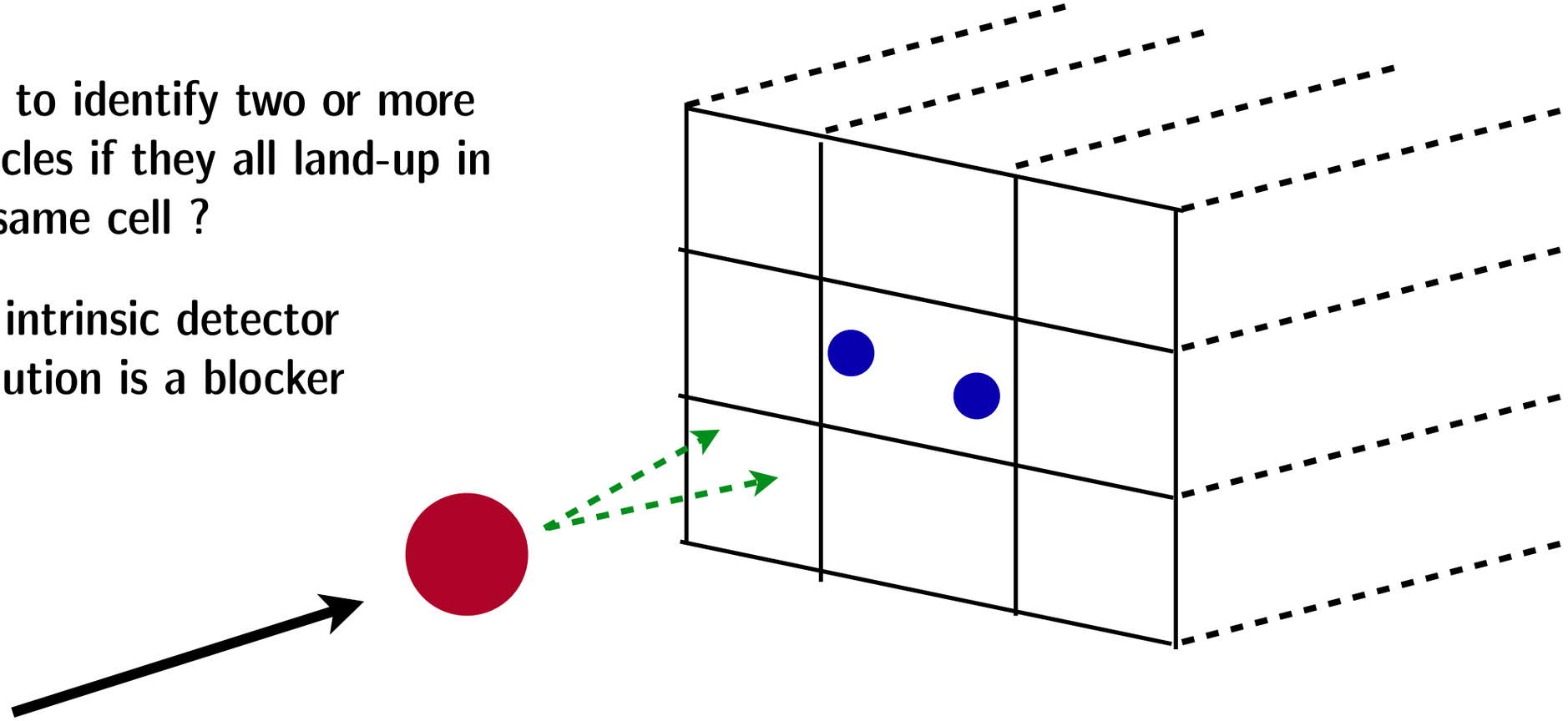
Ref : DOI:10.1109/ICASSP.2019.8682354. Corpus ID: 145823404

When do intrinsic calorimeter sizes are limiting factors ?

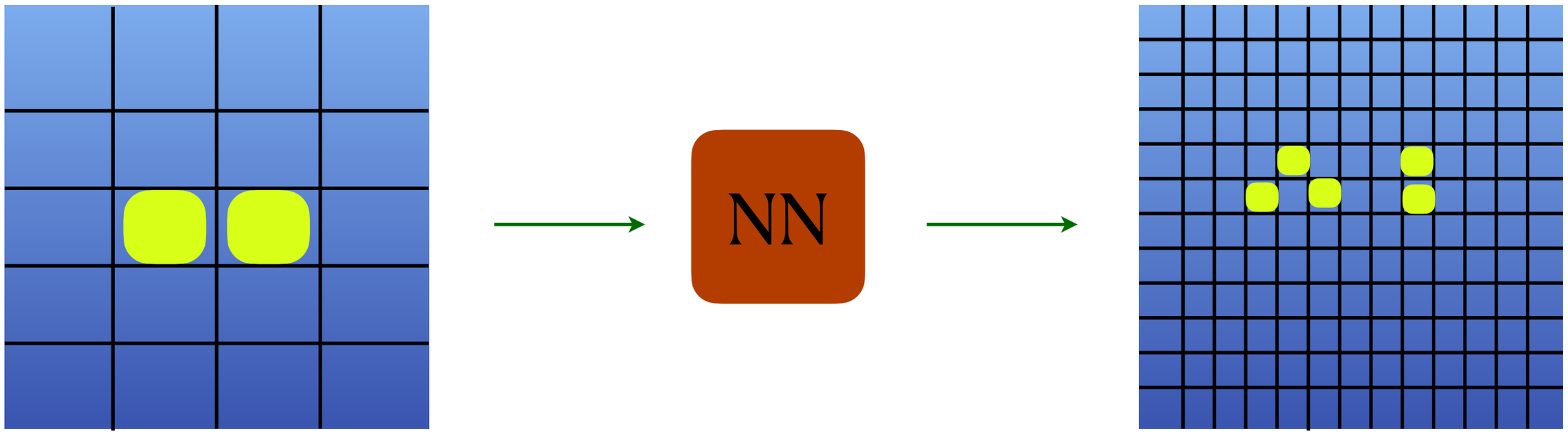


How to identify two or more particles if they all land-up in the same cell ?

The intrinsic detector resolution is a blocker



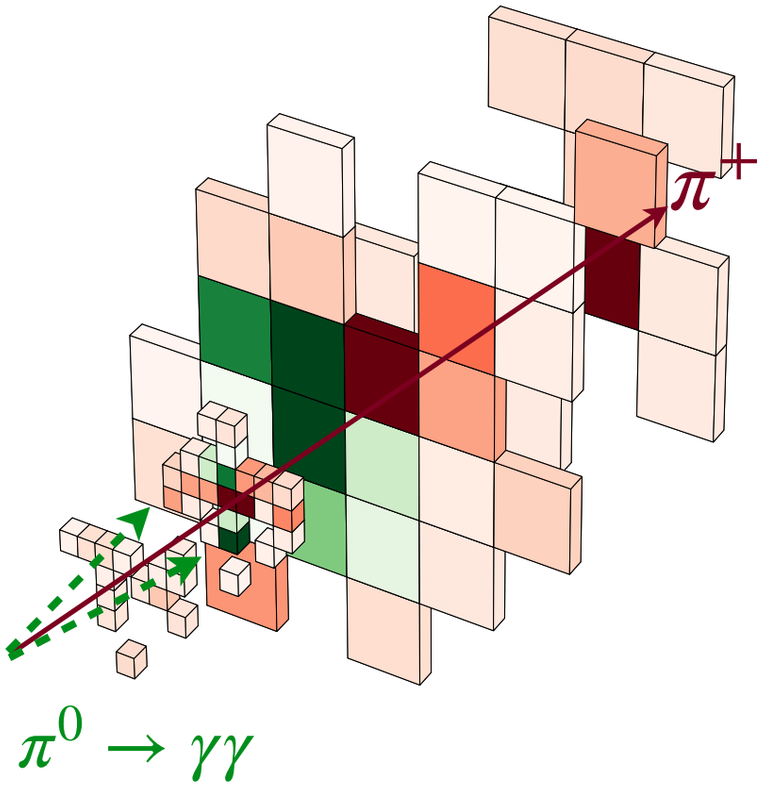
How ML can help us to overcome the problem ?



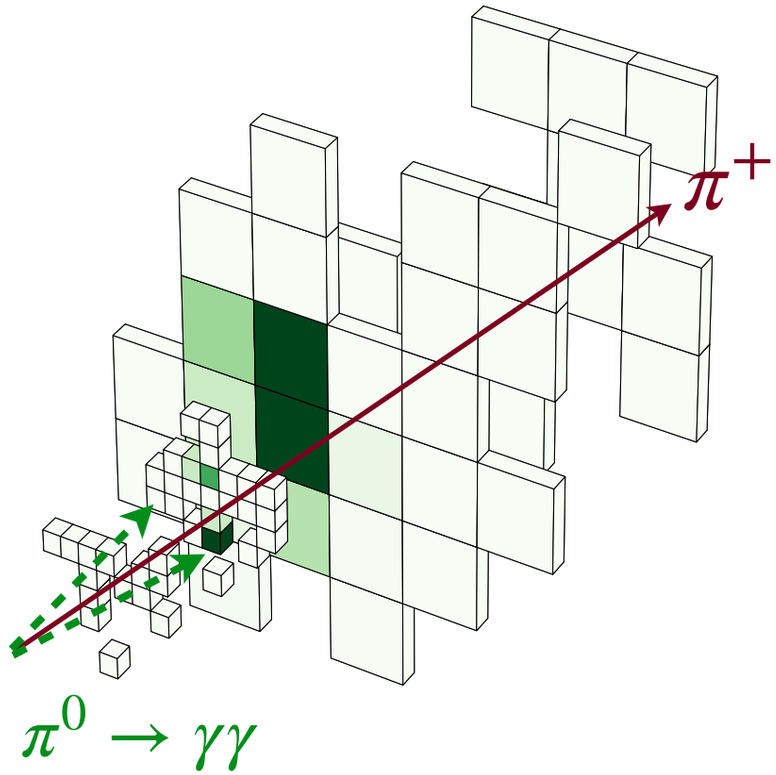
Input Target

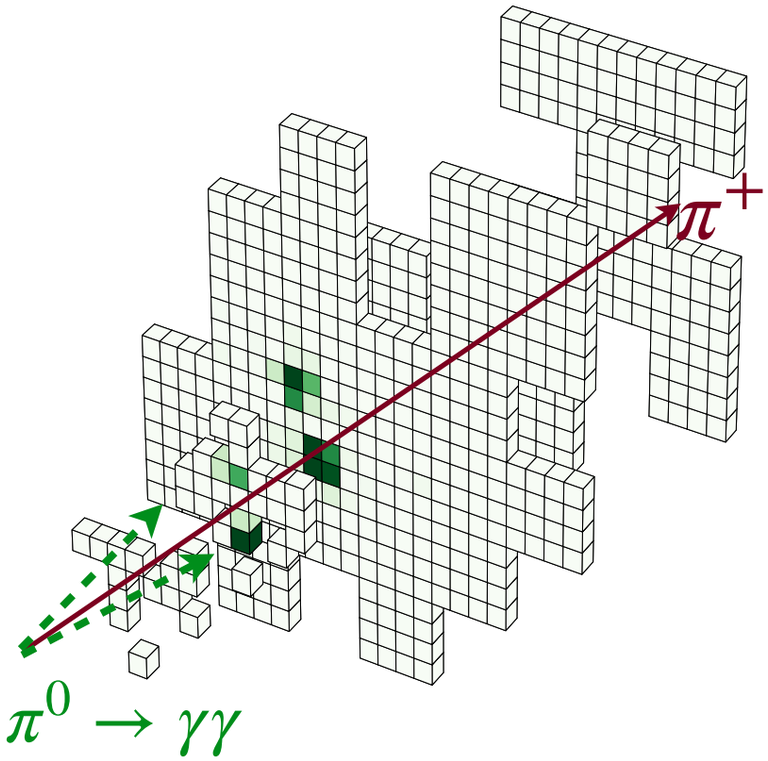
- Train a NN to teach high resolution shower pattern, starting from the low granularity one.
- Use it to predict shower profile for a real-life low-resolution measurement.
- This helps us to decode the shower pattern obtained from LR detector (and recover some physical information.)

A case for calorimeter super-resolution



8 X 8 Low Res detector

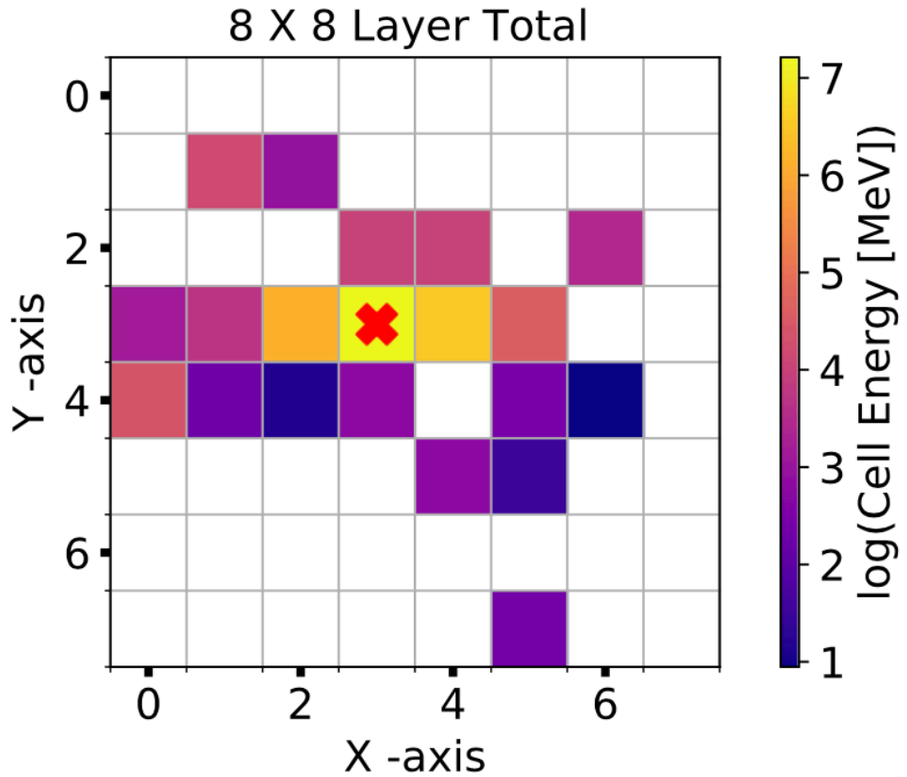


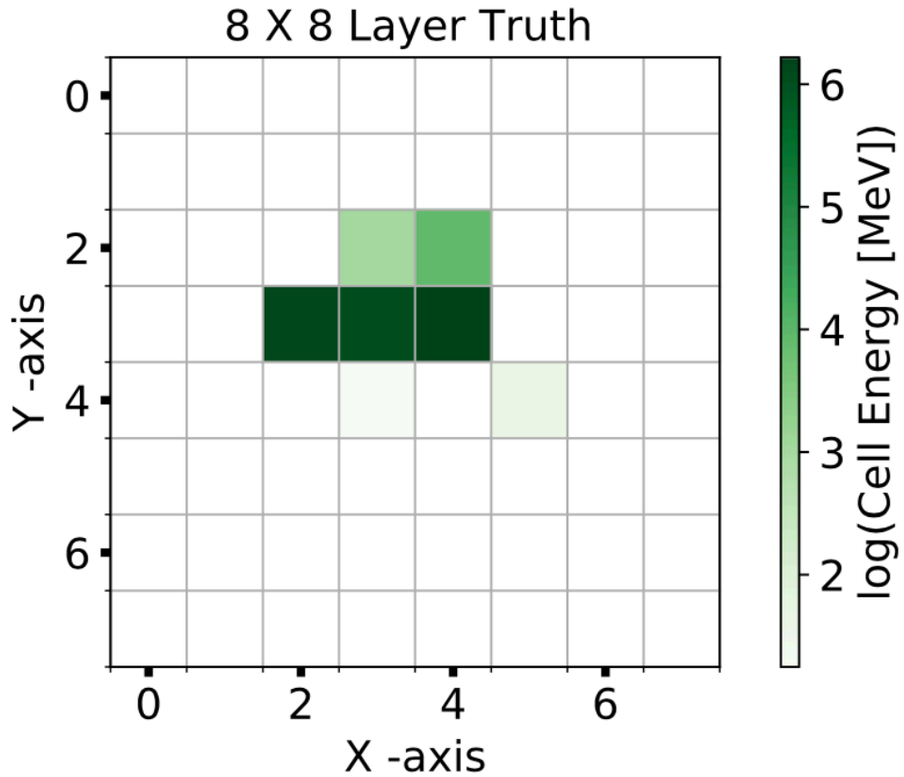


32 X 32 High Res detector

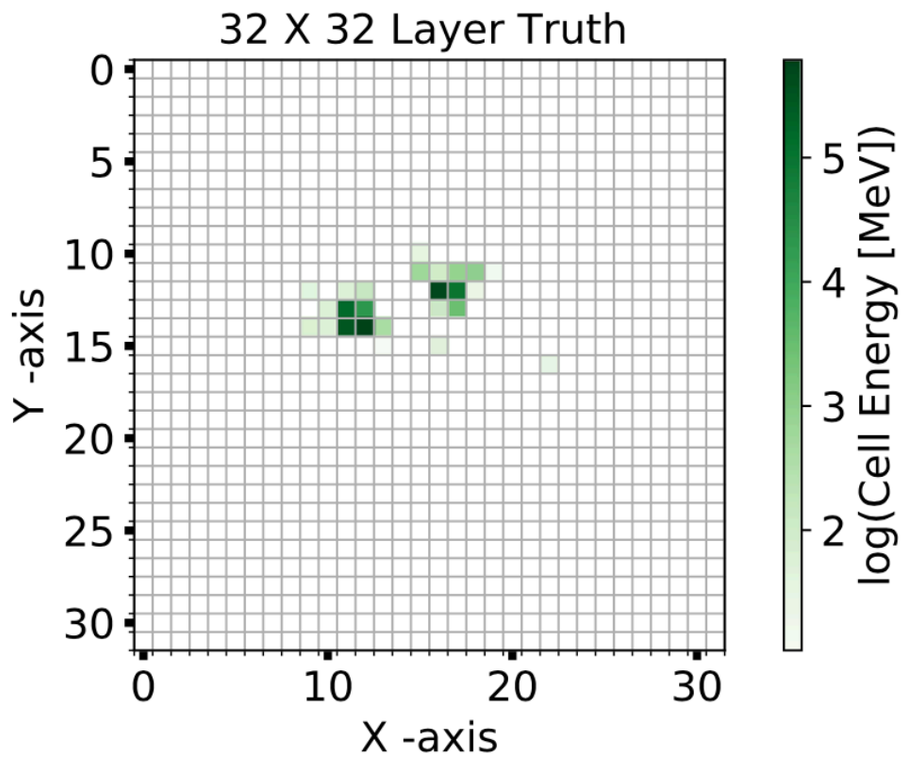
A higher resolution calorimeter has the ability to capture multi-prong decay pattern in showers.

What we aim to learn ?



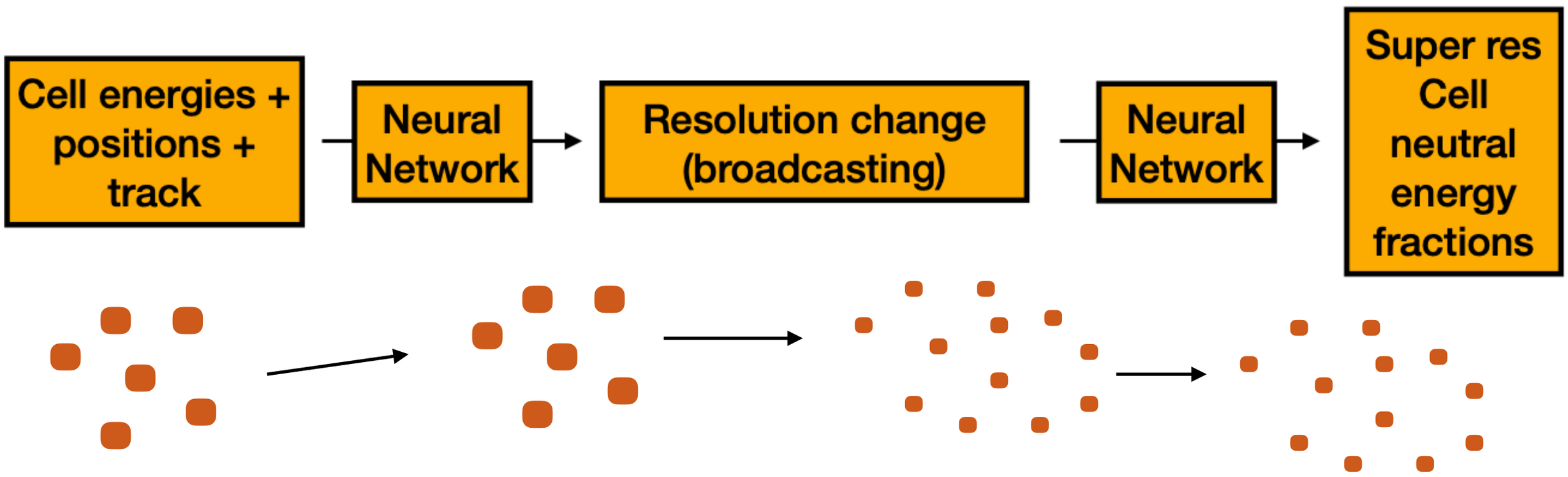


Separating the pi+ and two photons is a challenging experimental task



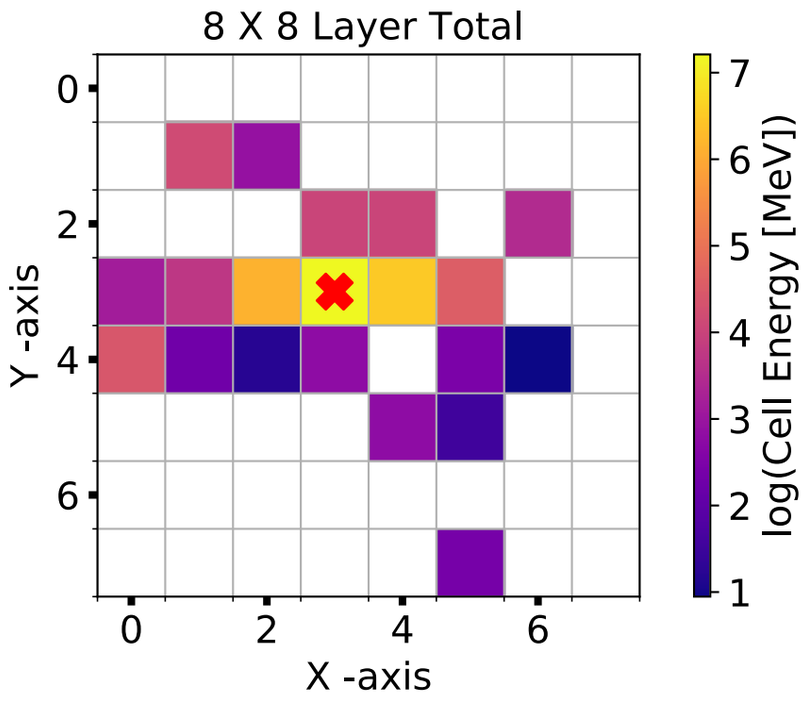
The proposed NN structure

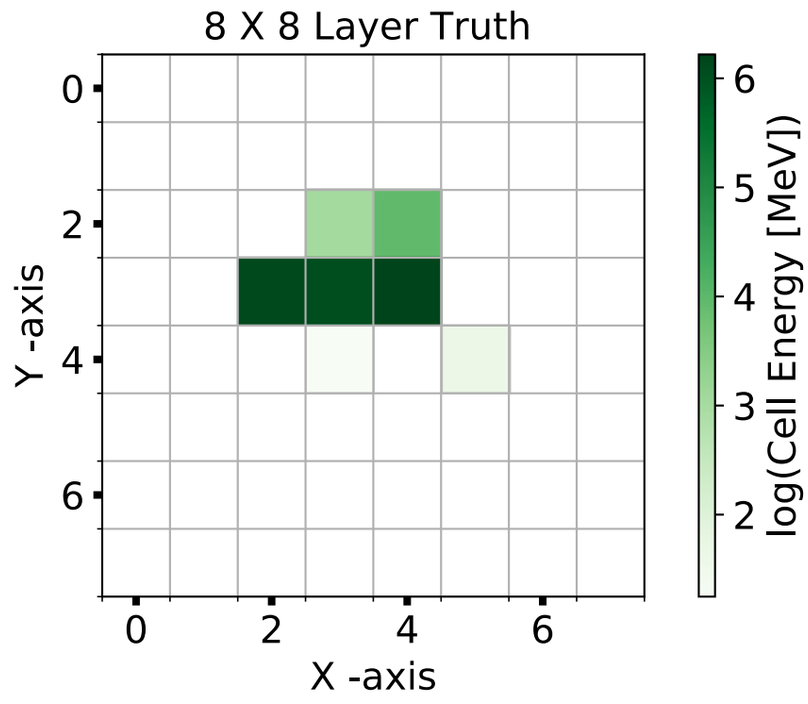
Common network structure for super-resolution task

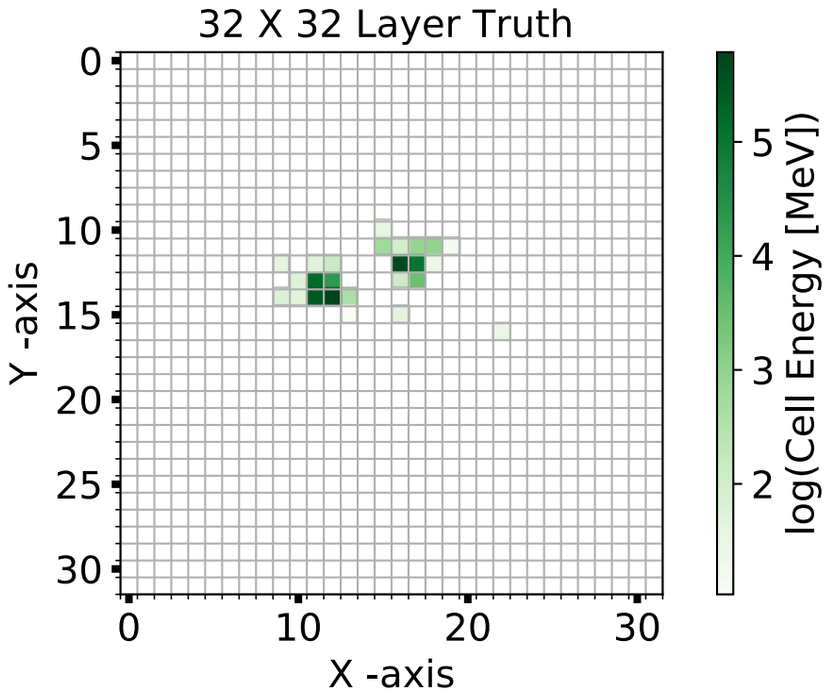


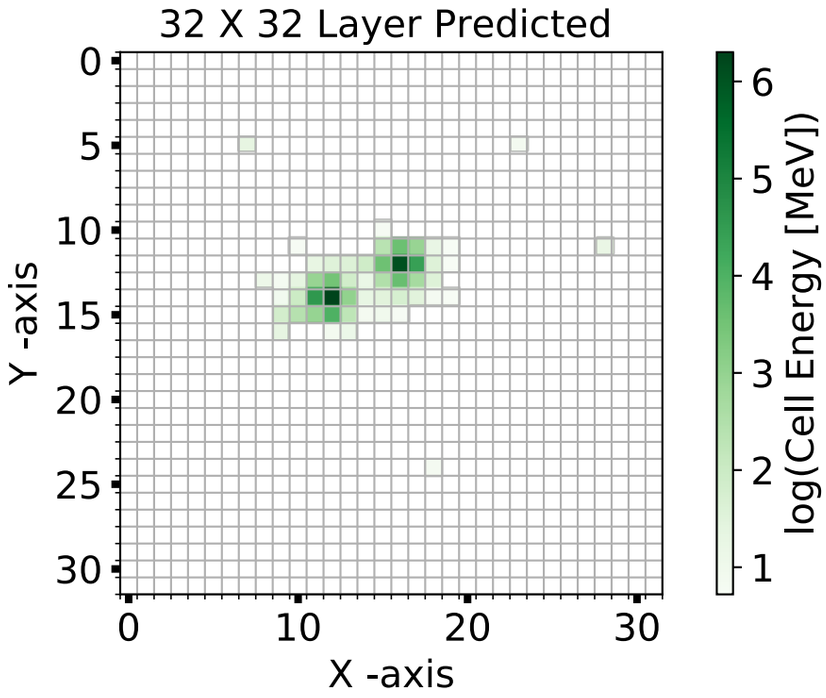
$$L_{event}^{\text{super-res}} = \frac{1}{E_{tot}} \sum_{c} E_c \sum_{s=0}^{us^2} (f_t^{sc} - f_d^{sc})^2$$

An event display for super-res prediction

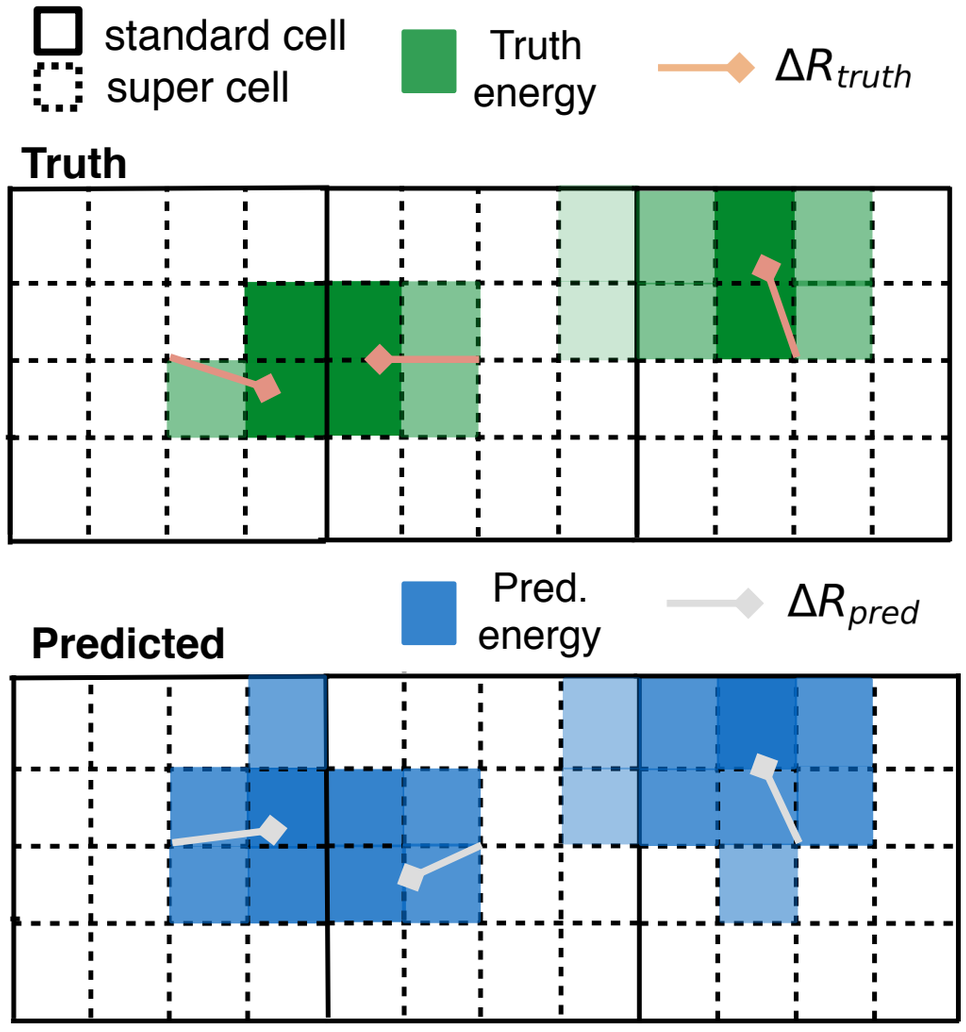




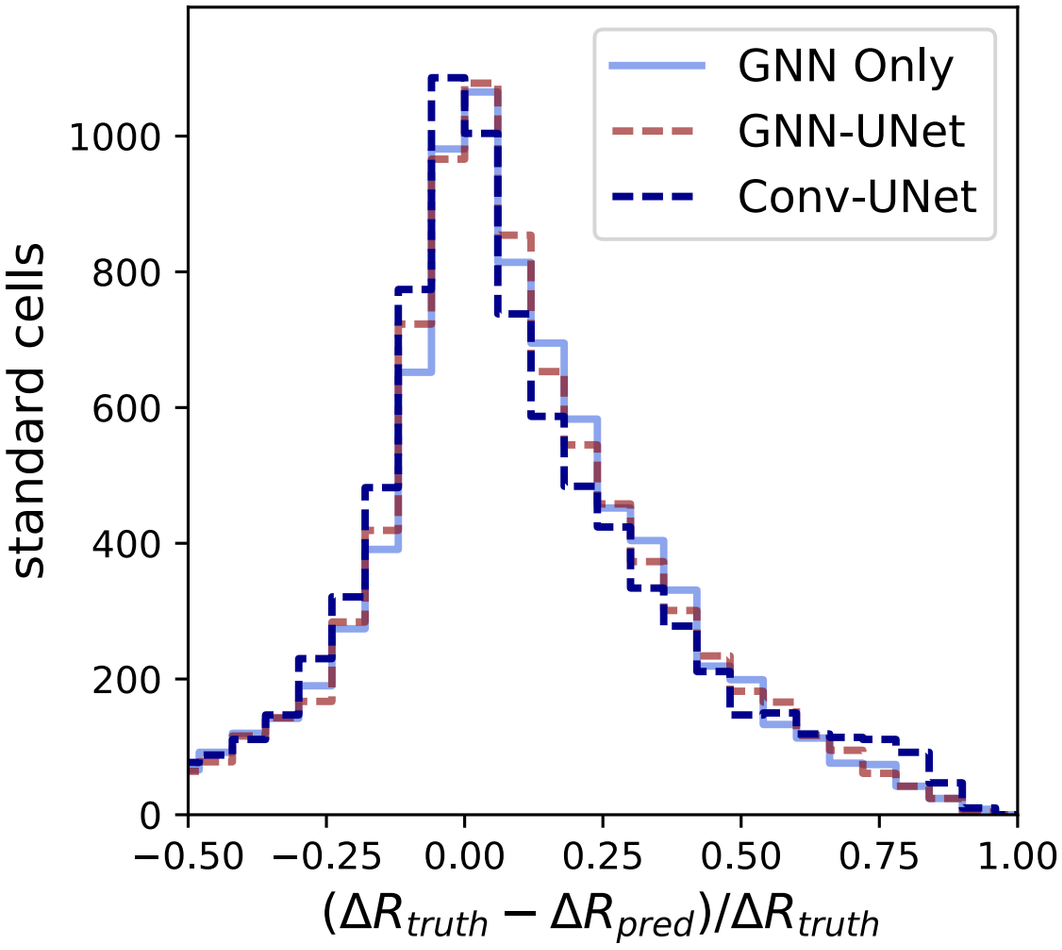




A closer look to super-res prediction

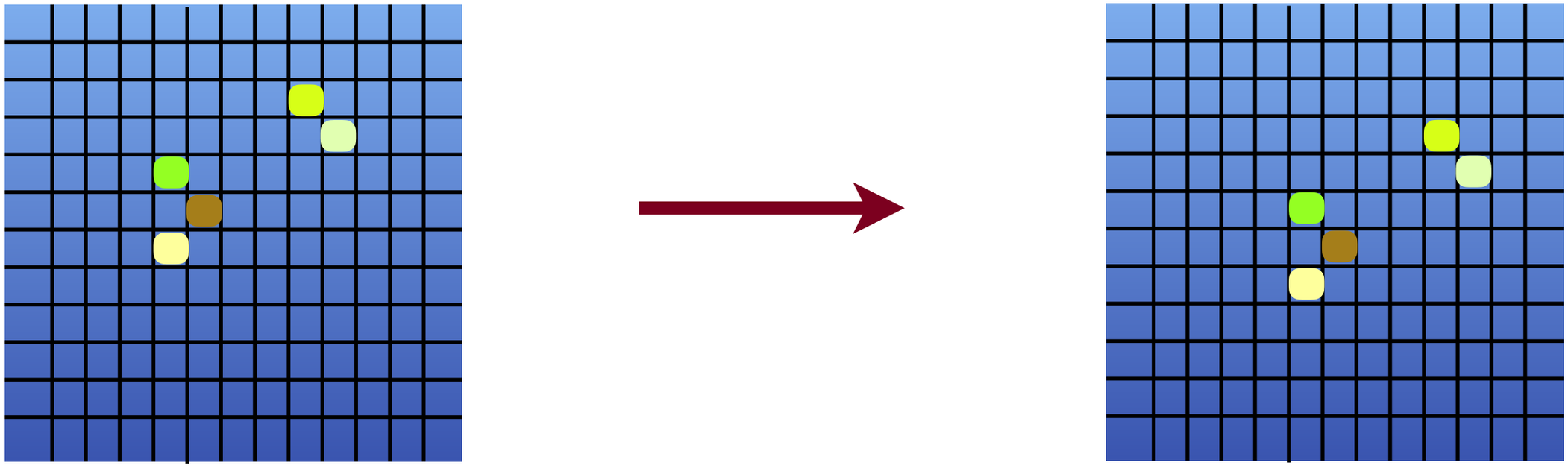


The approximate position of the barycenter is correctly predicted

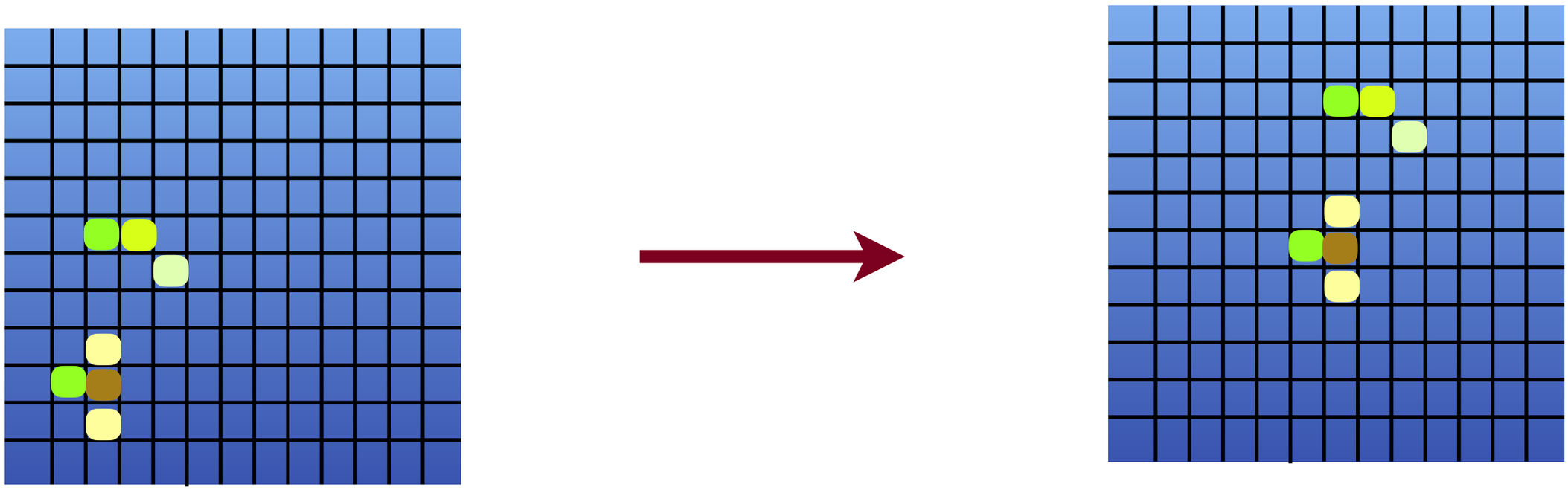


The statistical performance on A test sample

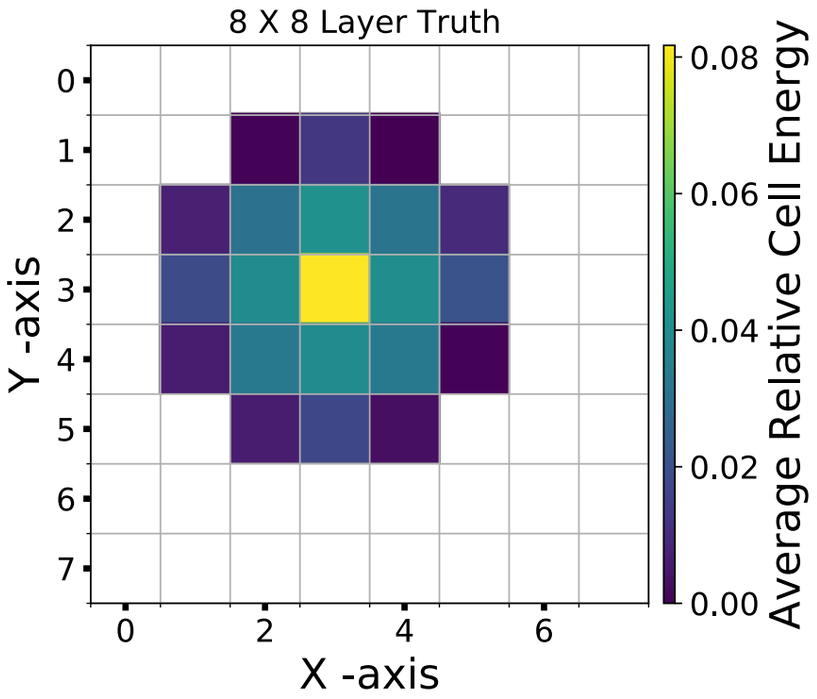
A statistical quantification of two-prong decay

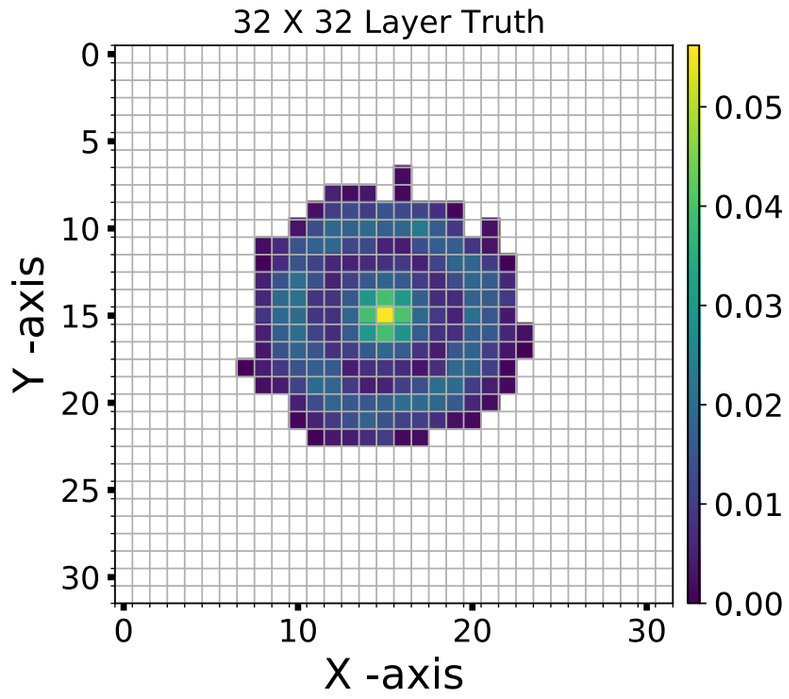


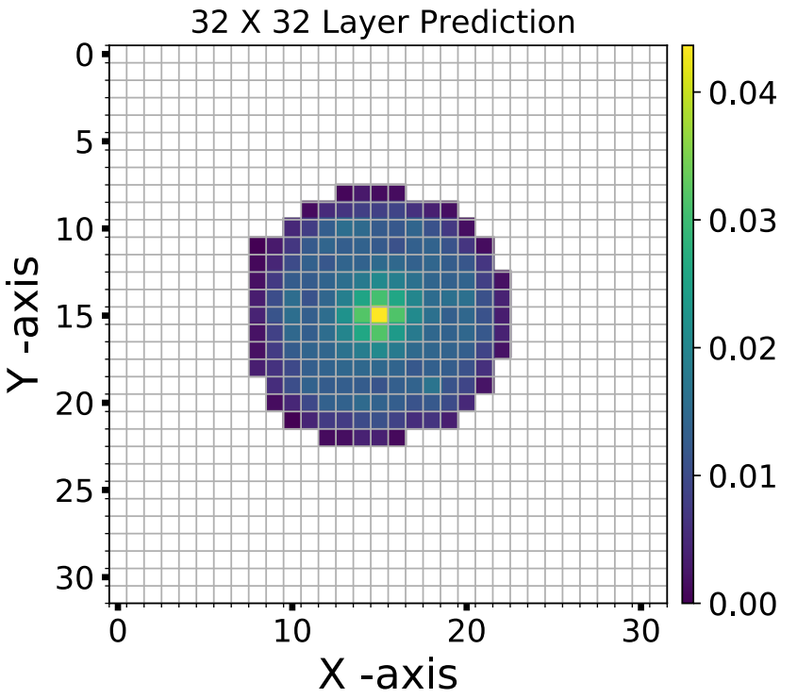
Centered around highest energetic cell



An average of many centered images

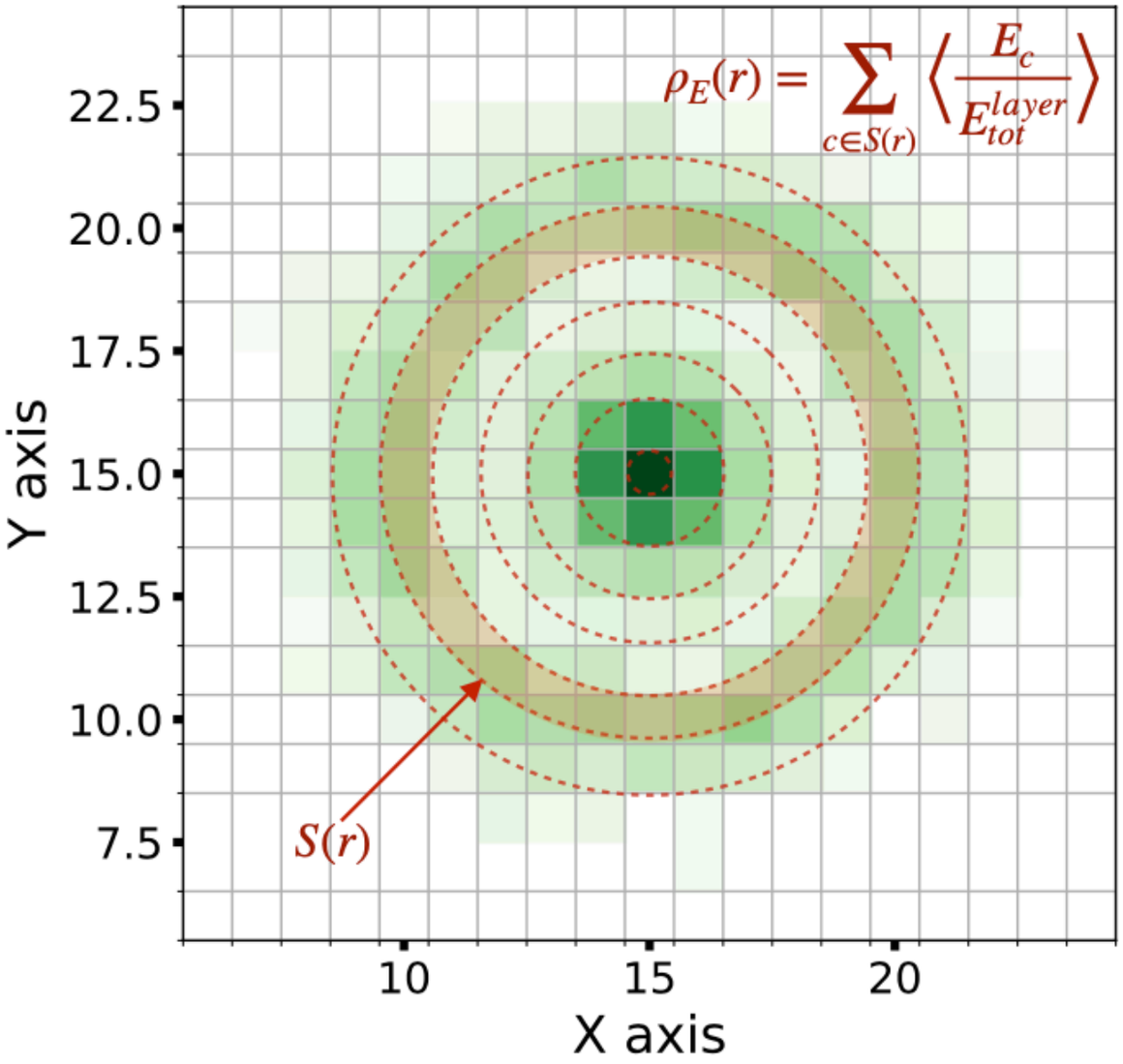


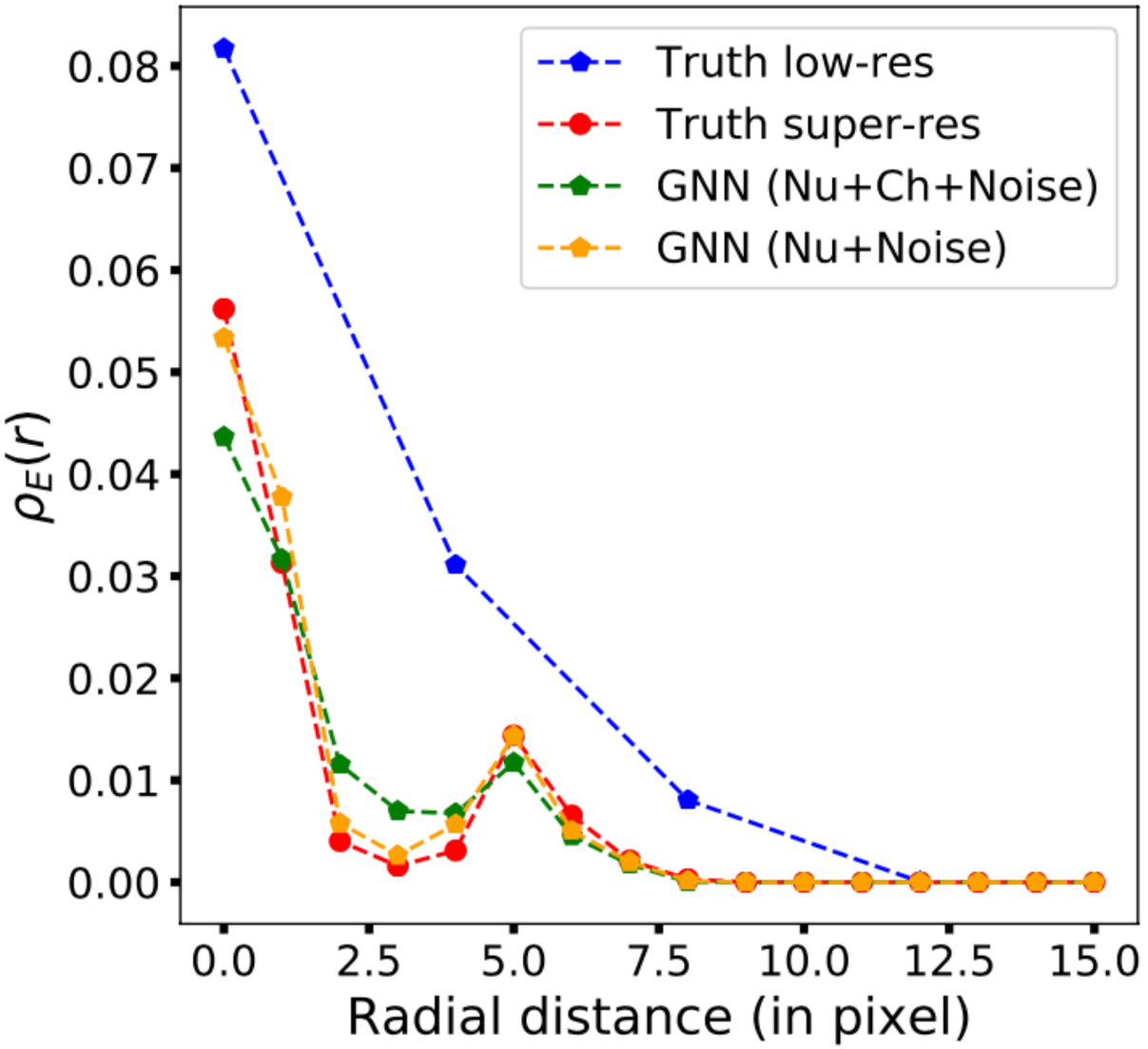




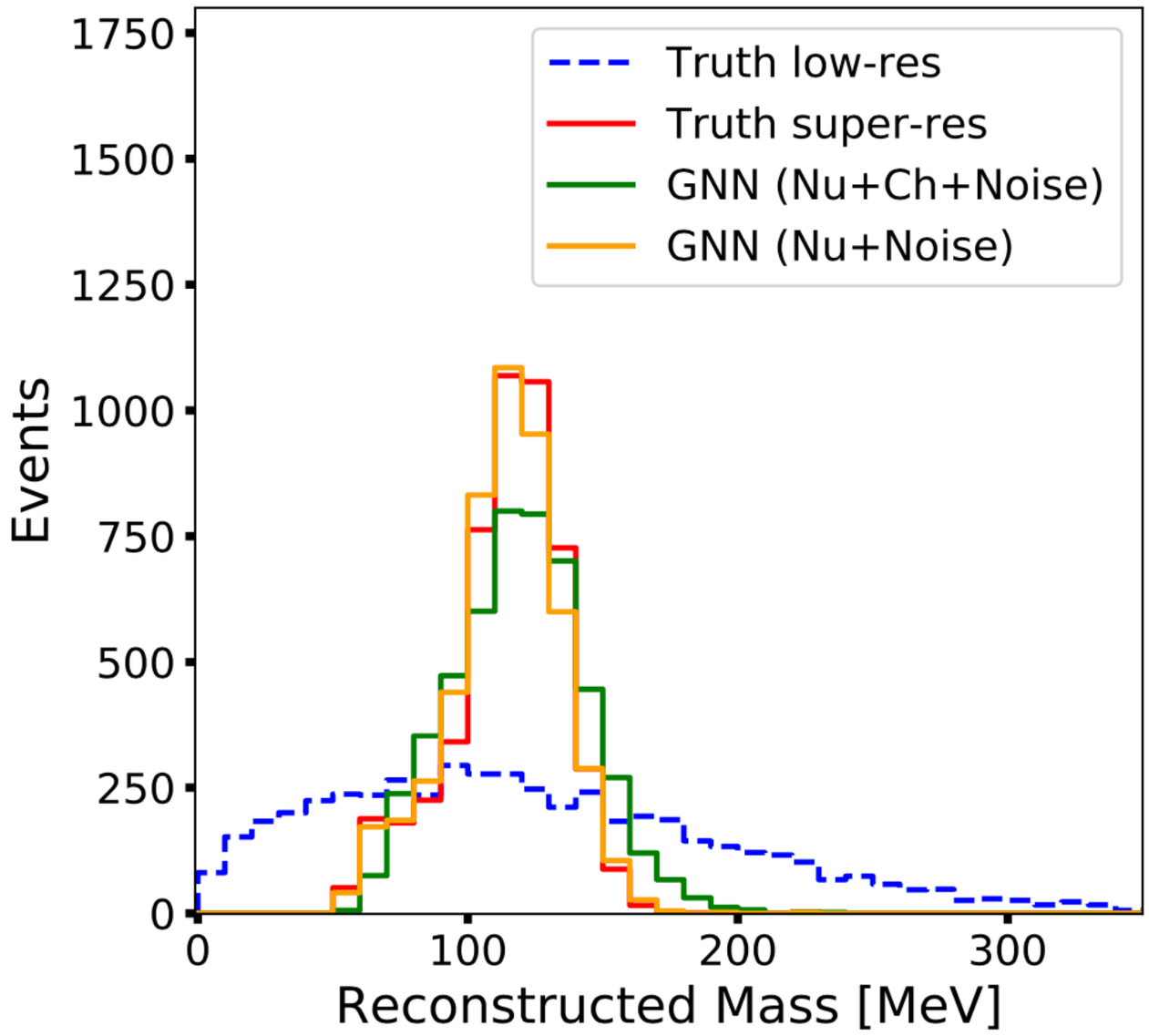
A secondary rim is present in super-resoluted image, but absent in the low-res

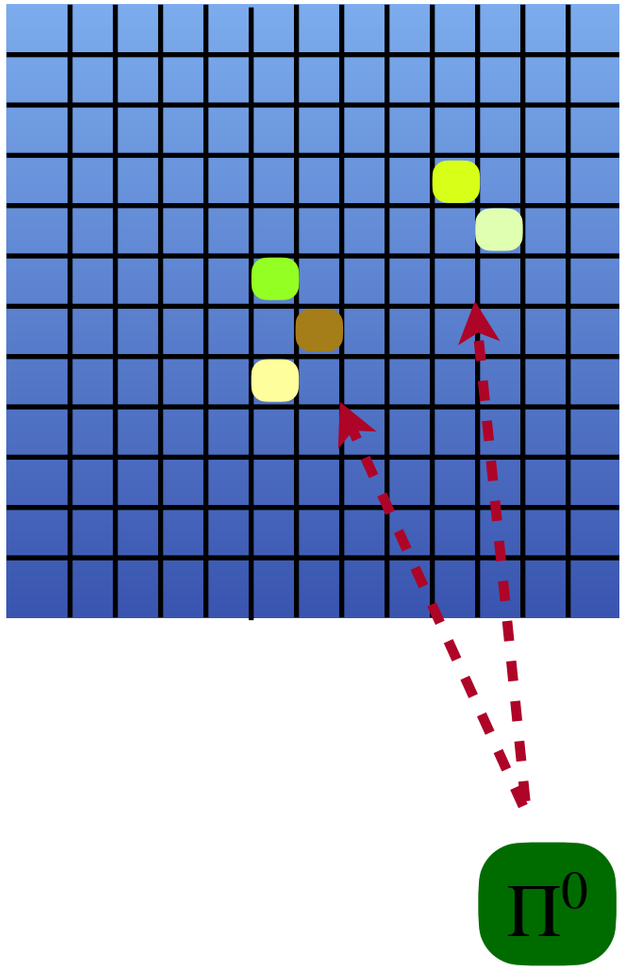
The integrated profile





The mass distribution





Invariant mass from reconstructed 4-vectors.

Summary

☑ We have demonstrated that a suitable ConvNet, Graph, Deepset architecture gives descent energyfraction estimation for the generalized case :

Input –> Variable Resolution + Noise + Track, Output –> Real resolution.

☑ The algorithm actually succeeds in yielding a complete image of neutral energy profile of the layers.

☑ The trained NN is able to learn and predict the noise pattern.

A network trained on topoclusters has better performance on the topoclusters.

☑ These ML based algorithms are shown to improve the energy and direction estimation over existing PFlow algorithm, in case of overlapping charged and neutral pions.

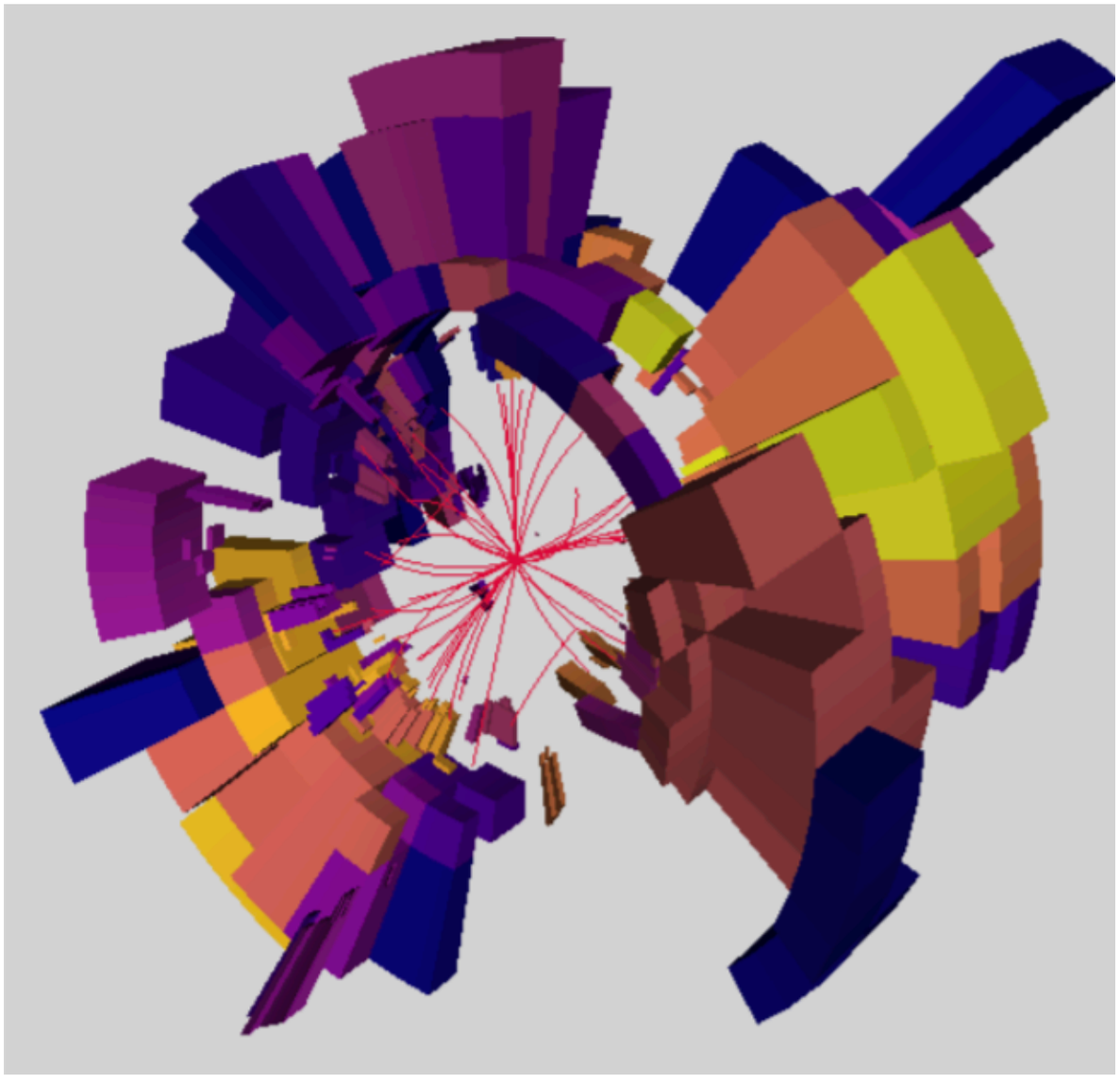
3 to 5 times resolution improvement obtained at low energy regime.

☑ Demonstrated the applicability of super-resolution techniques for calorimetric study.

☑ The ML algorithm at its current form is applicable to ATLAS as it is based on topoclusters.

Will perform further studies in complex environment before implementation.

And what's next in this direction ?



THANK YOU !!

Backup

Detector parameters & noise

Detector	Absorber	Scintillator	Subdetector (Legth)
ECAL	Lead **1.2**	Liquid Argon **4.5**	ECAL1 (**3 X_0**) ECAL2 (**16 X_0**) ECAL3 (**6 X_0**)
HCAL	Iron **4.7**	Plastic organic **1.0**	HCAL1 (**1.5 λ_{int}**) HCAL2 (**4.1 λ_{int}**) HCAL3 (**1.8 λ_{int}**)

$X_0 = 3.9$ cm

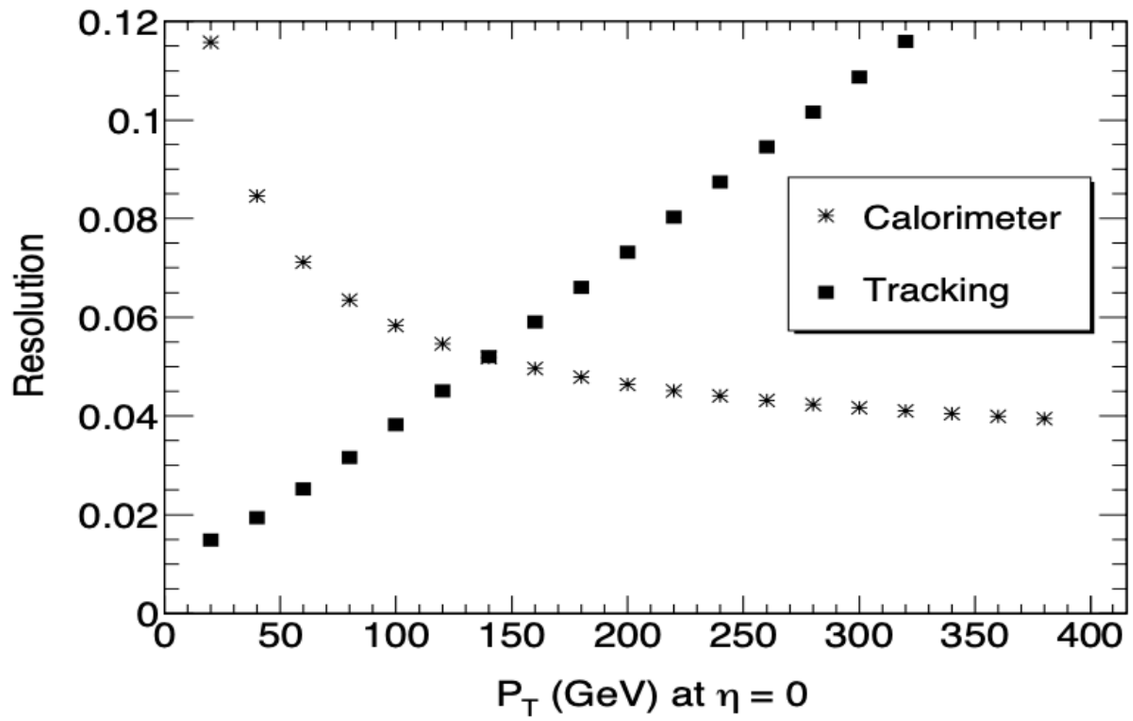
$\lambda_{int} = 17.4$ cm

Particles are generated in four different energy ranges [2-5], [5-10], [10-15] & [15-20] GeV.

In these energy regimes tracker has better resolution compared to calorimeters.

Noise is added at per cell level
(used current ~ ATLAS values which includes PU)

Detector Layer	Res. (HG)	Res. (LG)	Noise [MeV] (cf)
ECAL1	64×64	32×32	13 (4)
ECAL2	32×32	8×8	34 (16)
ECAL3	32×32	8×8	17 (16)
HCAL1	16×16	8×8	14 (4)
HCAL2	16×16	8×8	8 (4)
HCAL3	8×8	8×8	14 (1)

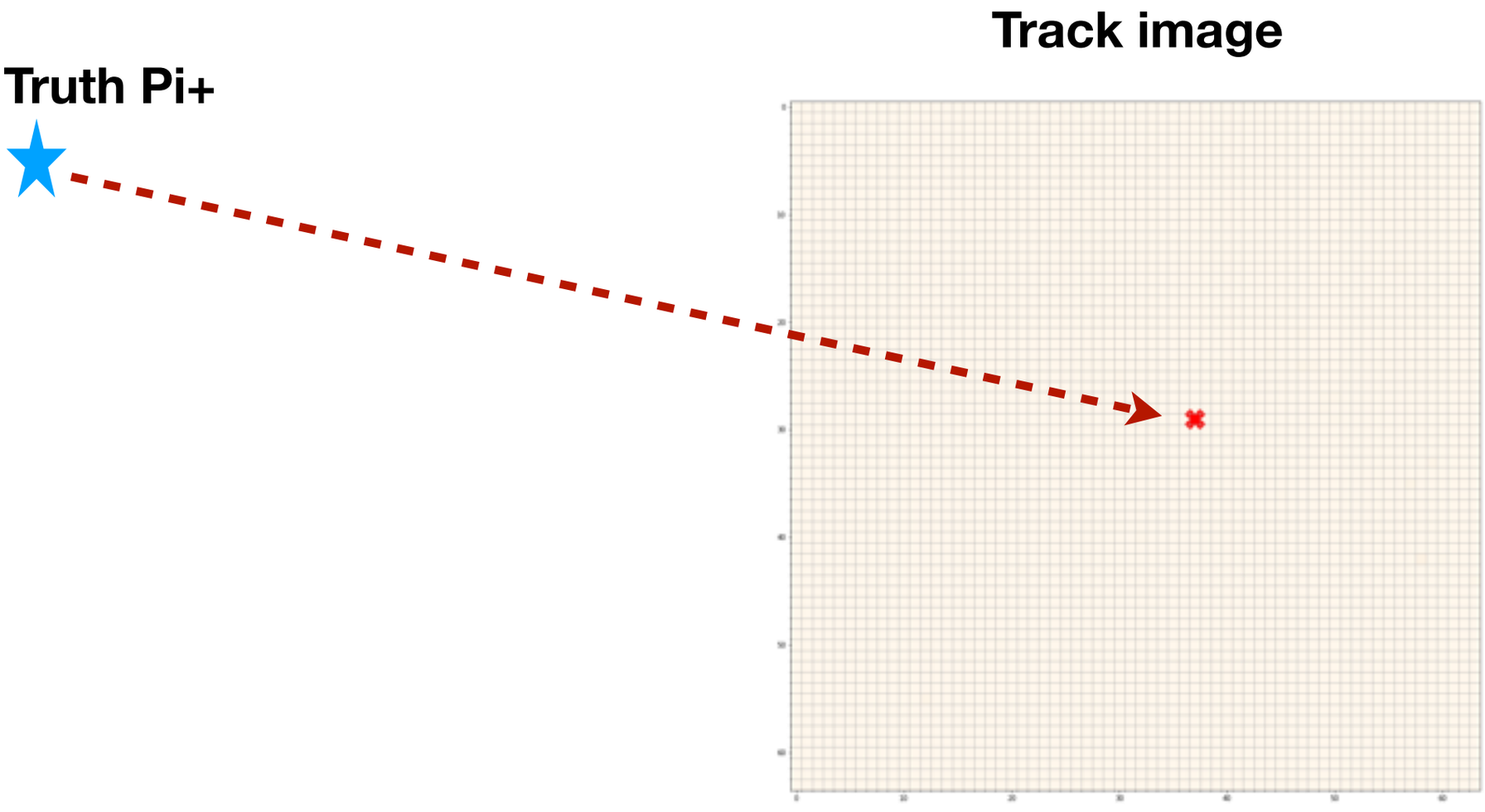


Introducing Track image

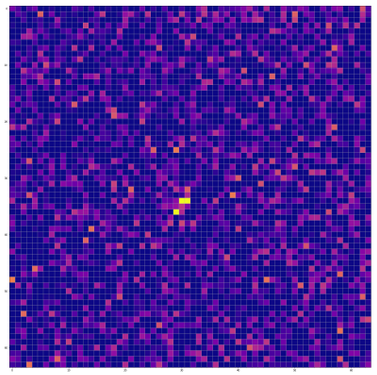
Forming the track :

smear the truth Pi+ momenta : $\frac{\sigma(p_T)}{p_T} = 5 \times 10^{-4} \times p_T$ **[GeV]**

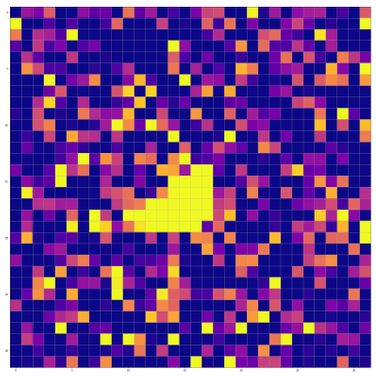
Keep the direction of Pi+ fixed (no magnetic field)



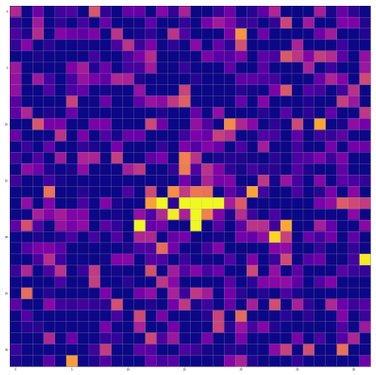
What does an event look like ?



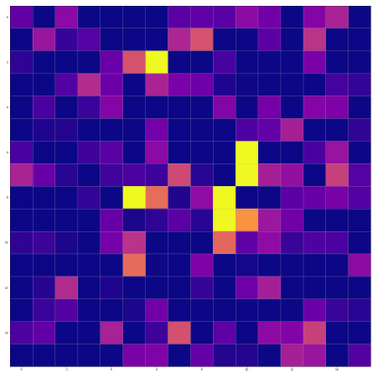
ECAL-1



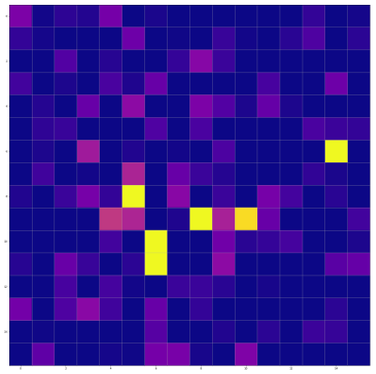
ECAL-2



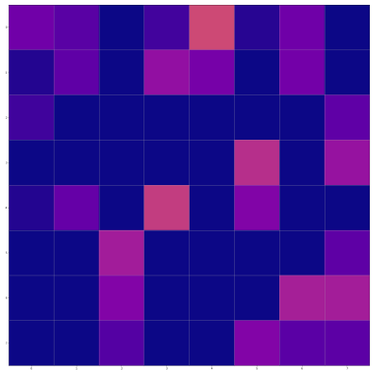
ECAL-3



HCAL-1



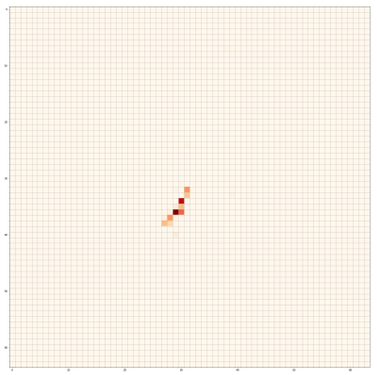
HCAL-2

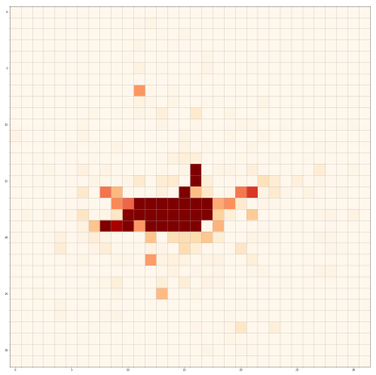


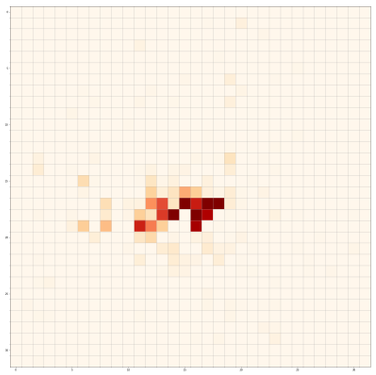
HCAL-3

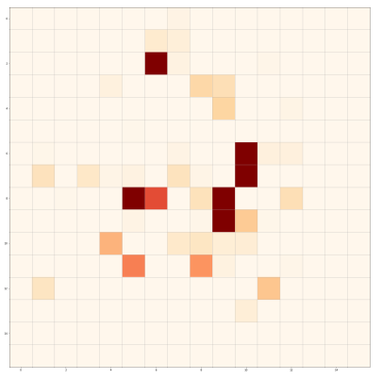
Energy distribution in Calorimeter layers

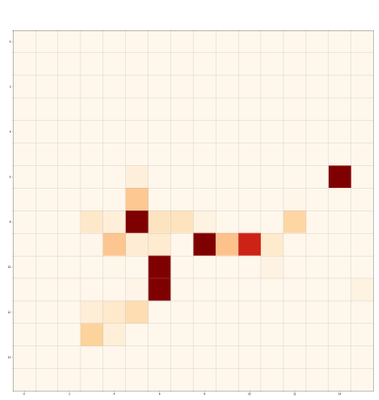
Beam direction ($E \in [15, 20]$ GeV)

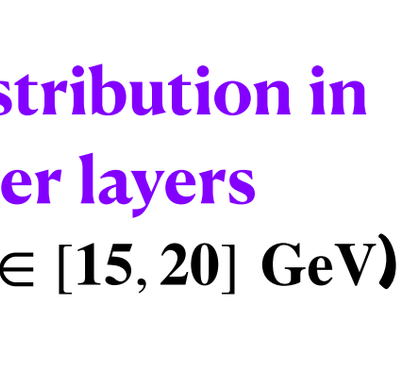


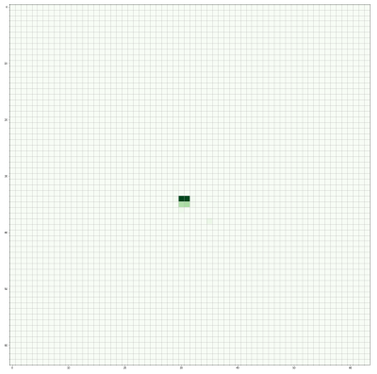


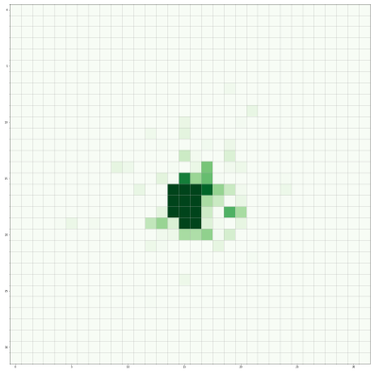


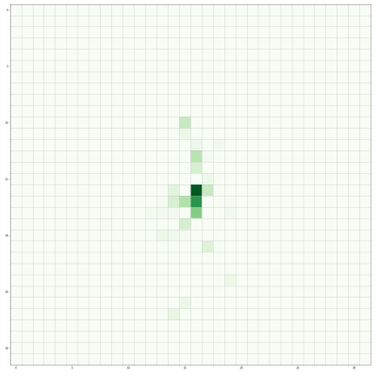


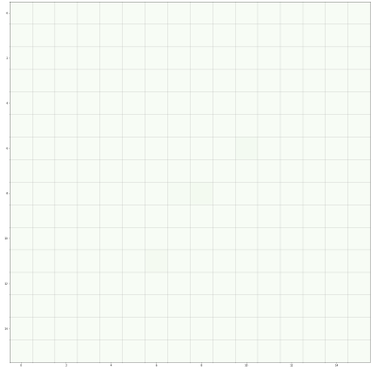


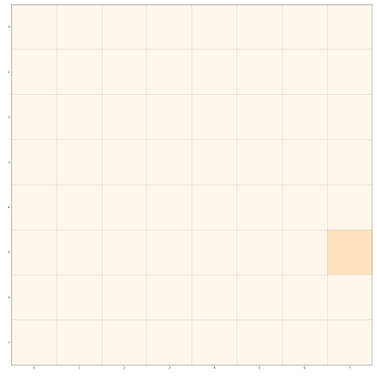






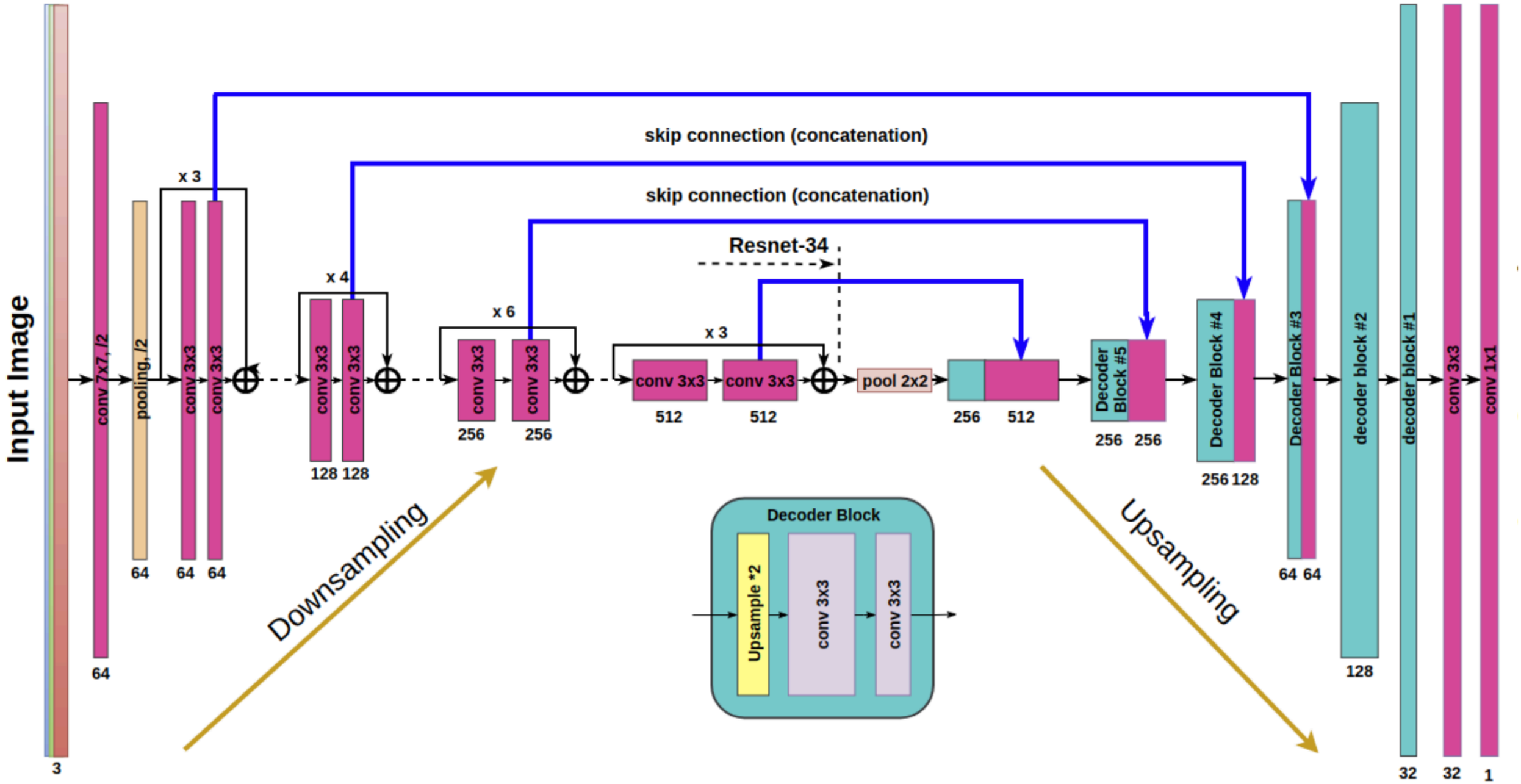






The architecture (cPFlow + Res-UNet)

Image from : arXiv 1806.05182



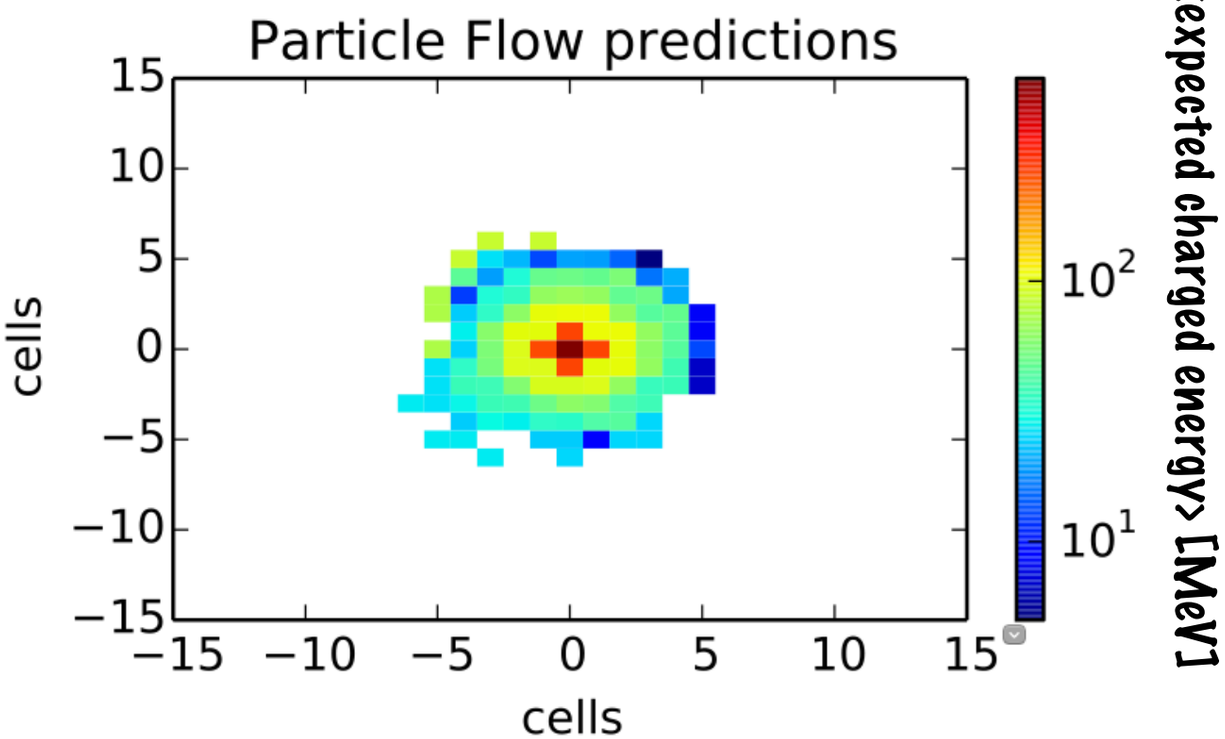
The Parametrized ParticleFlow (pPFlow)

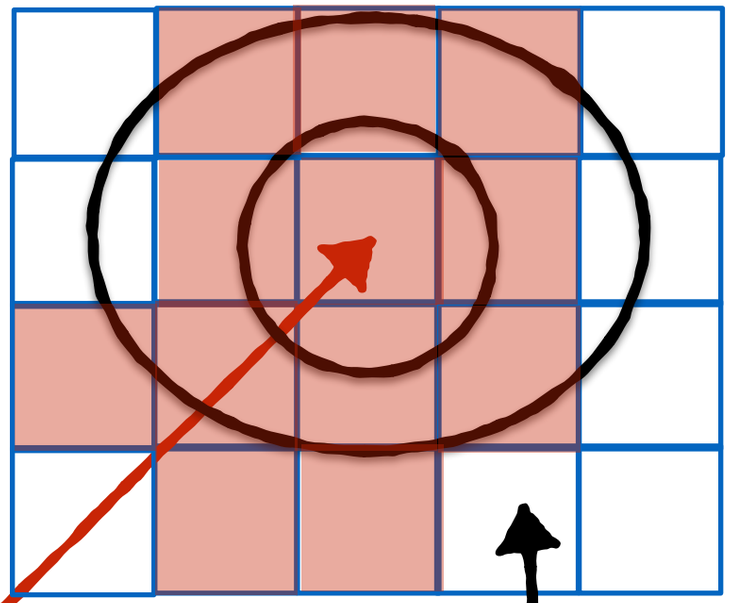
1. Group together the cells from their energy —-> topoclustering.
   First identify the seed cells ($\frac{E}{\sigma} \geq 5$) and merge with it neighboring cells with $\frac{E}{\sigma} \geq 2$
   Needed to remove the noise and cluster together cells fired by a common source (pi+, pi0)

2. Parametrize the expected energy given the momentum of the track (measured from the inner tracker, the layer of the "first" nuclear interaction).

3. The expected energy is subtracted from the topocluster and the remaining energy is considered as neutral energy

Example of the parametrization in the first calorimeter layer





Subtract expected energy in circle within the topocluster. The leftovers energy is considered as the neutral component

This method is used as a baseline for comparison