NuGraph for Wirecell Clustering

Wirecell Meeting

Avinay Bhat
07/30/2026

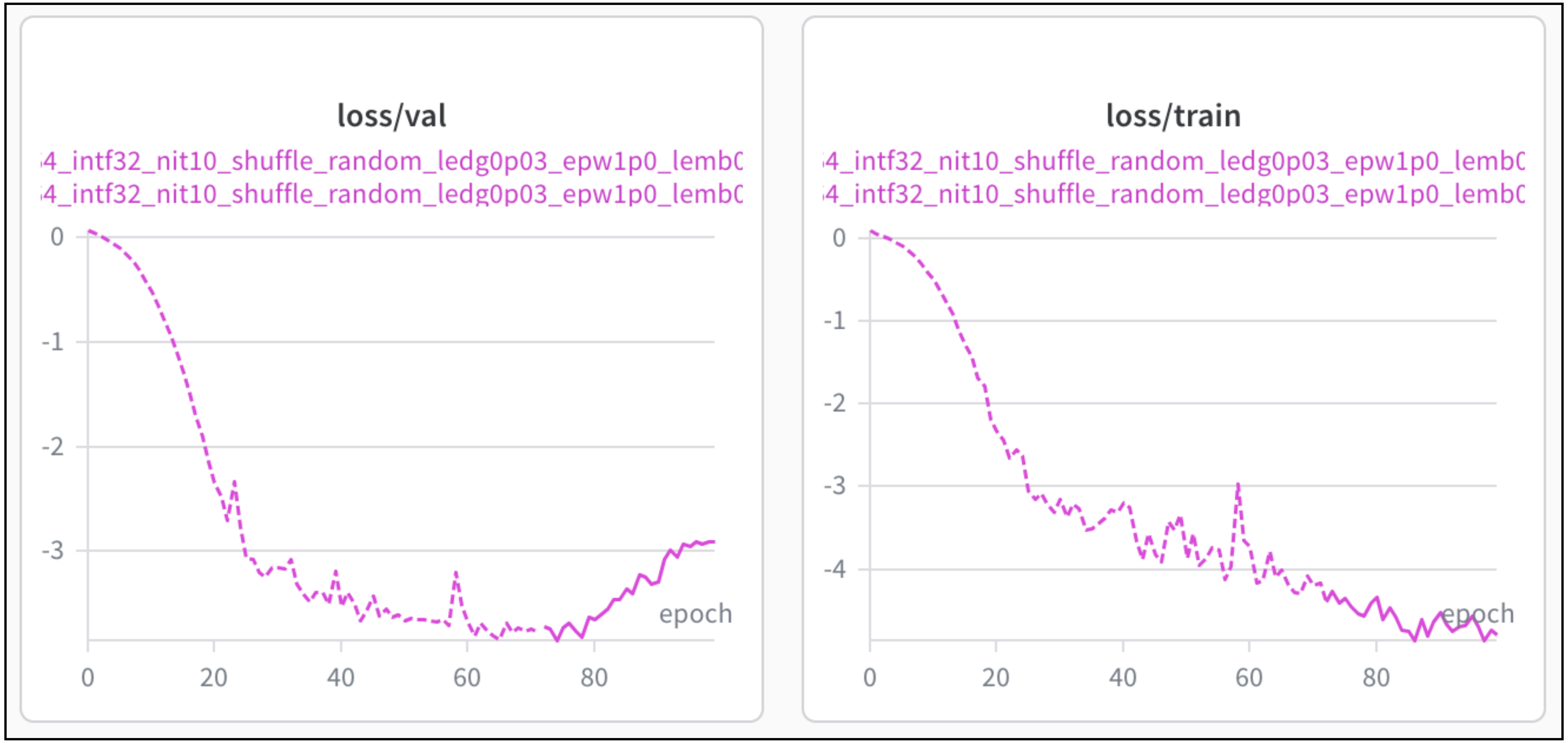
Training Dataset

- 352,613 simulated SBND events
- Wire-Cell 3D imaging/Clustering results
    - Merged 2D measurements -> NuGraph "Hit" Node
    - 3D Blobs -> NuGraph "SP" Node
- Truth-matched labels:
    - Hit-level semantic labels (ν / cosmic)
    - SP-level instance IDs (where available)
- Heavy class imbalance:
    - Many cosmic-only events
    - Sparse neutrino activity

Training Configuration

- Model: NuGraph Wirecell
- Features:
  - Minimal hit feature set (in_features = 4)
  - nexus features: 64
  - interaction features: 32
  - Message-passing iterations: 10
- Optimization:
  - AdamW + OneCycleLR
  - Learning rate: 3×10^{-4}
- Losses:
  - Semantic loss (Recall-weighted)
  - Instance embedding loss ($\lambda = 0.3$)
  - Edge loss ($\lambda = 0.03$)
- Distributed training on Sophia (DDP)

Training Results



- The training spikes are likely from OneCycleLR + weighted sampling hitting hard batches
- Validation is smoother

Precision, Recall and F1 score

Precision = Purity

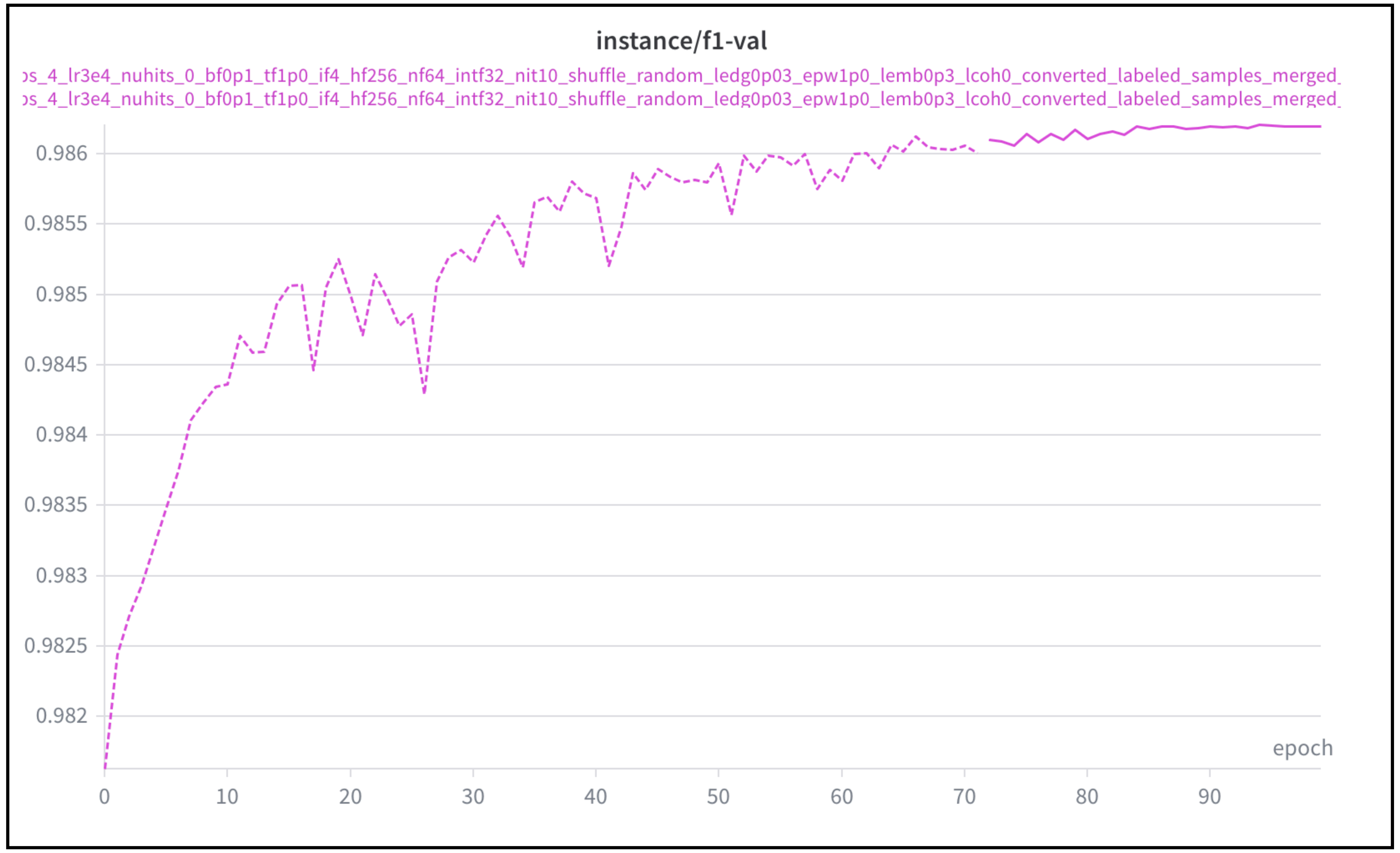
Recall = Efficiency

$$\text{Precision} = \frac{TP}{TP + FP}$$

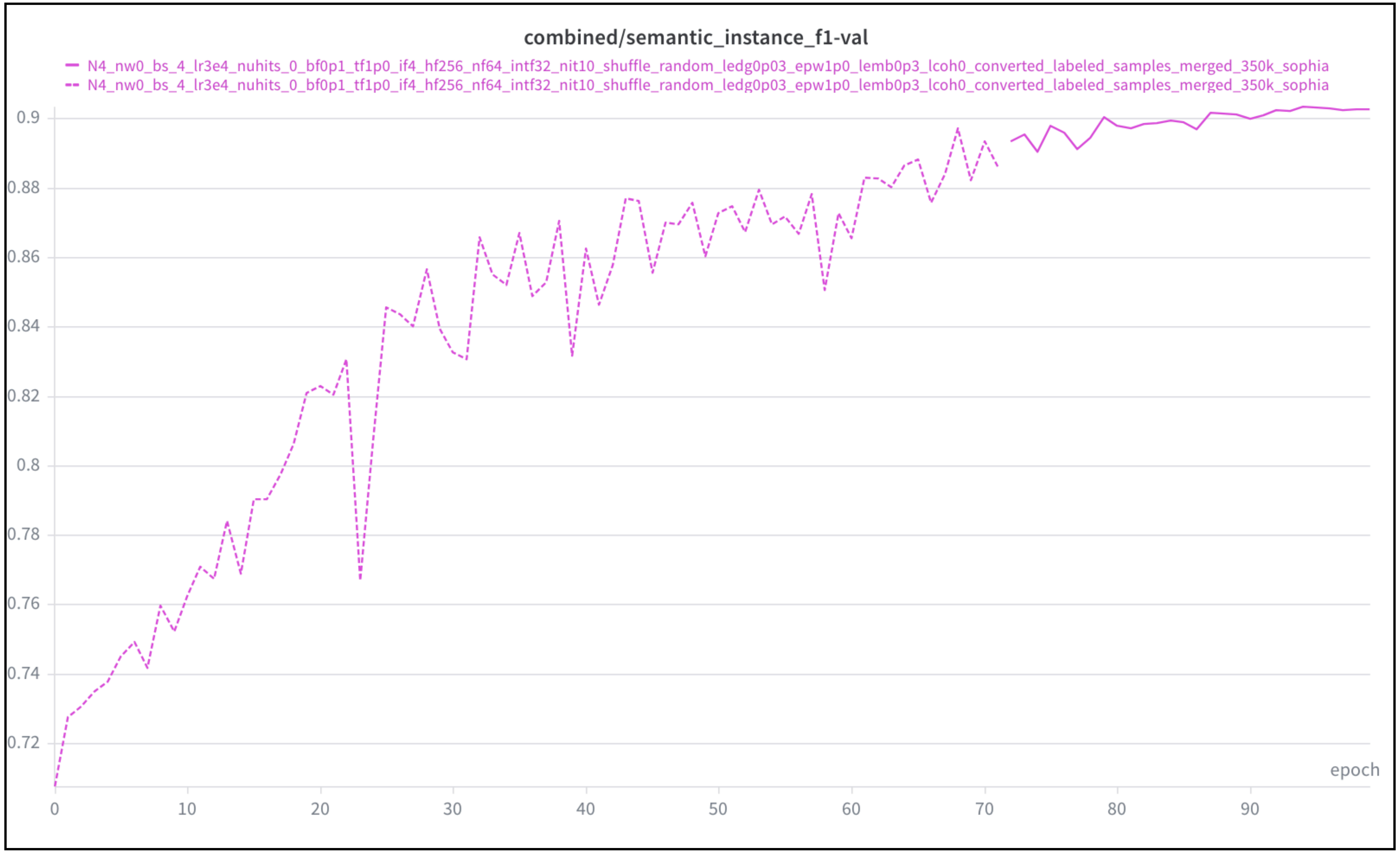
$$\text{Recall} = \frac{TP}{TP + FN}$$

$$\text{F1 Score} = \frac{2}{\frac{1}{\text{Precision}} + \frac{1}{\text{Recall}}} = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}$$

Validation Metrics



Validation Metrics



Three Tier Validation Strategy

- General held-out test sample: Evaluated 5,000 events from the merged ~350k-event BNB+cosmics training campaign.
- External NCπ^0 simulation: Evaluated all 1,255 test events from the independent 8,368-event BNB+cosmics NCπ^0 sample, without retraining or retuning thresholds.
- Real SBND data: Evaluated 19 run/subrun/event selections from Lynn’s Gen-1 NC sideband, corresponding to 38 APA records.
- The 19 data events were selected from a 94-event sideband; approximately 40–50% were expected to be NC, with CC and cosmic contamination intentionally accepted.
- This sequence tests ordinary generalization, performance on the photon-rich target topology, and finally simulation-to-data domain transfer.

Which data did I choose for testing model performance?

- NC interactions—and especially NCπ^0 final states—are central to the photon-shower clustering motivation.
- The sample contains real detector noise, dead/noisy channels, reconstruction imperfections, cosmics, and calibration effects absent or imperfectly modeled in simulation.
- It is an NC-enriched demonstration sample, not an unbiased sample for measuring NC selection efficiency or purity.

Real Data Production

- ART ROOT → WireCell NPZ → NuGraph HDF5

- Input files were reconstructed SBND BNBLight data from:
  - data_MCP2025B_05_DevSample_bnblight_v10_06_00_05_reco2_sbnd
  - data_MCP2025B_05_DevSample_Run1Rolling_bnblight_v10_06_00_05_reco2_sbnd

- Files reside below:
  - /pnfs/sbn/data_add/sbn_nd/poms_production/data/MCP2025B_05/v10_06_00_05/DevSample/reco2/bnblight/
  - /pnfs/sbn/data_add/sbn_nd/poms_production/data/MCP2025B_05/v10_06_00_05/DevSample_Run1Rolling/reco2/bnblight/

- Official upstream reconstruction used reco2_data.fcl.

- On the SBND GPVMs, the final submitted stage was img-clus, using the custom FCL:
  - /exp/sbnd/data/users/abhat/nugraph_data_sideband_lookup/wcls-img-clus-data-sptpc2d.fcl

- This FCL changed the WireCell input products from simulation simtpc2d tags to the real-data Reco1:sptpc2d Gaussian, Wiener, Wiener-summary, and bad-mask products.

HDF5 Conversion on EAF

- The 38 resulting rec-apa*.npz files were copied from PNFS to EAF scratch using FNAL bearer-token/XRootD access.

- Each NPZ was validated to contain ctpc_*, blobs, points, and ppedges.

- Produced a 156 MB inference-only HDF5 with 19 unique RSEs × two APAs.

Frozen NuGraph Model Applied to Data

- Used the best-joint-f1.ckpt from the ~350k-event Sophia training.

- NuGraph4 configuration:
  - 4 input features
  - 256-dimensional hit representation
  - 10 message-passing iterations
  - SP features enabled; vertex features disabled
  - Joint semantic, edge, and embedding objectives
  - lambda_edge = 0.03, lambda_embed = 0.3

- Training ran on six Sophia DGX nodes with eight A100 GPUs per node: 48 MPI-launched distributed ranks.

- Operating points were frozen from MC:
  - Semantic neutrino-score threshold: 0.612
  - Same-instance edge threshold: 0.30
  - Minimum cluster size: 2

- Candidate graphs use reconstruction information only: mapped reconstruction edges/ppedges, geometry kNN-radius edges (k=8, 80 mm), and reco-cluster neighbors (k=4).

- The real-data HDF5 explicitly records inference_only=1, has_semantic_truth=0, and has_instance_truth=0.

Semantic Results for NCPi0 ν Events (Simulation)

- Derived metrics for Semantic threshold: ν score >= 0.612
  - ν precision 0.8689
  - ν recall 0.8115
  - ν F1 0.8392
  - cosmic F1 0.9902
  - accuracy 0.9815
  - macro F1 0.9147
  - weighted F1 0.9812

Semantic Confusion Matric

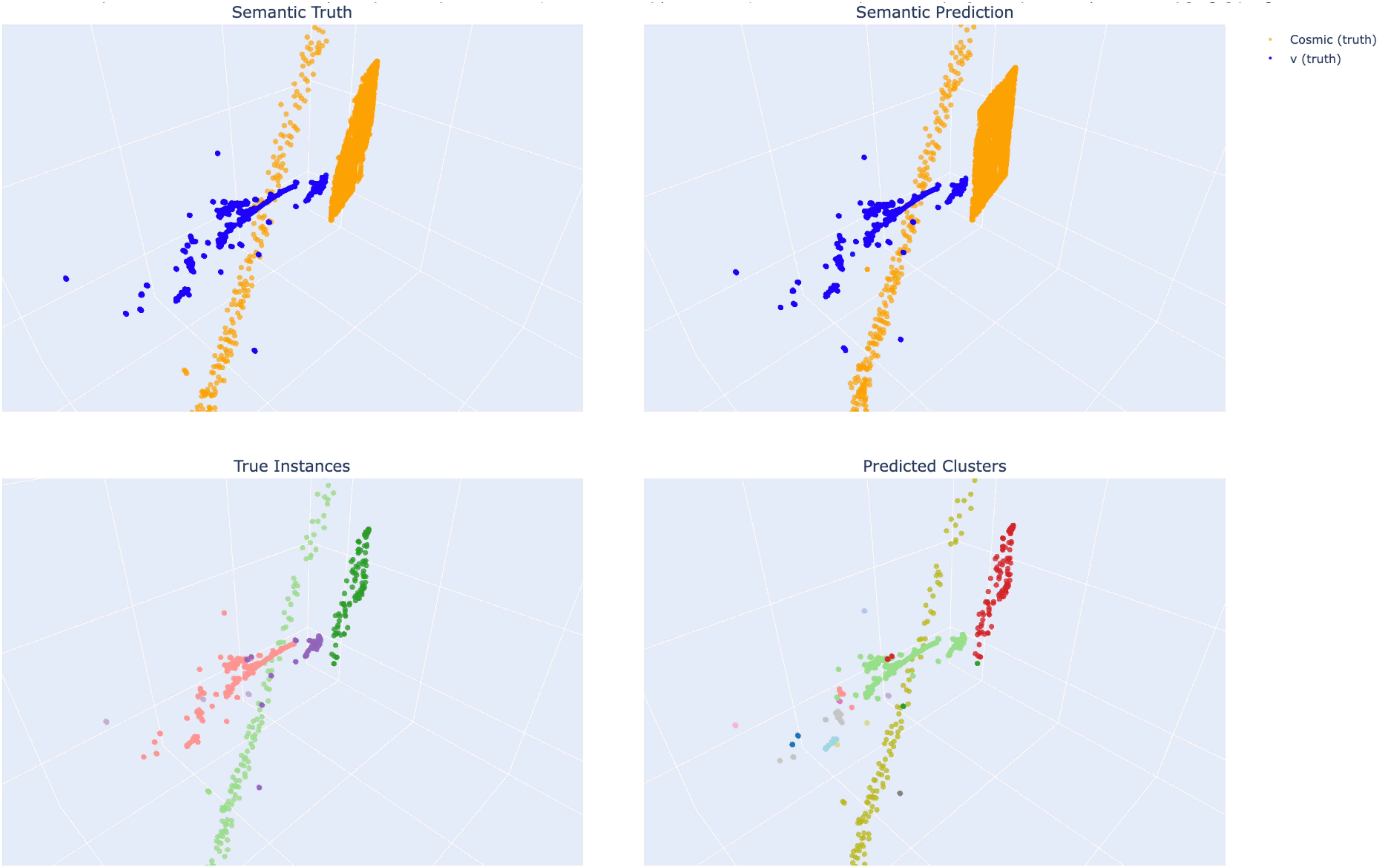
	Predicted ν	**Predicted Cosmic**
True ν	158,517	36,832
True Cosmic	23,913	3,055,513

- Event-level ν detection with permissive event cut:
  - pred_nu_hits >= 1, pred_nu_frac >= 0
    - ν-event precision 0.8804
    - ν-event recall 0.9022
    - ν-event F1 0.8911
    - accuracy 0.8295

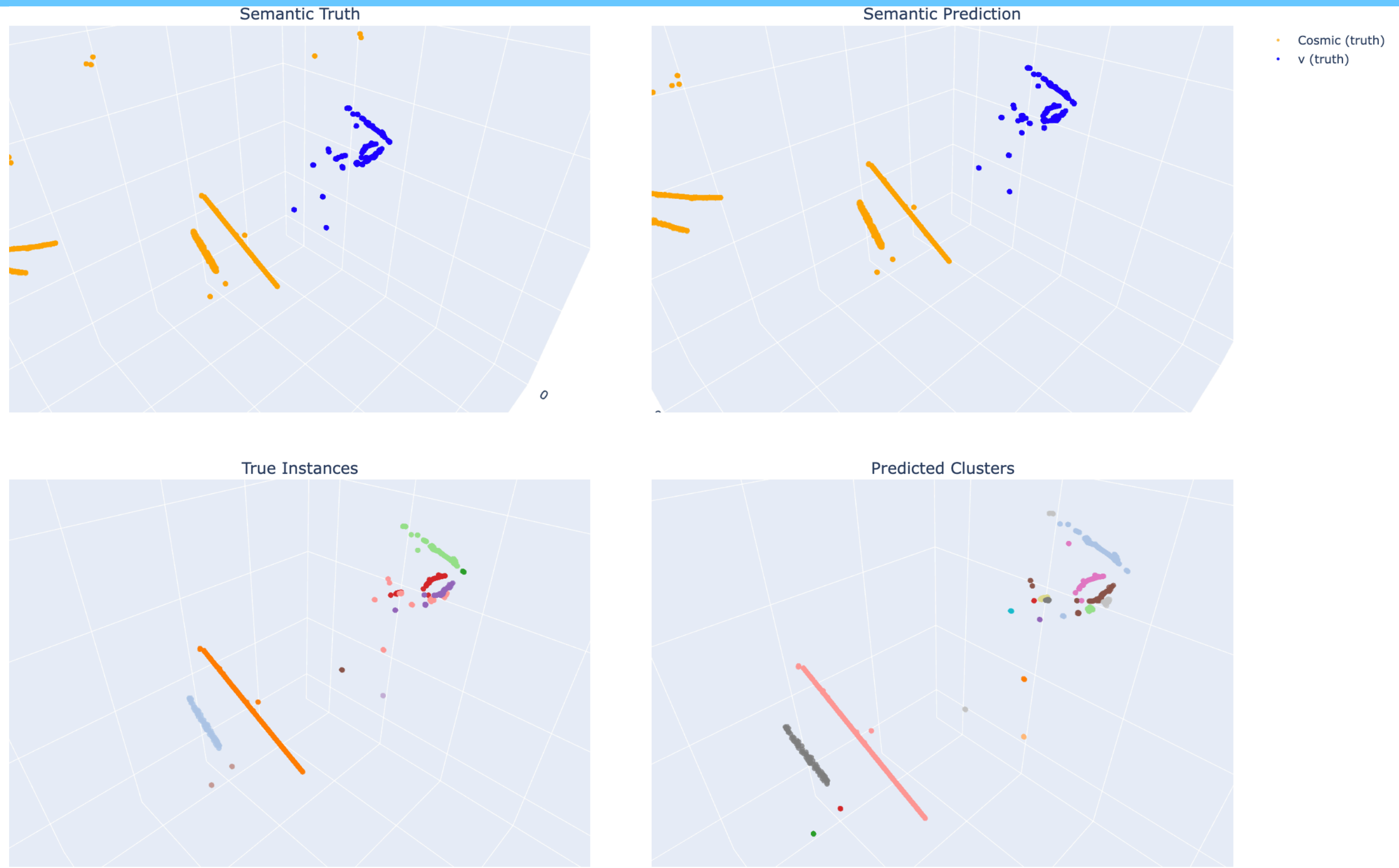
Instance Results for NCPi0 ν Events (Simulation)

- Task: predict whether two spacepoints belong to the same physical particle
- Positive class: same-instance edge
- Threshold: p(same) ≥ 0.3
- SP edge classification:
  - ROC-AUC 0.9975
  - PR-AUC 0.9963
  - same precision 0.9741
  - same recall 0.9972
  - same F1 0.9855
  - edge accuracy 0.9860
- SP instance clustering:
  - Graphs used 1,254
  - Total labeled SP 1,911,373
  - Weighted ARI 0.8553
  - Weighted purity 0.9347
  - Weighted completeness 0.9232

Visual Inspection of NCPi0 ν Performance



Visual Inspection of NCPi0 ν Performance



Results on NC enriched SBND data

- The 38 APA graphs contain 123,835 SPs; descriptively, 20.4% were classified neutrino-like, 79.0% cosmic-like, and 0.6% unmapped.
- Through-going cosmic tracks and localized neutrino-like activity remain topologically well separated.
- Major tracks and shower cores are generally coherent, with fragmentation concentrated around dense vertices, crossing tracks, and diffuse shower edges.
- These results support successful qualitative domain transfer, but not yet a quantitative data-performance claim. The next defensible step is expert hand-scanning or comparison with an independent reconstruction/selection label.

Back Up Slides