Status report on

DNNROI sigproc

Hokyeong Nam
Chung-Ang University

Work Completed & Next Steps

DNN SP Evaluation

- Track evaluation
  - Resolve the bias in tradition ROIs and DNN SP
    → Modified the function for calculating total charge
- Shower evaluation
  - Develop dedicated metrics for shower-like event
    → Introducing PCA-based vertexing at evaluation step

Optimize DNN performance

- Train w/ large dataset
  - Sergey found time offset between truth and reco waveform
    → Added time offset line into the configuration code → Sergey still experienced time offset
- Rebining
  - Checked Train-Val loss with UNet, but need to perform evaluation

Work Completed & Next Steps

Memory Profiling

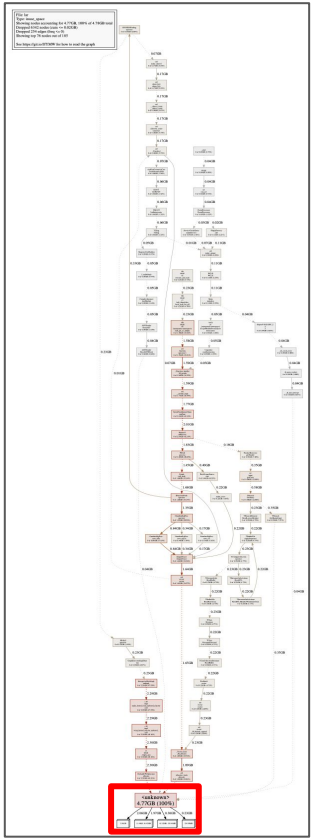
- GPerfTools
  - Less informative because some binary file name is missing
- Valgrind

Implement various NNs and compare their performance

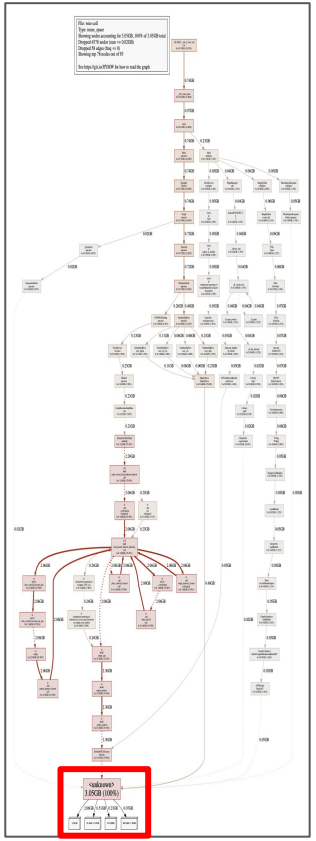
Model	PyTorch	wirecell-dnn	to-ts script	Performance	Resource
MobileNetV2	O	Yet	O	Checked	Checked
MobileNetV3	O	Yet	Yet	Yet	Yet
Transformer	Yet	Yet	Yet	Yet	Yet

- Develop metric for fair performance comparison between NN models

Memory Profiling - GPerfTools



❏ LArWC



❏ WC standalone

```
setup dunesw v10_04_07d01 -q debug:e26;
```

- Set up dunesw with qualifier of the debug:e26
- The missing binary warning and “unkonwn” node still appeared

- Even standalone simulation displayed same issue

Memory Profiling - Valgrind

```
Retagger: tagging trace set: dnnsp3 with 2560 traces, 1600 summary
Retagger: tagging trace set: dnnsp2 with 2560 traces, 1600 summary
./profile_valgrind.sh: line 19:  1116 Killed                  valgrind --tool=massif --massif-out-file=${OUTPUT_PREFIX}.out --stacks=yes --log-file=${VALGRIND_LOG} $PROGRAM -n1 -c ${FH
ICL_FILE} -s ${INPUT_FILE}
Done.
Massif output file: massif_output.out
Valgrind log file:  valgrind.log
```

- Given the heavy runtime, executed it in the background using tmux
- When I checked, the job had been killed

```
[hnam@dunegpvm11 ~]$ dmesg | grep -i kill
[211319.072909] gmain invoked oom-killer: gfp_mask=0x140cca(GFP_HIGHUSER_MOVABLE|__GFP_COMP), order=0, oom_score_adj=0
[211319.072992]  oom_kill_process.cold+0xb/0x10
[211319.074264] [1048848] 15629 1048848   167410      793   135168        0             0 gsd-rfkill
[211319.074742] oom-kill:constraint=CONSTRAINT_NONE,nodemask=(null),cpuset=/,mems_allowed=0,global_oom,task_memcg=/user.slice/user-14860.slice/session-550.scope,task=massif-amd64-li,pi
d=647650,uid=14860
[211319.074811] Out of memory: Killed process 647650 (massif-amd64-li) total-vm:14316984kB, anon-rss:9034152kB, file-rss:8kB, shmem-rss:0kB, UID:14860 pgtables:23832kB oom_score_adj:0
```

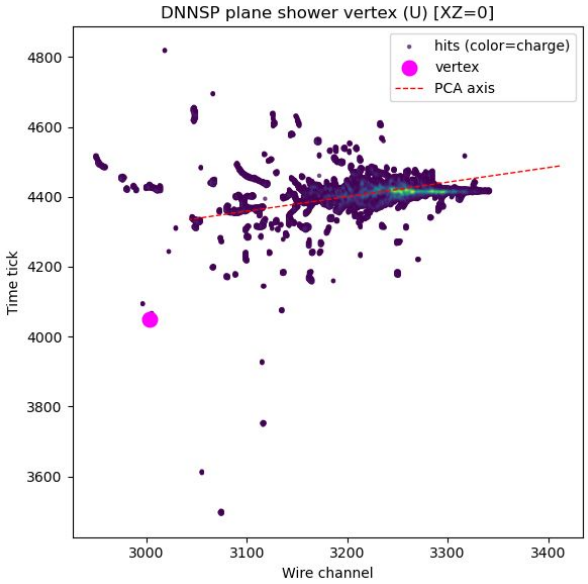
- According to the kill log, the job was killed due to an OOM error

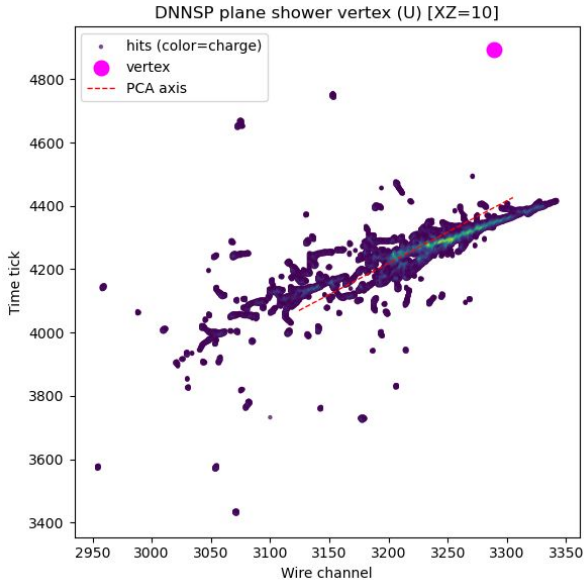
PCA-based Shower Vertexing

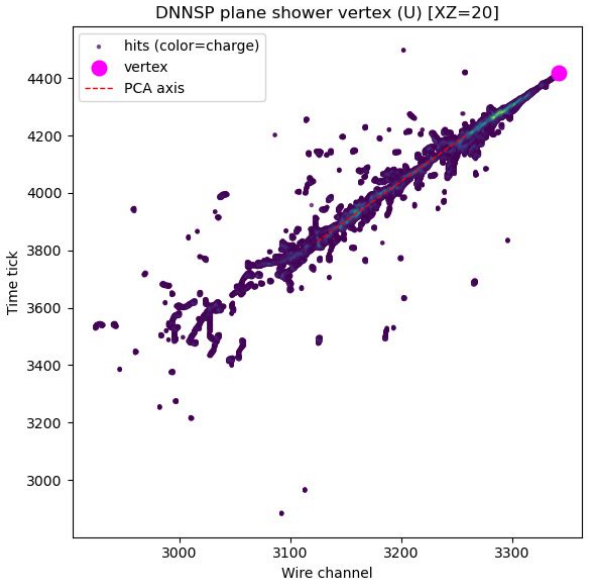
Strategy

- Charge-Weighted PCA: build a covariance matrix weighted by hit charges and take its leading eigenvector as the principal axis
- Charge Threshold: only consider hits above a minimum charge to suppress noise
- Local Total Charge Threshold: scan sorted hits for the first whose ±time tickwindow neighborhood's total charge exceeds a threshold
- Fallback: if no neighborhood passes the sum test, choose the hit with the highest projection as the vertex

PCA-based Shower Vertexing

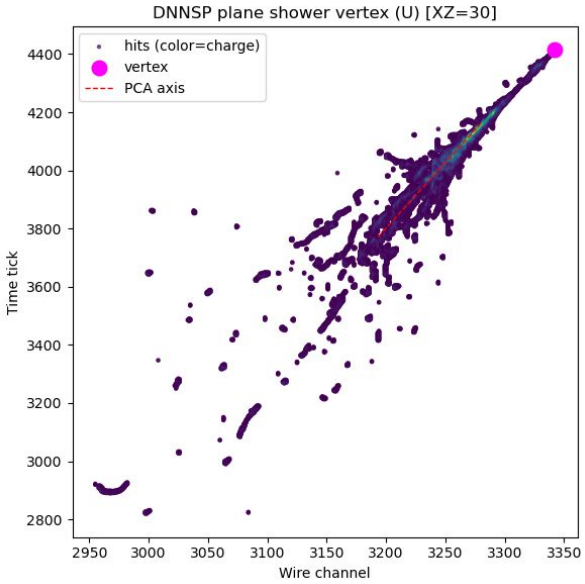


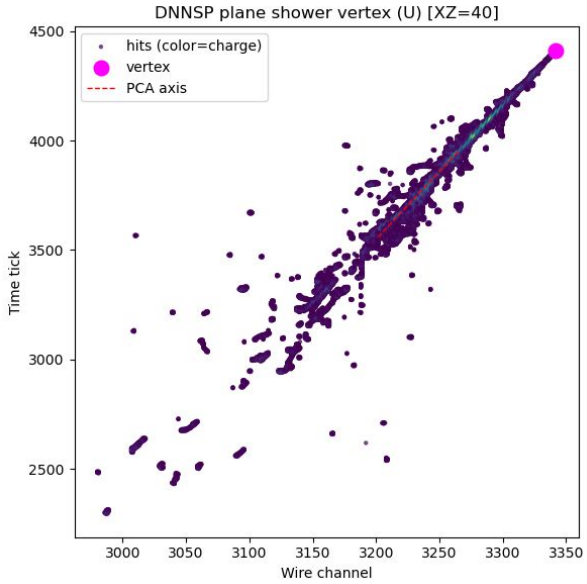


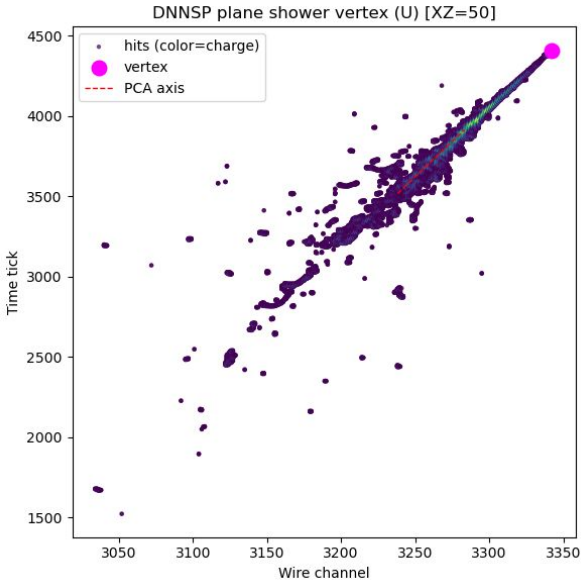


- Single electron shower event varying XZ angles
- Energy: 5 GeV

PCA-based Shower Vertexing

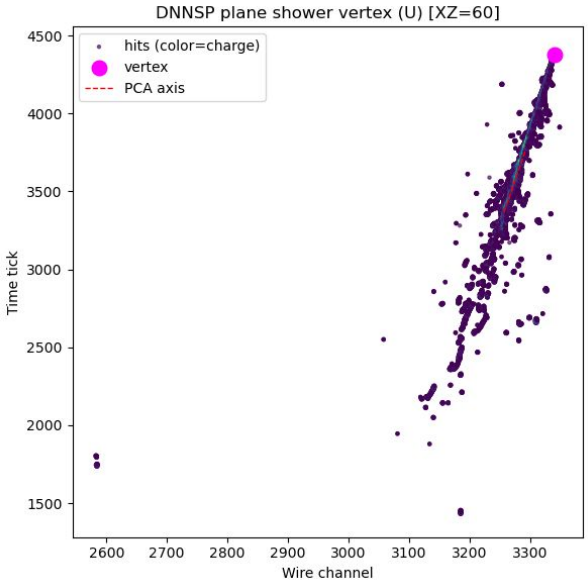


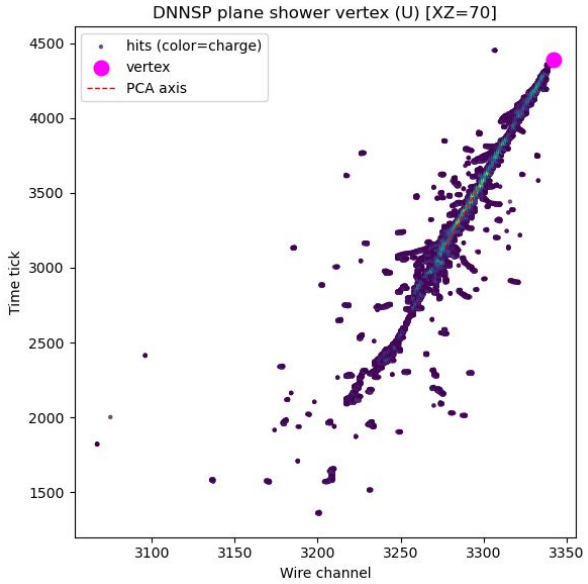


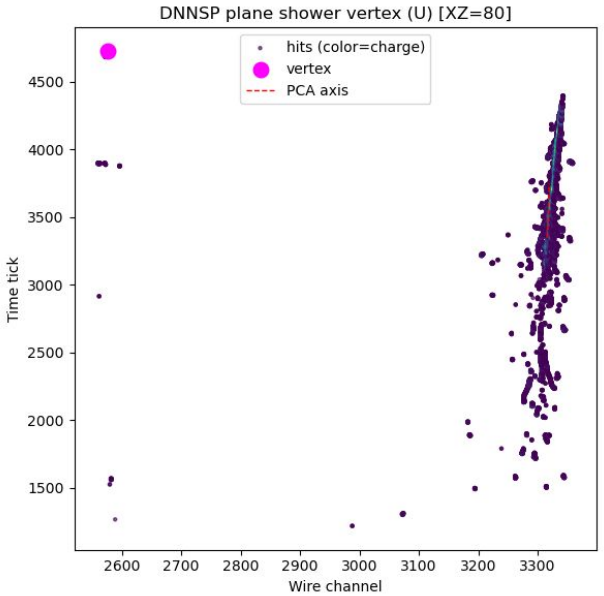


- Single electron shower event varying XZ angles
- Energy: 5 GeV

PCA-based Shower Vertexing

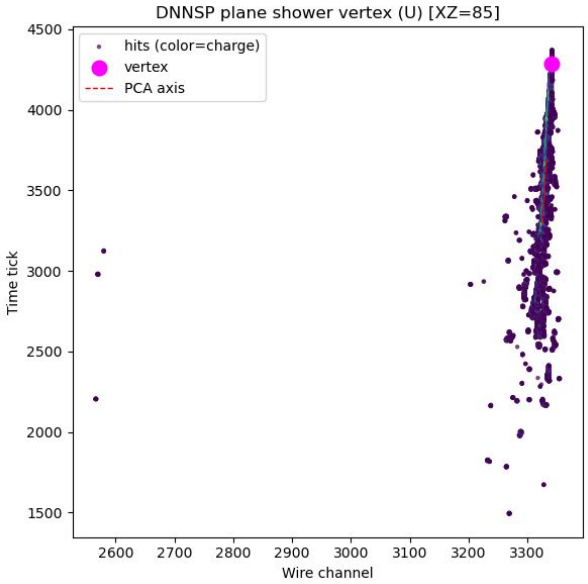






- Single electron shower event varying XZ angles
- Energy: 5 GeV

PCA-based Shower Vertexing



- PCA-based method sometimes failed due to outliers, specifically when the XZ angle is low
- Further validation is required with samples have different shower energy

Back Up

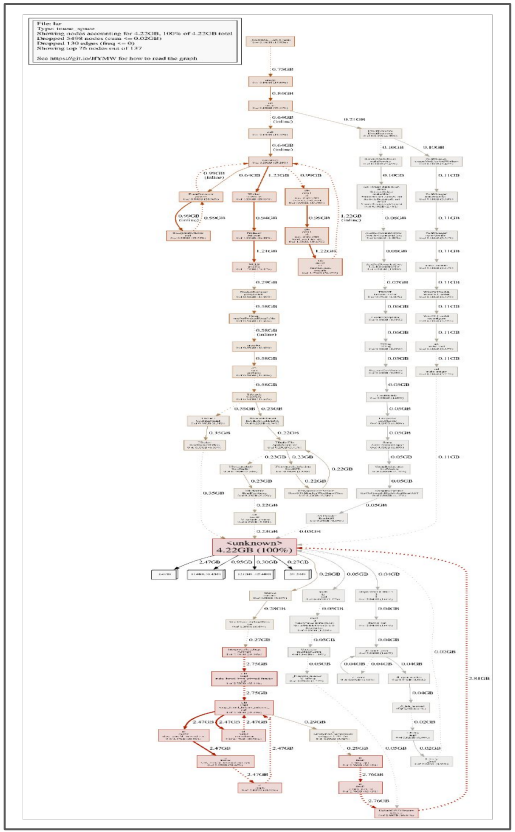
Profiling Memory usage for PD-HD data processing

```
Dumping heap profile to mem.prof.0033.heap (26182 MB allocated cumulatively, 2331 MB currently in use)
Dumping heap profile to mem.prof.0034.heap (27662 MB allocated cumulatively, 2858 MB currently in use)
Dumping heap profile to mem.prof.0035.heap (30621 MB allocated cumulatively, 3913 MB currently in use)
Dumping heap profile to mem.prof.0036.heap (31910 MB allocated cumulatively, 2741 MB currently in use)
Retagger: tagging trace set: dnnsp3 with 2560 traces, 1600 summary
Dumping heap profile to mem.prof.0037.heap (32934 MB allocated cumulatively, 1675 MB currently in use)
Dumping heap profile to mem.prof.0038.heap (33986 MB allocated cumulatively, 1764 MB currently in use)
Dumping heap profile to mem.prof.0039.heap (35010 MB allocated cumulatively, 1731 MB currently in use)
Dumping heap profile to mem.prof.0040.heap (36073 MB allocated cumulatively, 1879 MB currently in use)
Dumping heap profile to mem.prof.0041.heap (37097 MB allocated cumulatively, 1945 MB currently in use)
Dumping heap profile to mem.prof.0042.heap (38121 MB allocated cumulatively, 2064 MB currently in use)
Dumping heap profile to mem.prof.0043.heap (39145 MB allocated cumulatively, 2114 MB currently in use)
Dumping heap profile to mem.prof.0044.heap (40582 MB allocated cumulatively, 3221 MB currently in use)
Dumping heap profile to mem.prof.0045.heap (41812 MB allocated cumulatively, 2752 MB currently in use)
```

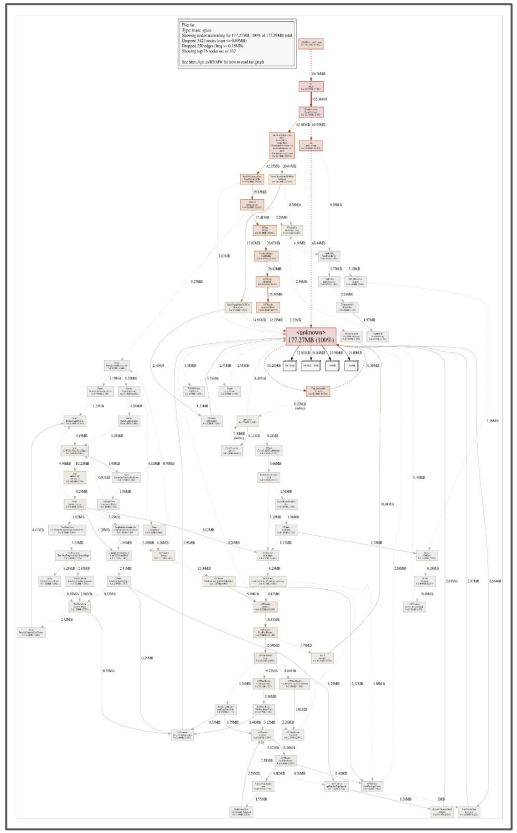
```
Program:    /cvmfs/larsoft.opensciencegrid.org/products/art/v3_14_04/slf7.x86_64.e26.prof/bin/lar
Config file:my_standard_reco_stage2_calibration_protodunehd_keepup_dnnroi.fcl
Input file: ../../data/stage1/run027673/np04hd_raw_run027673_0000_dataflow0_datawriter_0_20240704T050545_reco_stage1.root
Profiler:   /exp/dune/data/users/hnam/demo/gperftools/install/lib/libprofiler.so.0
TCMalloc:   /exp/dune/data/users/hnam/demo/gperftools/install/lib/libtcmalloc.so.4
Generating PDF reports using /exp/dune/data/users/hnam/demo/go/bin/pprof
Some binary filenames not available. Symbolization may be incomplete.
Try setting PPROF_BINARY_PATH to the search path for local binaries.
CPU profile report saved as cpu_profile.pdf
```

- Total # of 106 mem.prof file are generated
- Due to the warning, some of the nodes are appeared with name of “unknown”

Profiling Memory usage for PD-HD data processing



❑ 97th file



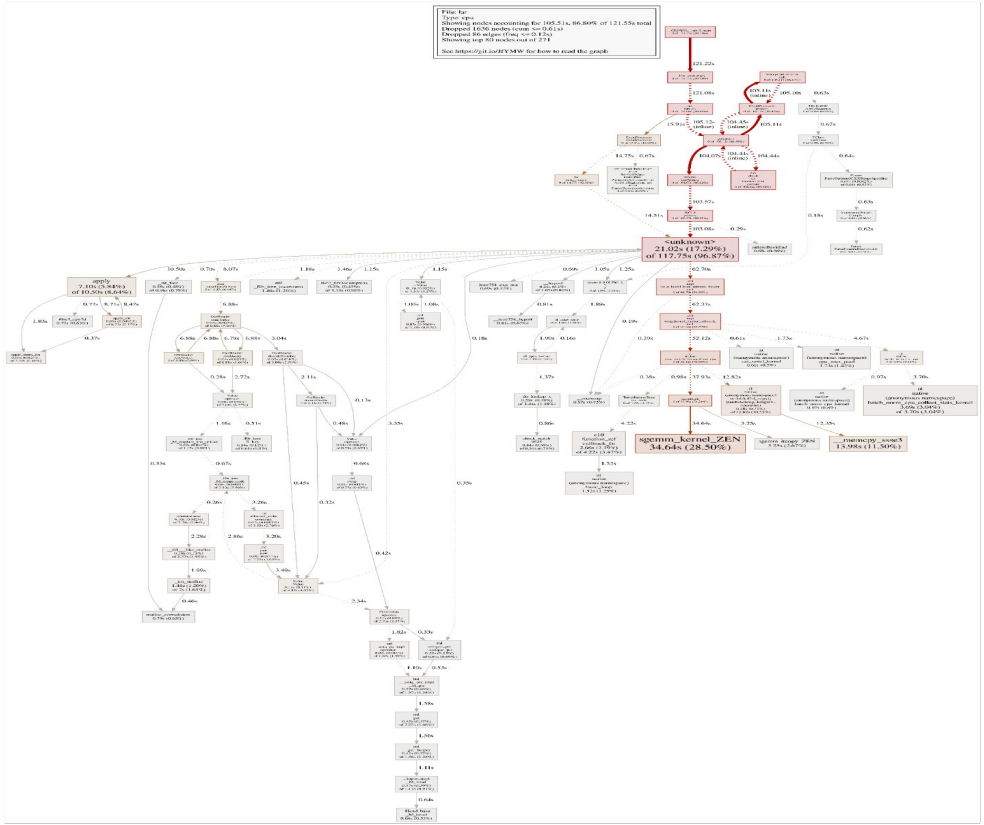
❑ 106th file

- Dumped into the pdf file
  - middle of dnn: 97th
  - at the end: 106th

Next

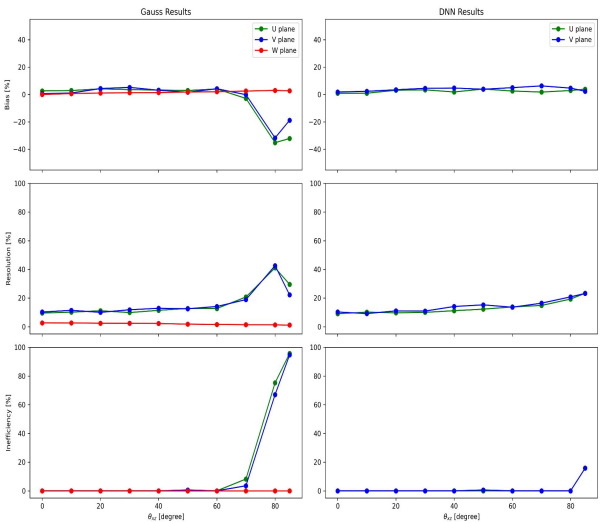
- How to solve the warning?
- How to identify crucial nodes?

Profiling CPU usage for PD-HD data processing

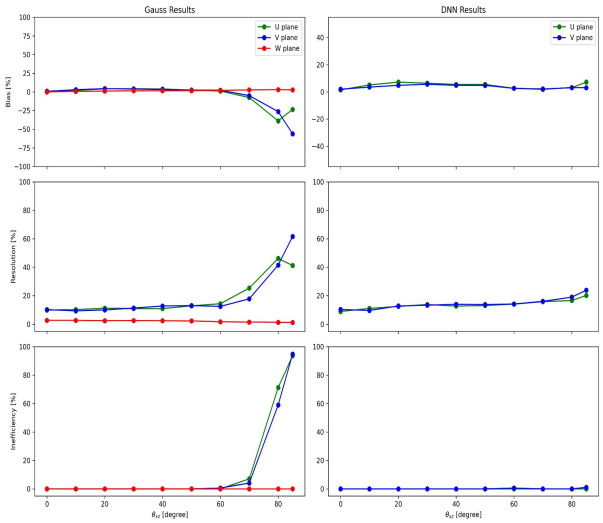


❑ cpu_profile

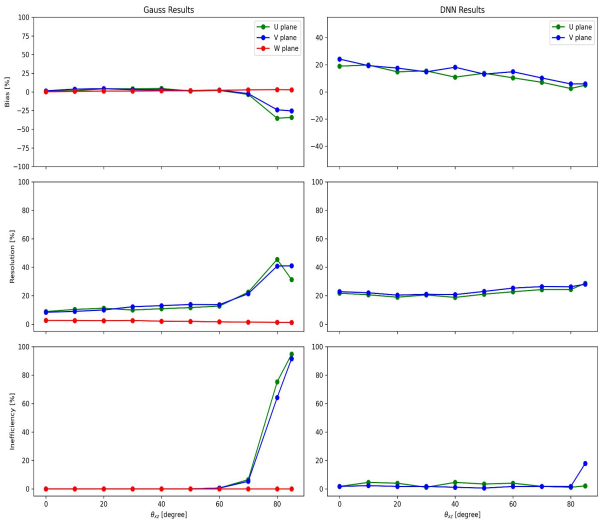
DNN ROI evaluation w/ UNet and MobileNetV2



❑ unet-cosmic390-newwc-depofluxsplat-pdhd.ts



❑ UNet trained by Hokyeong (rebin: 10)



❑ MobileNetV2 trained by Hokyeong (rebin: 10)

- The UNet I trained showed consistency with the model have been used
- The MobileNetV2 has bias about 10 - 20% across the Theta XZ angles