

# Status report on **DNN SigProc in PD-HD**

Hokyeong Nam Chung-Ang University

### **Outline**

Comparison of DNN Models

- DNN-ROI Performance Evaluation
  - Single Track Events
  - Single Shower Events

• Time/Memory Usage Measurement

Summary

# **Model Comparison - Network Architectures**

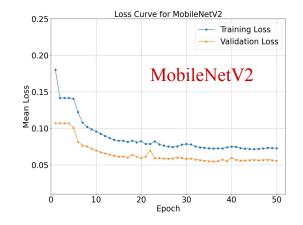
Model	Encoder	Decoder	Skip Connection	Downsampling Depth	Activations	
					Encoder	Decoder
UNet	Convs	4 Convs	Yes	1/16	ReLU	ReLU
MobileNetV2-UNet	Depthwise separable	2 Convs	None	1/32	ReLU6	ReLU
MobileNetV3-UNet	Depthwise separable + SE	4 Convs	Yes	1/32	h-swish + ReLU	ReLU
Transformer-UNet	Convs + Transformer bottleneck	4 Convs	Yes	1/16	ReLU	ReLU
					GELU (transformer)	

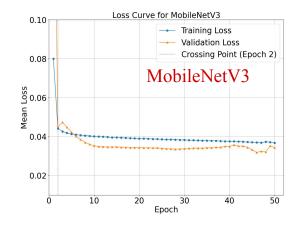
- Training dataset: 590 cosmic-ray events
- Optimizer: SGD (Stochastic Gradient Descent)
- Learning rate: 0.1
- Early stopping: Enabled

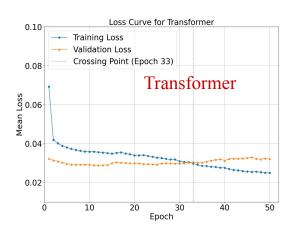
- Train/Val split: 0.9/0.1
- Loss: BCELoss (Binary Cross-Entropy Loss)
- Number of epochs : 50
- Output activation function: Sigmoid

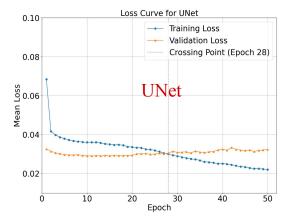
Training was carried out on the WC Cluster using an NVIDIA GeForce RTX 4090 GPU (24 GB)

# **Model Comparison - Train vs Val loss**









- Rebin factor was set to 10 during training
- Training losses at the selected checkpoints:
  - UNet: 0.029 (Epoch 28)
  - MobileNetV2: 0.072 (Epoch 50)
  - o MobileNetV3: 0.044 (Epoch 50)
  - o Transformer: 0.029 (Epoch 32)
- Among the models, UNet and Transformer reached the most stable convergence with the lowest final losses

#### **DNN-ROI Performance Evaluation**

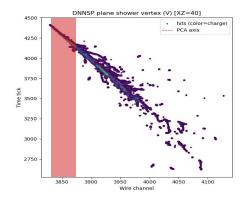
• For track events, three metrics are used: Bias, Resolution, and Inefficiency

$$Bias = 100 \times \left( \left| \frac{Q_{reco}}{Q_{truth}} \right| - 1 \right) \quad Resolution = 100 \times \frac{RMS\left( \frac{Q_{reco}}{Q_{truth}} \right)}{\left| \frac{Q_{reco}}{Q_{truth}} \right|} \quad Inefficiency = 100 \times \frac{Number of bad channels}{Number of valid truth channels}$$

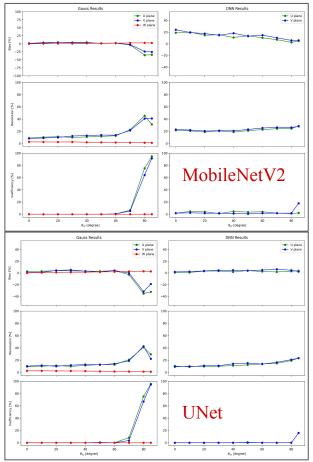
- For shower events, a charge profile based on vertex information was added as the fourth metric:
  - $\circ$  Sum the charge along the shower direction up to 42 wire channels ( $\approx 1-2$  radiation lengths)
  - Compare the reconstructed-to-truth ratio charge ratio

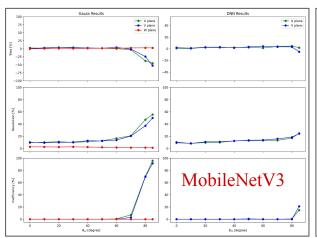
$$Q_{method} = \sum_{w \in W} Q_{method}(w) \qquad \qquad R_{cprofile} = \frac{Q_{reco}}{Q_{truth}}$$

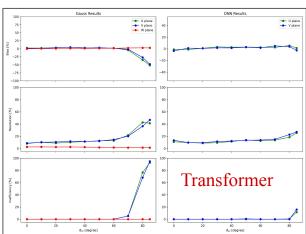
- Samples were generated with
  - Detector configuration: ProtoDUNE Horizontal Drift (PD-HD)
  - o XZ angle: 0°, 10°, 20°, 30°, 40°, 50°, 60°, 70°, 80°, 85°
  - o Shower energies: 100 MeV, 500 MeV, 1 GeV, 2 GeV, 3 GeV, 5 GeV
  - o Software: WCT standalone (Tracks), LAr-WCT (Showers)



# **Single Track Evaluation**

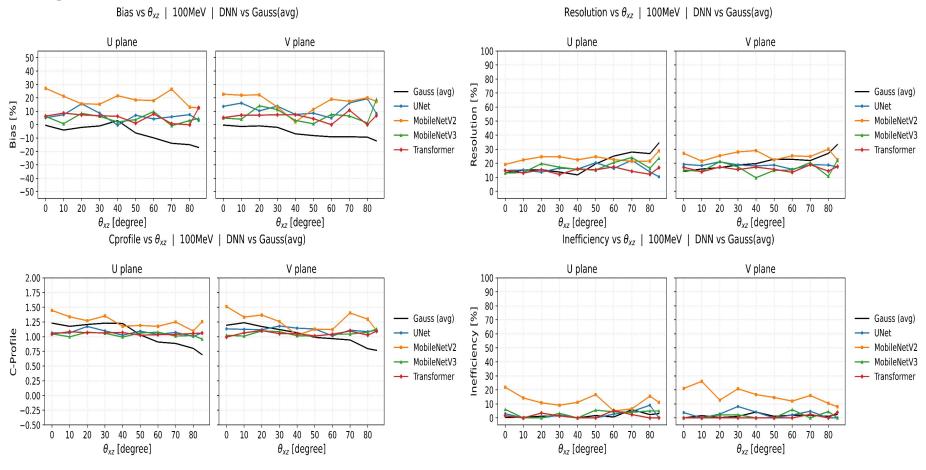






- All simulations were performed using the normal APA (2nd)
- MobileNetV2 exhibits ~20% bias at low angles
- The lower performance of MobileNetV2 is likely due to:
  - Fewer convolution blocks in the decoder
  - Absence of skip connections
- Other models demonstrate comparable performance

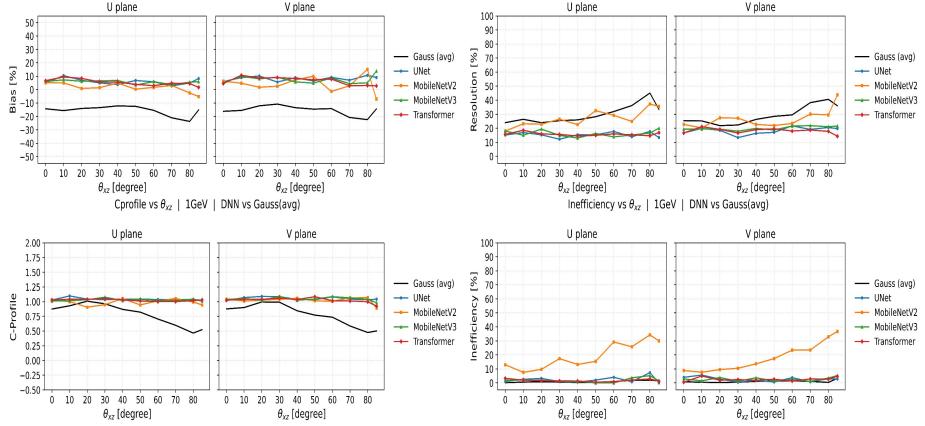
# **Single Shower Evaluation (100 MeV)**



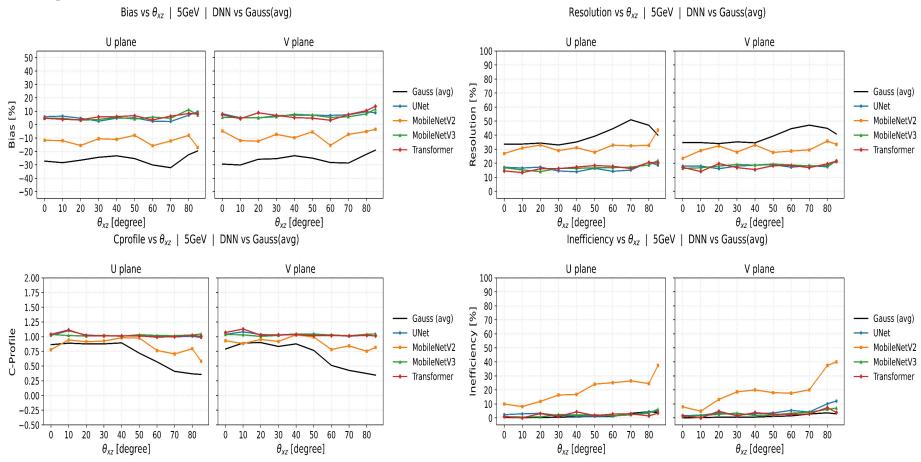
# **Single Shower Evaluation (1 GeV)**



Resolution vs  $\theta_{xz}$  | 1GeV | DNN vs Gauss(avg)



# **Single Shower Evaluation (5 GeV)**

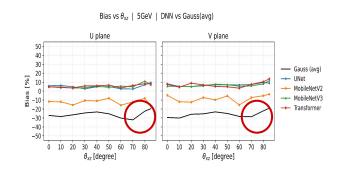


# **Single Shower Evaluation**

Evaluation plots for 500 MeV, 2 GeV and 3 GeV can be found on slides 19-21

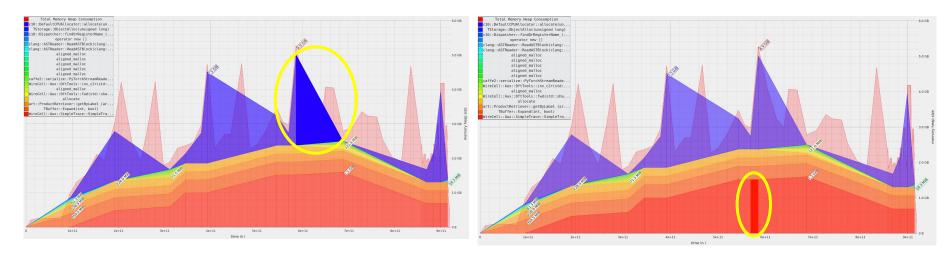
#### As the electron energy increases:

- Traditional ROI
  - Bias ↓, Charge Profile ↓, Resolution ↓, Inefficiency neary constant
- DNN-ROI (except MobileNetV2)
  - Bias ↑, Charge Profile ↑
  - $\circ$  Resolution  $\uparrow$  ( $\leq$  1 GeV), then flattens
  - Inefficiency  $\uparrow$  ( $\leq 500 \text{ MeV}$ ), then flattens

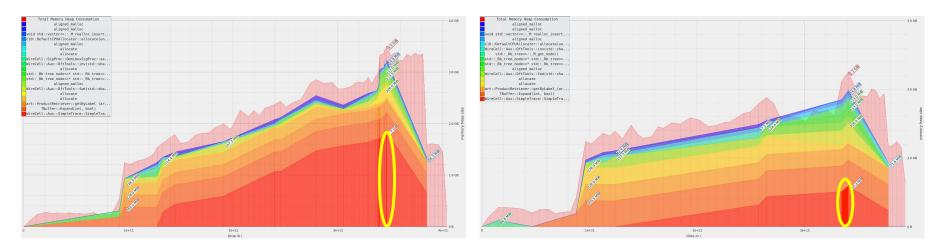


## **Remaining Questions**

- Why does the Traditional ROI appear to improve at high angles as the energy increases?
- Why does the inefficiency of the Traditional ROI remain close to zero across all energies and angles?

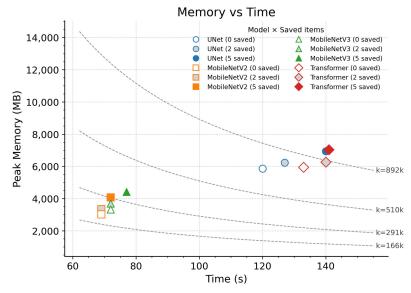


- Valgrind profiling was performed on the WC Cluster (using an AMD Ryzen Threadripper 7970X, 32 cores, 64 threads)
- DNN SigProc with **UNet** shows a peak memory usage of 4.9 GiB ( $\approx 5.26$  GB)
- Two main contributors to memory consumption (highlighted in the yellow circle):
  - Blue region (2.5 GiB): *libtorch\_cpu.so* (2.3GiB)
  - Red region (1.4 GiB): save\_data (858 MiB) and save\_mproi (343 MiB) in OmnibusSigProc



- Valgrind profiling was performed on the WC Cluster
  (using an AMD Ryzen Threadripper 7970X, 32 cores, 64 threads)
- With MobileNet, *libtorch\_cpu.so* consumes **less than 50 MiB**
- The main remaining factor of memory consumption (highlighted in the yellow circle):
  - Red region (1.7 GiB): save\_data (972 MiB) and save\_mproi (440 MiB) in OmnibusSigProc
- When *save\_data* is deactivated for unnecessary operations (right plot):
  - Red region (565 MiB): save\_mproi (440 MiB) in OmnibusSigProc

save_data	Model	Peak Memory (MB)	Time (s)	
	UNet	6942	140	
tight_lf, loose_lf,	Transformer	7032	141	
cleanup_roi, wiener, gauss	MobileNetV2	4071	72	
	MobileNetV3	4410	77	
	UNet	6225	127	
	Transformer	6267	140	
wiener, gauss	MobileNetV2	3343	69	
	MobileNetV3	3687	72	
	UNet	5863	120	
	Transformer	5945	133	
none	MobileNetV2	3012	69	
	MobileNetV3	3324	72	



- Benchmark was performed on dunegpvm03
  (AMD EPYC Processor, 4 vCPUs)
- The *lar* process utilized only a single vCPU
- DNN-ROI with MobileNetV3 consumes 1.5 GB more memory than Traditional ROI (see slide 24)

 $Data: np04hd\_raw\_run027673\_0000\_dataflow0\_datawriter\_0\_20240704T050545.hdf5$ 

# **Summary**

#### DNN-ROI Performance Evaluation

- UNet, MobileNetV3, and Transformer exhibited similar performance, all outperforming the Traditional ROI
- The charge profile indicated that the Traditional ROI degrades as the angle increases
- The underperformance of MobileNetV2 is due to weaker design of its decoder architectures

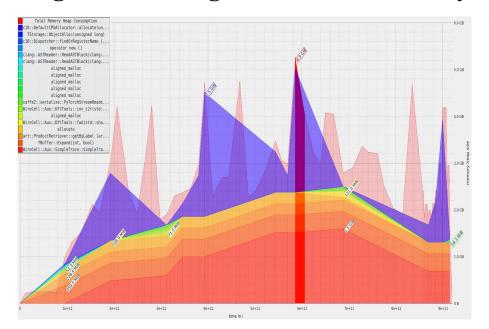
#### Time/Memory Consumption

- Deactivating *save\_data* function can reduce memory usage by approximately 1 GB
- The UNet encoder architecture is the primary contributor (~3 GB) to memory usage

Considering both performance and computational constraints, MobileNetV3 provided the best overall balance

# Back Up

# Valgrind Profiling - UNet Peak Memory



```
 25 GB: c18:aloc_cpulus igned longl (in )cvmfs/assift operaciencegrid org/products/littorchiv2 1_1bit inux44bit+3.18-2.17-e36/lib/libc18.sol

 25 GB. c0::Delaut/CAVInctor: alreate/uniqued long cont in /confelence operations products library (2.1 library 40t-1302) 1-426/b/(600.01)

 23 GB: (1): Stragetrol: Storagetrol (1): Stragetrol: see bete size L; (1): Syniat conts. (1): Allocator: bool in it verifylands associaceard anabody statistics (1): 114 (au. 54): +311-217-256 foliators could be seen as a final control of the contro

 2.3 Gib at Personase at detail: empty generic or (200 Amy Arefor, cito-Alexator), cito-Dispatch Ryset, cito-Scalar/lype, cito
                                                             * 23 GB striktalcerpty generic/10-freePerior, clin Blocker*, clin Dispatotie; Set, clin Scalaringe, citicoptionalco) (in primisianant operacience prior operator (in 116) (in 

 23 GB; st:detail empty, quirtif chrayRef
 diff ScaleType, bod, difficult solves
 in inversions region improducts libbrorive 2.1 Inform Mit + 3.04-2.07 Applications on purpose and products libbrorive 2.1 Inform Mit + 3.04-2.07 Applications on purpose and pur

 2368 stristskenty out@b.km/delo.@bostosio.@bostosio.@bostosio.@bostosio.@bostosio.in tyrektrut.comsiecomi rapadytsilitetrit@1 lidina68b1-318217436b8mt ou sil

 2.1 Gib. strative-energi quicitis-hardefor, cill-soptioni or, cill-sopti or, cill-sopti or, cill-sopti or, cill-soptioni or, cill-soptioni or, cill-soptioni or, 
                                                                                                                                   * 21 GB clit ingluing lenet fundar unbused or call (20 Operaturisme)*, 200 Object horizones (20 Armyleico, (20 
                                                                                                                                                    * 2.168 st. recently menory furnet reducate id this interfect, all this interfect, 
                                                                                                                                                                         *2.168; Olt indows lenel furbin unbook occall (10 Occabrillene)* (10 Occabrillene)* (10 Occabrillene)* (10 Invanishes)* (1
                                                                                                                                                                                         * 2.1 GB at accept, menor forms; call (10 km/p6/o, (10 cofondo, (10 cofondo, (10 cofondo, (10 cofondo) (10 cofondo o) (in c

 2.1 Gill: attempty(c1): Arrayled-o, c10: TensorOptions, c10: optional-o) (in yourfs)lasset operationaginal organizats (ibstrative 2.1 lbs/inunitable 3.10-2.17-e25 (ibs/inunitable 3.10-2.17-e25) (ibs/inunitable 3.10-2.17-e25)

 2.1 Gib stratile (anomous namesase) comute columni first Person const. (1) Aranfel co. (1) Aranfel co. (1) Aranfel co. (1) Aranfel co. (2) Aranfel co. (3) Aranfe
                                                                                                                                                                                                                                          • 2168 strative skn and though our quiet Tear and Life 

 2.1 Gib strative sins con/2 feward quilat Peror crosts, at Peror crosts, at Peror crosts, Othersyleto, Others
                                                                                                                                                                                                                                                                                * 2.1 GK (Cliniquing level furby valued <- call (Oxtoperatoriene*, Oxtoperatoriene*, Oxtoperatoriene*,

 2.1 Gill at: sps: siou consist invased collists ferror consist, at: Terror consist, at: Terro

 2.1 Gib structus tim consist lieus contis, at leur contis, cit/eur contis, cit/anylefe>, cit/optonie> contis, cit/anylefe>, cit/optonie> cit/anylefe>, cit/an
                                                                                                                                                                                                                                                                                                                                * 1168 dibingt was level fundor whosel <a color="block">million dispersance (in the color of the
                                                                                                                                                                                                                                                                                                                                                      * 21 SE at positivo consistatifat invarioresia at invarioresia (circianyario), citic appealo consis. (dictanyario), citic tanyario) in jovenistano apresionegrido gyrodustation (vol. 1 latinus Alire 1 10 21 a 25 libitios di que sel

 2.1 Gibt at native; convolution regroup backerisht; Tensor consts, at Tensor consts, at Tensor consts, at making-convincence consts in hornest prescriberary approducts (interpretable and prescriberary approach and products (interpretable and prescriberary approach and prescribe
                                                                                                                                                                                                                                                                                                                                                                                      * 21 GB at trackier connotion fat Senar consts, at Tenar consts, Citi-copional-consts, Citi-copional-consts, Citi-copional-consts, Citi-copional-consts, Citi-copional-consts, Citi-copional-copional-copional-copional-copi
                                                                                                                                                                                                                                                                                                                                                                                                      * 2.1 Gibt (cl.) implaces joined incinc untones exceeded Chiperaturisment of Chiperatu
                                                                                                                                                                                                                                                                                                                                                                                                                      * 1168 at i.us; conculor redispath (i.d. Dispath) pist, at Tenor credi, at Tenor credit, at Tenor
                                                                                                                                                                                                                                                                                                                                                                                                                                       * 21 SE both: adopted: Variableing clarentons correspond: controllation (2) Disablifeins d. at Feore cross. at

 2.1 Gife citi-implimate based from unbased function conflicts Operator Remain (citi-Operator Remain citi-Operator Remain (citi-Operator Remain Citi-Operator Remain 

 2.1 GB cld: Dispatcher collisional cld: Operatorised le const., structure > 0 const. (cone insulf for combinered operatorised compositionals) 2.1 bit in usbate 3.03.2 17 425 (billation) could

 2.1 GB toda it intervete Statempl runing list sectors $6 in Joints less it operations products (blanch $2.1 lb Linux $60+3.19.2.17-2.5 (b) list on the less it in the less it is a second or control of the list 
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                             2.1 GB: turth lit interpreter Statemach multichiector <> 6 for itumisfare of uperscience and proportions (library) via 1.1 Lb Linux (bit + 3.19-2.17-e) Silla (district coulse)

 245.4 MB: at:detail:empty qual;10:4mayRef-cx, c10:7ensorOptions constál in Journslands Expression opproducts/literative 2.1 bibliouxFait+3.00 2.17 e26/follotoch quusal

 1177 MB at native resize bytes quict!! Storagetry! unsigned long! (in confishered) operacienceprid organisats/fishorth/V2 1 th/(num64bit+110-217-e26/fish)/borth/quiss)

                             + 51.3 MB cafe2 sension: Prior Streen Reader (perfectorists): cocil basic strop-consts) in (normalized operations) organized literatury 2.1 bit incut 45+1.14-2.12-4.56 bit board operations and a cafe sension of products in the cafe sension of the
             • 1.4 GB: alocate (aloc traits/c464)

 1.4 GiB: M allocate (sti vectoch:378)

 14 GiB: IV create storage (sd vectoch:395)

 1.4 GB: Vector base (stl vector h 332)

 1.4 GB; vedor (st) vedor h5981

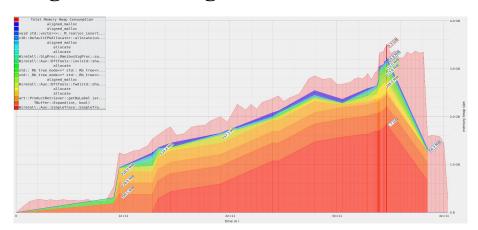
 14 GB: WinCell-Aux SimpleTaxe: SimpleTaxeInt. int, std:-vector-o-consts! (SimpleTaxe.cox8)

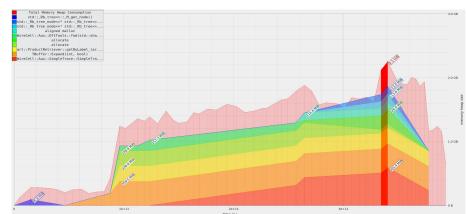
                                                                                                                   + 883 MB Mintel-Subtranships are identifications, of meteros, in other rounds, otherway, other conflicted conditional (Immunisher condition)
                                                                                                                   > 343.3 MB: VireCell:Sights::OnnibusSights: save monolatic vector-s-4, std:vector-s-4, int, std:multirep-s-const4.) (OrnibusSightscoss:NE)

 IZLI MB: VinGelt-Sinho: OnnibuSuPro: save rolativesture-6, abt mestor-6, int. abt mestor-6, iDminusinho.cox612

                                                                                                                   > 73.2 MB: Construct-o-(st) construct is 129
```

# Valgrind Profiling - MobileNetV2 Peak Memory





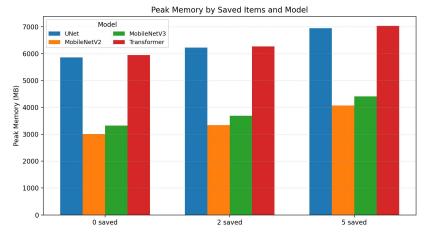
#### ▼ 3.3 GiB: Snapshot #78 (peak)

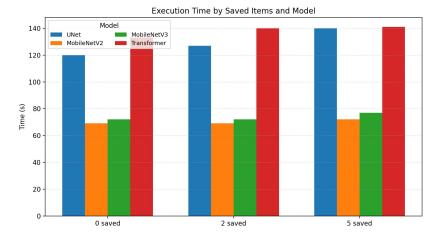
- ▼ 1.7 GiB: allocate (new allocator.h:137)
- 1.7 GiB: allocate (alloc\_traits.h:464)
  1.7 GiB: M allocate (stl vector.h:378)
  - ▼ 1.7 GiB: M create storage (stl vector.h:395)
  - ▼ 1.7 GiB: Vector base (stl vector.h:332)
  - 1.7 GiB: vector (stl vector.h:598)
  - ▼ 1.7 GiB: WireCell::Aux::SimpleTrace::SimpleTrace(int, int, std::vector<> const&) (SimpleTrace.cxx:8)
  - > 972.7 MiB: WireCell::SigProc::OmnibusSigProc::save data(std::vector<>&, std::vector<> const&, std::vector<> k, std::vector<> const&, std::vector<> k, std::vector<> k, std::vector<> const&, std::vector<> k, std::vector<> const&, std::vector<> const&, std::vector<> k, std::vector<> const&, std::vector<> con
  - > 440.7 MiB: WireCell::SigProc::OmnibusSigProc::save mproifstd::vector<>&, std::vector<>&, int, std::multimap<> const&) (OmnibusSigProc:cxx:763)
  - > 228.9 MiB: WireCell::SiaProc::OmnibusSiaProc::save\_roilstd::vector<>&, std::vector<>&, int. std::vector<>&) (OmnibusSiaProc.:cxx:612)
  - > 109.9 MiB: Construct<> (stl construct.h:119)

#### 2.1 GiB: Snapshot #48 (peak)

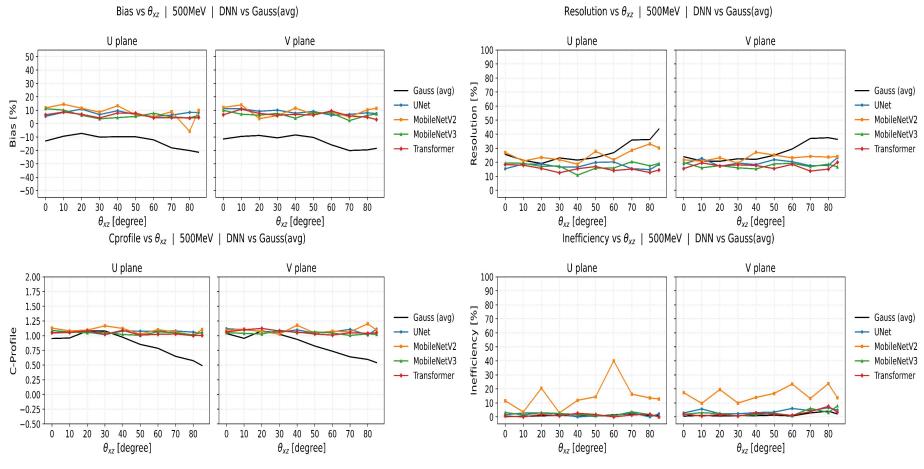
- 565.4 MiB: allocate (new allocator.h:137)
  - 565.4 MiB: allocate (alloc traits.h:464)
  - ▼ 565.4 MiB: M allocate (stl vector.h:378)
  - 565.4 MiB: \_M\_create\_storage (stl\_vector.h:395)
    - ▼ 565.4 MiB: Vector base (stl vector.h:332)
    - 565.4 MiB: vector (stl vector.h:598)
    - 565.4 MiB: WireCell::Aux::SimpleTrace::SimpleTrace(int, int, std::vector<> const&) (SimpleTrace.cxx:8)
      - > 455.5 MiB: WireCell::SigProc::OmnibusSigProc::save\_mproi(std::vector<>&, std::vector<>&, int, std::multimap<> const&) (OmnibusSigProc.cxx:763)
    - > 109.9 MiB: Construct<> (stl construct.h:119)

save_data	Model	Peak Memory (MB)	Time (s)	
	UNet	6942	140	
tight_lf, loose_lf,	Transformer	7032	141	
cleanup_roi, wiener, gauss	MobileNetV2	4071	72	
	MobileNetV3	4410	77	
	UNet	6225	127	
wiener gaues	Transformer	6267	140	
wiener, gauss	MobileNetV2	3343	69	
	MobileNetV3	3687	72	
	UNet	5863	120	
none	Transformer	5945	133	
Hone	MobileNetV2	3012	69	
	MobileNetV3	3324	72	

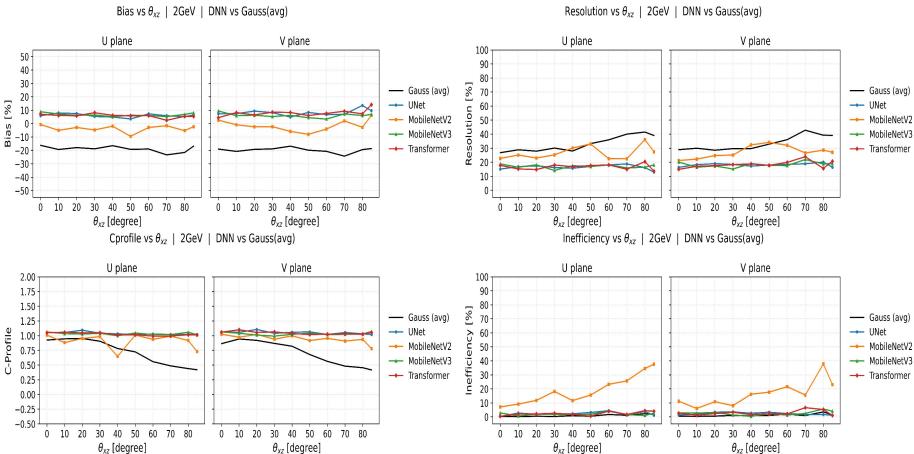




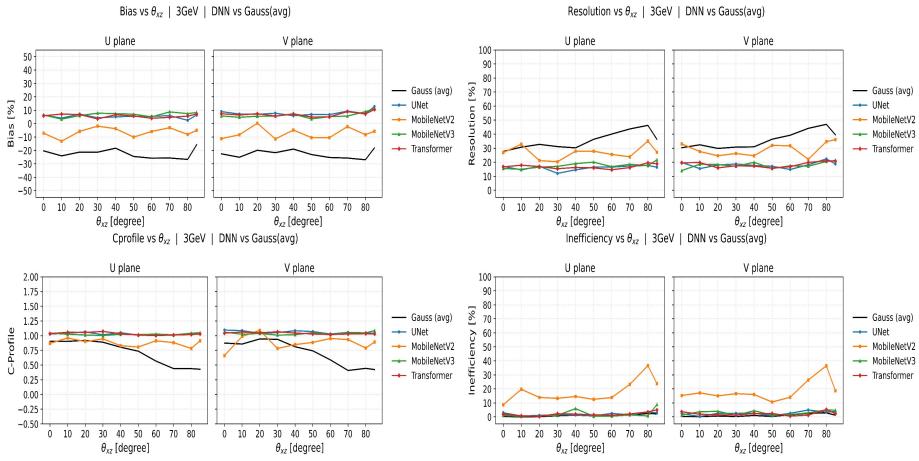
# **Single Shower Evaluation (500 MeV)**



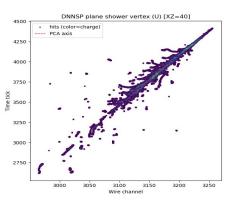
# **Single Shower Evaluation (2 GeV)**

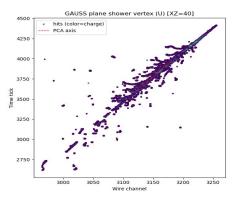


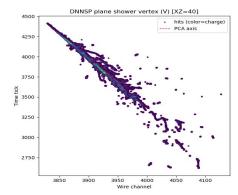
# **Single Shower Evaluation (3 GeV)**

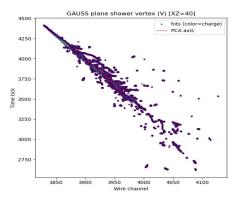


# Angle between reco and true direction (E = 5 GeV, XZ = 40)





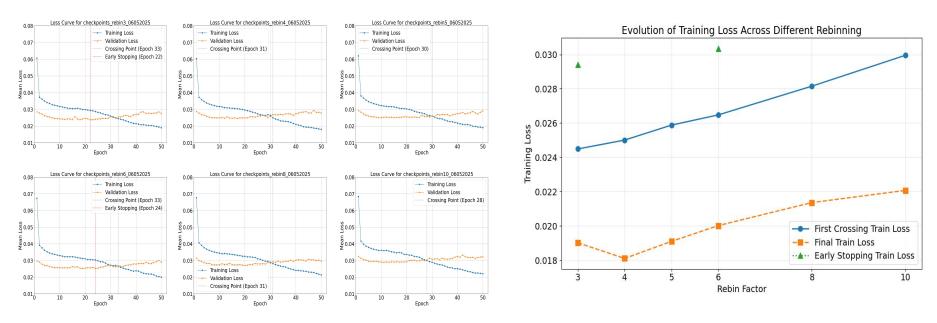






	Angle diff			
	Gauss (deg.)	DNN (deg.)		
U plane	0.74	0.82		
V plane	0.52	0.51		
W plane	1.22			

# **DNN** Training with rebining - UNet



- Re-trained the UNet with rebinning in time tick
- High resolution in time (y-axis) contributes better training loss:  $0.0299 \rightarrow 0.0245$
- Loss: Binary Cross-Entropy, Epoch = 50, Learning Rate = 0.1, Momentum = 0.9, train vs val split = 90:10

# Memory and Time consumption on the dunegpvm

Server	WCT	Resource	DNN ROI	Mem (MB)	Time (s)	Mem Ratio	Time Ratio
dunegpvm	dunesw	None	None	1841.24	59.67	1.00	1.00
dunegpvm	dunesw	CPU	UNet	7399.78	155.31	4.02	2.60
dunegpvm	dunesw	CPU	MobileNetV2	4575.95	97.3	2.49	1.63

- PyTorch Model file PD-HD data: run026763\_0008, run027673\_0000, run027673\_0001, run027673\_0002, run028588\_0019
  - UNet: unet-cosmic390-newwc-depofluxsplat-pdhd.ts (Jay)
  - MobileNetV2: CP49.ts (from my training, rebin: 10)
- MobileNetV2 showed a more efficient balance between memory and time compared to UNet