LAMBDA • HSE

Reconstruction of 3D Shower Structures for Neutrino Experiments

Vladislav Belavin
Ekaterina Trofimova
Andrey Ustyuzhanin

2021

Problem statement

LAMBDA • HSE

Clusterization task

Data:

$X - objects$

$X^l = \{x_i\}_{=1}^{l} - training\ sample$

Find:

$Y - cluster\ labels$

$a : X \rightarrow Y - clustering\ algorithm$

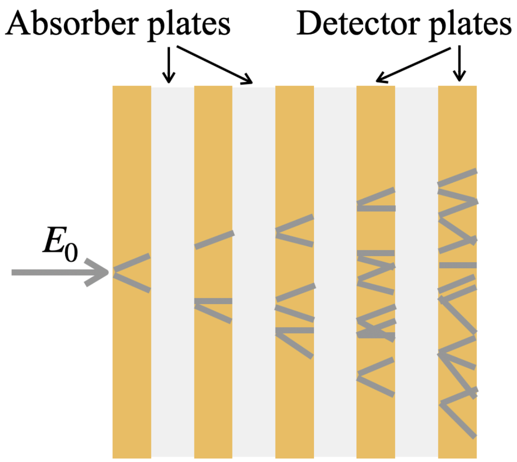
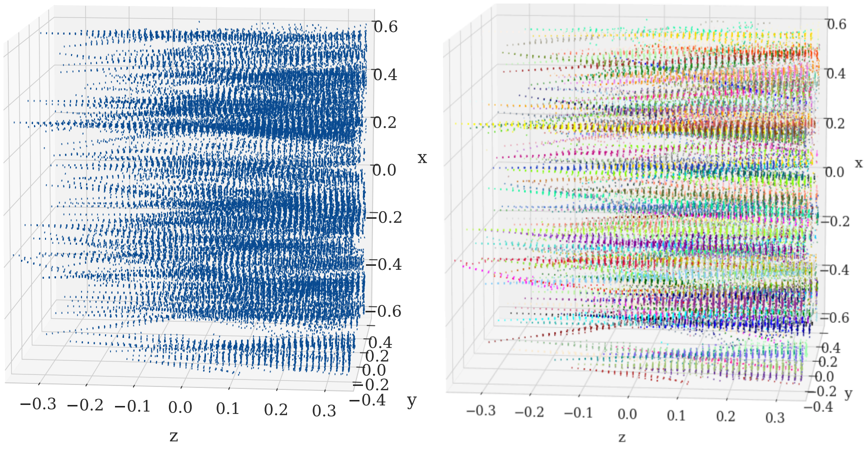


Figure 1. Scatch of EM shower development: neutrino decomposition products interactions with the emulsion films of detector.



(a) (b)

Figure 2. EM showers in a emulsion brick example: (a) row case (b) recovered showers case

Related work

B. Hosseini. *Search for Tau Neutrinos in the $\tau \to e$ Decay Channel in the OPERA experiment (2015)*

- only one shower with known origin
- $\sigma(\Delta E/E) \sim 0.21$

A.Ustyuzhanin, S. Shirobokov, V. Belavin, A. Filatov. *Machine-Learning techniques for electro-magnetic showers identification in OPERA datasets' (2017)*

- only one shower with unknown origin
- $\sigma(\Delta E/E) \sim 0.27$

Aim

To create an algorithm of overlapping EM showers reconstruction that allow physicists to restore the kinematics of the initial reaction, that is the starting point of lifting the veilon physics beyond the standard model

Objectives

1. To create a specific NN layer that propagate the information of initial particles to other detected points

2. To construct a new clasterization algorithm avoiding broken showers

3. To regress the Energy of initial particles

Data description

- 170 emulsion bricks (fig.2), 50-300 showers per brick (fig.3)
- 17 715 showers (fig. 4)

Each base-track is described by x, y, z *coordinates and* θ_x, θ_y *direction*

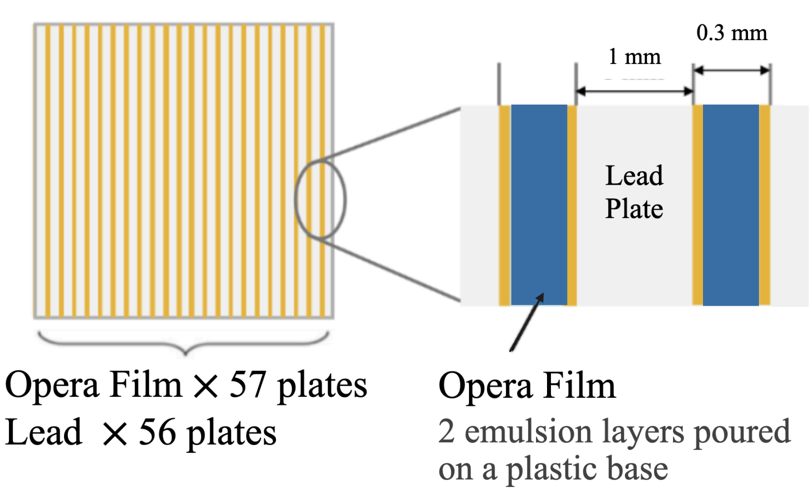


Figure 2. Sectional emulsion brick.

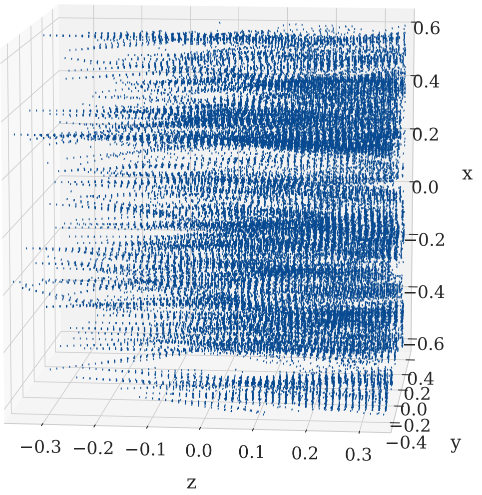


Figure 3. EM showers in emulsion brick example.

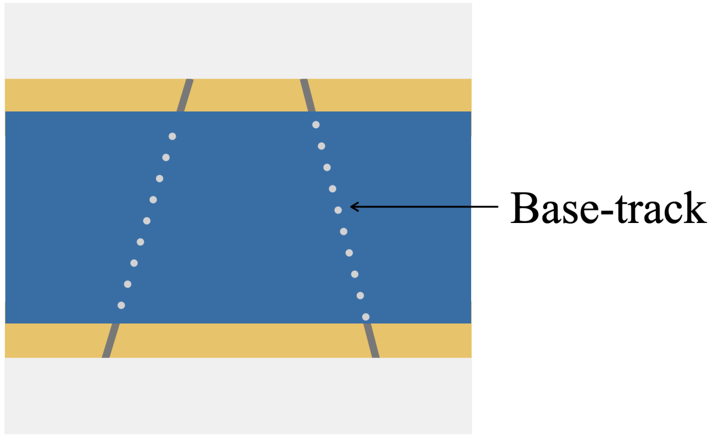


Figure 4. Base-track in emulsion film definition.

Metrics

LAMBDA • HSE

Energy resolution	$\Delta E/E$
Quality of initial position reconstruction	$MAE\ over\ x_0, y_0, z_0$
Quality of direction reconstruction	$MAE\ over\ \theta_x, \theta_y$
Ratio of recovered showers	$\dfrac{recovered\ showers}{total\ showers}$
Track classification (indirect metric)	$ROC - AUC$

Proposed Method (1). Preprocessing

The result of data preprocessing is a directed graph whose vertices correspond to tracks. Two nodes are connected with the edge iff:

1. one of the base-tracks lies in the cone of 16 mrad with origin in another base-track;
2. integrated distance (fig. 5) between tracks < threshold.

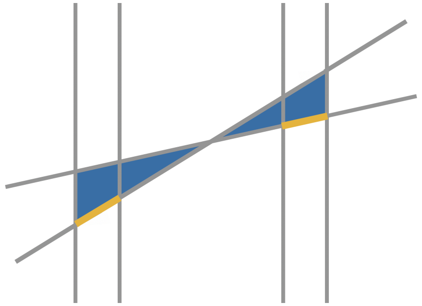


Figure 5. Integrated distance (blue area).

Language: Python

Link: https://github.com/ketrint/Electromagnetic_showers_reconstruction

Proposed Method (2). Edge Classification

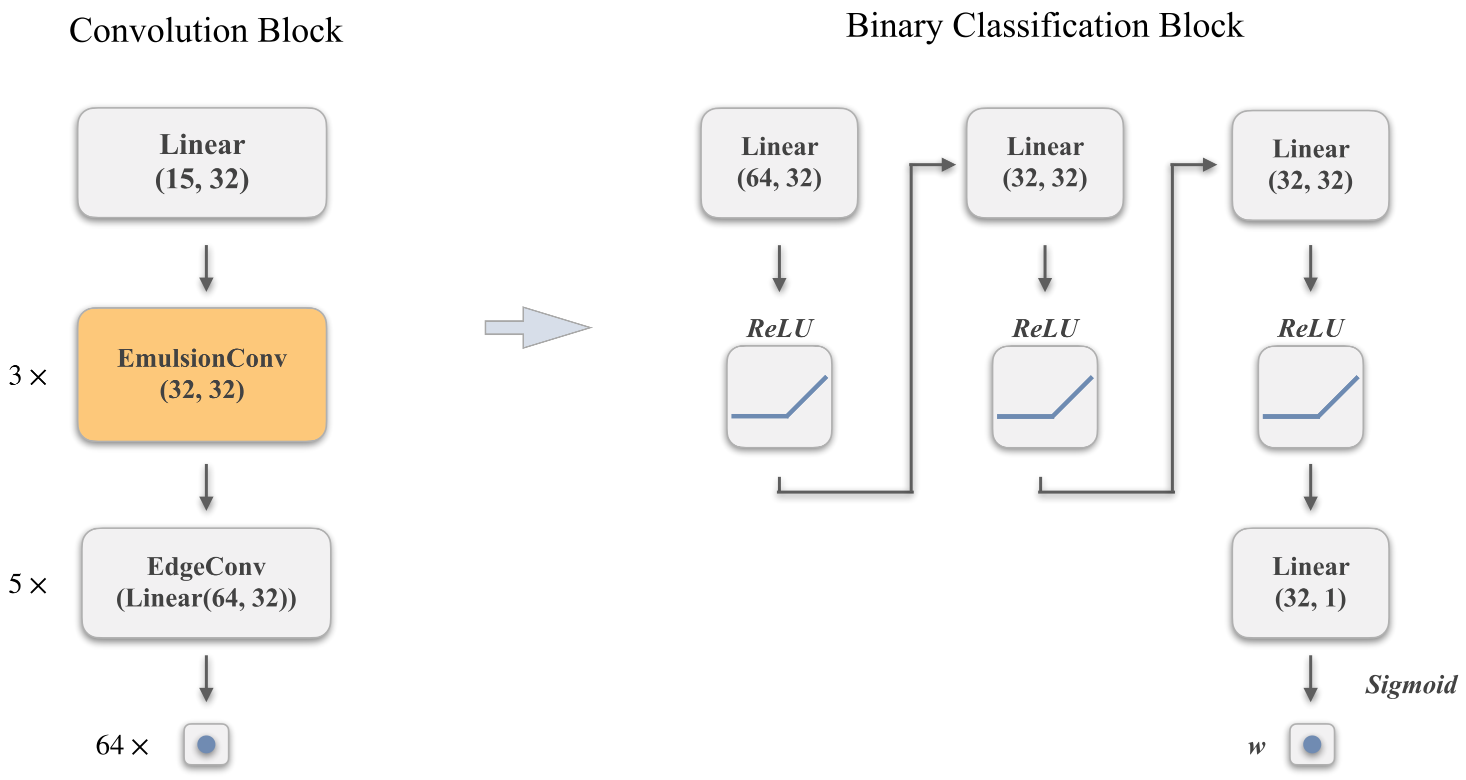


Figure 6. Edge Classification Neural Network Architecture.

Convolution block

EmulsionConv

applies the masks for activating only those edges that connect the base-tracks from subsequent emulsion films during the propagation

EdgeConv

maximum aggregation function

Binary Classification block

- If edge connects nodes from the same showers we mark it with weight ‘1’
- If edge connects nodes from different showers we mark it with weight ‘0’

Algorithm 1 EmulsionConv algorithm

Require: graph $(V, E) = \{v_h, e_{vh}\}$; M, U – neural networks

Ensure: updated graph $(V, E) = \{v_h, e_{vh}\}$

1: Group edges $E = \{(v_i, w_i)\}_{i=1}^{E}$ based on unique z_{w_i}. There would be 57 groups (for each emulsion layer) $\{g_k\}_{k=1}^{57}$
2: **for** each group g in increasing order of z **do**
3:     **for** each (v, w) in g **do**
4:         $m_{vw} = M(h_v, h_w, e_{vw})$
5:     **end for**
6:     $m_w = \mathrm{Agg}\{m_{vw}\}_{w \in N(v)}$
7:     $h_w = U(h_w, m_w)$
8: **end for**

Proposed Method (3). Clusterization

LAMBDA • HSE

Algorithm EWSCAN

Input: graph, representing a set of base-tracks;
$\min_{cl}$ – regulates the minimum cluster size at splitting;
threshold $\in [0, 1]$ – regulates the aggressiveness of splitting into clusters.
Output: a set of clusters
Method:

1. Set a score from edge classifier as pairwise distances between basetracks:

$$d(basetrack_i, basetrack_i) = w_i.$$

2. Define mutual reachability distance with parameter ‘k’:

$$d_{mreach_k} = max\{core_k(basetrack_i), core_k(basetrack_j), w_i\},$$

where $core_k(basetrack_i)$ is a distance from $basetrack_i$ to the k-nearest neighbor.

3. Construct minimum spanning tree, i.e. tree with the lowest sum of edges weights.

Repeat: delete edges with the highest weight

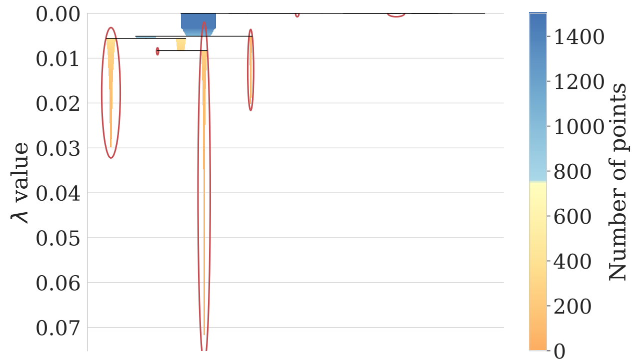


Figure 7. HDBSCAN condensed tree.

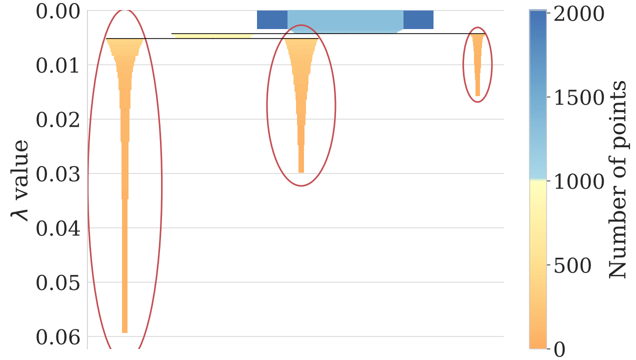


Figure 8. EWSCAN condensed tree.

Experiments

1. Architecture search

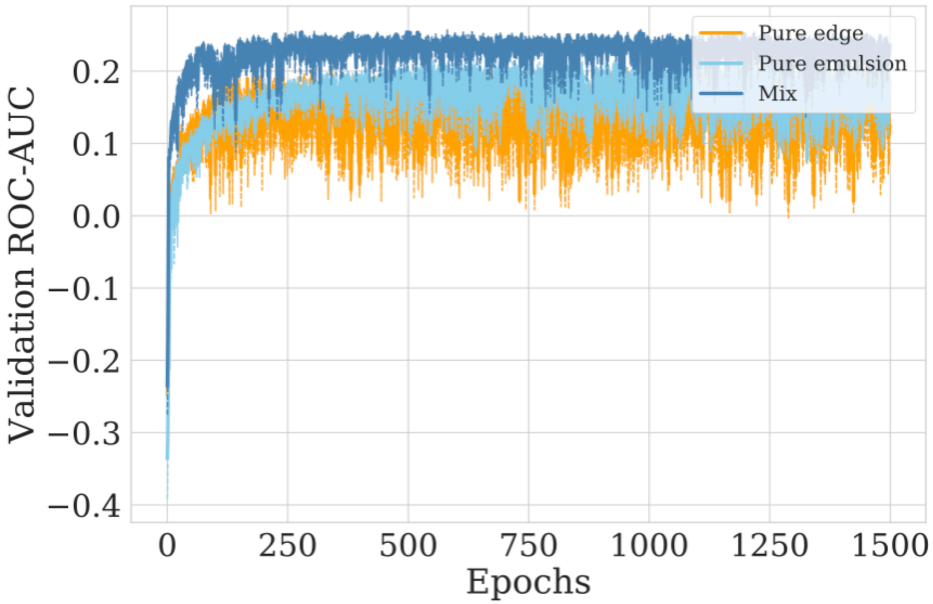


Figure 9. Validation training curves for three networks.

2. Clusterization

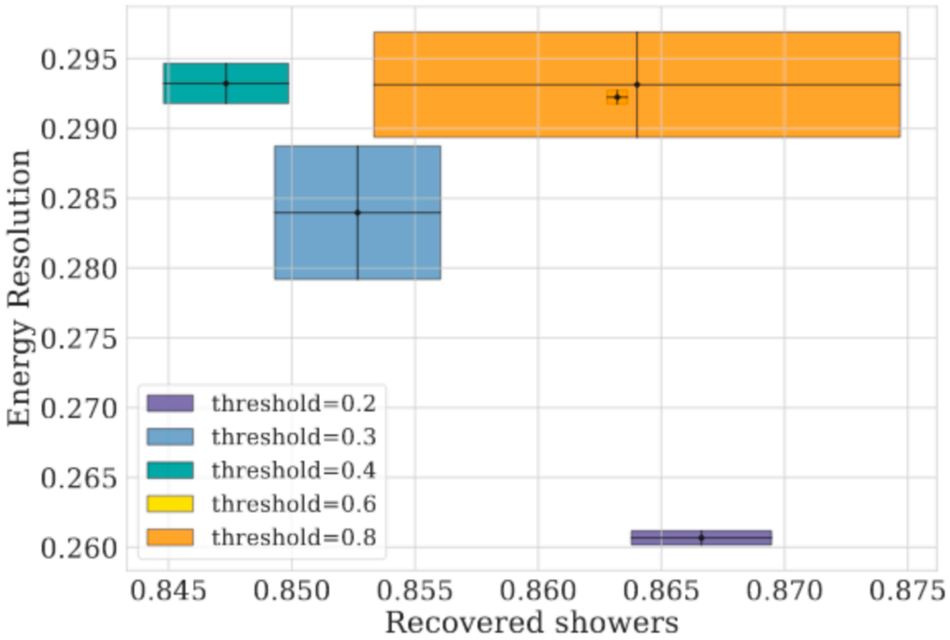


Figure 10. Grid search of optimal parameters for clustering.

Results

Table 1. Comparison of clusterization algorithms after Convolution block. Average Results over 40 bricks.

	EWSCAM			HDBSCAN
		Our		
Network / Metric	Pure Edge	Pure Emulsion	Mix	Edge+Emulsion
Recovered Showers, %	76.76 ± 1.19	81.05 ± 1.03	**81.77** ± 1.75	71.23 ± 6.49
Stuck Showers, %	16.67 ± 2.21	11.24 ± 1.31	11.89 ± 3.08	17.17 ± 5.56
Broken Showers, %	1.09 ± 0.06	1.11 ± 0.07	1.17 ± 0.21	6.76 ± 2.64
Lost Showers, %	5.48 ± 1.20	6.61 ± 0.45	5.18 ± 1.61	4.84 ± 1.77
MAE_x, μm	0.3104 ± 0.0006	0.3092 ± 0.0022	0.3107 ± 0.0010	0.3087 ± 0.0028
MAE_y, μm	0.2436 ± 0.0014	0.2448 ± 0.0016	0.2432 ± 0.0019	0.2431 ± 0.0022
MAE_z, μm	0.2083 ± 0.0021	0.2076 ± 0.0022	0.2070 ± 0.0023	0.2050 ± 0.0027
MAE_{tx}, rad	0.0126 ± 0.0004	0.0121 ± 0.0003	0.0136 ± 0.0013	0.0139 ± 0.0019
MAE_{ty}, rad	0.0127 ± 0.0003	0.0122 ± 0.0005	0.0131 ± 0.0008	0.0131 ± 0.0008

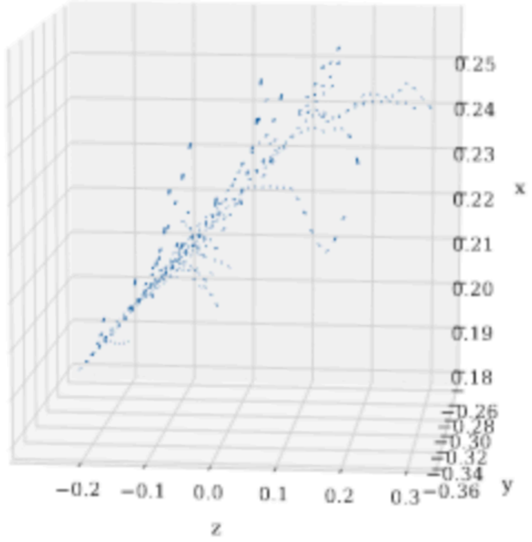


Figure 11. Cluster example.

Conclusions

1. Algorithm for multiple shower identification is developed
2. Algorithm works for 50-300 showers in a brick
3. 92% of showers recovered
4. The initial particle parameters are reconstructed