

Target dataset definition

- Dataset presentation/description:
 - Pixelated / wire TPC readouts
 - + optical
 - Each has:
 - Particle bomb (MPVMPPR)
 - True 3D information:
 - Pixel
 - PMTs on walls diagonal to pixel pads
 - G4 steps
 - Pre-detsim data + labels
 - Post-detsim data + labels
 - Wire
 - PMTs behind wire plane
 - Pre-detsim data + labels (3D)
 - (Mid detsim) Pre-response data + labels (2D)
 - 2D and 3D correspondence
 - Optical for both
 - Post-response waveforms + instance-separated waveforms

Overarching goals

Key challenges and AI tasks:

- Panoptic segmentation
- 2D $\leftrightarrow$ 3D correspondence
- TPC+Optical
 - E.g. flash matching but we don't have anything concrete yet

Potential AI solutions:

- Baseline: SPINE reconstruction
- FM: Panda reconstruction

Necessary tools:

- Setting up common eval code. This includes:
 - Finalizing metric definitions (all T1L)
 - Finalizing data loaders (Omar)
- Finalize and de-slop data viewers (all T1L)
- Documentation:
 - On methods & evaluation rubric & metric definitions (Junjie)
 - Dataset definition (Omar)
 - Generation instructions: (Patrick)

- Patrick's container + get full pipeline running

Timeline and prospects

- Expect to have dataset done by workshop start
- Tasks finished by end of workshop:
 - SPINE (Francois) and Panda (Sam) trained or training:
 - Just panoptic segmentation
 - & 2D $\leftrightarrow$ 3D for SPINE (& Panda..)
 - All relevant tools finished and packaged for downstream use.

Items that require all of T1-LArTPC:

- **metric definitions**
 - Quad chart for this
 - We give finalized list of metrics, consolidate and finalize ?
 - We should make slides explaining all metrics (desc. & visually)
- **data viewers**

Whether there is any specific T1/T2 joint topic worth scheduling?

- joint discussion on multi-generator simulations & reco studies/benchmarks using T2 generators and T1 simulators would be beneficial
 - Quad charts for what it means to have several generators and full simulation & reconstruction for them

What knowledge/tools/workflows we should make sure are transferred while the relevant US/Japan collaborators are together

- Eval code/backbone/system/whatever
- How to construct and use data loaders correctly
 - Incl. data format, structure etc.
- General foundation model introduction?
- General machine learning knowledge?