ZDC Simulation with ML

Wen-Chen Chang

2024/12/6

* Date: Dec 6 (Friday), 2024

* Time: 10:00-11:00 AM at Taiwan (GMT+8)

* Zoom link:

https://cern.zoom.us/j/66342263280?pwd=DBemHUOnO6QIiQyU5y2WbeaEaBGcyT.1

CaloChallenge 2022: A Community Challenge for Fast Calorimeter Simulation

https://arxiv.org/abs/2410.21611

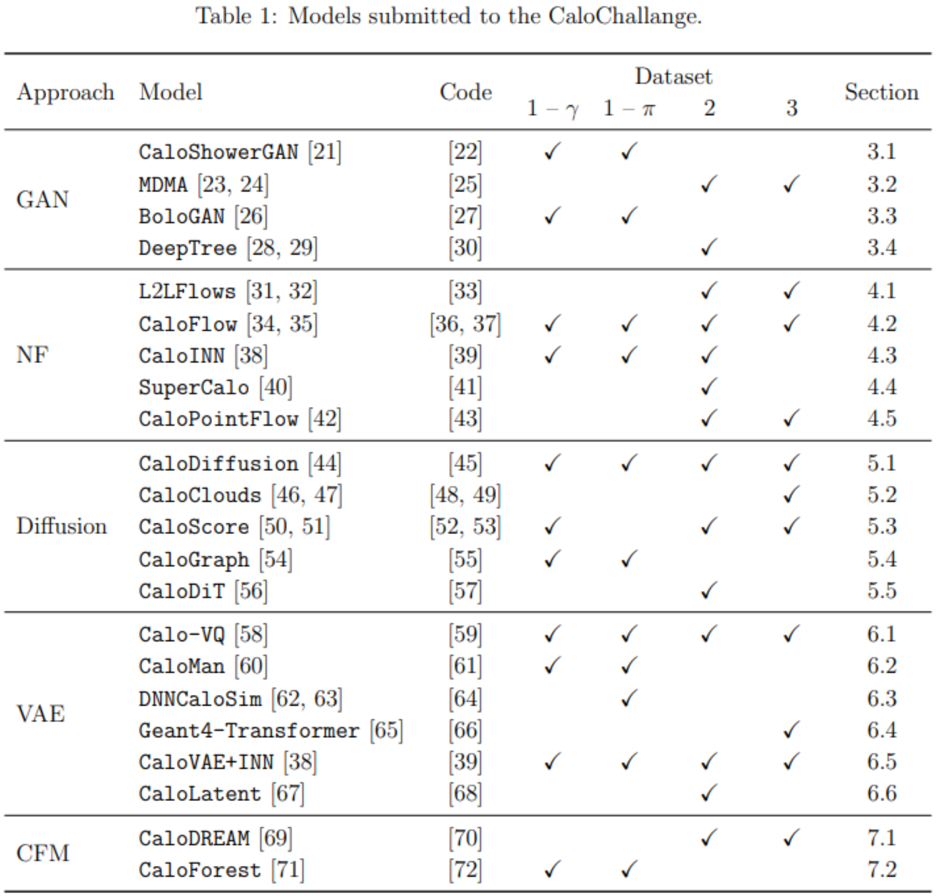
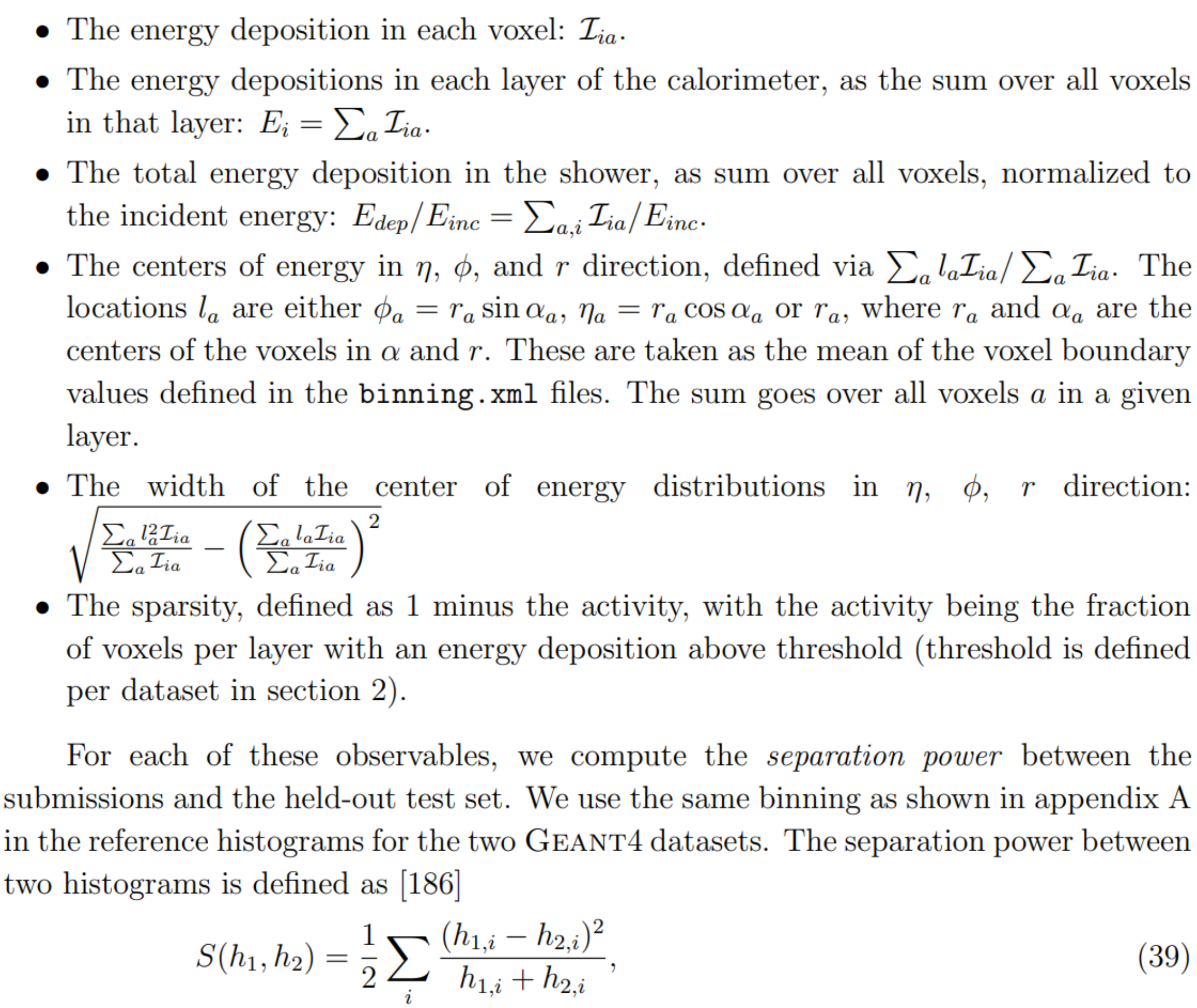


Table 1: Models submitted to the CaloChallange.

Approach	Model	Code	Dataset				Section
			$1-\gamma$	$1-\pi$	2	3	
GAN	CaloShowerGAN [21]	[22]	✓	✓			3.1
	MDMA [23, 24]	[25]			✓	✓	3.2
	BoloGAN [26]	[27]	✓	✓			3.3
	DeepTree [28, 29]	[30]			✓		3.4
NF	L2LFlows [31, 32]	[33]			✓	✓	4.1
	CaloFlow [34, 35]	[36, 37]	✓	✓	✓	✓	4.2
	CaloINN [38]	[39]	✓	✓	✓		4.3
	SuperCalo [40]	[41]			✓		4.4
	CaloPointFlow [42]	[43]			✓	✓	4.5
Diffusion	CaloDiffusion [44]	[45]	✓	✓	✓	✓	5.1
	CaloClouds [46, 47]	[48, 49]				✓	5.2
	CaloScore [50, 51]	[52, 53]	✓		✓	✓	5.3
	CaloGraph [54]	[55]	✓	✓			5.4
	CaloDiT [56]	[57]			✓		5.5
VAE	Calo-VQ [58]	[59]	✓	✓	✓	✓	6.1
	CaloMan [60]	[61]	✓	✓			6.2
	DNNCaloSim [62, 63]	[64]		✓			6.3
	Geant4-Transformer [65]	[66]				✓	6.4
	CaloVAE+INN [38]	[39]	✓	✓	✓	✓	6.5
	CaloLatent [67]	[68]			✓		6.6
CFM	CaloDREAM [69]	[70]			✓	✓	7.1
	CaloForest [71]	[72]	✓	✓			7.2

Metric: High-level Features (Histograms)

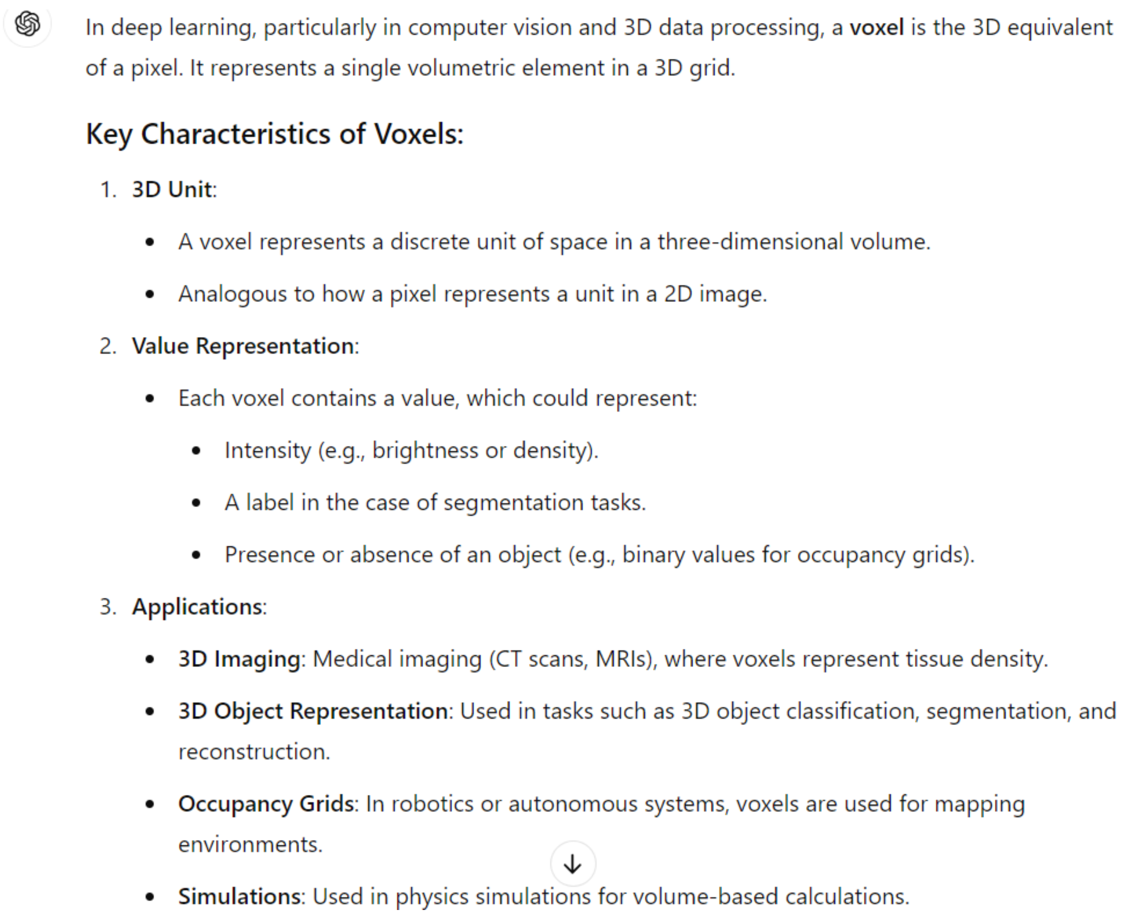


- The energy deposition in each voxel: $\mathcal{I}_{ia}$.
- The energy depositions in each layer of the calorimeter, as the sum over all voxels in that layer: $E_i = \sum_a \mathcal{I}_{ia}$.
- The total energy deposition in the shower, as sum over all voxels, normalized to the incident energy: $E_{dep}/E_{inc} = \sum_{a,i} \mathcal{I}_{ia}/E_{inc}$.
- The centers of energy in η, ϕ, and r direction, defined via $\sum_a l_a \mathcal{I}_{ia} / \sum_a \mathcal{I}_{ia}$. The locations l_a are either $\phi_a = r_a \sin\alpha_a$, $\eta_a = r_a \cos\alpha_a$ or r_a, where r_a and α_a are the centers of the voxels in α and r. These are taken as the mean of the voxel boundary values defined in the `binning.xml` files. The sum goes over all voxels a in a given layer.
- The width of the center of energy distributions in η, ϕ, r direction: $\sqrt{\frac{\sum_a l_a^2 \mathcal{I}_{ia}}{\sum_a \mathcal{I}_{ia}} - \left(\frac{\sum_a l_a \mathcal{I}_{ia}}{\sum_a \mathcal{I}_{ia}}\right)^2}$
- The sparsity, defined as 1 minus the activity, with the activity being the fraction of voxels per layer with an energy deposition above threshold (threshold is defined per dataset in section 2).

For each of these observables, we compute the *separation power* between the submissions and the held-out test set. We use the same binning as shown in appendix A in the reference histograms for the two GEANT4 datasets. The separation power between two histograms is defined as [186]

$$S(h_1, h_2) = \frac{1}{2} \sum_i \frac{(h_{1,i} - h_{2,i})^2}{h_{1,i} + h_{2,i}}, \tag{39}$$

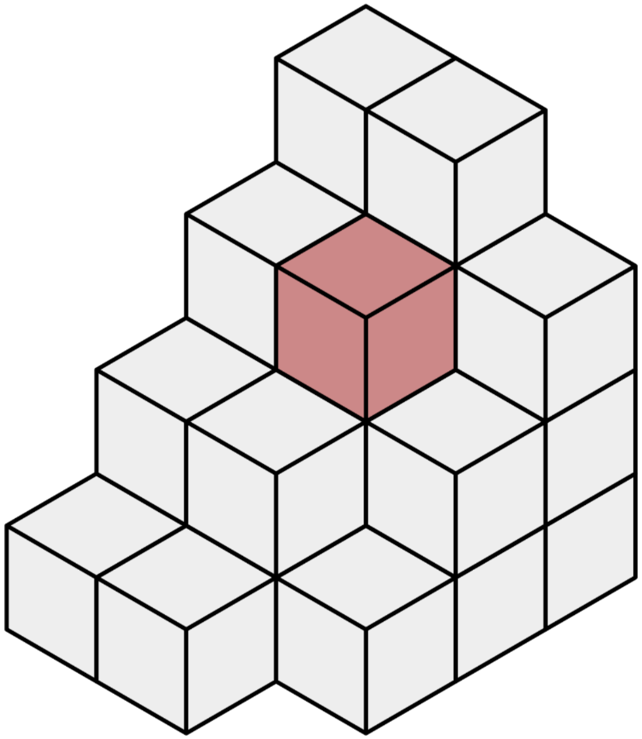
Voxel



In deep learning, particularly in computer vision and 3D data processing, a **voxel** is the 3D equivalent of a pixel. It represents a single volumetric element in a 3D grid.

Key Characteristics of Voxels:

1. **3D Unit**:
   - A voxel represents a discrete unit of space in a three-dimensional volume.
   - Analogous to how a pixel represents a unit in a 2D image.
2. **Value Representation**:
   - Each voxel contains a value, which could represent:
     - Intensity (e.g., brightness or density).
     - A label in the case of segmentation tasks.
     - Presence or absence of an object (e.g., binary values for occupancy grids).
3. **Applications**:
   - **3D Imaging**: Medical imaging (CT scans, MRIs), where voxels represent tissue density.
   - **3D Object Representation**: Used in tasks such as 3D object classification, segmentation, and reconstruction.
   - **Occupancy Grids**: In robotics or autonomous systems, voxels are used for mapping environments.
   - **Simulations**: Used in physics simulations for volume-based calculations.



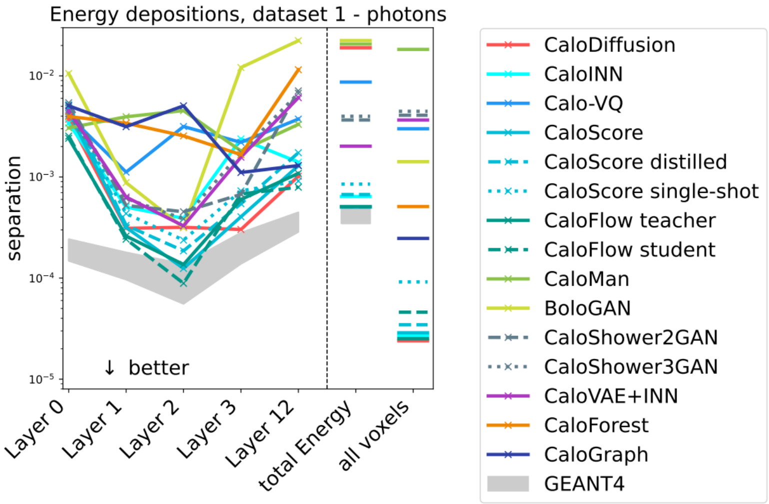


Figure 32: Separation power of energy depositions with threshold at 1 MeV.

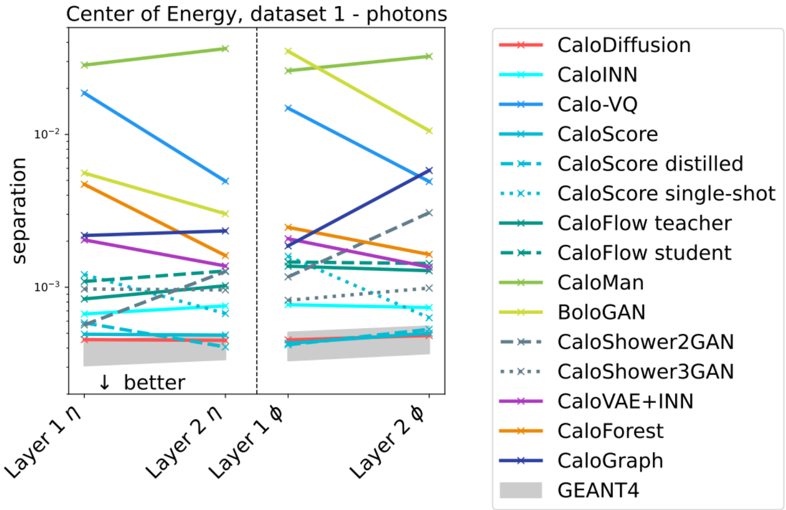


Figure 33: Separation power of centers of energy with threshold at 1 MeV.

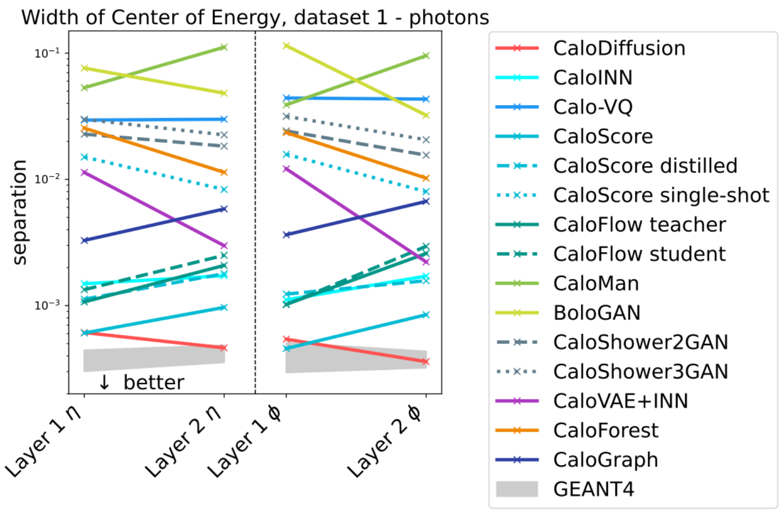


Figure 34: Separation power of widths of centers of energy with threshold at 1 MeV.

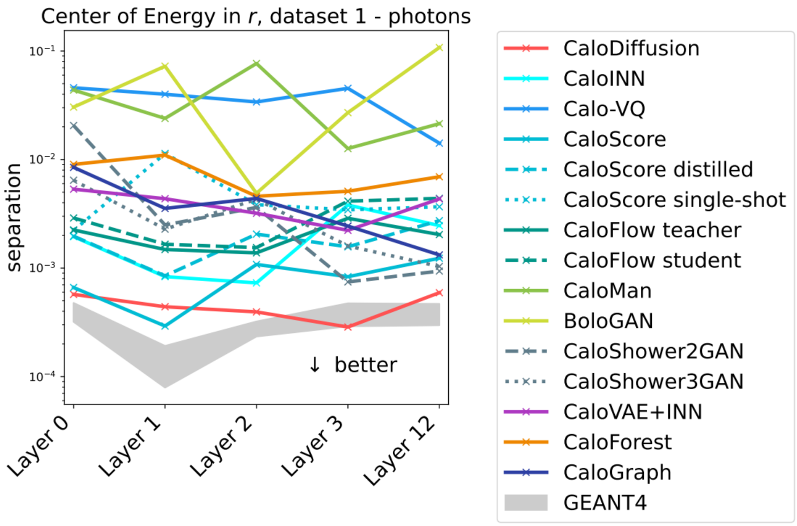
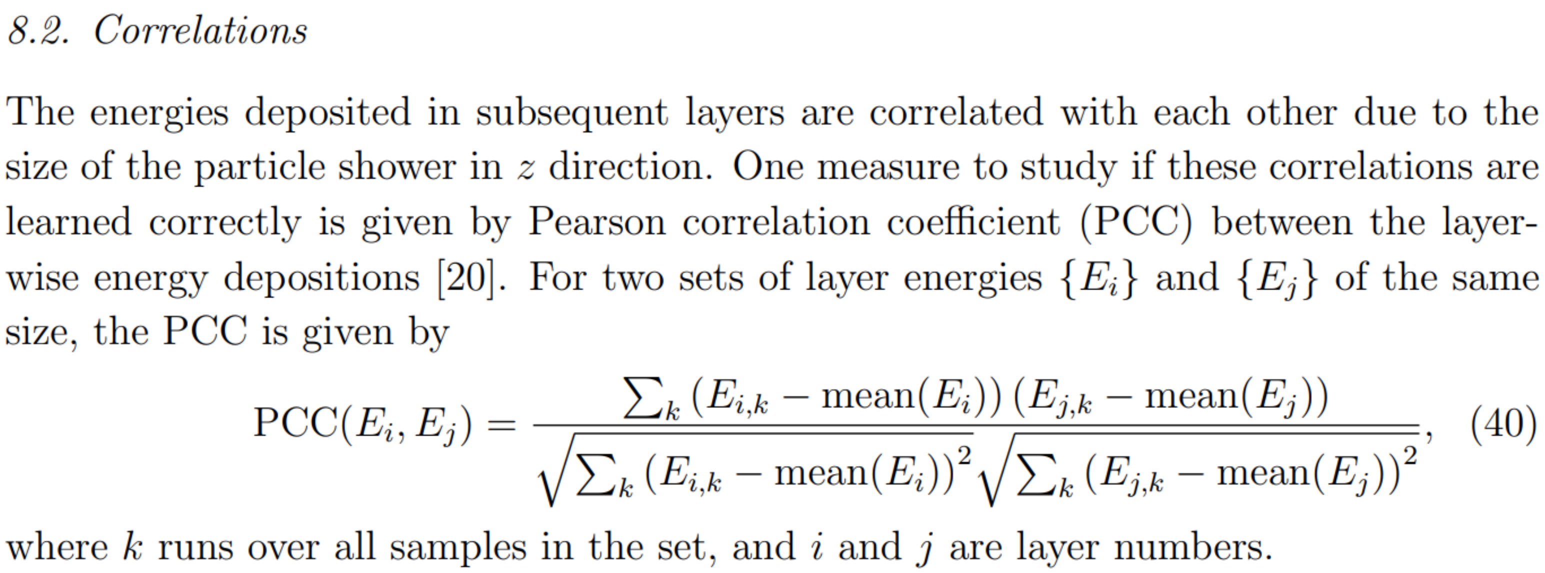


Figure 35: Separation power of centers of energy with threshold at 1 MeV.

Metric: Pearson Correlation Coefficient (PCC)



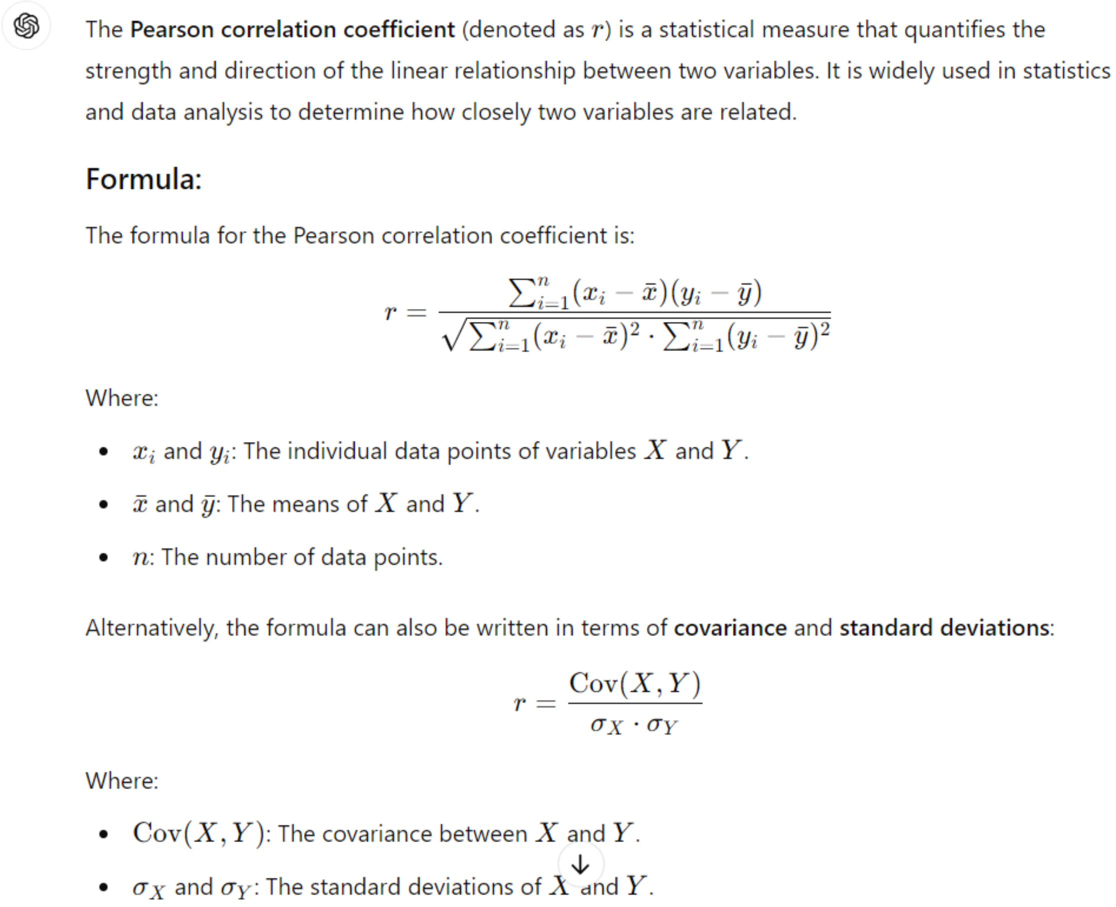
8.2. Correlations

The energies deposited in subsequent layers are correlated with each other due to the size of the particle shower in z direction. One measure to study if these correlations are learned correctly is given by Pearson correlation coefficient (PCC) between the layer-wise energy depositions [20]. For two sets of layer energies $\{E_i\}$ and $\{E_j\}$ of the same size, the PCC is given by

$$\mathrm{PCC}(E_i, E_j) = \frac{\sum_k \left(E_{i,k} - \mathrm{mean}(E_i)\right)\left(E_{j,k} - \mathrm{mean}(E_j)\right)}{\sqrt{\sum_k \left(E_{i,k} - \mathrm{mean}(E_i)\right)^2}\sqrt{\sum_k \left(E_{j,k} - \mathrm{mean}(E_j)\right)^2}}, \quad (40)$$

where k runs over all samples in the set, and i and j are layer numbers.

Pearson Correlation Coefficient (PCC)



The **Pearson correlation coefficient** (denoted as r) is a statistical measure that quantifies the strength and direction of the linear relationship between two variables. It is widely used in statistics and data analysis to determine how closely two variables are related.

Formula:

The formula for the Pearson correlation coefficient is:

$$r = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2 \cdot \sum_{i=1}^{n}(y_i - \bar{y})^2}}$$

Where:

- x_i and y_i: The individual data points of variables X and Y.
- $\bar{x}$ and $\bar{y}$: The means of X and Y.
- n: The number of data points.

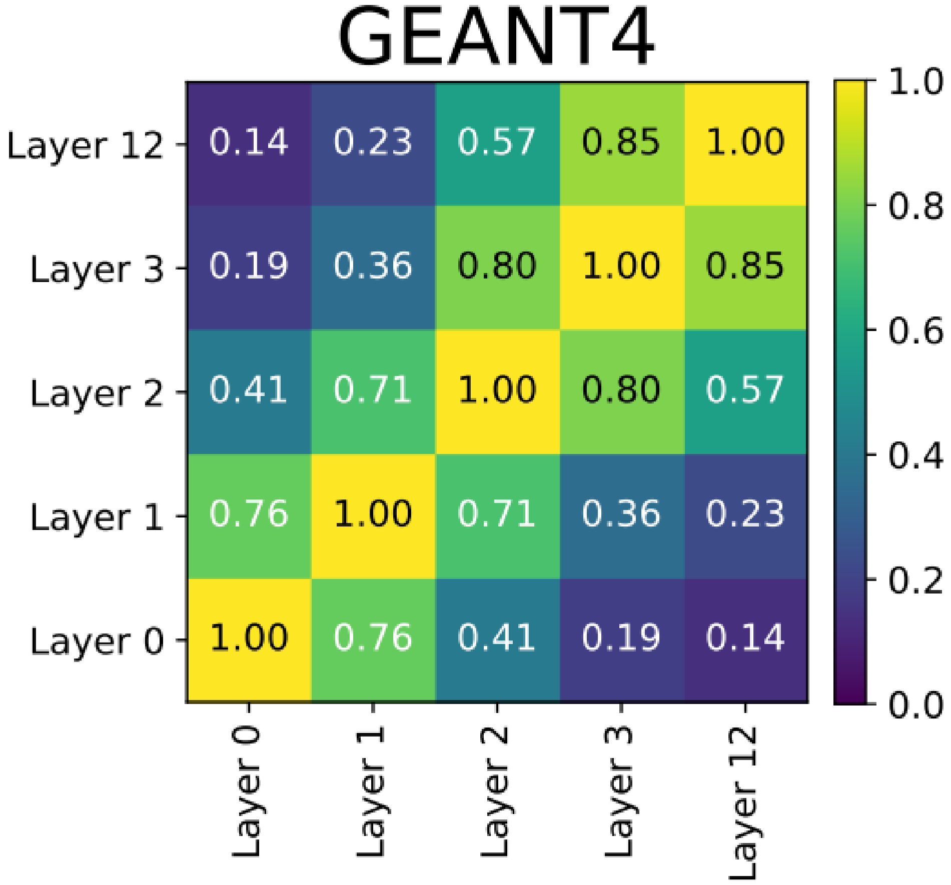
Alternatively, the formula can also be written in terms of **covariance** and **standard deviations**:

$$r = \frac{\mathrm{Cov}(X, Y)}{\sigma_X \cdot \sigma_Y}$$

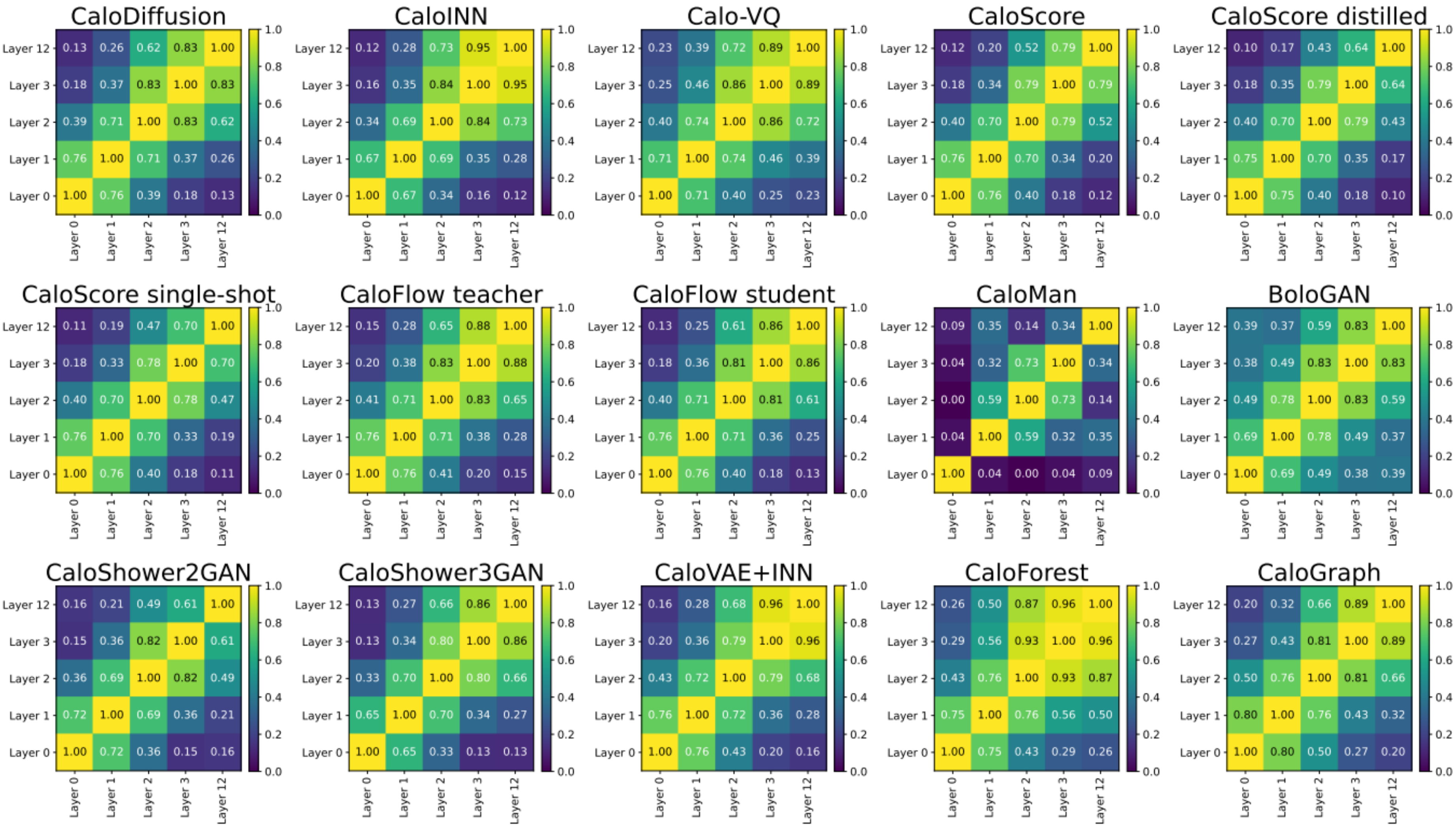
Where:

- $\mathrm{Cov}(X, Y)$: The covariance between X and Y.
- σ_X and σ_Y: The standard deviations of X and Y.

Pearson Correlation Coefficient (PCC)



	Layer 0	Layer 1	Layer 2	Layer 3	Layer 12
Layer 12	0.14	0.23	0.57	0.85	1.00
Layer 3	0.19	0.36	0.80	1.00	0.85
Layer 2	0.41	0.71	1.00	0.80	0.57
Layer 1	0.76	1.00	0.71	0.36	0.23
Layer 0	1.00	0.76	0.41	0.19	0.14



We now move on to investigate the correlations between the energies deposited in the layers in figure 38. Overall, most of the submissions reproduce the pattern induced by GEANT4 well, but there is a noticeable tendency of models to overestimate the correlation between layers 3 and 12, as seen in the top right corners. Some models based on GANs and VAEs, which had higher separation powers, also seem to have a harder time reproducing these correlations.