Contribution ID : **82** Type : **Talk**

Procedural generation of tomographic datasets for machine learning

Wednesday, 9 September 2026 14:30 (30)

Industrial computed tomography allows for non-destructive inspection of complex components, making it an interesting domain of applications for machine learning methods. In this context, a large bottleneck for training robust models is data acquisition. Already time-intensive task of obtaining data using real-world CT scanners is further exacerbated by the requirement of creating accurate labels.

We solve this problem by implementing a pipeline for the generation of massive synthetic datasets, and using only some examples of real-world data.

Our data pipeline uses parametrized CAD geometry to generate thousands of randomized objects from which we derive both ground-truth labels and attenuation data further fed into tomographic simulations.

The simulation code is GPU-accelerated allowing us to substantially reduce the time needed to generate the datasets.

To reproduce effects visible in real data, we implement a polychromatic beam, and simulate artefacts such as non-uniformity of the beam and detector matrix response. We also apply custom beam hardening corrections for our reconstructions.

We use the synthetic data in semi-supervised training of a W-net model for the specific task of segmentation of ionization chambers. We show, that semi-supervised approach allows for effective training with only partial labeling (as low as 25%-50% of the data). We then validate performance of segmentation models trained using simulations on the real-world data.

Primary author(s) : ĆWIEK, Arkadiusz (National Centre for Nuclear Research); KLIMASZEWSKI, Konrad (NCBJ); MROZOWSKI, Mikołaj (NCBJ); OGONOWSKI, Aleksander; Mr WIĘCEK , Daniel (NCBJ)

Presenter(s) : MROZOWSKI, Mikołaj (NCBJ)

Session Classification : Medical and Industrial Imaging

Track Classification : Machine Learning in Physics