Data Augmentations for Nuclear Feature Extraction in Semi-Supervised Contrastive Machine Learning

Jordan Stomps[1], Paul Wilson[1], Ken Dayman[2], Michael Willis[2], James Ghawaly[2], and Dan Archer[2]

[1] *University of Wisconsin–Madison*
[2] *Oak Ridge National Laboratory*

Abstract

Persistent radiation monitoring can be used as a powerful tool for detecting movements of nuclear material in a variety of use cases and nuclear nonproliferation scenarios. Existing gamma-ray detection systems can collect large volumes of data that can potentially be used in machine learning algorithms for anomaly detection or classification. However, the domain expertise and/or computational costs required to label sufficient radiation data (i.e. identify constituent nuclides) for machine learning may be prohibitive. Semi-supervised machine learning alleviates the cost of labeling by learning from the limited attributed data and a larger unlabeled corpus. One method, contrastive learning, learns patterns in a self-supervised manner by using a set of data augmentations to perturb data in ways that should not alter the inferred labels and enforces maximum agreement between pairs of samples. Whereas contrastive learning is traditionally conducted on images, where valid transformations are maturely developed and intuitively understood, this work endeavors to design and apply valid data augmentations for nuclear radiation data based on the underlying physics. That is, appropriate transformations should reflect realistic radiation detector physics and maintain classification information for radiation signatures. A non-exhaustive set of six transformations are presented, ranging from channel resampling, masking, and nuclear interactions to perturbing the signal-to-background ratio, detector resolution, and gain shift. These augmentations are tailored for physical measurements, rather than just simulated data. Demonstration is conducted using radiation measurements from sodium iodide detectors deployed around the High-Flux Isotope Reactor and the Radiochemical Engineering Development Center at Oak Ridge National Laboratory. These transformations are intended to be used in a contrastive learning framework trained to identify anomalous spectra, fine-tuned using a set of manually characterized samples. The ideal result is a model that reduces the burden of labeling training data while still utilizing measurements taken, reflecting value in unlabeled data.

1 Introduction

The continuous analysis of radiation data collected in and around nuclear fuel cycle facilities can aid in facility operations by verifying normal operations and be a potential safeguards capability by persistent monitoring. Radiation detectors can be used to measure signatures produced by nuclear material. Traditional techniques that require manual inspection by a subject-matter expert (SME) to determine the presence of specific nuclear materials will become intractable for measurement techniques with high sampling rates, distributed sensor networks, or multimodal data fusion. Maintaining resource efficiency will require methods that can analyze data with reduced computing and domain costs without sacrificing detection accuracy.

Modern machine learning (ML) methods like neural networks have shown powerful capabilities to model complex and nonlinear, real-world scenarios but require large volumes of labeled (i.e. well characterized) data to train accurate classifiers. In nuclear nonproliferation, and radiation detection specifically, large volumes of data could be collected but labeling the amount necessary for training a neural network may be infeasible. Semi-supervised machine learning (SSML) techniques attempt to solve this challenge by learning knowledge obtained from (often limited amounts of) labeled data while still utilizing the larger unlabeled data corpus for information on the broader statistical data distribution [1].

The Multi-Informatics for Nuclear Operating Scenarios (MINOS) project has created a large multimodal dataset at Oak Ridge National Laboratory (ORNL) using a network of sensors distributed around High-Flux Isotope Reactor (HFIR), a research reactor, and Radiochemical Engineering Development Center (REDC), a radiochemical processing facility. These sensors collected measurement modalities capable of detecting and characterizing nuclear facility operations including (but not limited to) radiation, seismo-acoustic, and electromagnetic [2]. The MINOS radiation dataset exhibits properties amenable to SSML applications. This paper serves as an iterative demonstration of SSML capabilities for nuclear nonproliferation using radiation measurements collected by MINOS.

1.1 Contrastive Learning

Contrastive learning, or self-supervision more broadly, is a set of methods that train a model by distinguishing between similar and dissimilar representations of data classes. These representations allow a model to learn patterns in data distributions that maximize distinction between discriminative classes. Researchers at Google have introduced a contrastive learning framework that has been shown to produce state-of-the-art performance on image classification datasets: the Simple Framework for Contrastive Learning of Visual Representations (SimCLR) [3]. The framework utilizes mature augmentations that have been designed for image classification. Each transformation exploits certain invariant properties of an image to maintain labeling information. Care must be taken that the transformation does not occlude labeling information but merely perturbs the image. SimCLR involves four pipeline components:

- **Data augmentation** for transforming samples using a set of augmentation types/rules.
- **Base encoder** that conducts representation learning.

- **Projection head** to move from representation to contrastive learning and measuring similarity.
- **Contrastive loss function** that maximizes agreement between positive (similar) samples and ultimately results in the predictive, or discriminative, model task.

2 Data Augmentations

The first step in a self-supervised pipeline is data augmentation, which employs transformations that will be used to learn representations relevant for downstream tasks, like classification. While appropriate data augmentations might be mature for image classification, valid transformations for spectroscopic data must be developed. If physically viable transformations can be constructed, then self-supervision can be used to process unlabeled data and create a classifier with low labeling cost. Detailed here are transformations designed to naturally aid in robust contrastive learning based on physical principles of nuclear radiation detection.

A set of manually labeled, one-minute material transfer spectra from MINOS consisting of seven transfer classes is used to demonstrate augmentations: unirradiated ^{237}Np targets used for ^{238}Pu production, irradiated ^{237}Np containing ^{238}Pu, unirradiated Cm targets used for ^{252}Cf production, irradiated Cm containing ^{252}Cf, ^{225}Ac, activated metals, and spent fuel [4].

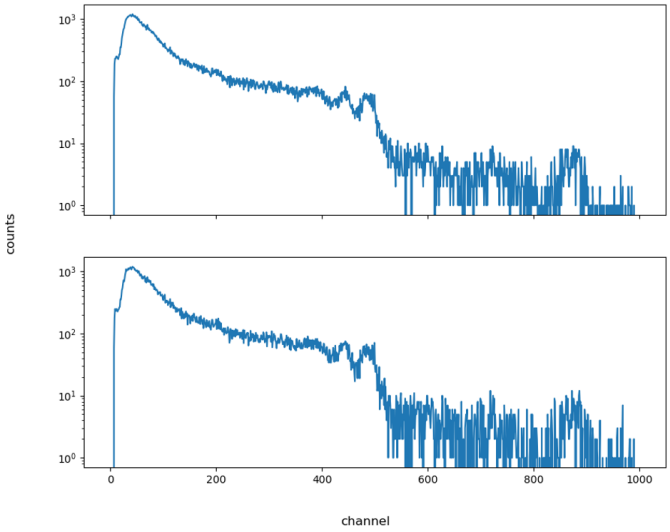

Figure 1: Channel resampling augmentation example: original transfer spectrum of activated metal (top) compared to an augmented version (bottom).

2.1 Channel Resampling

Arguably the simplest transformation involves statistically resampling each spectrum from a Poisson distribution. The count magnitude in each energy bin of the spectrum serves as the expected value for a Poisson distribution that is randomly sampled to generate a new count

magnitude. This means the new spectrum is similar in shape and magnitude yet fully transformed. See Figure 1 as an example.

2.2 Different Backgrounds

Here, a characteristic event signature is superimposed onto a different background distribution. Learning to recognize the underlying background continuum may help a model distinguish it from event signatures.

It is necessary for this augmentation to first estimate a background distribution before transforming the spectrum. A naïve implementation would take a measurement where an event was not be present (evaluated by a SME) as some notional estimate of background. This approach is limited in that it does not guarantee the absence of a radiological event or transfer without the determination by a SME or some other analysis procedure. An alternative method, baseline estimation and denoising using sparsity (BEADS) [5], does not require additional evaluations besides supplying the spectrum to be augmented. BEADS uses a method of matrix decomposition—essentially a low-pass and high-pass filter—to separate a spectral measurement into spectral peaks, a smooth baseline, and statistical noise. This system is sufficient for a method of rough background estimation and requires less computation compared to the analysis of multiple, manually inspected background spectra.

If samples from the background corpus are on average representative of a typical background response, the downstream model should be able to learn from this augmentation. For automatically generated augmentations, the model will randomly select a spectrum from the background corpus, remove the existing background from the spectrum in question, and add the library background to the separated peaks. See Figure 2 as an example.

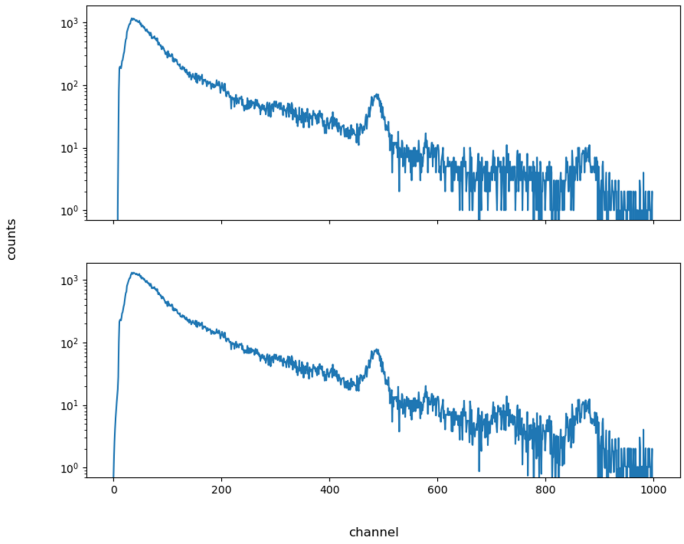

Figure 2: Different background augmentation example: original transfer spectrum of ^{252}Cf (top) compared to an augmented version (bottom).

2.3 Signal-to-Background

As a detector moves away from a source (or vice versa), the count-rate intensity of radiation associated with the source decreases on the order of $1/r^2$ (r being the radius, or distance, between the detector and the source). Two sodium iodide (NaI) detectors could have similar detector responses, but the unique circumstances of their measurement may result in varying background distributions or signature intensities, especially if they are different distances to the source. Therefore, perturbing spectra by manipulating their signal-to-background ratios could amount to the largest variation between realistic spectral measurements. Signal-to-background operates in almost the same manner as Section 2.2 above, except that the chosen baseline estimate is perturbed prior to being added to the event signatures. The spectrum is scaled by some ratio, r, randomly upscaling or downscaling by up to two-times the original background spectrum. See Figure 3 as an example.

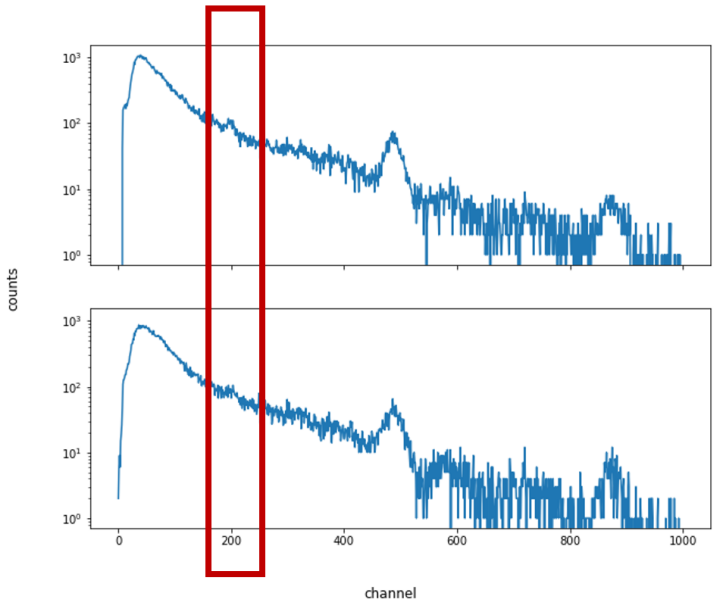

Figure 3: Signal-to-background augmentation example: original transfer spectrum of an unirradiated Np target (top) compared to an augmented version (bottom). An area of noticeable difference is highlighted in red.

2.4 Nuclear Interactions

This augmentation involves artificially introducing detector specific phenomena, peaks, and distributions associated with radiation signatures present or absent from the original spectrum. The model first uses BEADS to isolate peaks from the baseline distribution in a spectrum. If an augmentation inserts a radiation signature that is irrelevant to the characteristic signatures of a label class, this encourages the model to learn that this optional signature is not needed for a classification decision if it is present in a test spectrum.

For this analysis, the list of possible peak insertions consists of radiation events that occur in MINOS but are not transfers: the potassium-uranium-thorium (KUT) continuum (^{40}K: 1460 keV, ^{232}Th: 2614 keV), ^{238}U: 609 keV, ^{214}Bi: 1764 and 2204 keV, ^{214}Pb: 295 and 352 keV, and ^{41}Ar: 1294 keV. Using the most prominent peak found using a peak fitting algorithm, a linear relationship between peak width and energy as well as peak amplitude and energy is found. The artificial peak is inserted into the spectrum along with the approximated background near the peak shoulders and lower magnitude escape peaks if the inserted peak has an energy greater than 1.022 MeV. See Figure 4 as an example. While it is possible to also subtract a peak from a spectrum, this is ignored when randomly selecting augmentations and generating spectra by the model.

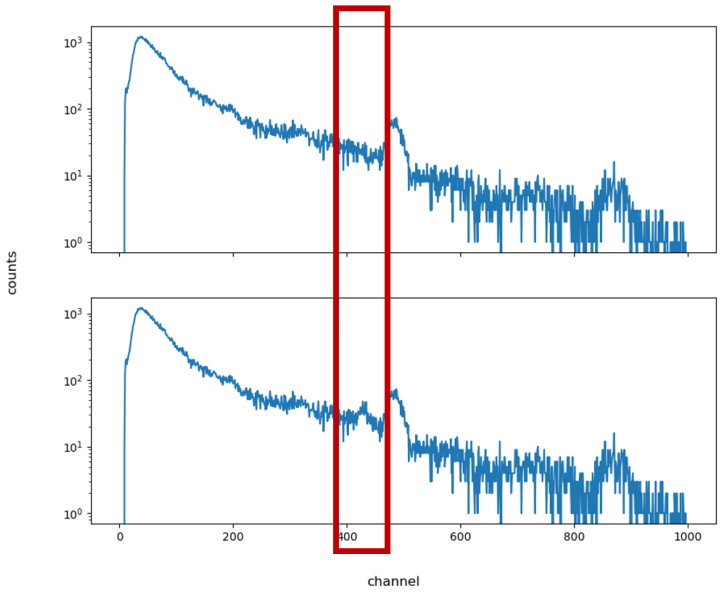

Figure 4: Nuclear interactions augmentation example: original transfer spectrum of an irradiated Np target (top) compared to an augmented version with an ^{41}Ar 1294 keV photopeak inserted (bottom). The area of noticeable difference is highlighted in red.

2.5 Detector Resolution

Detection efficiency can affect photopeak energy resolution leading a peak to appear sharper or wider in a spectrum. This is typically measured using full-width half-maximum (FWHM). Altering detector resolution would be valuable because it simulates mild variations in real-world detector applications. Manipulating the resolution of an existing peak requires identifying a peak of interest and accurately measuring the peaks Gaussian parameters. Lacking user input, the model randomly selects a peak in the spectrum and its region of interest (ROI). First, the peak is fit to a Gaussian distribution, and a peak width (i.e. standard deviation) is estimated. The peak width is multiplied by a scalar, and a new peak is randomized using channel resampling and inserted using the new peak fit parameters. See Figure 5 as an example.

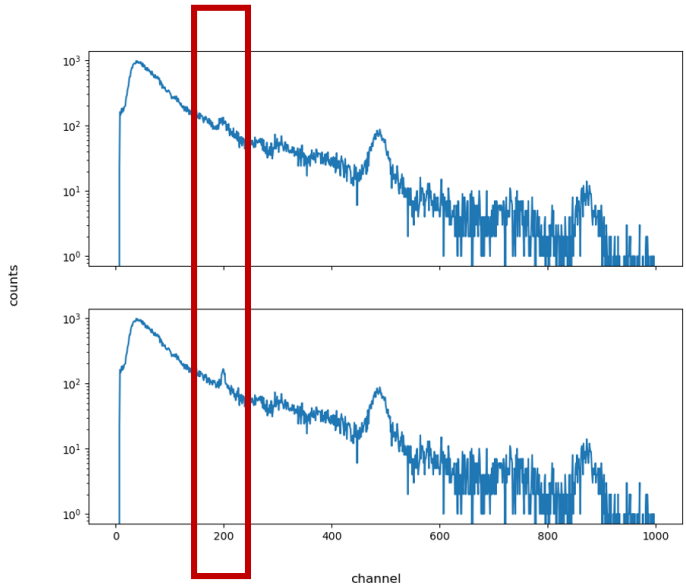

Figure 5: Resolution augmentation example: original transfer spectrum of ^{225}Ac (top) compared to an augmented version with a perturbed 200 keV photopeak (bottom). The area of noticeable difference is highlighted in red.

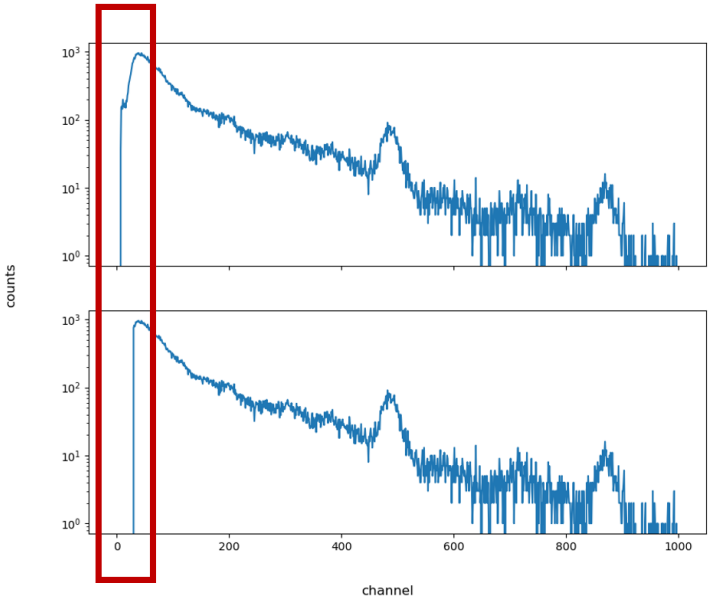

Figure 6: Masking augmentation example: original transfer spectrum of spent fuel (top) compared to an augmented version (bottom). The area of noticeable difference is highlighted in red.

2.6 Masking

Like in image classification, a detector's spectrum could be masked in the analysis, but care must be taken to select an appropriate mask size. Larger masking regions could mean masking classification information. Conversely, too small of a masked region and it becomes a meaningless augmentation for representation learning. Of relevance would be a low-energy mask, common in detection systems that ignore high count-rate, low energy distributions to avoid overwhelming the spectrum with a dominant, negligible feature. The default block applied here randomly ranges from 0 to a maximum of 100 bins. See Figure 6 as an example.

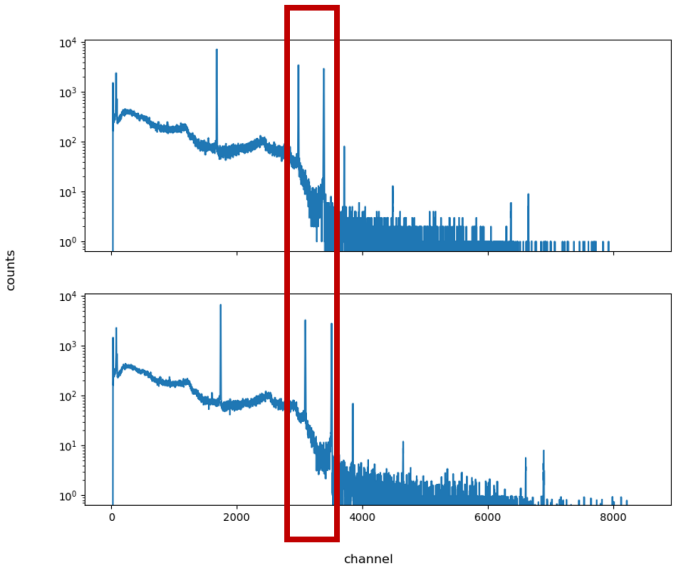

Figure 7: Gain shift augmentation example: original laboratory measurement spectrum of $^{137}Cs^{60}Co$ (top) compared to an augmented version (bottom). An area of noticeable difference is highlighted in red.

2.7 Gain Shift

If the calibration in pulse-height discrimination is off, the entire spectrum can be shifted by several units of energy (as determined by the multi-channel binning). Mild gain shift is typical for long-running measurements, so calibration often occurs over the course of a measurement sequence. If gain shift is introduced as a data augmentation, it may make a contrastive learning model more robust to these shifts in energy and encourage tolerance for uncertainty in photopeak centroids. Gain shift is applied as a linear resampling (i.e. rebinning with a linear interpolation). This scaling results in a behavior and shape similar to gain shift observed in a detection system, linear amplifier, and discriminator. See Figure 7 as an example. Note that Figure 7 perturbs a laboratory measurement of $^{137}Cs^{60}Co$, emphasizing that these augmentations are not specialized for an individual dataset but generalized to principles of gamma spectroscopy.

3 Using Augmentations in Semi-Supervised Contrastive Learning

Take Figure 8 as an example contrastive learning framework inspired by SimCLR. Given an instance of a spectrum, s, apply a transform $\tau \in \mathcal{T}$ to get an augmented instance $\tau(s) = \tilde{s}$. These are fed into a model to produce a higher-dimensional representation that captures more abstract patterns in s: $h = g(\tau(s))$. Then, $f(h) = f(g(\tau(s)))$ uses the representation to output a projection for contrastive learning or a prediction, z. Because $\{\tilde{s}_1, \tilde{s}_2\}$ were created by transforming s, all three should share the same label because they all contain the same intrinsic labeling information. The optimization task of contrastive learning is then to ensure that the representations, h, agree for similar (i.e. positive) pairs. Negative pairs from dissimilar samples should have maximum disagreement.

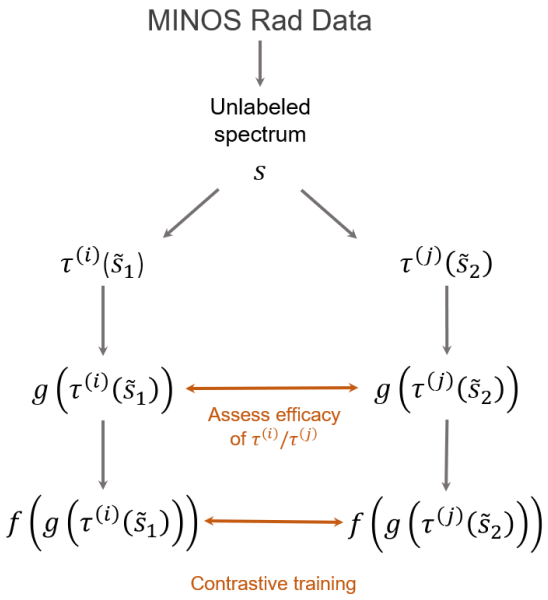

Figure 8: A model framework for contrastive learning using MINOS data. A contrastive learning model that uses two transformed instances of an original radiation spectrum, s. Using a representation model $g(\tau(\tilde{s})) = h$ and a predictive model $f(h) = z$, positive pairs are trained to maximize agreement (i.e. contrastive learning). This pipeline is similar to the one used by SimCLR [3] for image classification.

A natural extension to the above framework involves semi-supervised applications of unlabeled and labeled data. The next iteration of SimCLR [6] uses a method of task-agnostic training with unlabeled data and task-specific training with labeled data to fine-tune and scale the final model. The final framework should utilize unlabeled data for contrastive learning, maximizing the predictive capability from similar samples, and use labeled samples—however few—to ensure the representations learned do indeed typify the distinct characteristics between classes (e.g. different types of nuclear material transfers).

4 Conclusion

It has been demonstrated that the inclusion of unlabeled data improves classification accuracy when the availability of labeled data is limited. Here, the question is posited in the inverse direction: given a self-supervised model that acts upon unlabeled data, how does the introduction of labeled data improve accuracy? One challenge scenario attempts to learn representations useful for classifying different types of shielded radiological material transfers. A large, unlabeled corpus like from MINOS could be used to contrastively learn meaningful representations for radiation data utilizing the augmentations described here. Then, a realistically small number of labeled spectra could be used to train a classifier on representations to classify specific nuclear materials versus other shielded radiological material transfers. In this way, the framework transfers knowledge from the unlabeled data, guided by labeled samples, towards nuclear material detection and characterization. This future work could establish valuable capabilities that harness monitoring data in a resource-efficient manner.

Acknowledgements

This work is supported by the U.S. Department of Energy's National Nuclear Security Administration under Award Number DE-NA0003921.

References

[1] A. Singh, R. Nowak, and J. Zhu, "Unlabeled data: Now it helps, now it doesn't," in *Advances in Neural Information Processing Systems*, vol. 21, Curran Associates, Inc., 2008.

[2] A. D. Nicholson, D. E. Archer, I. Garishvili, I. R. Stewart, and M. J. Willis, "Characterization of gamma-ray background outside of the High Flux Isotope Reactor," *Journal of Radioanalytical and Nuclear Chemistry*, vol. 318, pp. 361–367, Oct. 2018.

[3] T. Chen, S. Kornblith, M. Norouzi, and G. Hinton, "A Simple Framework for Contrastive Learning of Visual Representations," June 2020.

[4] K. Dayman, J. Hite, R. Hunley, N. S. V. Rao, C. Geulich, M. Willis, J. Ghawaly, D. Archer, and J. Johnson, "Tracking Material Transfers At A Nuclear Facility With Physics-informed Machine Learning And Data Fusion," Institute of Nuclear Materials Management, 2021.

[5] X. Ning, I. W. Selesnick, and L. Duval, "Chromatogram baseline estimation and denoising using sparsity (BEADS)," *Chemometrics and Intelligent Laboratory Systems*, vol. 139, pp. 156–167, Dec. 2014.

[6] T. Chen, S. Kornblith, K. Swersky, M. Norouzi, and G. Hinton, "Big self-supervised models are strong semi-supervised learners," *arXiv preprint arXiv:2006.10029*, 2020.