ISBN: 978-3-95450-247-9 ISSN: 2673-5490 doi: 10.18429/JACoW-IPAC2024-THPG84

3D VISUALIZATION AND ANALYSIS OF NEUTRON SCATTERING DATA IN THE CONTROL ROOM

M. Kilpatrick*, D. Bruhwiler, E. Carlin, B. Gur, R. Nagler, B. Nash, RadiaSoft LLC, Boulder, CO, USA
C. Hoffmann, Z. Morgan, A. T. Savici, M. Tucker, B. Vacaliuc
Oak Ridge National Lab, Oak Ridge, TN, USA
A. Kuhn, J. Mensmann, P. Messmer, M. Nienhaus, S. Roemer, D. Tatulea
NVIDIA, Berlin, Germany

Abstract

Neutron scattering experiments have undergone significant technological development through large area detectors with concurrent enhancements in neutron transport and electronic functionality. Data collected for neutron events include detector pixel location in 3D, time and associated metadata, such as sample orientation, neutron wavelength, and environmental conditions. Working with single-crystal diffraction data, live streaming from the TOPAZ detector, we are developing both interactive and automated 3D analysis of neutron data by leveraging NVIDIA's Omniverse technology. We have implemented machine learning techniques to automatically identify Bragg peaks and separate them from diffuse backgrounds and analyze the crystalline lattice parameters for further analysis. A novel CNN architecture has been developed to identify anomalous background from detector instrumentation for dynamical cleaning of measurements. Our approach allows scientists to visualize and analyze data in real-time from a conventional browser, which promises to improve experimental operations and enable new science. We have deployed a cloud based server, leveraging Sirepo technology, to make these capabilities available to beamline users in the control room.

BACKGROUND IDENTIFICATION

The experimental samples are often held within an Aluminum sample holder at Topaz. This sample holder gives rise to a background signal that should be removed in order to distinguish it from the signal coming from the sample itself. The data from the Aluminum sample holder shows a ring-like structure in q-space. We here describe our method to detect and remove these Aluminum rings. We want to ensure that only the rings are removed and not the Bragg peaks.

The first step to identification of the Aluminum rings is to normalize the data. We define the normalized intensity data as follows:

$$W_N = \frac{\ln W + 1}{C} \tag{1}$$

where W is the unnormalized intensity and C is a constant we will choose such that the most frequently occurring voxel intensity is at a value of $W_N = 1$. This allows us to use a consistent intensity cut across all data in which the same Aluminum sample holder was used.

* kilpatrick@radiasoft.net

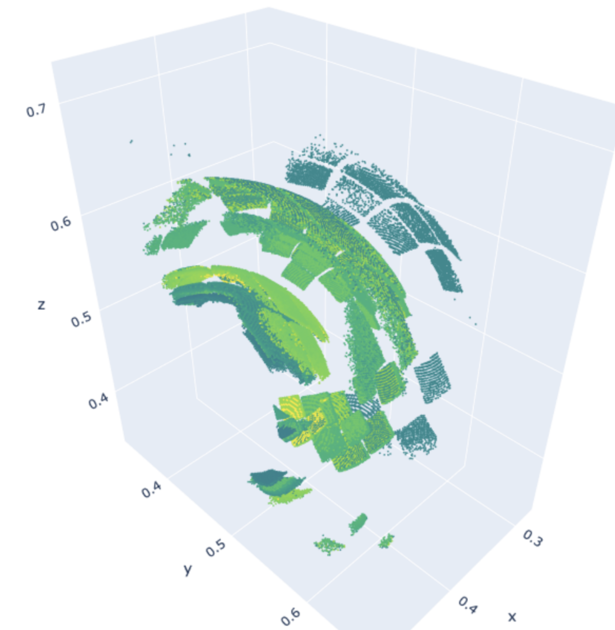

Figure 1: A 3D visual of the identified clusters from a Aluminum dataset.

Applying this intensity normalization to data with the Aluminum sample holder and with or without an additional sample, we see that the sample contains higher intensities, which can be cut out of the data with an intensity cut. Once we have applied this normalization to a sample, we have found that the aluminum background can be found at $2.08 < W_N < 3.25$ in units of normalized intensity. We have investigated how to best identify the voxels from the aluminum background. We have identified a set of DBSCAN hyperparameters that can identify the aluminum rings, though we do not plan to integrate this into the final workflow since the computation time can take many minutes and is not suitable for data streaming. However, since it is able to identify the background from a range of datasets we can use it to create a labelled dataset for training a neural network.

As we look at a 3D representation of the aluminum rings, we see pairs of layers that are perpendicular to a vector emitting from the center of the q-space. Knowing that the background should form a segmented sphere in q-space, we can identify the structure in a radius vs intensity plot, such as Fig. 1. The background has a near vertical distribution which

ISBN: 978-3-95450-247-9 ISSN: 2673-5490 doi: 10.18429/JACoW-IPAC2024-THPG84

shows that is occupies a small radius for each layer. There are some points that have a angular distribution that correspond to identified background that does not correspond to the aluminum background. In Fig. 1, we see the 3D representation of the aluminum background and the corresponding radius vs normalized intensity plot. It is straight forward to see that the angular distrubutions is the later plot correspond to spurious background in the former plot. We are looking into methods to reduce these spurious background points from our labelled datasets. Methods such as a t-SNE [1] or PCA analysis [2], perform well with aluminum background alone, but are not able to isolate the background from all sample types. Once we have a labelled dataset we can move on to training a neural network to identify and remove the aluminum background.

In addition to the cut based method above, we have also developed a weight based method for the labelling of the Aluminum rings. We bin the voxels of a given sample in concentric spheres which shows the large peak in the number of voxels at finite radii which is a standard attribute of the Aluminum background. Utilizing this knowledge we define a scale factor, see Eq. (2), that is the maximum gradient in a window that rolls along the interated radius that we have defined, see Fig. 2.

$$S_{rings} = \frac{max(\nabla R) \cdot \bar{R}}{1 - bin} \tag{2}$$

This value is then normalized by the mean in the window and weighted to prioritize bins closer to the origin. We can then put this into a DBSCAN clustering algorithm and use the scale factor to prioritize clustering the rings. After clustering the data we decompose the voxels positions, normalized intensities, and scale factor into a 1D representation. Using a PCA analysis, we use the normalized voxel intensity, scale factor, and 3 dimensional location to further separate out the correct labelled voxels from the false positives. A cut, approx. 10^{-2}, can then be applied that removes most of the falsely identified voxels, see Fig. 2.

Now that we have two methods to label voxels as Aluminum rings, we need to determine which performs the best. We are currently working to summarize the efficiencies of each method to finalize the method to use for labelling.

3D U-net Aluminum Ring Prediction Once we have a robust method for automatic identification of Aluminum rings within a sample, we can add the labelling to each sample for training within a neural network. Since the aluminum rings in the reciprocal space are inherently three dimensional, we have developed a three dimensional U-net [3] architecture for training and predicting this background process, see Fig. 3. A 3D U-net is used for learning dense volumetric segmentation from sparse annotations and includes a contraction path down to a latent layer and expansion path to the same size as the input. It is possible that the latent layer could include some relevant physical information about the background process and could provide some insight to scientists at ORNL.

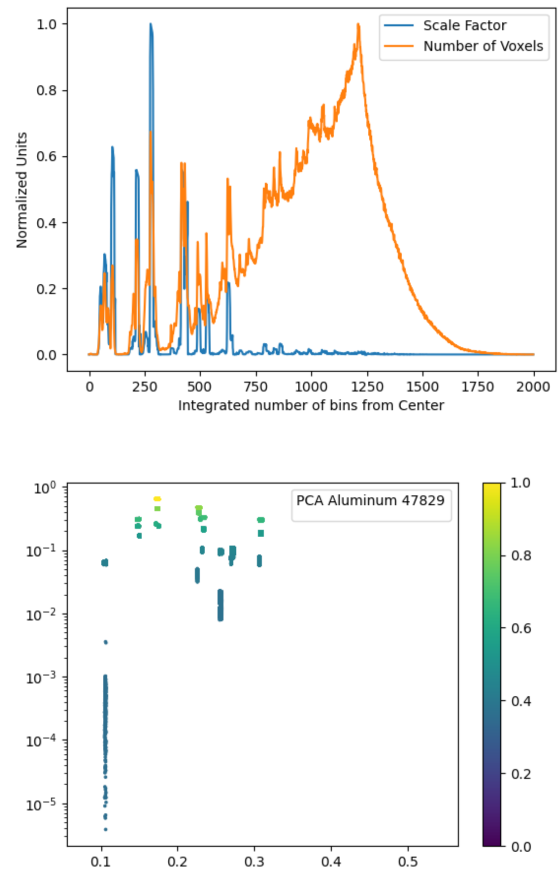

Figure 2: Example of the radius defined scale factor for a sample of Aluminum data (left) and the PCA analysis after clustering (right).

Initially there was difficulties constructing a valid U-net model, due to not enough memory when trying to create the simplest network. This is caused by the 1000^3 resolution of each voxel dataset. We did not want to reduce the resolutions since it's not guaranteed that the aluminum ring structure is resolution invariant. Due to this we developed a framework to take the entire voxel dataset and apply a robust data splitting framework. First, we isolate the innermost 500^3 space of the dataset since the Aluminum background is concentrated towards the origin of the reciprocal space. Second, we divide the volume into a subset of smaller cubes that can be feed into the NN. We have allowed for a varied approach for model testing where a cube can be any divisor of 500. After extensive testing, our best performing model that was able to be trained and evaluated within a singe GPU was trained and evaluated on 100^3 cubes. We have provided methods to input the full voxel dataset, split out the cube subsets, evaluate them with our model, and combine them back into the original sample after removing the identified voxels from the prediction.

ISBN: 978-3-95450-247-9 ISSN: 2673-5490 doi: 10.18429/JACoW-IPAC2024-THPG84

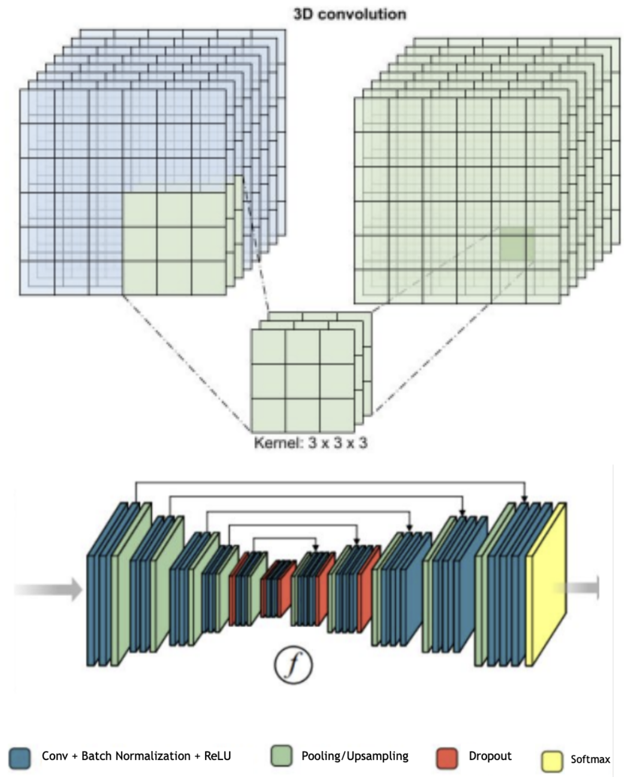

Figure 3: Diagram of three dimentional convolutions for a possible convolutional layer (left) and a standard model definition for a U-net (right).

These methods have been developed into a module for use in the streaming framework in a similar manner to the UB matrix analysis methods. While we put extensive work in correctly identifying the Aluminum background with DBSCAN, the clustering method takes several minutes to run on a single sample. The 3D U-net can predict the Aluminum background voxels on the order of 5 seconds for each sample. With the improved speed of this prediction we will be able to actively remove the background during data streaming and detector operation.

In Fig. 4, we show an example of the prediction of the 3D U-net where each voxel is evaluated to give a discriminator. We still need to robustly fine tune the discriminator output, but a conservative cutoff of 0.05 provided a consistent prediction of the Aluminum background. Now that we have a robust method to provide a ground truth label to the U-net model training. This trained model can then be used to provide a prediction for each voxel that is part of the background. We are able to remove these voxels and display the cleaned dataset to the user using NVIDIA Omniverse, see Fig. 5. Omniverse is a service that enables developers to easily integrate Universal Scene Description (OpenUSD) and RTX rendering technologies into existing software tools and simulation workflows for building AI systems [4].

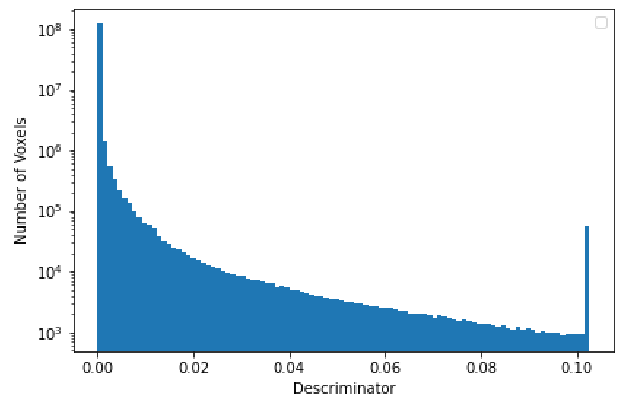

Figure 4: An example distribution of the voxel discriminator output from the U-net.

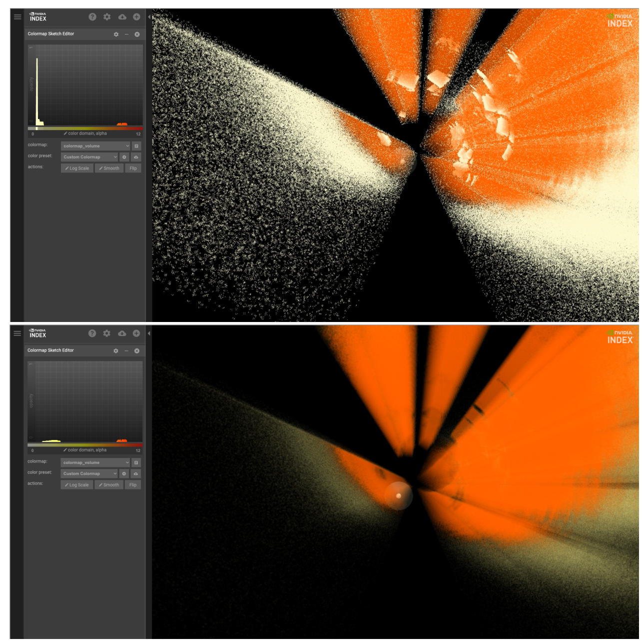

Figure 5: An example of identifies aluminum background viewed in Omniverse. On the left is the entire dataset showcasing the background and on the right is the same dataset with the background removed.

CONCLUSION

The structured aluminum background of the TOPAZ experiment at SNS is a well-known background for analysis. Our investigation has shown two viable methods for identifying this background to provide a ground truth label for machine learning. We have developed a novel 3D U-net architecture to identify and remove the structured aluminum background from neutron scattering datasets. These models and in-parallel visualization on NVIDIA's Omniverse can provide accelerated feedback to users within the control. A user will be able to view, predict, and remove this background during live operation of the detector.

ISBN: 978-3-95450-247-9 ISSN: 2673-5490 doi: 10.18429/JACoW-IPAC2024-THPG84

REFERENCES

[1] L. v. d. Maaten and G. Hinton, "Visualizing Data using t-SNE," *J. Mach. Learn. Res.*, vol. 9, no. 86, pp. 2579–2605, 2008, http://jmlr.org/papers/v9/vandermaaten08a.html

[2] I. Jolliffe, "Principal Component Analysis," *Int. Encycl. Stat. Sci.*, pp. 1094–1096, 2011, doi:10.1007/978-3-642-04898-2_455

[3] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional Networks for Biomedical Image Segmentation," *arXiv e-prints*, 2015, doi:10.48550/arXiv.1505.04597

[4] IndeX, https://developer.nvidia.com/nvidia-index