Data Reconstruction Using Deep Neural Networks for Particle Imaging Neutrino Detectors

François Drielsma

on behalf of the SLAC ML group

drielsma@slac.stanford.edu

April 7, 2020

U.S. DEPARTMENT OF ENERGY
Office of Science

SLAC NATIONAL ACCELERATOR LABORATORY

ML for LArTPCs | SLAC Group

SLAC NATIONAL ACCELERATOR LABORATORY

At SLAC, research supported by DoE ECA grant (**K. Terao**):

- **Deep-learning-based data reconstruction chain for liquid argon time-projection chambers**
- μBOONE, pDUNE, ICARUS, ArgonCube 2x2, DUNE
- Disclaimer: no physics results in this talk

Group consists of three **scientists**, three **postdocs**, three **grad students**

T. Usher
ICARUS

F. Drielsma
ICARUS

Q. Lin
ICARUS

L. Domine
ICARUS

P. Tsang
pDUNE

R. Itay
μBOONE

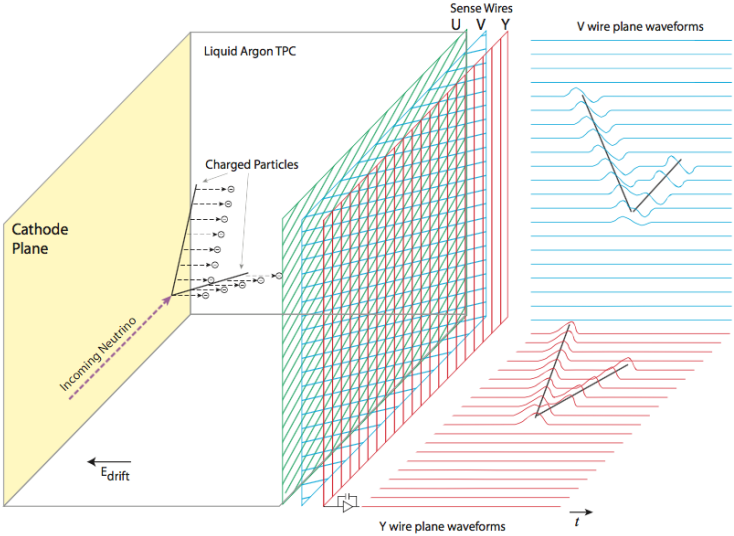
D.H. Koh
ICARUS

P. Cotes de Soux

Liquid Argon TPCs

High resolution **picture** of ionizing particle trajectories

- **mm**-scale spatial res.
- **MeV**-scale energy res.
- Scalable
- Dense
- High repetition rate



Liquid Argon TPCs

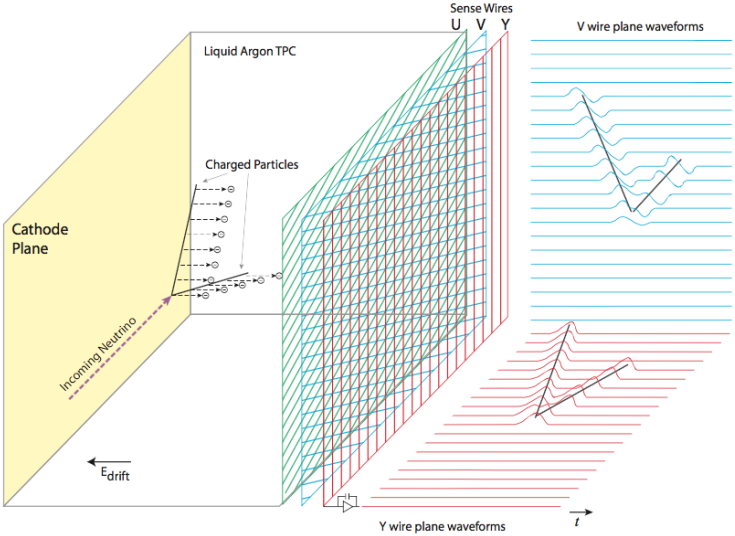
High resolution **picture** of ionizing particle trajectories

- **mm**-scale spatial res.
- **MeV**-scale energy res.
- Scalable
- Dense
- High repetition rate

Goal: measure ν_e appearance with unprecedented accuracy

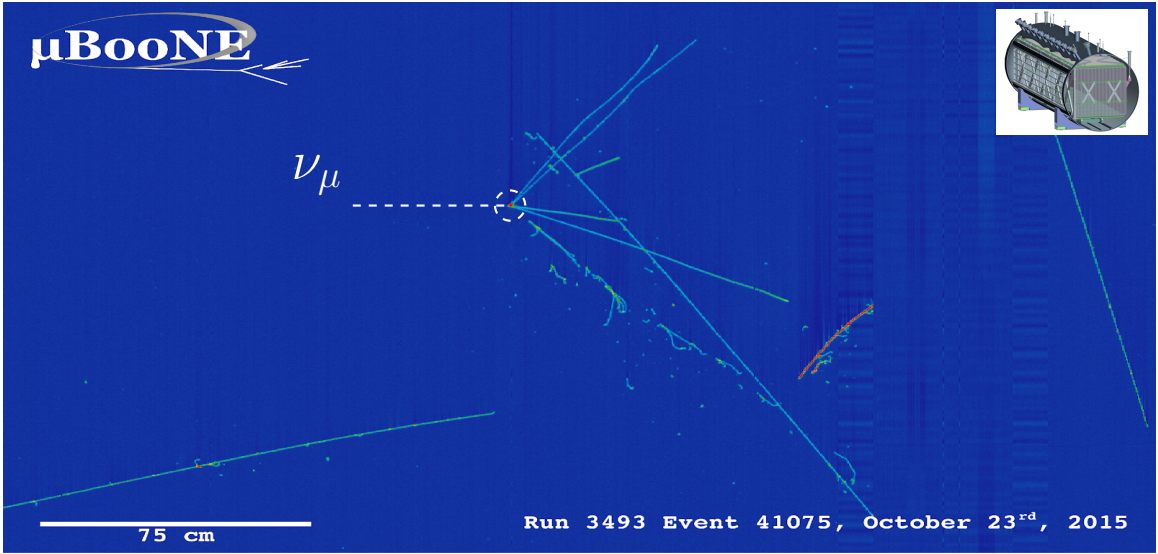
- Track–shower separation
- $e - \gamma$ separation

$\rightarrow$ **Flavor separation**



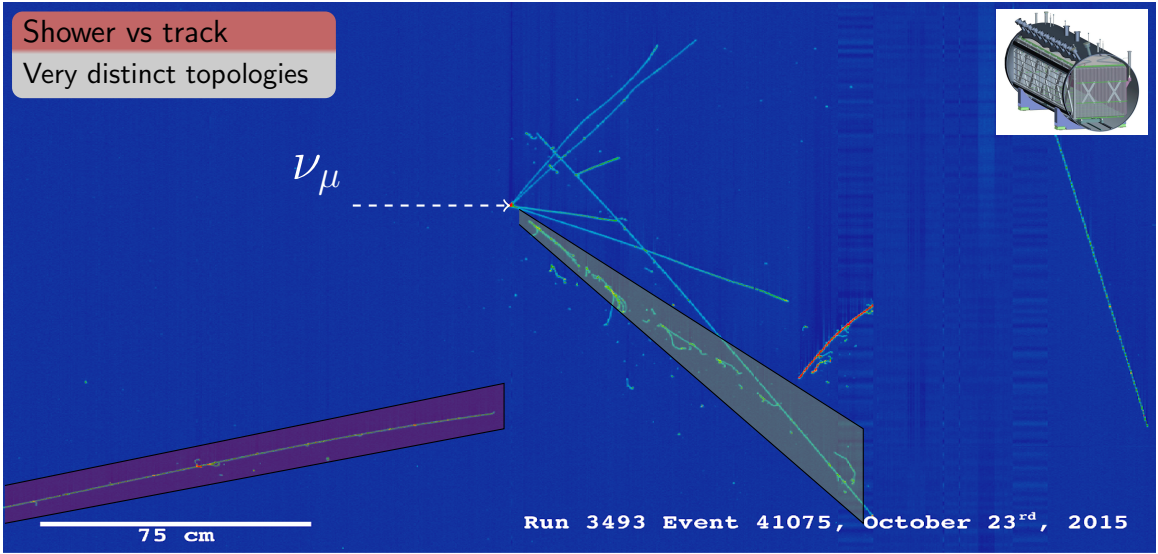
Liquid Argon TPCs

Single 2D projection



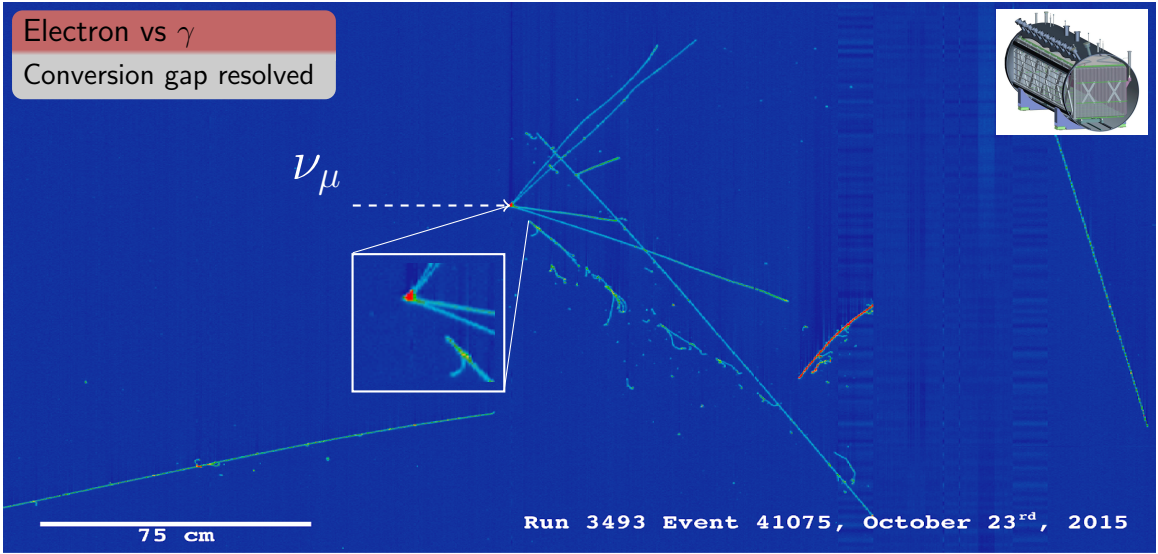
Liquid Argon TPCs

Single 2D projection

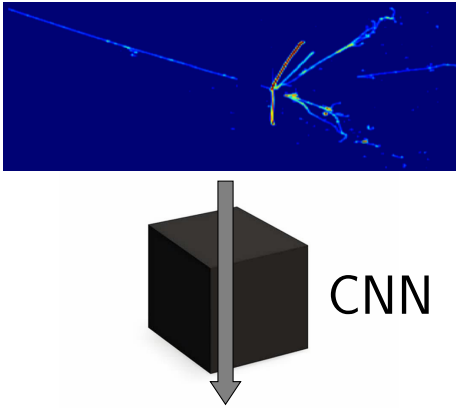


Liquid Argon TPCs

Single 2D projection



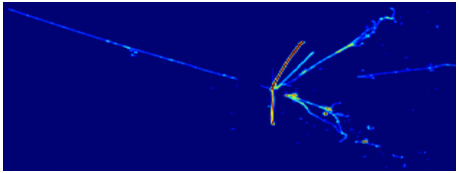
Neutrino Imaging Detectors | Whole-image analysis

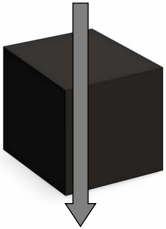


High-level Physics variables

- Interaction class (ν_e/ν_μ, CC/NC)
- Neutrino energy

Neutrino Imaging Detectors | Whole-image analysis





High-level Physics variables

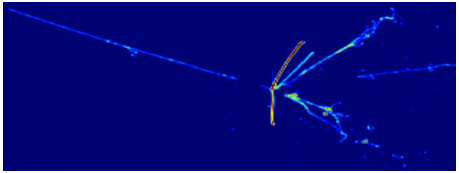
- Interaction class (ν_e/ν_μ, CC/NC)
- Neutrino energy

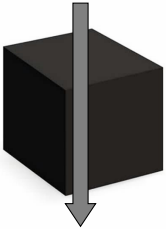
Technically the simplest approach, if it works...

Problems expected in high-res LArTPCs

- Complex topologies, huge phase space
- What if things fail ? Why ?

Neutrino Imaging Detectors | Whole-image analysis





CNN

High-level Physics variables

- Interaction class (ν_e/ν_μ, CC/NC)
- Neutrino energy

Technically the simplest approach, if it works...

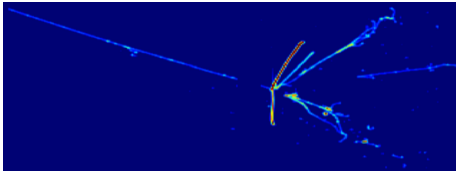
Problems expected in high-res LArTPCs

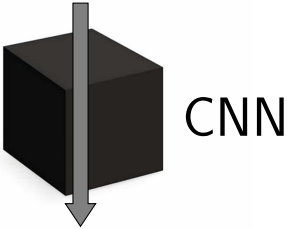
- Complex topologies, huge phase space
- What if things fail ? Why ?

Could we enforce physics principles:

- Key features like vertex
- dE/dx, particle ID

Neutrino Imaging Detectors | Whole-image analysis





High-level Physics variables

- Interaction class (ν_e/ν_μ, CC/NC)
- Neutrino energy

Technically the simplest approach, if it works...

Problems expected in high-res LArTPCs

- Complex topologies, huge phase space
- What if things fail ? Why ?

Could we enforce physics principles:

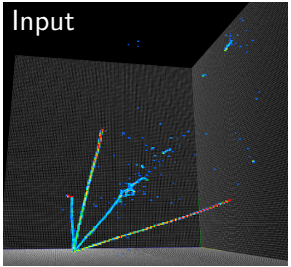
- Key features like vertex
- dE/dx, particle ID

$\rightarrow$ **Yes ! That's our research !**

What are the optimal network architectures for our data to maximize the physics output ?

Hierarchical feature extraction | Big Picture

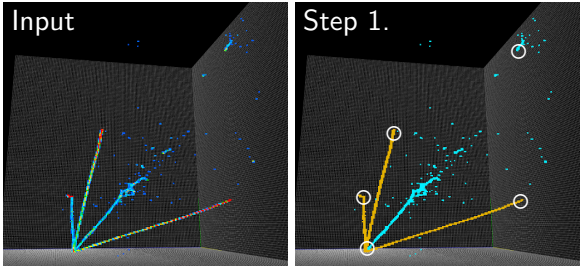
Enforce extraction of **hierarchical physics features**



Hierarchical feature extraction | Big Picture

Enforce extraction of **hierarchical physics features**

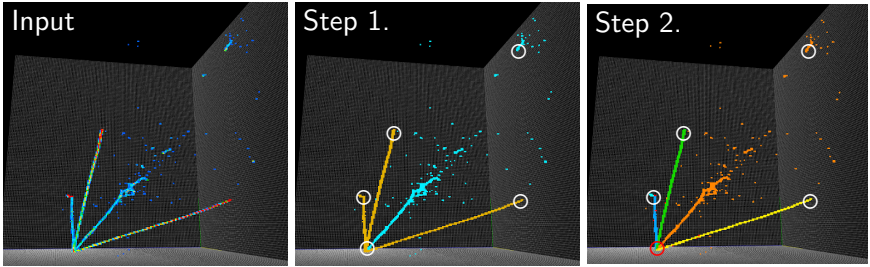
1. Pixel feature extraction + key points (particle start/end)



Hierarchical feature extraction | Big Picture

Enforce extraction of **hierarchical physics features**

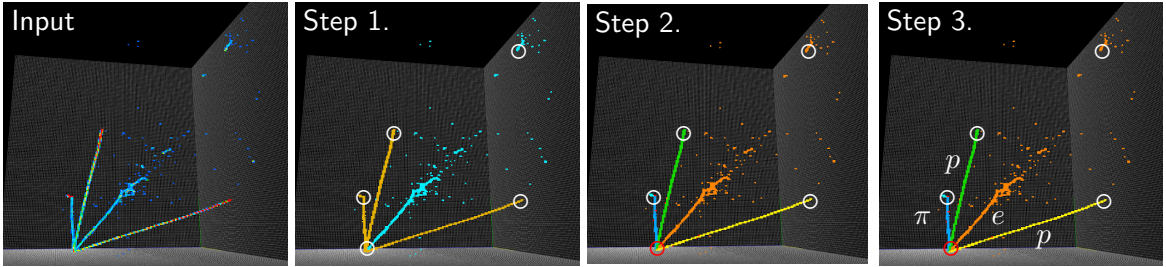
1. Pixel feature extraction + key points (particle start/end)
2. Vertex finding + particle clustering



Hierarchical feature extraction | Big Picture

Enforce extraction of **hierarchical physics features**

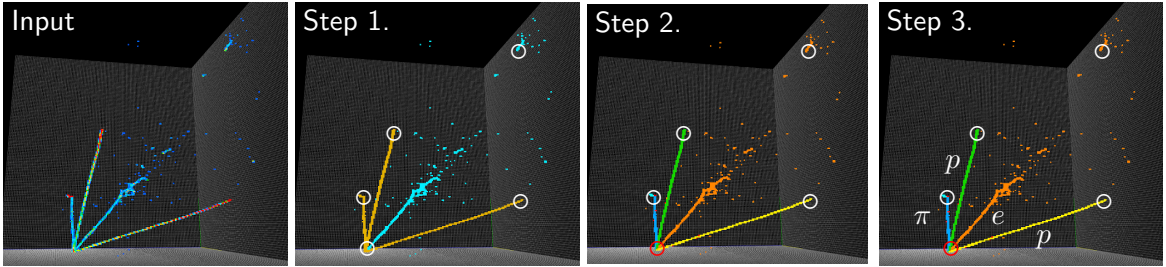
1. Pixel feature extraction + key points (particle start/end)
2. Vertex finding + particle clustering
3. Particle type + energy/momentum



Hierarchical feature extraction | Big Picture

Enforce extraction of **hierarchical physics features**

1. Pixel feature extraction + key points (particle start/end)
2. Vertex finding + particle clustering
3. Particle type + energy/momentum
4. Interaction (particle flow) reconstruction

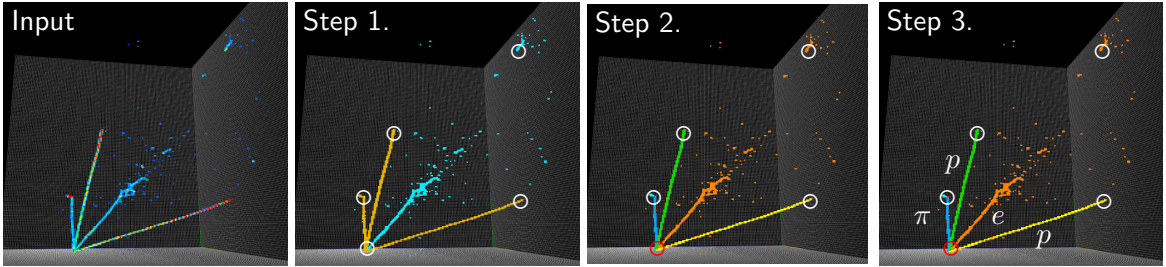


Hierarchical feature extraction | Big Picture

Enforce extraction of **hierarchical physics features**

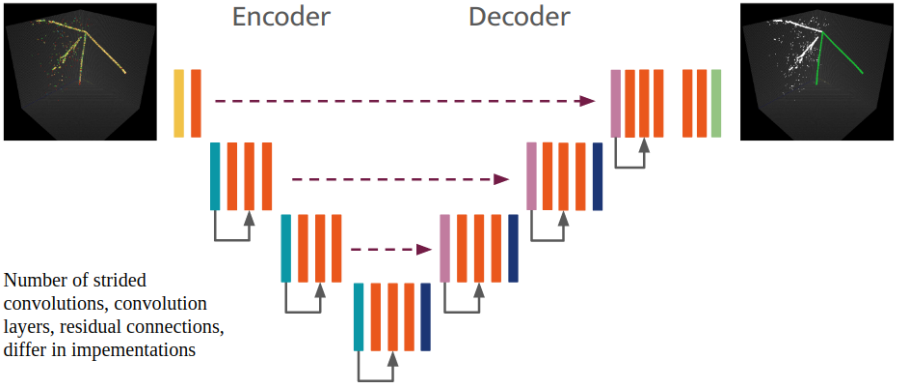
1. Pixel feature extraction + key points (particle start/end)
2. Vertex finding + particle clustering
3. Particle type + energy/momentum
4. Interaction (particle flow) reconstruction

Make it for 2D/3D data + whole chain trainable



Pixel feature extraction | Architecture

UNet + Residual connections + Sparse convolution → **Sparse UResNet**



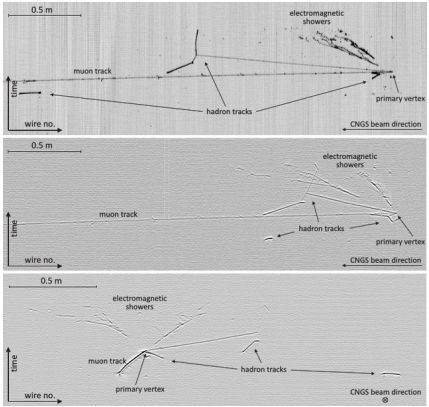
Number of strided convolutions, convolution layers, residual connections, differ in impementations

arXiv:1903.05663

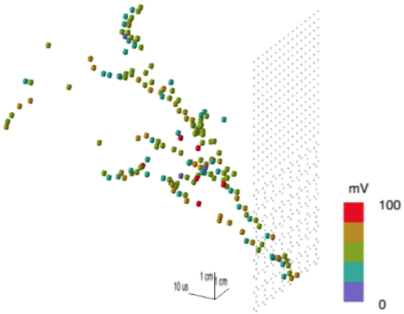
L. Domine

Pixel feature extraction | Input

Wire Data
Pixel Data
→ Image —UNet→ Types / Points / Fragments —GNN→ Particles —GNN→ Interactions → Physics

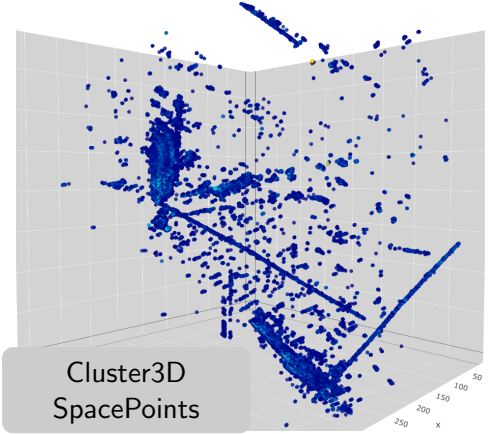


ICARUS, arXiv:1210.5089



LArPix, arXiv:1808.02969

Pixel feature extraction | Space points



Algorithms to go from to 2D to 3D:

- BNL's WireCell
- T. Usher's Cluster3D

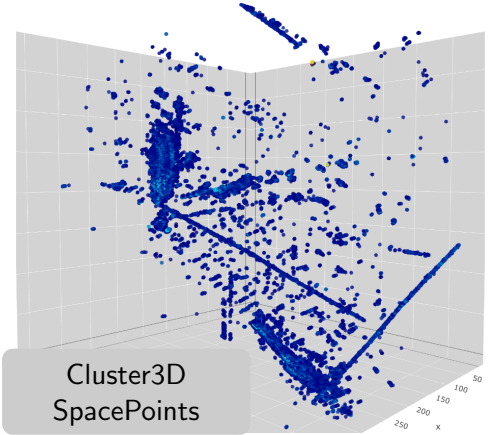
Cluster3D designed for high efficiency, relies on downstream space point solver

- Traditional likelihood-based
- **Semantic segmentation to discriminate against "ghost" points**

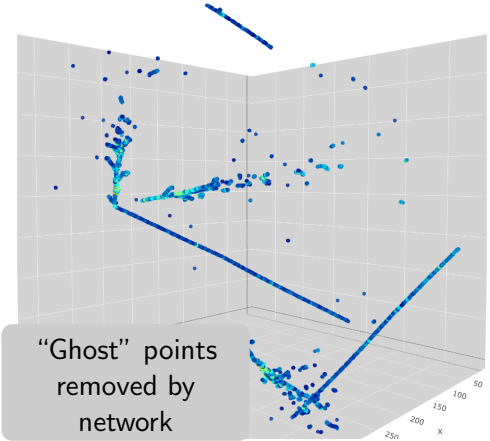
← **ICARUS simulation** on 2.3^3 m^3 region

T. Usher, P. Tsang, L. Domine

Pixel feature extraction | Deghosting

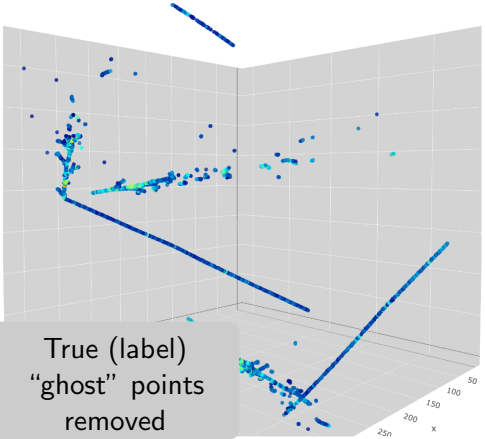


Cluster3D SpacePoints

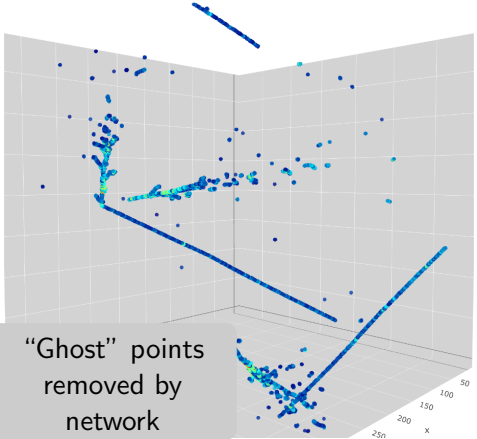


"Ghost" points removed by network

Pixel feature extraction | Deghosting

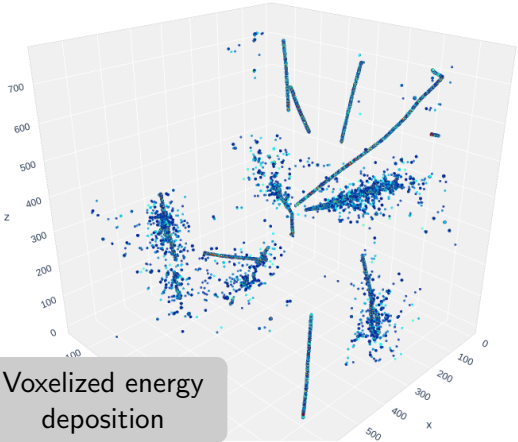


True (label) "ghost" points removed

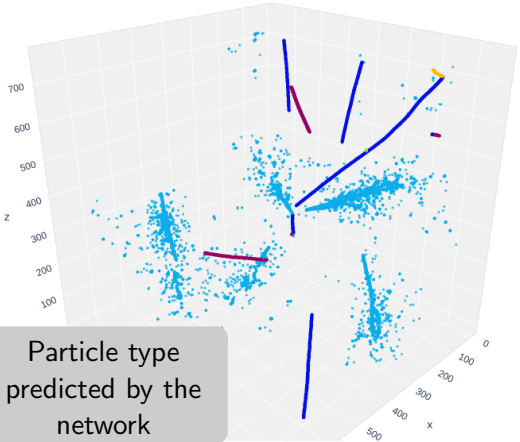


"Ghost" points removed by network

Pixel feature extraction | Particle type



Voxelized energy deposition



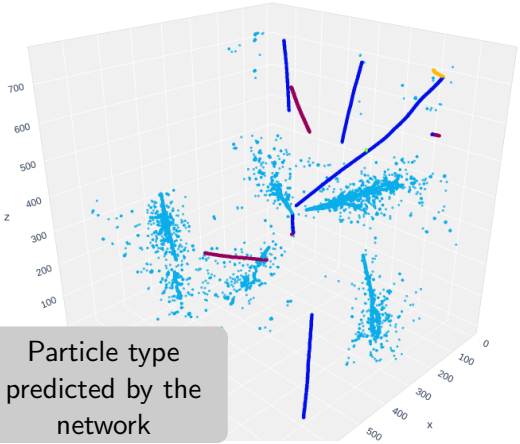
Particle type predicted by the network

Pixel feature extraction | Particle type

Particle type identification accuracy:

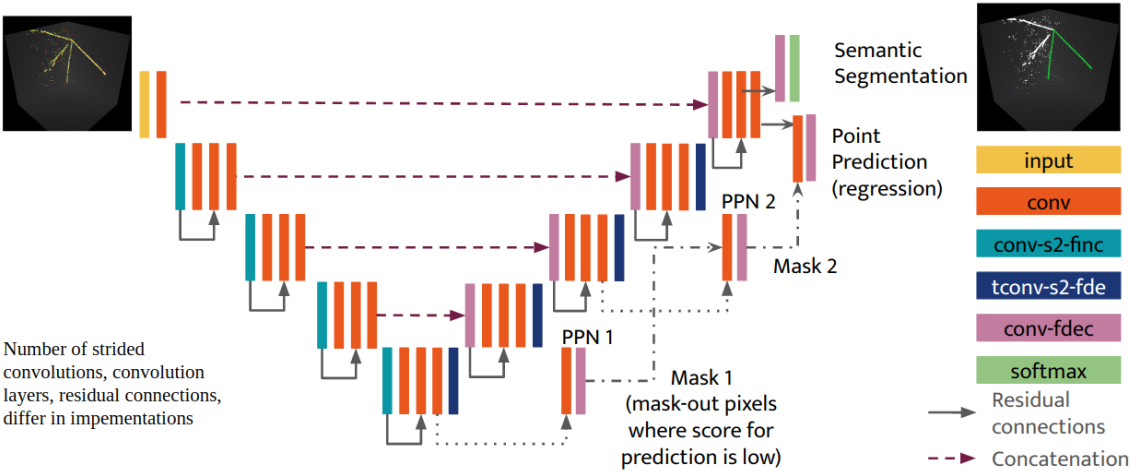
Particle type	Voxel fraction	Accuracy
HIP	17 %	98.2 %
MIP	34 %	99.4 %
Showers	47 %	99.2 %
Delta rays	1 %	96 %
Michel	1 %	94.7 %
Total		**99 %**

Network adapted to very sparse data,
see paper for details: arXiv:1903.05663



Particle type predicted by the network

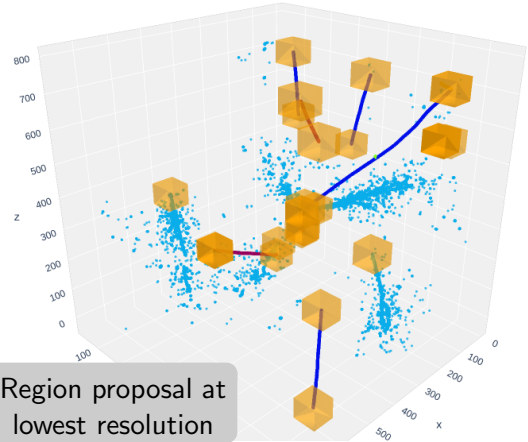
Pixel feature extraction | Point proposal



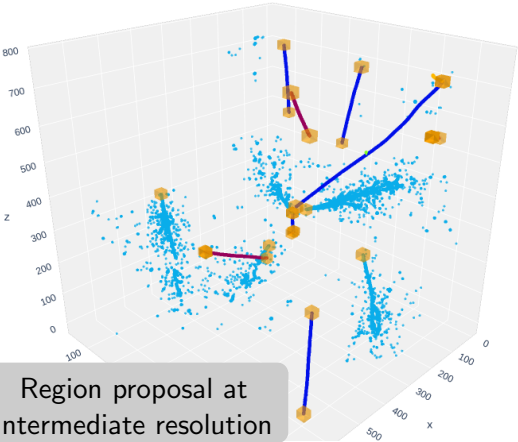
doi.org/10.5281/zenodo.1300713

L. Domine

Pixel feature extraction | Point proposal



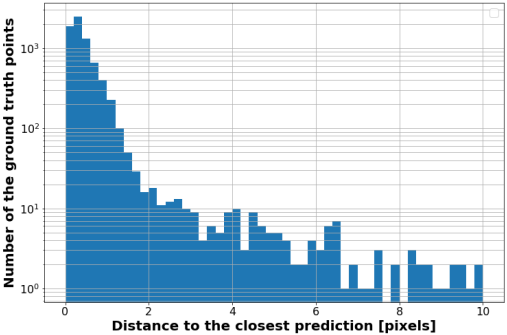
Region proposal at lowest resolution

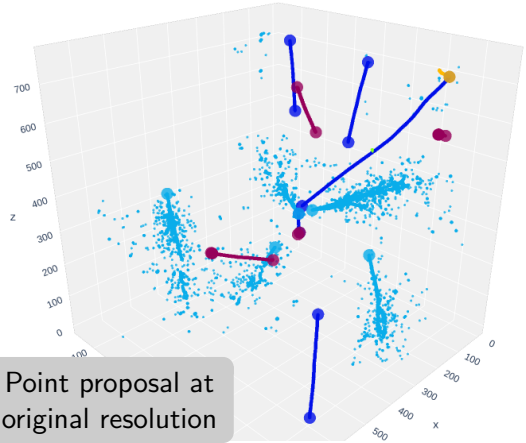


Region proposal at intermediate resolution

Pixel feature extraction | Point proposal

Point proposal efficiency (97 % < 10 px):



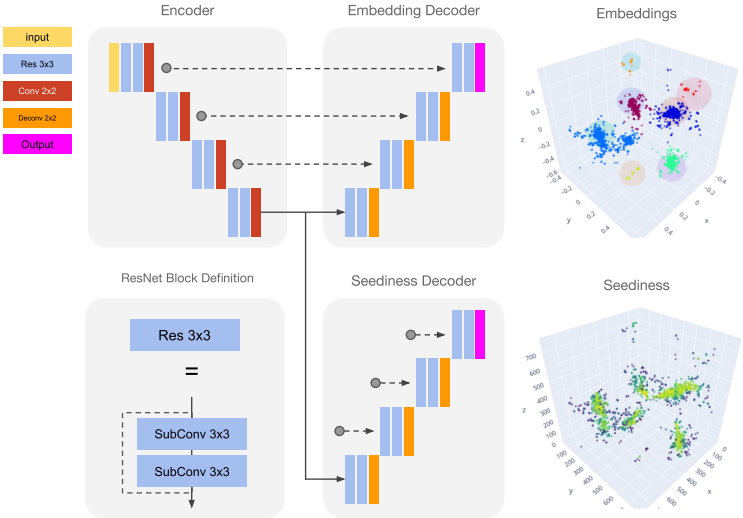


Point proposal at original resolution

doi.org/10.5281/zenodo.1300713

Pixel feature extraction

Fragment clustering

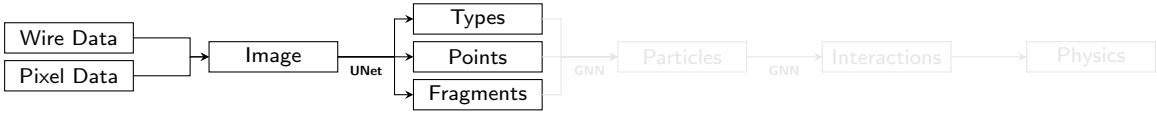


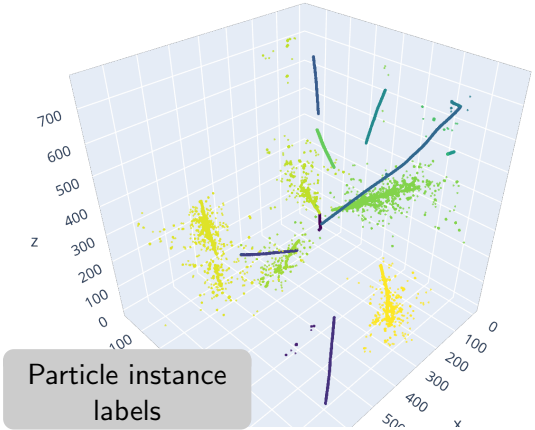
Network predicts 3 things:

- **Embedding**: space in which fragments are spatially separated
- **Seediness**: likelihood that a voxel is a cluster centroid in embedding space
- **Margin**: Size of the cluster in embedding space

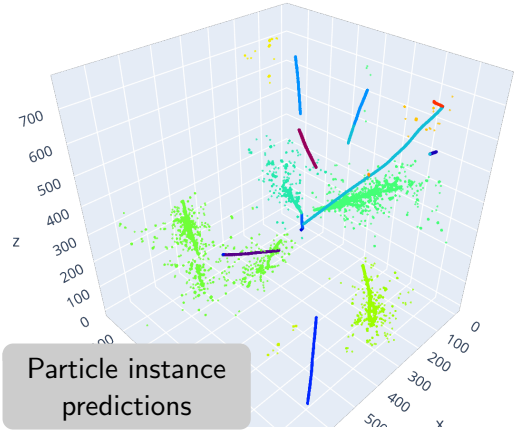
D. H. Koh

Pixel feature extraction | Fragment clustering



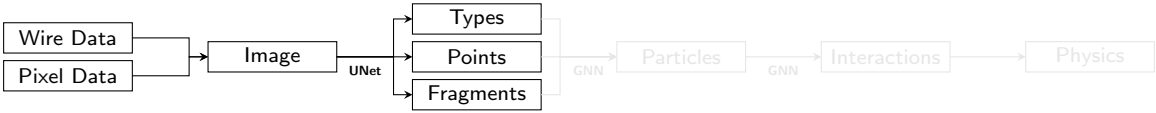


Particle instance labels

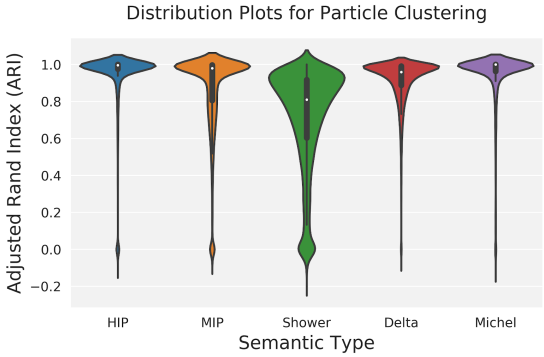


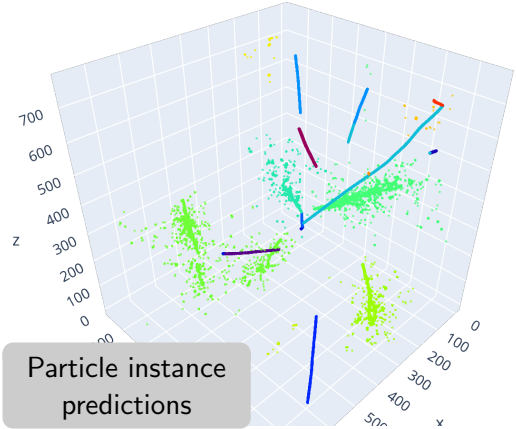
Particle instance predictions

Pixel feature extraction | Fragment clustering



Instance clustering accuracy (ARI):





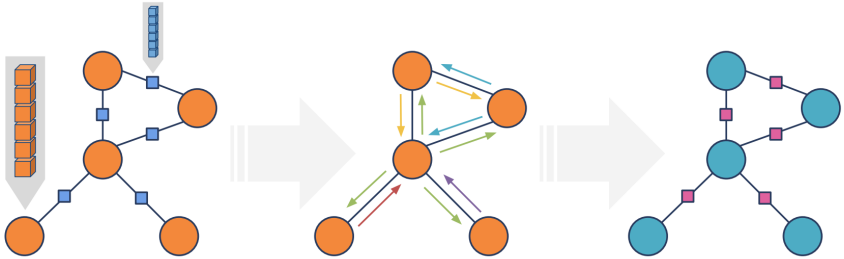
Particle instance predictions

Object clustering

CNN struggles with far detached elements such as **shower secondaries**.

Graphical Neural Networks (GNN) are ideal for this:

- Based on *nodes* and *edges*. Features propagate by *message passing* (MP)



Define a graph with node & edge level features.

Graph Convolution + Message passing

Updated node and edge features

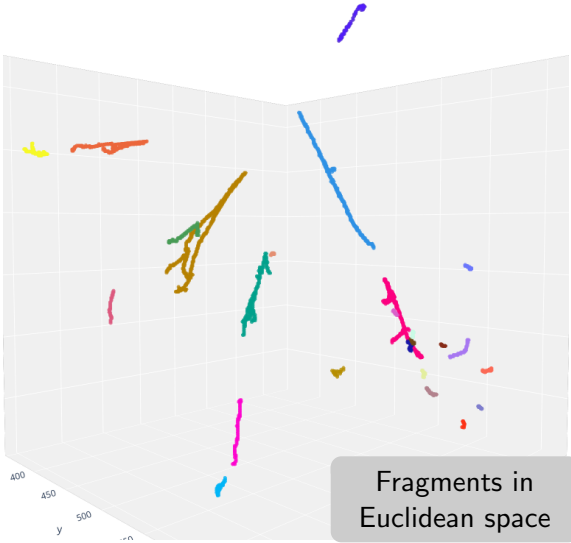
F. Drielsma, Q. Lin, P. Cotes de Soux

Shower clustering

Network input

From previous stages:

- Fragmented EM showers



Fragments in Euclidean space

Shower clustering

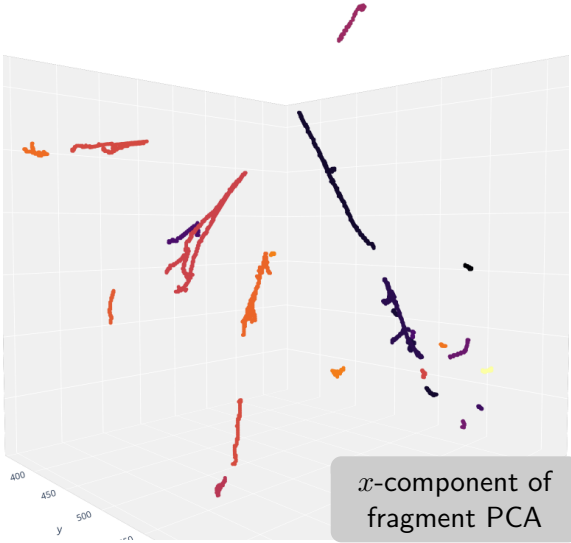
Network input

From previous stages:

- Fragmented EM showers

Node features:

- **Geometrical features**
- CNN autoencoder
- Aggregated embedding of voxels



x-component of fragment PCA

Shower clustering

Network input

From previous stages:

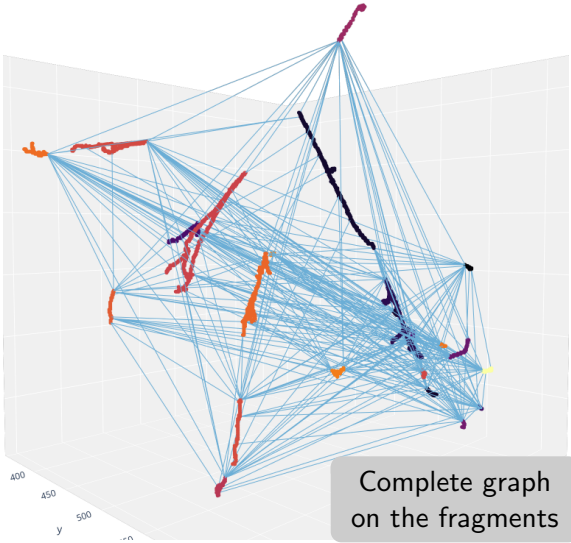
- Fragmented EM showers

Node features:

- **Geometrical features**
- CNN autoencoder
- Aggregated embedding of voxels

Input graph:

- Connect every node with every other node (**complete graph**)



Complete graph on the fragments

Shower clustering

Network input

From previous stages:

- Fragmented EM showers

Node features:

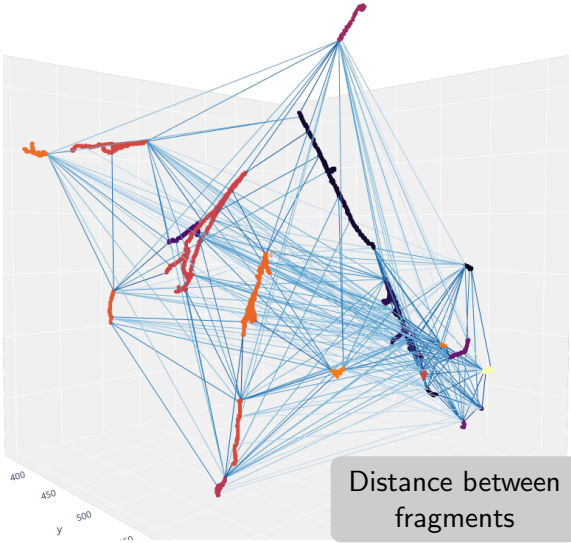
- **Geometrical features**
- CNN autoencoder
- Aggregated embedding of voxels

Input graph:

- Connect every node with every other node (**complete graph**)

Edge features:

- Displacement vector (+variations)
- Points of closest approach



Distance between fragments

Shower clustering

Edge classification

At each message passing:

- **Node update**

  $\boldsymbol{x}'_i = \boldsymbol{\Theta}\boldsymbol{x}_i + \sum_{j\in\mathcal{N}(i)} \boldsymbol{x_j} \cdot \boldsymbol{h}_\Theta(\boldsymbol{e}_{i,j})$

  NNConv: arXiv:1704.02901

- **Edge update**

  $\boldsymbol{e}'_{i,j} = \boldsymbol{\phi}_\Theta(\boldsymbol{x}_i, \boldsymbol{x}_j, \boldsymbol{e}_{i,j})$

  MetaLayer: arXiv:1806.01261

Shower clustering

Edge classification

At each message passing:

- **Node update**

  $\boldsymbol{x}'_i = \boldsymbol{\Theta}\boldsymbol{x}_i + \sum_{j\in\mathcal{N}(i)} \boldsymbol{x_j} \cdot \boldsymbol{h}_\Theta(\boldsymbol{e}_{i,j})$

  NNConv: arXiv:1704.02901

- **Edge update**

  $\boldsymbol{e}'_{i,j} = \boldsymbol{\phi}_\Theta(\boldsymbol{x}_i, \boldsymbol{x}_j, \boldsymbol{e}_{i,j})$

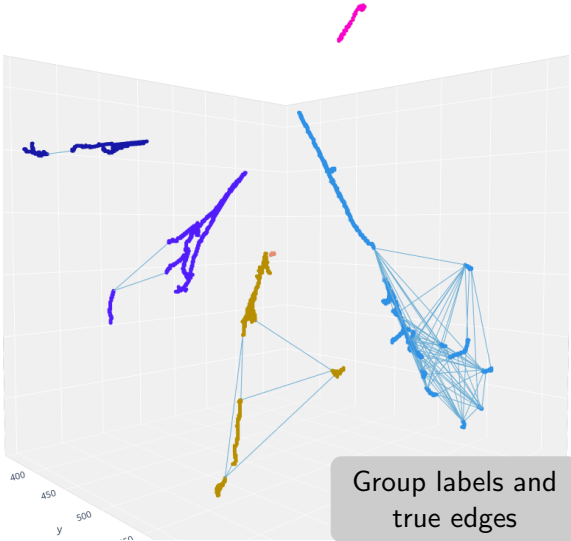
  MetaLayer: arXiv:1806.01261

After $n = 3$ node+edge updates:

- Edge binary classification

Target:

- Predict adjacency matrix $A_{ij} = \delta_{g_i,g_j}$ with $\boldsymbol{g}$ the true partition of the set
- Apply cross-entropy loss



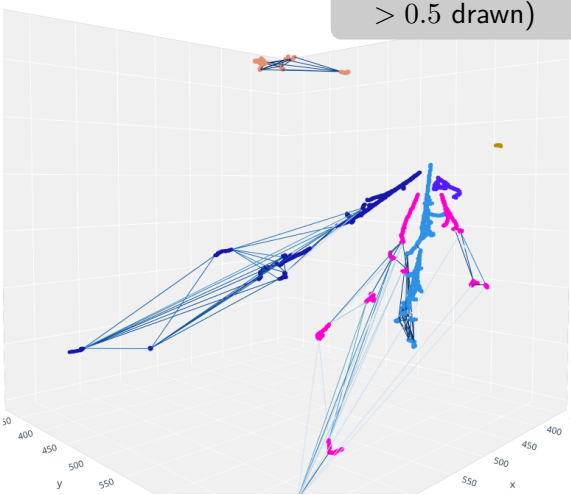
Group labels and true edges

Shower clustering

The network predicts a score matrix $\boldsymbol{S}$, estimate of the true adjacency matrix $\boldsymbol{A}$

- How to recover a set partition $\hat{\boldsymbol{g}}$?

Edge scores (only > 0.5 drawn)



Shower clustering

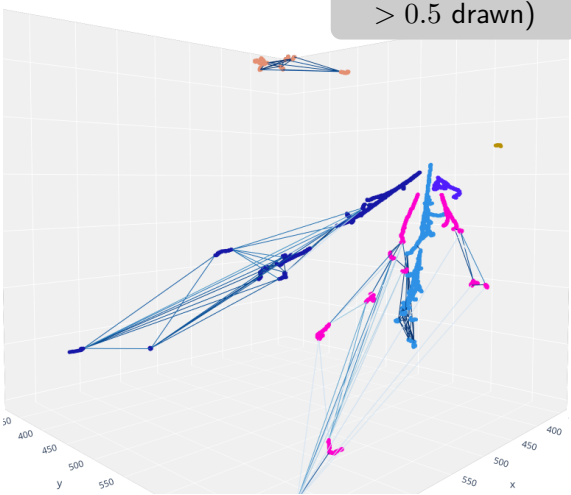
The network predicts a score matrix $\boldsymbol{S}$, estimate of the true adjacency matrix $\boldsymbol{A}$

- How to recover a set partition $\hat{\boldsymbol{g}}$?

We want the partition $\hat{\boldsymbol{g}}$ that minimizes the CE loss, given $\hat{A}_{ij} = \delta_{\hat{g}_i, \hat{g}_j}$, e.g.

$$\hat{\boldsymbol{g}} = \min_{\boldsymbol{g} \in G} \mathcal{L}_{CE}(\boldsymbol{S}, \hat{\boldsymbol{A}}(\boldsymbol{g}))$$

Edge scores (only > 0.5 drawn)



Shower clustering

The network predicts a score matrix $\boldsymbol{S}$, estimate of the true adjacency matrix $\boldsymbol{A}$

- How to recover a set partition $\hat{\boldsymbol{g}}$?

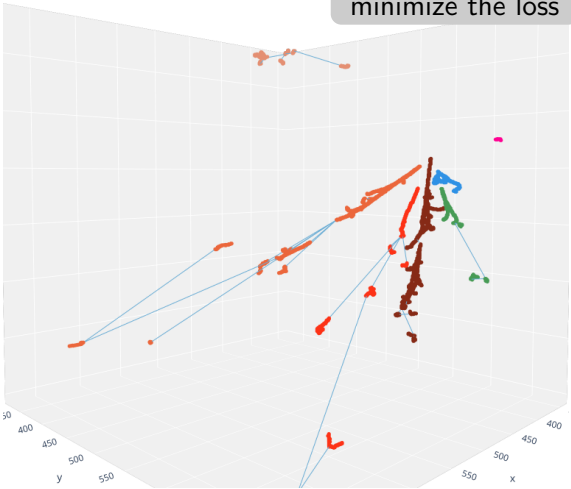
We want the partition $\hat{\boldsymbol{g}}$ that minimizes the CE loss, given $\hat{A}_{ij} = \delta_{\hat{g}_i, \hat{g}_j}$, e.g.

$$\hat{\boldsymbol{g}} = \min_{\boldsymbol{g} \in G} \mathcal{L}_{CE}(\boldsymbol{S}, \hat{\boldsymbol{A}}(\boldsymbol{g}))$$

G is the set of all possible partitions

- Bell number, huge ($B_{20} \simeq 5 \times 10^{13}$), cannot brute force, how to optimize?
- Build MST on edge scores
- Down select edges until score cannot be improved further

MST edges that minimize the loss



Shower clustering

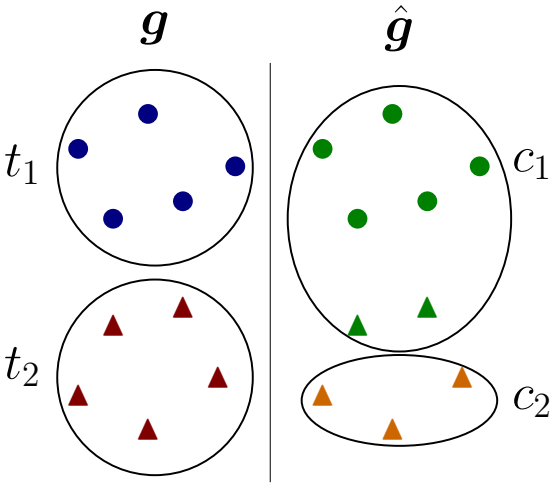
Metrics

Purity $= \frac{1}{N}\sum_{i=1}^{n_p} \max_j |c_i \cap t_j|$

- c_i predicted cluster
- t_j true cluster with highest count in c_i

Efficiency $= \frac{1}{N}\sum_{i=0}^{n_t} \max_j |c_j \cap t_i|$

- c_j pred. cluster with highest count in t_i
- t_i true cluster



Shower clustering | Metrics

Purity $= \frac{1}{N}\sum_{i=1}^{n_p} \max_j |c_i \cap t_j|$

- c_i predicted cluster
- t_j true cluster with highest count in c_i

Efficiency $= \frac{1}{N}\sum_{i=0}^{n_t} \max_j |c_j \cap t_i|$

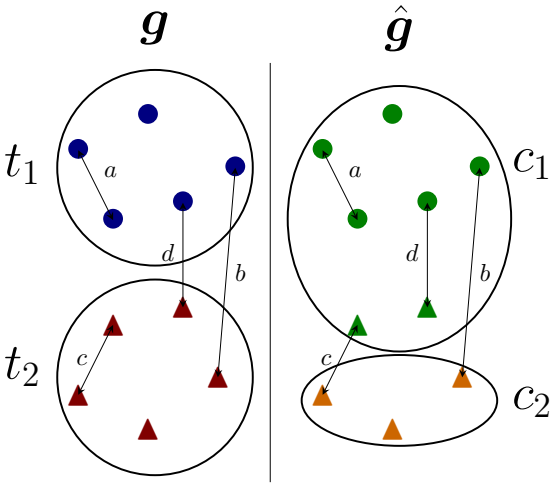
- c_j pred. cluster with highest count in t_i
- t_i true cluster

Adjusted Rand Index (ARI)

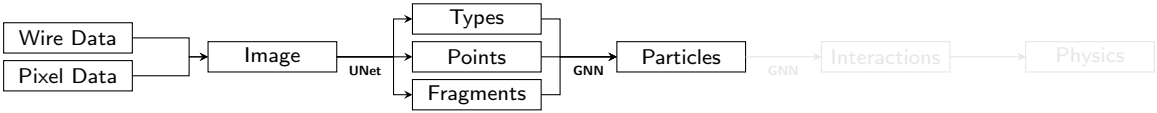
- Measure of overlap of prediction and truth, adjusted for random chance

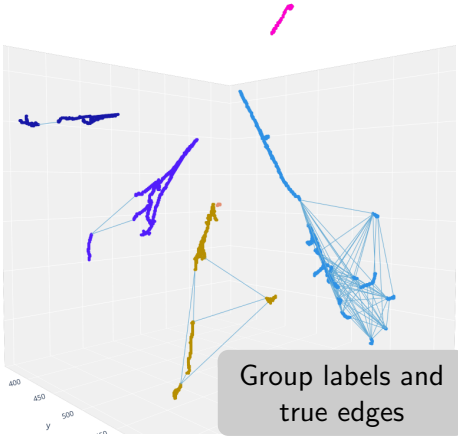
  RI $= \frac{a+b}{a+b+c+d}$

  ARI $= \frac{RI - E(RI)}{1 - E(RI)}$

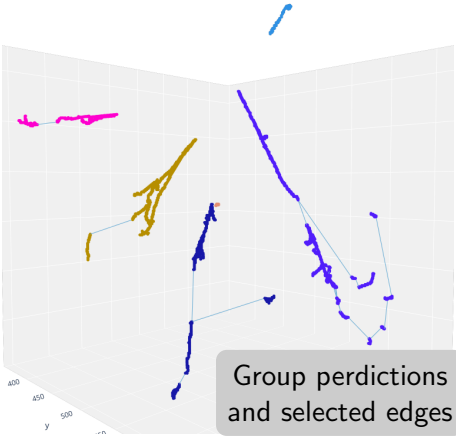


Shower clustering | Performance



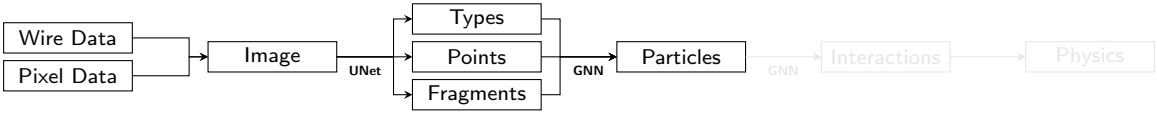


Group labels and true edges

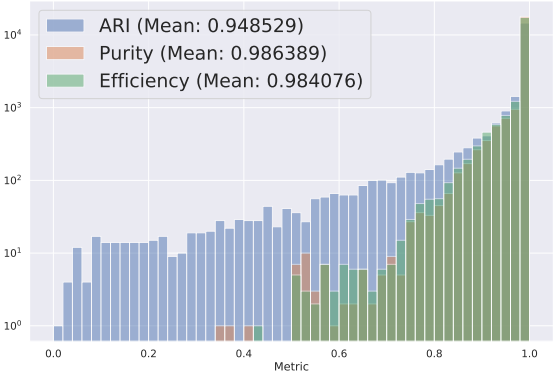


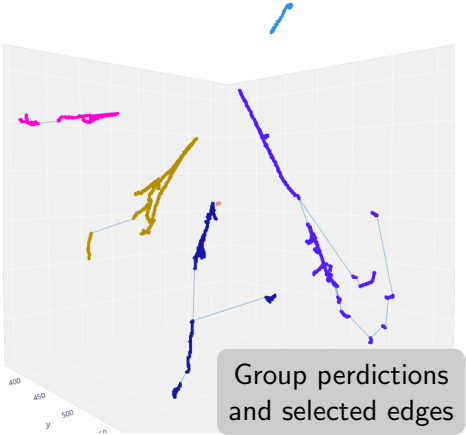
Group perdictions and selected edges

Shower clustering | Performance



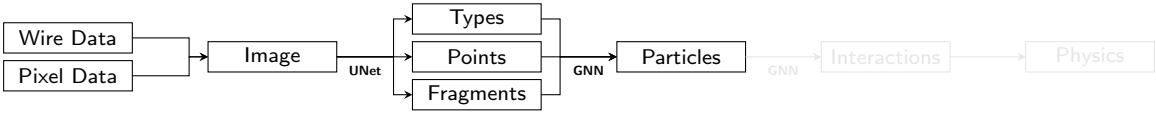
Shower clustering accuracy:





Group perdictions and selected edges

Shower Clustering | Start identification

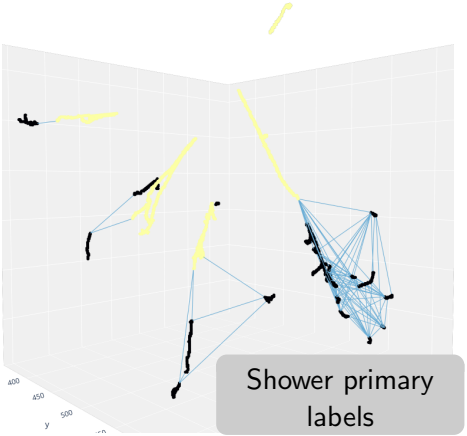


Identifying shower starts is a very useful task

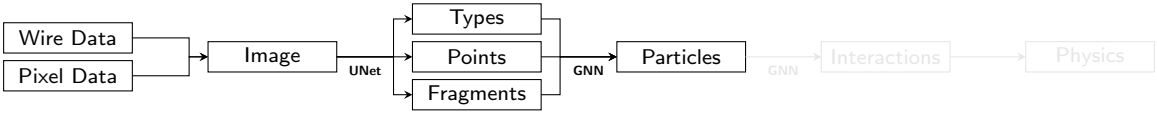
- Start point, direction
- Shower matching for π_0 reconstruction

Given a partition $\hat{\boldsymbol{g}}$:

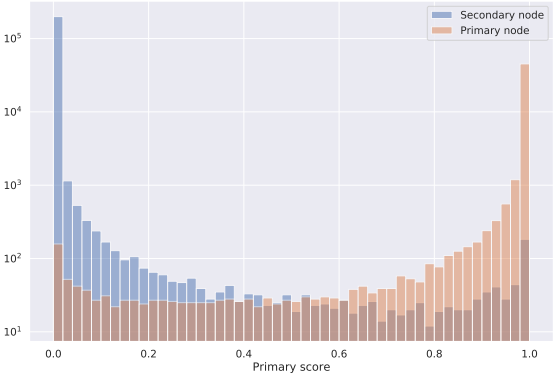
- Create a complete graph within each pred. group
- Predict which node comes first in time within each pred. group
- CE loss on primary labels, if only 1 true primary in pred. group

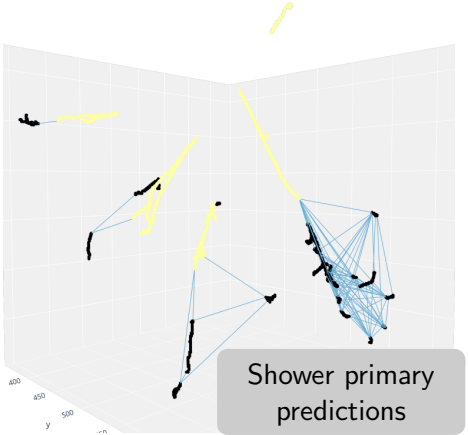


Shower Clustering | Start identification



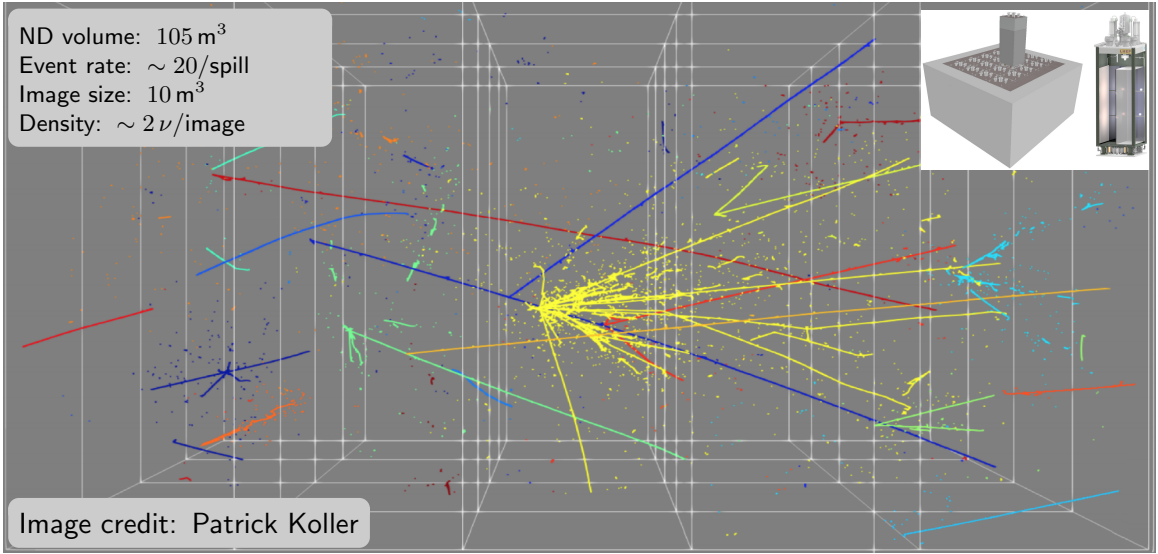
Shower primary accuracy (**98.3 %**):



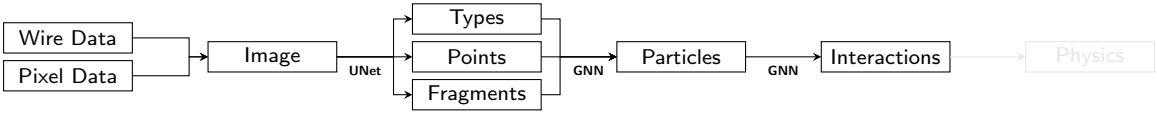


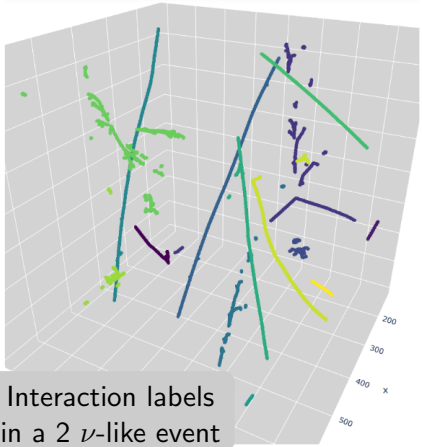
High rate LArTPCs

DUNE ND



Interaction Clustering | Performance





Interaction labels in a 2 ν-like event

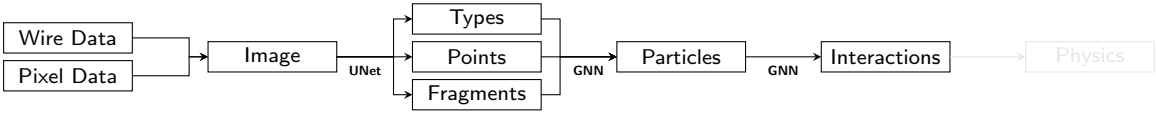
The shower clustering task can be extended to **interaction clustering**:

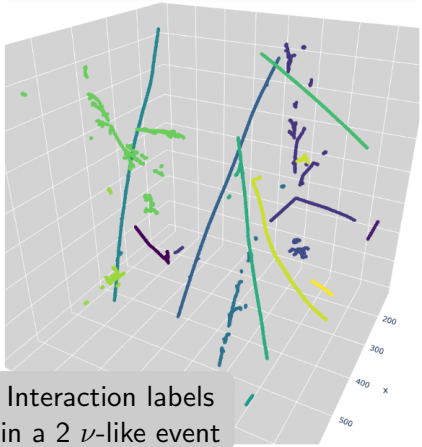
- Interaction group = particles that originate from the same vertex
- Fragment ID $\rightarrow$ Particle ID
- Particle ID $\rightarrow$ Interaction ID

Crucial to:

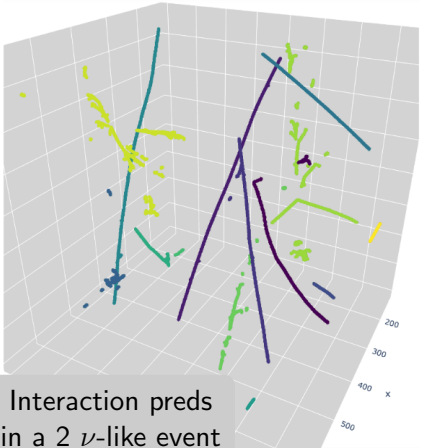
- Separate signal from background
- **Resolve pileup (DUNE ND)**

Interaction Clustering | Performance



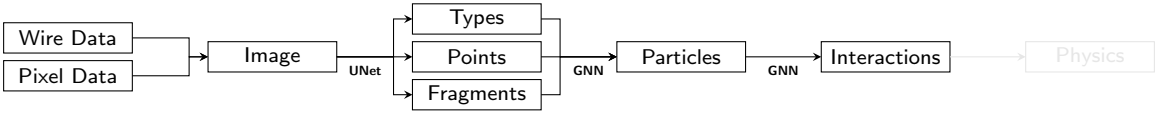


Interaction labels in a 2 ν-like event



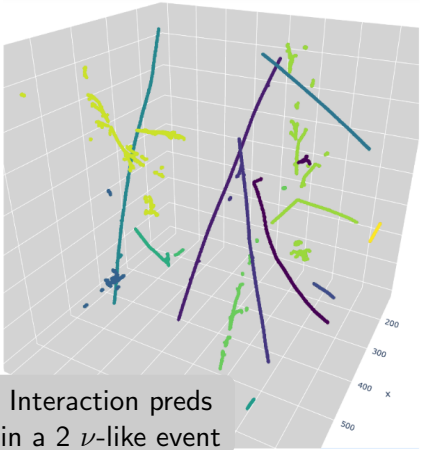
Interaction preds in a 2 ν-like event

Interaction Clustering | Performance



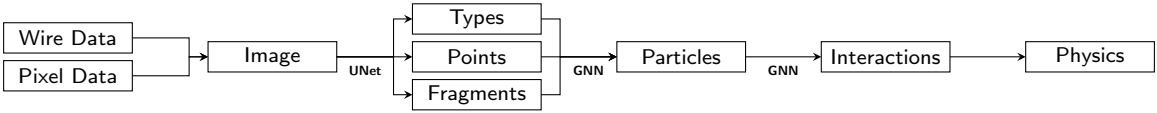
Interaction clustering performance:

Metric	# of ν-like	Mean Score
Efficiency	1	98.8 %
	2	97.6 %
	4	95.6 %
Purity	1	99.4 %
	2	99.3 %
	4	99.3 %
ARI	1	95.6 %
	2	93.2 %
	4	88.0%



Interaction preds in a 2 ν-like event

Summary



ML Reconstruction Chain for LArTPCs:

- Trend in neutrino detection: high-resolution particle imaging
- Resulting analysis trend: computer vision → **Machine Learning**
- LArTPC images too information rich to be reduced to simple variables in one pass
- **Hierarchical feature extraction** very successful so far

Areas we will work on not covered in this talk:

- Data vs simulation domain discrepancy
- Error propagation

Please email for more details or if you want to participate !

Back-up slides

Liquid Argon TPCs

Sizes of current and future LArTPCs:

- μBOONE: 100 t (10x2.5x2.5 m^3)
- pDUNE: 200 t (6x6x6 m^3)
- ICARUS: 400 t (2x(20x3x3) m^3)
- ArgonCube 2x2: 10 t (4x(0.67x0.67x2) m^3)
- DUNE-ND: 150 t (35x(1x1x3) m^3)
- DUNE-FD: 40 kt (4x(12x12x60) m^3)

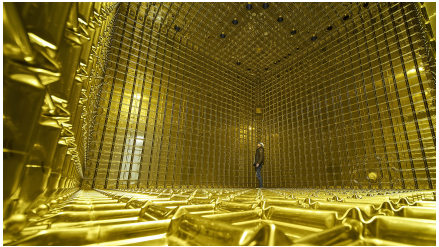
Numbers

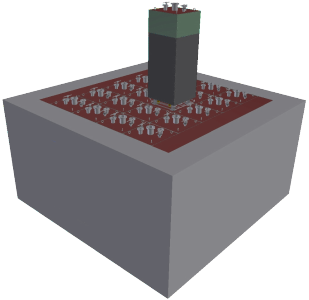
Some example numbers (for ICARUS):

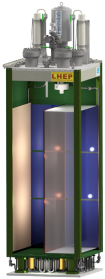
- Wire pitch: 3 mm
- Angle between planes: 60°
- Drift field: 500 V/cm
- Drift velocity: ~ 0.15 cm/μs
- TPC time resolution: $0.4\,\mu$s (< 1 mm)
- PMT coverage: $\sim$2 %
- Scintillation light: 20 % prompt (6 ns), 80 % late (1.5 μs)
- Photon yield: 24000/MeV

Liquid Argon TPCs

Pictures







Submanifold Sparse Convolutions

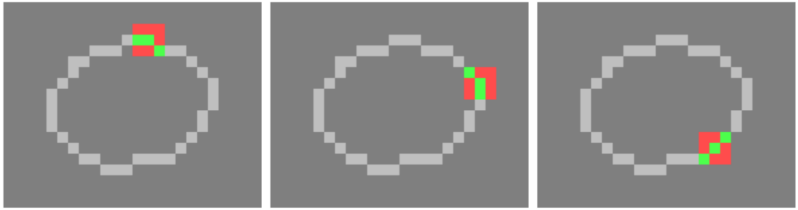
1. Resources waste of dense convolutions on sparse data
2. Dilation problem
   - One nonzero site leads to 3d nonzero sites after 1 convolution
   - How to keep the same level of sparsity throughout the network?

https://arxiv.org/pdf/1711.10275.pdf

Sparse CNNs

In more details, two new operations:

- Sparse convolutions (SC)
  - Discards contribution of non-active input sites
  - Output site active if at least one input site is active
- Sparse submanifold convolutions (SSC)
  - Output size = Input size
  - Output site active iff center of receptive field active
  - Only compute features for active output sites



https://arxiv.org/pdf/1711.10275.pdf

Point proposal | Loss

PPN outputs **voxel location**, **position within voxel** and **point class**

Three components to the point proposal loss:

- Pixel classification loss at each of three depth (pixel contains point or not)

$$\mathcal{L}_{\text{class},i} = \frac{1}{N_i} \sum_{k=1}^{N_i} y_k \log(p_k) + (1 - y_k) \log(1 - p_k)$$

- L^1 distance from true point at highest resolution **on active voxels**

$$\mathcal{L}_{\text{dist}} = \frac{1}{N_3^*} \sum_{k=1}^{N_3^*} \min_j ||\vec{p}_i - \vec{q}_j||$$

- Particle type loss at highest resolution **on active voxels**

$$\mathcal{L} = \frac{1}{N_3^*} \sum_{k=1}^{N_3^*} \sum_c y_{k,c} \log(p_{k,c})$$

Fragment clustering | Loss

Instance Segmentation:

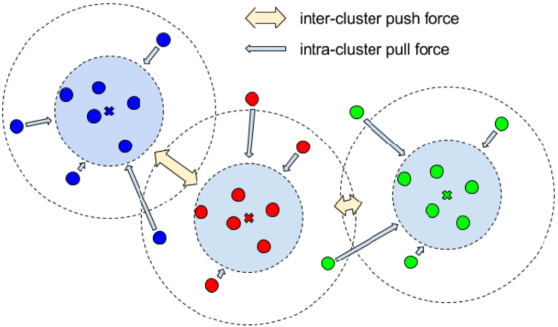
- Three component loss: pull together points that belong to the same cluster, keep distance between clusters, and regularization

$$L = \alpha L_{var} + \beta L_{dist} + \gamma L_{reg},$$

$$L_{var} = \frac{1}{C}\sum_{c=1}^{C}\frac{1}{N_c}\sum_{i=1}^{N_c}[\max(0, \|\mu_c - x_i\| - \delta_v)]^2$$

$$L_{dist} = \frac{1}{C(C-1)}\sum_{\substack{c_A,c_B=1\\ c_A\neq c_B}}^{C}[\max(0, 2\delta_d - \|\mu_{c_A} - \mu_{c_B}\|)]^2$$

$$L_{reg} = \frac{1}{C}\sum_{c=1}^{C}\|\mu_c\|$$

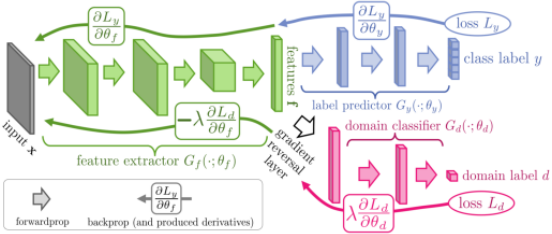


arXiv:1708.02551

Domain discrepancies | GAN

What can we do about imperfect simulation ?

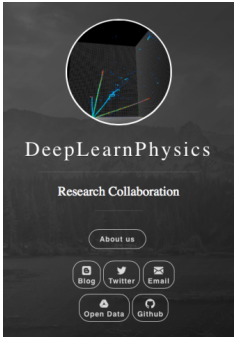
- Issue: the signal distribution learned by the algorithm may be different in two domains!
- Mitigation techniques in ML domain ?
  - Can try CNN to locate where it is
  - Can try CNN to fix the discrepancy
  - Can try a training technique to minimize the effect



Maximize the loss to discriminate data vs. simulation, feature extractors are penalized to key on simulation specific information

Domain-Adversarial Training of Neural Networks: J. Mach. Learn. Res. 17 (2016)

Open Source Development | Highlights



DeepLearnPhysics: Collaboration for ML technique R&D

- Open simulation sample (used throughout this talk)
  - Open real data ? Soon ! (3D prototype R&D at SLAC)
- Open source container (Singularity)
- Open source code (GitHub)
  - All the code used to make this talk is available

→ **Reproducible results** !

- Readers have reproduced arXiv:1903.05663