Augmented Signal Processing in Liquid Argon Time Projection Chambers with a Deep Neural Network

- Software Effort to Improve LArTPC Reconstruction
- *JINST* 16 (2021) 01, P01036
LArTPC is the central detector in many next-gen neutrino experiments, e.g. DUNE, SBN etc.

- Calorimetry and rich topology information

LArTPC wire-readout measures induced charge $\otimes$ response

$$M(t', x') = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} R(t, t', x, x') \cdot S(t, x) dt dx + N(t', x')$$

Signal Processing (SP) of LArTPC resolves charge from the original measurement:

$$S(\omega_t, \omega_x) \sim \frac{F(\omega_t, \omega_x) \cdot M(\omega_t, \omega_x)}{R(\omega_t, \omega_x)} \quad \text{IFT} \quad S(t, x)$$

Current state of the art SP algorithm based on 2D deconvolution:

*JINST 13 P07006 (2018)*
Remaining Challenges in Current Signal Processing (SP)

- “Prolonged Track” – weak signal
- “Tear Drop” - distorted waveform
- Noisy dots - noise

![Graph showing efficiency after signal processing](image-url)
Need to improve “ROI” (Region of Interest) finding

Information from Other Planes: Utilizing plane redundancy linked by geometry information

Machine Learning:
- Quickly cover more phase space;
- Utilizing more complicated correlations

<table>
<thead>
<tr>
<th>Decon. for charge</th>
<th>ROI:</th>
<th>SP result:</th>
</tr>
</thead>
<tbody>
<tr>
<td>Waveform → charge, dense</td>
<td>Hit finding, sparsify</td>
<td>Sparse, charge</td>
</tr>
</tbody>
</table>

Decon. w/ Tight LF
Decon. w/ Loose LF

Waveform data

Noise Filtering + 2D decon.

Decon. Waveform

ROI

SP Result
ROI finding as semantic segmentation

2D semantic segmentation – U-Net
• arXiv:1505.04597 (MICCAI 2015)

U-Net: auto encoder-decoder + skip connections
• Output is sparsely connected components
• Input and output are similar at leading order

Pytorch implementation:
https://github.com/HaiwangYu/Pytorch-UNet
1, make time slices

Using other plane information projected to the target plane: utilizing plane redundancy

- Inspired by Wire-Cell ‘3D Imaging’.
  - arxiv:1803.04850

- Key algorithm “Fast projection” realized using “RayGrid” tools:

2, Matching active wires (with initial ROI) from multiple planes

3, On target plane, tag 3-Plane matched ROIs (MP3) or 2-Plane matched ROIs (MP2)
Input and label for DNN - ProtoDUNE Single-Phase simulation

Input candidates

tight/loose: deconvolution with tight/loose low frequency filter

label was made by rasterization of charge deposits – 2D gaussian response

Haiwang for CPAD 2021
ProtoDUNE Single-Phase

• NP04 experiment at CERN
• full-scale prototype for the first far detector module of DUNE
  • next-generation, long-baseline neutrino oscillation experiment

Sketch of a ProtoDUNE-SP APA from arxiv:1706.07081

major components of the ProtoDUNE-SP TPC from arxiv:1706.07081
2D convolution based simulation for LArTPC

- *JINST 13 P07006 (2018)*
- Software stack:
  - LArSoft for the generator and Geant4 simulation
  - Wire-Cell Toolkit for the detector response

Average waveform: data vs. MC

**U plane**
- MC scaling: 1.6
- Tick: [570, 970]
- Ch: [8050, 8200]

**V plane**
- MC scaling: 1.6
- Tick: [570, 970]
- Ch: [8950, 9200]

**W plane**
- MC scaling: 1.6
- Tick: [570, 970]
- Ch: [9500, 9750]
Efficient network – 500 samples already provide good performance – 6 min/epoch

- Platform: I9-9900K, 32 GB memory, Nvidia GTX 2080 Ti 11GB, Samsung 970 500GB NVMe SSD
- Sample: 500 APA Planes using cosmic generator (450 training, 50 validation)
- Loss: cross entropy
- Optimizer: Stochastic Gradient Descent with momentum
Example ROI finding event displays on simulated data

- **Truth**
- **Ref.**
- **DNN-ROI no Geometry**
- **DNN-ROI with Geometry**

**straight-line prolonged tracks**

**cosmic track with a section featuring a large projection angle**

**an interaction vertex of a charged pion with many activities**
Reference SP

ProtoDUNE DATA
run: 5145
subRun: 1
event: 26918
V plane

3/19/21
Haiwang for CPAD 2021
ProtoDUNE-SP Data
run: 5145
subRun: 1
event: 26945
V plane

ProtoDUNE: how and why
L. Manenti: X02.00001
and other ProtoDUNE Talks
Evaluation with simulation

- DNN ROI finding has higher efficiency and purity than the reference ROI finding especially at large angles.
- More effective on tracks have asymmetric angles induction planes
- DNN w/o MP information performs slightly better than Ref. but much worse than DNN w/ MP information

Pixel based evaluation:
\[ \text{eff} = \frac{0 + 3 + 0}{0+3+4} = \frac{3}{7} \]
\[ \text{purity} = \frac{0 + 3 + 0}{0+4+1} = \frac{3}{5} \]
Established initial machinery for Deep-Learning in Wire-Cell Toolkit

https://github.com/WireCell/wire-cell-toolkit

- Data preparing in LarSoft/Wire-Cell
- Training with python
- Production with C++
DNN ROI finding algorithm provides a software improvement on LArTPC reconstruction
• based on the the hardware feature - redundant wire plane
• combines domain knowledge with machine learning
• significant improvement on both simulation and ProtoDUNE-SP data


Wire-Cell: https://lar.bnl.gov/wire-cell/
Backup
RayGrid fast projection

Non-orthogonal coordinate system

\[ r_{ij}^{lm} = r_{00}^{lm} + j w^{lm} + i w^{ml} \]

Projection on the "n" plane

\[ p_{ij}^{lmn} = (r_{ij}^{lm} - c^n) \cdot \hat{p}^n \]
WireCell Detector Response Simulation

Ramo’s theorem
\[ i = e \vec{v} \cdot \vec{E}_v = e \vec{v} \cdot (-\nabla \phi) \]

Response matrix:
- Field response from Garfield simulation
- Electronic response

2D convolution – realized by FFT and IFFT

Add inherent electronics noise
- M. Diwan. *Basic mathematics of random noise part ½*

Refer: *JINST 13 P07006 (2018)*

2D Conv.
\[
M(t', x') = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} R(t, t', x, x') \cdot S(t, x) dt dx + N(t', x')
\]

Inherent electronic noise spectrum for different lengths of wires, *JINST 13 P07006 (2018)*
projection of a minimum ionizing particle (MIP) track on one induction plane
Input Candidates for DNN - ProtoDUNE data

- tight_lf
- loose_lf
- MP2
- MP3
## Training

<table>
<thead>
<tr>
<th></th>
<th>Unit Time</th>
<th>Total Time</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Generator</strong></td>
<td>2 sec/event</td>
<td>0.3 hour/500 events</td>
</tr>
<tr>
<td><strong>G4</strong></td>
<td>23 sec/event</td>
<td>3.2 hour/500 events</td>
</tr>
<tr>
<td><strong>detector response, truth tagging and waveform preprocessing</strong></td>
<td>68 sec/APA</td>
<td>9.4 hour/500 APA</td>
</tr>
<tr>
<td><strong>Network training</strong></td>
<td>6 min/epoch</td>
<td>5 hour/50 epoch</td>
</tr>
<tr>
<td></td>
<td>(1 epoch: 1 iteration of 500 APA sample)</td>
<td></td>
</tr>
<tr>
<td><strong>Sum</strong></td>
<td></td>
<td>17.9 hour</td>
</tr>
</tbody>
</table>
Mean loss of image samples are also expected to vary

We validated this hypothesis using a bootstrapping method. 1000 test samples (each sample contains 50 images) were randomly chosen from a test data set. Mean loss value was calculated for each sample and the distribution is shown below. We can see that the sample mean loss has a Gaussian like distribution that varies from 0.0055 to 0.0075.
ProtoDUNE Data
run: 5145
subRun: 1
event: 26925
V plane
ProtoDUNE Data run: 5145 subRun: 1 event: 26925 V plane