



Wire-Cell ProtoDUNE Local Meeting

Yujin Park
Chung-Ang University
Sep02, 2026 (Wed)

Wire-Cell DNNROI

8.1 Held-out comparison of the distilled and quantized models

Table 3 reports the held-out test metrics for the compact model in its successive forms: the standalone baseline trained without distillation, the distilled students of the three knowledge-distillation strategies of section 6, and the two INT8 QAT-distilled models. Every model in the table has the identical MobileNetV3 U-Net architecture, evaluated on the same 60-event held-out test set; for each model the checkpoint with the lowest validation loss is used.

Table 3. Held-out test metrics (60 events) for the compact MobileNetV3 U-Net: the standalone baseline, the distilled students of the three distillation strategies, and the two INT8 QAT-distilled models. All rows share the same network architecture and therefore the same inference cost.

model	distillation	Dice	ROI eff.	ROI pur.
MobileNetV3 U-Net	— (baseline, no KD)	0.891	0.687	0.976
MobileNetV3 ← Transformer	Strategy A, response-only	0.911	0.747	0.949
MobileNetV3 ← Nested U-Net	Strategy B, feature KD	0.909	0.736	0.956
MobileNetV3 ← Transformer	Strategy C, bottleneck KD	0.912	0.761	0.927
MobileNetV3 ← Transformer	Strategy C, INT8 QAT-KD	0.893	0.727	0.920
MobileNetV3 ← Nested U-Net	Strategy B, INT8 QAT-KD	0.898	0.698	0.963

Knowledge distillation improves the compact model substantially. Relative to the no-KD baseline, the selected production recipe (Strategy C, the Transformer teacher with bottleneck feature distillation) raises the Dice score from 0.891 to 0.912 and, more importantly, the ROI efficiency from 0.687 to 0.761, a gain of more than seven percentage points in the recall of true ROIs. The two feature-distillation strategies (B and C) improve on response-only distillation (A), with the Transformer-bottleneck tap of Strategy C the strongest. The improvement comes at a modest cost in ROI purity, from 0.976 to 0.927; this trade is acceptable because the downstream reconstruction is limited by recall: a missed ROI is unrecoverable, whereas a spurious one can be rejected by later stages. For reference, the large teacher networks reach a validation-set Dice of about 0.92 to 0.93, so the distilled student closes a large fraction of the student–teacher gap while keeping the compact model’s small size and fast inference.

Selected Best Model: Strategy C (MobileNetV3 ← Transformer, Bottleneck KD)

- Achieves the highest Dice score and ROI efficiency
 - The drop in ROI purity is an acceptable trade-off.
-
- While Strategy C looks the best model, qualitative justification alone leaves ambiguity in the trade-off.
 - Adopted the F_β **score** to rigorously balance ROI purity and ROI efficiency and determine the best model

Wire-Cell DNNROI

F_β score:

- A fundamental evaluation metric in machine learning that combines Precision and Recall, balancing their trade-off using a weighting parameter β
- $\beta = 1$: Balances Precision and Recall equally (Standard F_1 Score).
- $\beta < 1$ ($\beta = 0.5$) Prioritizes precision(purity)
- $\beta > 1$ ($\beta = 2.0$): **Prioritizes recall(efficiency)**

$$F_\beta = (1 + \beta^2) \frac{\text{Precision} \times \text{Recall}}{(\beta^2 \times \text{Precision}) + \text{Recall}}$$

Xin

run	Dice	eff_roi	pur_roi	F2	Rank
baseline	0.8900	0.6870	0.9760	0.7302	6
A-UNet	0.9110	0.7520	0.9440	0.7839	2
A-Nested	0.9090	0.7530	0.9360	0.7835	3
A-Tx	0.9110	0.7470	0.9490	0.7803	4
B	0.9085	0.7359	0.9495	0.7706	5
C	0.9118	0.7609	0.9273	0.7892	1

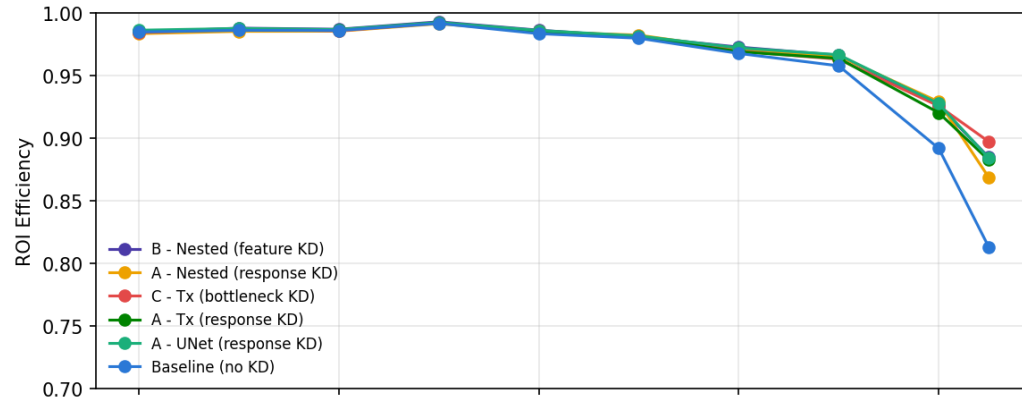
Yujin

run	Dice	eff_roi	pur_roi	F2	Rank
baseline	0.9069	0.7179	0.9689	0.7570	6
A-UNet	0.9101	0.7345	0.9594	0.7708	5
A-Nested	0.9099	0.7444	0.9503	0.7781	4
A-Tx	0.9116	0.7562	0.9365	0.7865	2
B	0.9068	0.7545	0.9347	0.7848	3
C	0.9129	0.7592	0.9408	0.7897	1

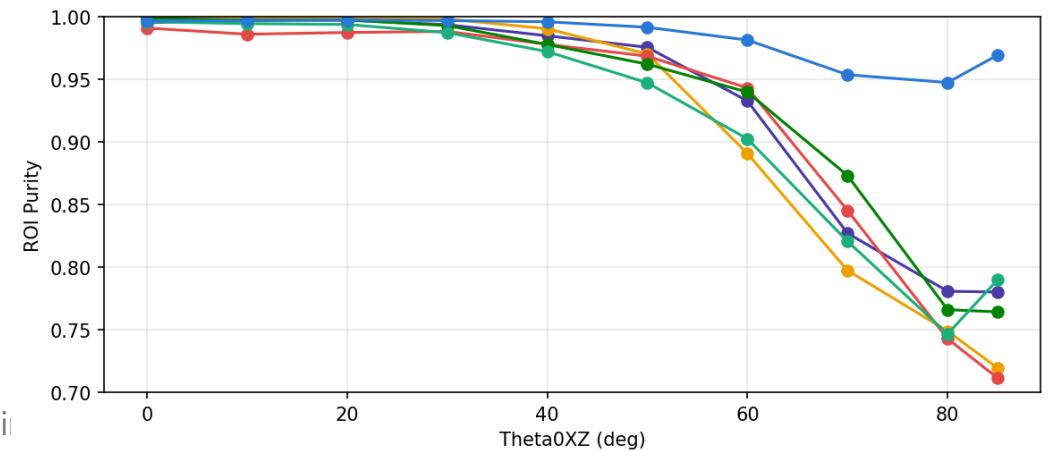
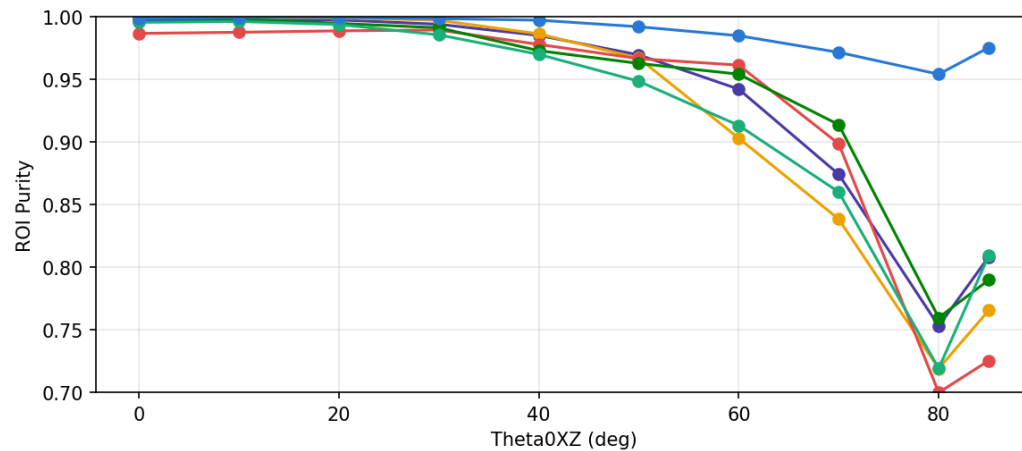
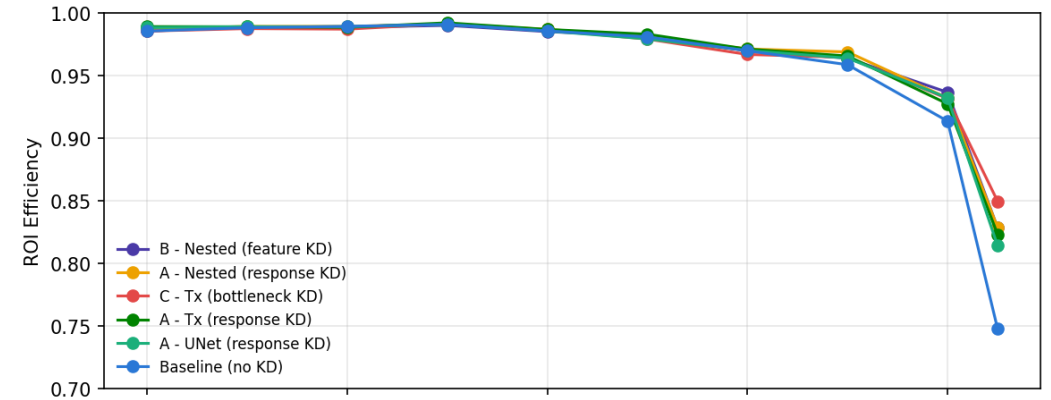
Wire-Cell DNNROI

- LArWC simulation with 1GeV muon tracks events
- anode 1
- 10 events per angle

U plane



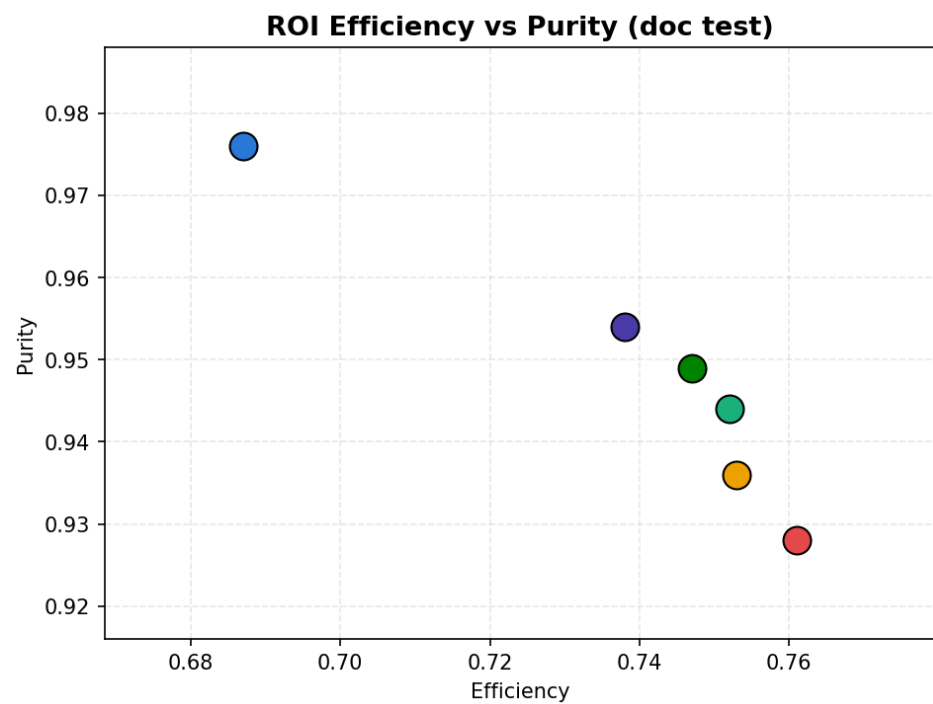
V plane



Backup



xin



yujin

