Plans for next LArTPC dataset

Brett + Haiwang Brett Viren

July 12, 2023

ML task and data sets

Learn domain translation of triplet of images (3-plane (U,V,W) ADC).

Three domains

- D1 simulation with q1D response, cosmics (kinematics)
- D2 simulation with 2D response, cosmics + AR39 decays (kinematics)
- D3 real detector 3D response, cosmics + AR39 decays (trigger), coherent noise

Common: ProtoDUNE-SP (3-plane, multi-APA), ADC, incoherent noise spectrum

Two translations

- D1 $\rightarrow$ D2 (sim $\rightarrow$ sim) both domains known, few paired and non-paired metrics.
- D2 $\rightarrow$ D3 (sim $\rightarrow$ det) only if/when D1 $\rightarrow$ D2 is successfully, no paring, many variations, non-paired metrics.

Metrics

Comparisons given metric $M(\cdot)$

- $|M(\text{D1}) - M(\text{D1} \to \text{D2})|$ is not small
- $|M(\text{D1} \to \text{D2}) - M(\text{D2})|$ is small
- Use $|M(\text{D1}) - M(\text{D2})|$ to provide scale to "small"

Paired vs unpaired

For D1 $\to$ D2 we may exploit the artificial pairing for proof-of-concept metrics.

Examples

- Pixel level or waveform level (paired) or mean waveform (unpaired).
- Total signal-process charge, with conventional SP and SP+DNNROI.
- Plane-charge consistency (paired or unpaired via means).
- Others: "WireMod", ICARUS MPV dQ/dx, ...

Questions

D1+D2: matched **samples** or merely matched **distributions**

- incoherent noise
- muon kinematics

Of course, D3 will not be matched samples nor distributions.

- Adding Ar39 only to D2 provides an element of unmatched distribution.

D3: ADC pathology

Real ProtoDUNE-SP data has ADC-related pathology that are not simulated.

- Includes ADC “stuck codes” and “ledge effect”.
- Do we apply mitigation processing or task UVCGAN to learn pathology?

Schedule

D1+D2

No real show stoppers

- Near term time pressure for other things delay start by a week(?)

D3

Need to select runs rich in cosmic muons (no particle beam)

- Probably delay start on D3 until D1 $\rightarrow$ D2 shows initial promise.