ISBN: 978-3-95450-238-7 ISSN: 2226-0358 doi:10.18429/JACoW-ICALEPCS2023-TUPDP116

MACHINE LEARNING BASED SAMPLE ALIGNMENT AT TOPAZ

M. Henderson, J. Edelen*, M. Kilpatrick, I.V. Pogorelov, RadiaSoft LLC, Boulder, USA
S. Calder, B. Vacaliuc, ORNL RAD, Oak Ridge, USA
R.D. Gregory, G. Guyotte, C. Hoffmann, B. Krishna, ORNL, Oak Ridge, USA

Abstract

Neutron scattering experiments are a critical tool for the exploration of molecular structure in compounds. The TOPAZ single crystal diffractometer at the Spallation Neutron Source studies these samples by illuminating samples with different energy neutron beams and recording the scattered neutrons. During the experiments the user will change temperature and sample position in order to illuminate different crystal faces and to study the sample in different environments. Maintaining alignment of the sample during these processes is key to ensuring high quality data are collected. At present this process is performed manually by beamline scientists. RadiaSoft in collaboration with the beamline scientists and engineers at ORNL has developed a new machine learning (ML) based alignment software automating this process. We utilize a fully-connected convolutional neural network with dropout, configured in a U-net architecture, to produce sample segmentation masks from which we compute the sample center of mass. We then move the sample using a custom python-based EPICS IOC interfaced with the motors. In these proceedings we provide an overview of our ML tools and initial results aligning samples at ORNL.

INTRODUCTION

The TOPAZ instrument is a high-resolution neutron time-of-flight (TOF) Laue diffractometer for single-crystal diffraction located on a beamline at the Spallation Neutron Source (SNS) user facility at Oak Ridge National Lab (ORNL) [1]. Along with the diffractometer itself (which already features a variable number of individual neutron detectors), the TOPAZ beamline is host to a large array of instruments and hardware for executing experimental controls and maintaining desirable sample environments. Implementing these controls and collecting the data needed for users to achieve their scientific goals are the responsibility of a dedicated group of beamline operators with a high degree of expertise in experimental neutron science. Simplifying the workflows of these operators, and automating controls processes wherever possible, is a critical effort for maximizing the scientific output of user facilities like the SNS.

CONTROLS AT TOPAZ

In addition to the overall mechanical complexity of beamline hardware and instrumentation, operational logistics at TOPAZ include the navigation of two distinct operational modes and a rich but complex network of controls software.

Operational Modes

The operational modes in use at TOPAZ are an ambient-temperature (ambient) mode and a cryogenically-controlled (cryo) mode. These modes differ not only in the range of temperatures they represent within the sample chamber, but also in the hardware and controls available during their operation. Most notably, because the cryostream instrument used in cryo-mode enters the chamber from a port which typically hosts a diagnostic camera in ambient-mode, an alternative camera mounted to a side port (from which the sample arm is not clearly visible) must be used. Figure 1 demonstrates the difference in sample views provided in ambient- (top) and cryo-mode (bottom). Additionally, sample shields sometimes used in cryo-mode are visible in and can partially obscure the sample images.

Controls Network

Experimental controls at TOPAZ are implemented using Channel Access (CA) protocols within the EPICS framework. Like most instruments at the SNS, TOPAZ encompasses an extensive network of EPICS controls and process variables (PVs). This network covers sample arm motor positions, diagnostic camera controls, thermal conditions, neutron guide environments, detector settings, and much more [1]. In addition to low-level controls, the network also features several layers of abstracted controls mechanisms. These include mechanisms for security purposes, such as virtual motors used to validate proposed motor controls, and input-output controllers (IOCs) for simplifying workflows and automating controls processes.

Sample Alignment

One feature of the EPICS network at TOPAZ is an IOC for executing semi-automated sample alignment. This IOC automates the determination of motor positions, but requires a human operator to identify approximate sample centroids in diagnostic camera images. Operators uses a graphical user interface (GUI) to initiate an alignment state, and then must click a sample image in the GUI to identify the centroid. This process is repeated over several adjustments to motor positions, after which the center of the sample will have been aligned with the current beam position. Although this process is much faster than manually and incrementally updating motor positions until the sample centroid and beam are aligned, it still requires operators to expend time and attention on alignment which could be better spent on more complex tasks. In cases where samples must be re-aligned frequently (e.g., due to sample shifting during thermal variations), this expense becomes significant.

* jedelen@radiasoft.net

ISBN: 978-3-95450-238-7 ISSN: 2226-0358 doi:10.18429/JACoW-ICALEPCS2023-TUPDP116

IMAGE SEGMENTATION TASK

The human-in-the-loop portion of the current sample alignment process at TOPAZ can be thought of as an image segmentation task, in which an observer must correctly determine which pixels in an image contain the sample and which do not (see Fig. 1). In the context of a machine vision algorithm, this requires the computation of a sample mask, an image equal in dimensions to the sample image but with pixel values of either zero (not sample) or one (sample). Once a determination of the spatial distribution of the sample has been made by an observer, its centroid can easily be identified as the center of the distribution. For machine-computed sample masks, this is done by treating the mask as a discrete density array and computing its center of mass. Following identification of the sample center at each stage alignment, the observer must initiate the appropriate controls process either by clicking the sample center on-screen (for a human observer) or by setting the corresponding EPICS PVs (for an automated controls algorithm).

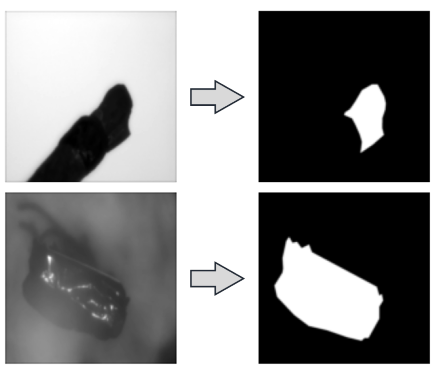

Figure 1: Examples of raw sample images from the TOPAZ beamline (left) and human-defined sapmle masks (right).

UNET MODEL

To meet the need for robust semantic segmentation of sample images from the TOPAZ beamline, we chose to develop an implementation of the popular convolutional neural network (CNN) architecture known as UNet [2]. Originally designed for semantic and instance segmentation of medical imagery, UNet performs strongly across a broad range of segmentation tasks and application domains.

UNet Architecture

The UNet architecture (see Fig. 2) is specifically designed for semantic and instance segmentation of images. It features a contraction path for extracting semantic information from an input, and an expansion path for reconstructing this semantic information into a spatially resolved format. The output of the network is a transformed representation which maps semantic (potentially including instance-level) information to the pixels of the input image.

Along a UNet's contraction path, an input is subjected to a series of standard two-dimensional convolutional units involving two two-dimensional convolution operations each followed by an optional dropout operation and ReLU activation. After each convolutional unit, max-pooling is performed to reduce the spatial dimensions of the data prior to the application of the next unit. The number of contraction units used is parameterized by the total number of UNet layers. At the bottom-most layer of the UNet, the input is represented as a compact feature-space vector which contains rich semantic information but no spatial information [2].

The feature-space representation produced by the contraction path is propagated through an expansion path along which transpose two-dimensional convolutions re-introduce spatial dimensionality. Although the transpose convolutions increase the spatial extent of the transformed data, they do not reproduce the spatially resolved features of the input. To retrieve this information, the outputs of each unit along the contraction path are forwarded to the corresponding expansion unit in the same layer, and are concatenated with the up-sampled data. Aside from the transpose convolutions, convolutional units along the expansion path are typically chosen to match those of the contraction path.

Hyperparameters of our UNet model including the number of layers, the number of output filters at each layer, the kernel size, and the stride were initially chosen to match those of the original UNet in [2]. These values, along with dropout and learning rates, were then used as tuning parameters for the model architecture, which was ultimately subjected to a hyperparameter tuning procedure.

Hyperparameter Tuning

Our UNet model was implemented in Python using the TensorFlow library [3], and we chose to use the Keras-Tuner optimization framework [4] to conduct Bayesian optimization of hyperparameters. Standard protocols like early-stopping and parameter constraints were used to reduce the significant computational costs of this effort. The resulting optimized UNet architecture features three layers, 3×3 regular convolutional kernels, 2×2 transpose convolutional and max-pooling kernels, $16 \cdot 2^{n_L - l}$ output filters at each layer (where n_L is the total number of layers and l is the layer number), dropout rates of 0.05, 0.05, and 0.1 (constant at each layer), and a learning rate of 1.44×10^{-3}.

Transfer Learning

An important benchmark for the success of our ML-based controls applications is the ability to rapidly deploy to new facilities without the need to train new models from scratch. In addition to our efforts at TOPAZ, we have been conducting similar work for HB2A, a beamline instrument at the High-Flux Isotope Reactor (HFIR) facility at ORNL [5]. To test the generalization abilities of our UNets, we performed additional rounds of training during which models initially trained on data from one beamline were either fed more data from their original beamline or the opposite one. This process is an example of transfer learning [6, 7], which has been shown to achieve improved performance through enhanced generalization. In our case, this process provided an

ISBN: 978-3-95450-238-7 ISSN: 2226-0358 doi:10.18429/JACoW-ICALEPCS2023-TUPDP116

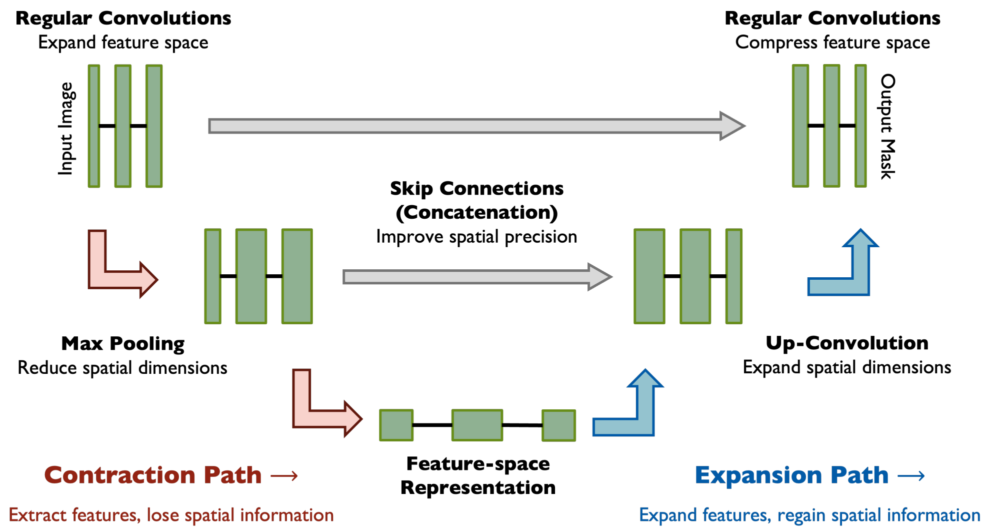

Figure 2: A depiction of the UNet architecture, with explanations of key components.

especially robust test of generalization given that TOPAZ sample images are taken with ordinary cameras, whereas HB2A sample images are taken with a neutron camera. We found that transferred models not only achieve similar training and validation losses compared to those trained on a single dataset, but ultimately outperformed those models on unseen test data for both beamlines.

UNCERTAINTY QUANTIFICATION

Accurate, robust uncertainty quantification (UQ) is critical for any ML application that will eventually inform real-world decisions [8, 9]. In our ML-based automated controls applications, uncertainties are used to determine, e.g., the need for human intervention during alignment procedures. The task of quantifying model predictive uncertainties is an area of study unto itself, and requires extensive consideration and testing for successful application.

Quantification Methods

We chose to test two popular statistical approaches for quantifying model prediction uncertainty: ensemble bagging and MC-dropout. In the ensemble method, an ensemble of ML models with the same architecture are randomly initialized and trained to produce similar but distinct models. Ensemble predictions are taken as the mean of individual model predictions, and uncertainty statistics are computed from the variances (see Fig. 3). In the MC-dropout method, a network is allowed to undergo random connection dropouts (which are usually disabled after training) so that predictions from a single model exhibit random variation. Several such predictions are made, and the mean and variance are used in much the same way as for ensemble bagging.

Although inherently stochastic ML models like variational networks and Bayesian neural networks (BNNs) can provide built-in UQ which benefits from enhanced interpretability and connection to model predictions [8], they also introduce significant development and operational costs and in our case would require extensive augmentation of the UNet model. On the other hand, statistical methods like the ones used here have been known to produced less accurate and potentially biased predictions of model performance [8, 9]. To verify the adequacy of statistical methods for this application, we have conducted thorough testing of each method against ground-truth model errors computed over a test dataset.

After initial rounds of UQ testing, we determined that an ensemble method approach was the most viable for use in our application. Despite initial concerns about potentially higher execution times in comparison to the MC-dropout method, we found that both methods required similar execution times for the same number of overall predictions (roughly 100-300 ms for 20 predictions). Although the ensemble method also requires a much more substantial investment of training time, we ultimately found that the resulting statistics showed much higher fidelity to the ground-truth error (see Fig. 4). With verification of these results in live testing (see Model Deployment), we are now also considering a combination of these methods which employs an ensemble of UNets each producing a set of MC-dropout predictions.

Ensemble Performance & Statistics

We randomly initialized and trained ensembles of 20 UNet models for 16 distinct combinations of training set (TOPAZ/HB2A), hyperparameter optimization (or lack of), and continued or transferred training (if any). Figure 3 shows an example sample image with the accompanying sample and uncertainty masks for sub-optimal (Ensemble 1) and optimal (Ensemble 2) ensembles. Computing ensemble variances in the sample mask and center of mass predictions over the test set, we were able to compare against the MSE metric relative to ground-truth masks. Doing so revealed that, especially for optimal ensembles configurations, the distributions of error estimates matched very well with the

ISBN: 978-3-95450-238-7 ISSN: 2226-0358 doi:10.18429/JACoW-ICALEPCS2023-TUPDP116

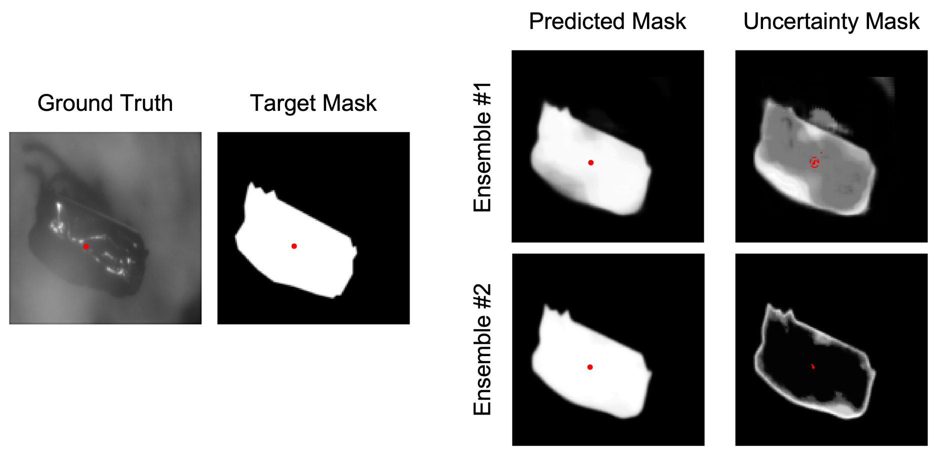

Figure 3: Uncertainty quantification through ensemble statistics for two ensembles with distinct training pipelines.

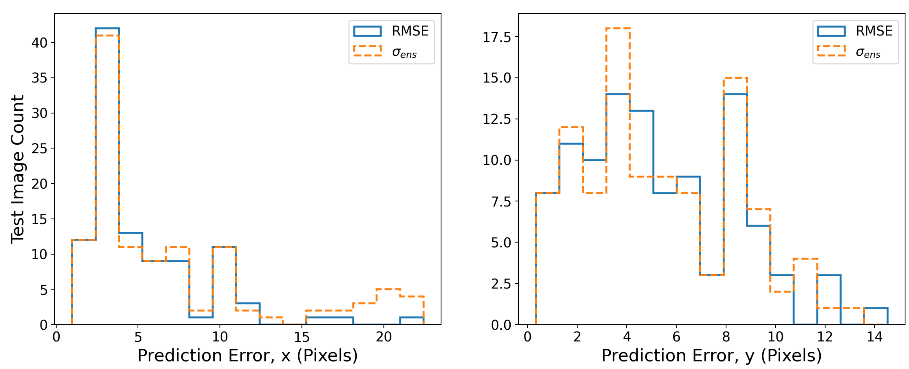

Figure 4: Distributions of RMS error and ensemble variances for an optimized model over a test set of image data.

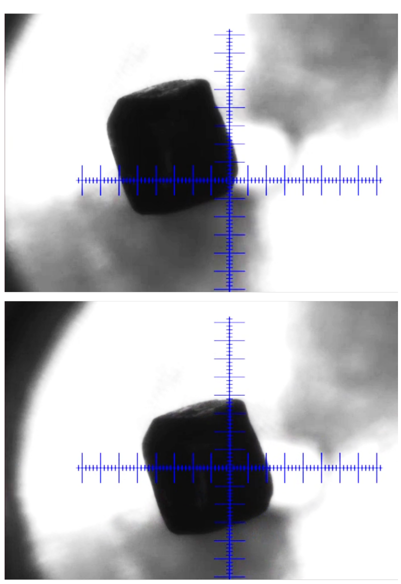

Figure 5: Sample images from TOPAZ before (top) and after (bottom) automated alignment using rscontrols.

true error distributions represented by the ground-truth MSE. Example distributions for errors in final centroid predictions over the test set (which are easier to visualize than error distributions for the sample mask predictions) are given in Figure 4 for a nearly-optimal ensemble. Distributions for optimal ensembles also match closely, but may appear trivial as errors are tightly clustered around 0 pixel-error.

The results of our ensemble UQ testing were highly promising, as ground-truth MSE cannot be computed during live operation and statistical measures of predictive uncertainty like ensemble variance tend to succeed at capturing aleatoric sources of uncertainty but do not always capture epistemic sources in ground-truth predictive uncertainties like the MSE [9]. Equipped with a robust method of evaluating real-time model predictions, we were able to confidently pursue automated alignment at TOPAZ in a monitored live-testing environment.

MODEL DEPLOYMENT

We performed iterative deployments of our ML-focused automated controls software at TOPAZ for testing under supervised conditions. These deployments and the resulting improvements to our software led to the creation and finally utilization of our new flexible EPICS-based controls framework, *rscontrols*, which is described in ”A Flexible EPICS Framework for Sample Alignment at Neutron Beamlines” published in these proceedings. This software allows us to quickly construct EPICS IOCs which embed our UNet

ISBN: 978-3-95450-238-7 ISSN: 2226-0358 doi:10.18429/JACoW-ICALEPCS2023-TUPDP116

models and controls process into a unified framework which can be initiated, operated, and run quickly and easily by beamline operators.

Through *rscontrols*, our integrated center of mass prediction tools were used to interact with the existing motor automation software at TOPAZ to perform a successful demonstration of fully-automated sample alignment (see Fig. 5). In this process, our UNet models were used to produce live sample masks from which centers of mass were computed. Then, the automation software replaced iterative clicks typically made by operators with *put* operations to the EPICS PVs associated with click positions. Throughout the process, our software monitored the progress of the motor automation IOC in order to make assignments to the EPICS PVs at the proper time.

CONCLUSION

We have successfully built and tested machine learning models for image segmentation to automate sample alignment processes at TOPAZ. One of the most important aspects of this work has been the quantification of model prediction errors in live settings given the absence of ground-truth data. Much of our early development efforts prior to live automation tests were dedicated to UQ, and the promising results achieved during offline testing were an absolutely necessary benchmark for attempting tests with real experimental equipment.

This work also solidifies that machine learning models for image segmentation have proven to be useful components in the further automation of sample alignment tasks at TOPAZ, despite the pre-existence of semi-automated controls software on the beamline. Using ML models, we have achieved a higher degree of automation and reduced the overall workload associated with sample for beamline operators. As we move forward with the automation of more complex controls tasks at TOPAZ, we remain committed to practicing safe, responsible ML development and application.

ACKNOWLEDGMENTS

This work is supported by the U.S. Department of Energy, Office of Science, Office of Basic Energy Sciences, SBIR and STTR Program under Award Number(s) DE-SC0021555.

REFERENCES

[1] L. Coates *et al.*, "A suite-level review of the neutron single-crystal diffraction instruments at oak ridge national laboratory", *Rev. Sci. Instrum.*, vol. 89, no. 9, p. 092 802, 2018. doi:10.1063/1.5030896

[2] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation", in *MICCAI 2015 Proceedings, Part III 18*, Münich, Germany, Springer, 2015, pp. 234–241. doi:10.1007/978-3-319-24574-4_28

[3] TensorFlow: Large-scale machine learning on heterogeneous systems, https://www.tensorflow.org/

[4] Keras, https://keras.io

[5] S. Calder *et al.*, "A suite-level review of the neutron powder diffraction instruments at oak ridge national laboratory", *Review of Scientific Instruments*, vol. 89, no. 9, p. 092 701, 2018. doi:10.1063/1.5033906

[6] R. Caruana, "Multitask learning", *Machine learning*, vol. 28, pp. 41–75, 1997. doi:10.1023/A:1007379606734

[7] S. Pan and Q. Yang, "A survey on transfer learning", *IEEE Trans. Knowl. Data Eng.*, vol. 22, no. 10, pp. 1345–1359, 2009. doi:10.1109/TKDE.2009.191

[8] Y. Gal, "Uncertainty in deep learning", Ph.D. thesis, University of Cambridge, 2016.

[9] E. Hüllermeier and W. Waegeman, "Aleatoric and epistemic uncertainty in machine learning: An introduction to concepts and methods", *Machine Learning*, vol. 110, pp. 457–506, 2021. doi:10.1007/s10994-021-05946-3