Enhancing Nuclear Material Storage Container Surveillance with Automation and Machine Learning Toolkits

Ross Lee, Steven Lukow, Joseph Hafen, David Grow, Jonathan Gigax

LA-UR-23-XXXXX
April 20, 2023

1 Abstract

Nuclear material storage containers at U.S. national laboratories must meet performance requirements set by the DOE to ensure the safety of workers, public, and the environment. As part of these requirements, container inspections must be performed routinely and are often in high-dose environments where maintaining as-low-as-reasonably-achievable dose requirements limit the number and frequency of surveillance. Additionally, these inspections are human-intensive tasks that require a greater-than-familiarity knowledge of each container's status by subject matter experts. Observing and documenting all anomalies of concern, such as dents and corrosion, both during and between inspections challenges the current surveillance paradigm. Improvements to the frequency and accuracy of inspections can be achieved with an autonomous system capable of detecting anomalous features of nuclear storage containers in real-time. Our implementations use a supervised machine learning approach with region-based convolutional neural network architectures (Mask R-CNN and Detectron2). Using both real and simulated container images, model performance is comparable to that achieved with the COCO dataset. Detection of all anomalous features, especially those not trained into supervised models, presents a larger challenge. Our approach leverages the use of two separate techniques: unsupervised machine learning models and sensor fusion for online automated training. We evaluate the improvements feasible from unsupervised anomaly detection to identify previously unseen defects and allow datasets to be entirely constructed from images of undamaged nuclear storage containers. A second improvement leverages supervised learning instance segmentation models on 2D images augmented with additional 3D scanning hardware to generate "ground truth" true-positive regions autonomously in a bench-top implementation. These techniques are combined to build a more general toolset that can better detect nuclear material storage container anomalies and improve container inspections overall.

2 Introduction

Short term nuclear material storage at Los Alamos National Laboratory (LANL) must be compliant with requirements outlined in Department of Energy (DOE) Manual 441.1-1 [1], hereafter Manual. The Manual prescribes a range of properties and performance criteria for container packages to ensure maximized containment of nuclear material. The SAVY-4000 container is the Manual compliant containerization package of choice at LANL, and is the principle barrier to radiation release in storage environments. Thus, the container must maintain performance requirements and structural integrity throughout its service life to ensure continued protection to the worker, public, and environment.

Inspections of storage containers performed in recent years have shown the appearance of corrosion occurring in a variety of containers, including the SAVY-4000 and older Hagan containers [2, 3, 4, 5]. Figure 1 provides examples of corrosion appearing in both container types. Under some conditions, corrosion has appeared on steel components near the filter of the container. The presence of corrosion has implications for many of the performance qualification requirements, notably the design qualification release rate. Pristine SAVY-4000s have an established release rate drop below the A2 limit for the most conservative storage conditions. There is currently little to no information on the performance for containers exhibiting corrosion.

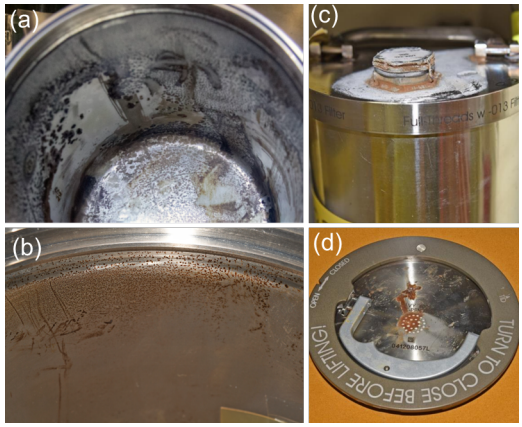


Figure 1: Images of corrosion appearing in (a) Hagan and (b) SAVY-4000 storage containers during service. External indicators of corrosion visible on certain (c) Hagan and (d) SAVY-4000 lids near the filter.

Moreover, structural damage to containers, in the form of dents, are caused routinely by operators handling containers. Minor dents are not a concern for pristine containers as the conditions that cause dents are significantly less catastrophic than a 12 foot drop test used to demonstrate compliance with Manual release rate limits. As a result, most containers remain in service with minor dents or surface damage. However, these defects often leave a net tensile stress on the inner wall and can exacerbate corrosion processes inside the container. Detecting their presence on containers and quantitatively evaluating damage to the container wall are both key pieces of information necessary for assessing container health following the appearance of corrosion.

A key activity prescribed by the Manual to ensure container reliability involves routine surveillance. This, however, is a human-intensive effort whose quality relies heavily on the operator's knowledge and familiarity with storage containers, in addition to its use history. Inspections can be further limited by other factors including time constraints placed on surveillance activities. The development of an analytical tool that combines subject matter expert classification knowledge and with a shareable data

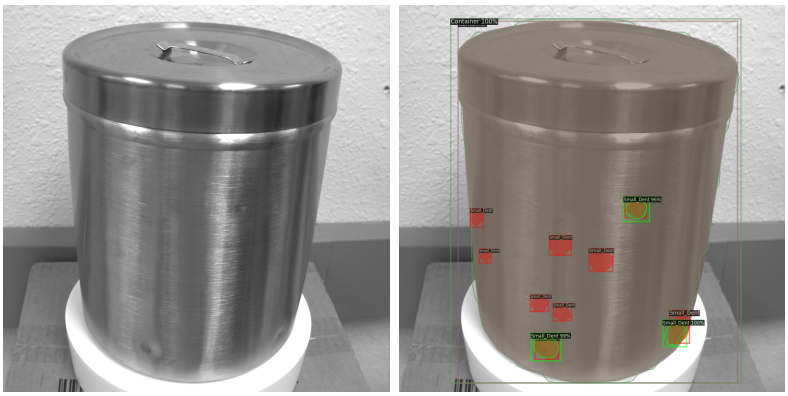
output is critical to improving the current surveillance paradigm. Our recent efforts have focused on the development of robust supervised and unsupervised machine learning (ML) models that can be deployed in the field to co-opt the experience and enhance data acquired during surveillance activities.

3 Related Work

Most recent computer vision approaches to nuclear material container surveillance can be loosely categorized as supervised, unsupervised, or a combination of the two.

Supervised approaches are widely used for image classification, object detection, and segmentation problems. In these types of problems, an expert-in-the-loop draws around instances of interest, annotating the data and thereby pinpointing an example of an object to be detected. A supervised approach can be implemented for segmenting features of interest on SAVY-4000 and Slip Lid containers alike. In the case of SAVY-4000 containers, we demonstrated undamaged container feature detection with the state-of-the-art Mask R-CNN implementation created by Meta AI via Detectron2 [6].

Mask R-CNN can also be used for detecting damage features on nuclear material storage containers. In previous work, we demonstrate Mask R-CNN for segmenting instances of dents and scratches on the surface of Slip Lid containers [7, 6]. This model utilized three different cameras to segment dents and scratches. While the report demonstrated reasonable (about 86%) precision, the composite image accuracy sorely lacked (at 40%) and demonstrated a need for expanded approaches to improve damage detection, particularly to eliminate false negatives.



(a) *(b)*

Figure 2: Figure (a) shows the raw image collected from a color camera. Figure (b) shows the ground truth instances of damage (red) and the detections performed by Detectron2 (green), underscoring the need to eliminate false negatives.

Unsupervised approaches have recently generated interest for their potential use in expanding detection capabilities. Because unsupervised approaches to computer vision problems require no annotation, the time from training to inference is often significantly shorter. These types of models can often be used for clustering or anomaly detection. Anomaly detection is of particular interest because the purpose of an algorithm is generally to train on observations that are typical and make inferences on images that are atypical. This is particularly attractive when the dataset is sparse, because unsupervised anomaly detection frameworks such as PatchCore or PaDiM require only about 100 images of pristine examples to train sufficiently.

4 Dataset Enhancement

One of the advantages to using the Mask R-CNN implementation in Detectron2 is the flexibility of image data that the implementation can incorporate. Because the Detectron2 Mask R-CNN implementation is built in PyTorch, the number of channels accessible can be easily changed to accommodate a custom channel from a pre-processing step. Ideally, a step would account for suspected areas of damage on the surface of a Slip Lid container and would only minimally add to the total training process. In this, we took inspiration from an unsupervised approach to processing the Slip Lid images before training on Detectron2 Mask R-CNN implementation.

4.1 Patch Distribution Modeling (PaDiM)

Patch Distribution Modeling (PaDiM) is a state-of-the-art unsupervised anomaly detection capable of detecting and localizing image anomalies. Because pre-trained CNN architectures can output relevant features, PaDiM can completely bypass conventional supervised neural network optimization practices to generate patch embedding vectors. Using the patch embedding vectors generated, PaDiM learns the image features of a given image dataset by assuming a multivariate Gaussian distribution and computing a covariance matrix given by

$$\Sigma_{ij} = \frac{1}{N-1}\sum_{k=1}^{N} = (x_{ij}^{k} - \mu_{ij})(x_{ij}^{k} - \mu_{ij})^{T} + \epsilon I$$

where Σ_{ij} represents the sample covariance matrix, N represents the number of normal training images, x_{ij}^{k} represents an embedding vector for an image k, and μ_{ij} the sample mean of the set of embedding vectors.

After training, the covariance matrix and mean can be saved as a model for deployment. To make inferences with this model, PaDiM uses the inverse of the covariance matrix to compute the Mahalanobis distances between the test image patch embeddings and the learned normal distribution, returning a map of patch anomaly scores given by

$$M(x_{ij}) = \sqrt{(x_{ij} - \mu_{ij})^{T}\Sigma_{ij}^{-1}(x_{ij} - \mu_{ij})}$$

where $M(x_{ij})$ is the matrix of Mahalanobis patch distances as a function of tested patch embedding. Interpolating the distances to the dimensions of the test image, this effectively generates an image map of anomaly scores [8]. Should the image be converted to gray scale, the heatmap could be used as a channel. If a threshold is set, this enables PaDiM to segment anomalous instances, however for the purposes of this work we choose to omit the segmentation step.

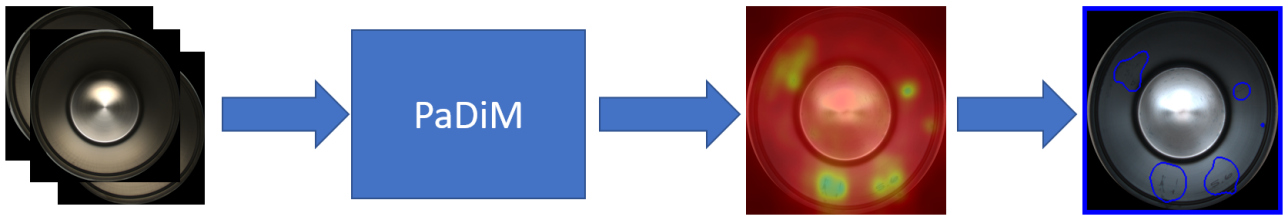


Figure 3: The images from left to right describe the PaDiM algorithm. PaDiM trains on pristine examples and computes anomaly heatmaps. Setting a threshold, it allows the option for automatic segmentation of anomalies.

4.2 PaDiM Pre-processing

PaDiM is an attractive candidate for an unsupervised pre-processing step because anomaly map images can be processed as a heatmap, converted to grayscale and used as an additional channel for an image dataset trained on Detectron2. An additional channel creates another aspect of the data the model conforms to when optimized. By concatenating an additional channel to the RGB images in a damaged Slip Lid container dataset, the newly enhanced image dataset provides additional information to a candidate CNN-based machine learning architecture.

Another advantage to using PaDiM for pre-processing is that it takes nominal time to train. A training session for a PaDiM model trained on a 200-image dataset requires roughly 2 minutes to extract the features and start computing anomaly maps. This is in direct contrast to the supervised Detectron2 implementation of Mask R-CNN, which is known to take significantly more time to train. Adding a pre-processing step is therefore negligible to the overall training time.

4.3 Deterministic Ground Truth Generation

Although significant progress has been made with respect to detecting anomalous features of interest within visual images, the limitation of using only one visual sensor leads to false detections in some cases. This can pose a challenge for an automated system compared to human performance, where we naturally use the many senses we have (sight, touch, smell, etc.) to form a more complete model of the world. The two most utilized senses for detecting physical anomalies on containers are sight and touch, which can be approximated using a visual camera and a depth-mapping system. An additional challenge comes from the container material itself, which is relatively reflective. Simple stereoscopic vision and time-of-flight (ToF) cameras are not well-suited for this application because of the reflective surface and sharp specular highlights, but other technologies such as structured light sensors with multiple exposure capabilities.

The LMI Gocator Snapshot sensor was selected because of its wide field of view, quick scan time, and software API. Previous work has shown its application on metallic containers to reconstruct an entire surface model, which is leveraged in this work. Indentations (dents) can be detected by calculating a local radius of curvature for each vertex on a reconstructed 3D mesh and isolating regions with a negative curvature. This process can be overlaid on a visual image to either generate ground truth detections or validate inferences performed by another detection method, like Detectron2.

5 Experiments

5.1 Unsupervised Training with PaDiM

One of two pre-processing steps was used for computing anomaly maps in an effort to further localize detections to regions of interest. The procedure for concatenating an anomaly channel to augment the dataset from three to four channels was as follows:

1. A PaDiM set to the PyTorch ResNet-18 backbone was trained on 225 undamaged 3-Qt Slip Lid containers with a NVIDIA A6000 GPU.
2. The PaDiM model was used to compute the anomaly score maps of an additional 309 images of both damaged and undamaged Slip Lid containers.
3. The images were scaled and converted to grayscale, then concatenated as an additional channel to each corresponding RGB Slip Lid image.

4. Detectron2 was then trained on the 4-channel enhanced image dataset and evaluated on a separate set of 40 Slip Lid containers.

The second procedure where the G channel was replaced with the anomaly map was as follows:

1. A PaDiM set to the PyTorch ResNet-18 backbone was trained on 225 undamaged 3-Qt Slip Lid containers with a NVIDIA A6000 GPU.
2. The PaDiM model was used to compute the anomaly score maps of an additional 309 images of both damaged and undamaged Slip Lid containers.
3. The images were scaled and used to replace the G channel for each image in the Slip Lid image dataset.
4. Detectron2 was then trained on the 3-channel modified image dataset and evaluated on a separate set of 40 Slip Lid containers.

5.2 Sensor Fusion Experimental Setup

A Universal Robotics UR5e robotic manipulator was used to position a sensor package integrating a LMI Gocator 3210 Snapshot sensor with a FLIR Blackfly S color camera above an inspection table. A software package was developed to integrate a Detectron2 model trained on container bodies and various types of damage with the robotic hardware used [6]. At each step in the process, the detected outputs are visible to the worker to ensure the procedure is being completed appropriately. The test procedure is as follows:

1. A worker places a container on the inspection table and starts the robot's movement sequence.
2. The robot uses the visual camera to locate the container by moving the end-effector in a planar motion at a fixed height.
3. Once the robot has successfully located the container, it changes to position the visual camera at a set distance from the side of the container and collects five images in a sequence and detects damage using a trained Detectron2 model.
4. The robot moves the end-effector vertically to collect multiple scans with the Gocator sensor at three exposure levels each.
5. A mesh-processing step is performed to register and merge the measured point clouds and convert the composite to a triangular mesh that can be evaluated for curvature.
6. Both the Detectron2 inferences of dents and the computed regions of negative curvature are superimposed on the original image to compare detection accuracy between each method.

6 Results

6.1 PaDiM

Performance							
Dataset	Container	Dent	Residue	Scratches	AP	AP50	AP75
4-channel PaDiM	94.4648	18.8316	77.5495	0.7633	47.9023	62.4847	50.7775
3-channel PaDiM	94.5780	14.9780	73.1683	1.9878	46.1780	62.2216	49.7031
Original	94.3988	16.6292	74.2244	1.3892	46.6604	62.0277	49.9312

Inferences were conducted according to the procedures outlined in 5.1 and 5.1 with varying degrees of success. Average precision (AP) scores were computed for each class and overall for Jaccard indices

(IoU) of 50% and 75%. The first unsupervised pre-processing technique outlined in 5.1 was slightly successful for the 4-channel system. While each overall AP metric improved slightly with the addition of a fourth channel from the unsupervised pre-processing of PaDiM, each was minute, ranging from a 0.7% - 2.6% increase in overall AP score. What is more pronounced is the effect pre-processing had on individual class AP scores for segmentation. Adding a fourth channel improved the 'Dent' class AP score by 13.2% from the original and about 4.5% for the 'Residue' class. Outlying this trend is a 45% decrease in scratch detection performance from the original.

The second pre-processing technique outlined in 5.1 was largely unsuccessful. For the 3-channel replacement method, the AP and AP75 scores actually declined slightly while the AP50 slightly improved. The 'Dent' class AP scores performed the worst of the three approaches, but performed the best on the scratches of the three approaches.

6.2 Sensor Fusion

The sensor fusion process outlined in 5.2 proved to be successful. Containers could be placed in an area roughly bounded by a 30cmx40cm rectangle before limitations related to the robot's workspace were observed. The workspace limitation was partially resolved by orienting the camera at an angle above the container for the localization step, which significantly increased the area the container could be placed because the minimum distance is bounded by the LMI Gocator's working distance. The detection confidence threshold was set to 50% to ensure a container could be detected from multiple perspectives with temporal stability while driving the robot, which lead to multiple false positive detections annotating workers' feet, chair legs, or even the table as "containers". This issue was solved by sorting the detections by confidence and only using the mask with the greatest confidence.

For the damage detection step, the confidence threshold was also set to 50%. False positives detected around the room were negated by overlaying the "container" mask with the highest confidence and only drawing detections on areas under that mask. Detection performance was excellent overall, and the system was run multiple times with multiple containers with novel damage that was not included in the training process. The entire procedure from the moment the robot begins moving to the final comparison between Detectron2 and the LMI Gocator takes 30 seconds. Once the container is localized on the container, other images and scans can be collected without restarting the whole procedure, significantly improving damage detection. An example of an image detected with the system is shown in Figure 4.

This process does have a couple limitations compared to manual annotation. The field of view of the LMI Gocator sensor used is more narrow than the visual camera, resulting in the accuracy of deterministic detections decreasing significantly towards the sides of the container tangent to the view direction. Work is being performed to improve the working field of view by using a larger Gocator Snapshot sensor and integrating multiple viewpoints. Another limitation lies with the current algorithm used to determine if a container is dented. It is not unusual to find protruding divots caused by an internal item or container, which are of similar interest. This could be solved by comparing the scanned point cloud to a known "pristine" one from the same perspective to measure deviations.

7 Conclusion and Future Work

Our experiments using an unsupervised pre-processing technique yielded mixed results, particularly between the dent and scratch classes. The first experiment, which used an anomaly map for a fourth

(a) *(b)* *(c)*

Figure 4: Figure (a) shows the raw image collected from a color camera. Figure (b) shows the detections performed by Detectron2 (blue), false positives proposed by Detectron2 (manually recolored to orange) and a false negative annotated manually after the detection step (red). Figure (c) shows overlaid detection masks (white) generated by the Gocator sensor.

channel yielded better segmentation for 'Dent' instances and poorer segmentation for 'Scratch' instances. This may be for a few reasons. First, there are far more instances of dents than there are of scratches in the training set. Scratches could also be harder to discern due to their narrow figures. Second, because the fourth channel by default is interpreted as an alpha channel by Detectron2 Mask R-CNN implementation, it is entirely possible that the image was too faded to interpret without significant modification.

This work also investigated the value of fusing the data from a standard visual imaging camera with the additional information provided by a high resolution depth camera. A test fixture was constructed using a UR5e robotic manipulator to position a camera and LMI Gocator sensor in front of a metal nuclear storage container analog to overlay detections, showing value as a consistent method for validating damage detections from current machine learning models. Although the LMI Gocator sensor is likely not the best tool to observe nuclear storage containers in storage due to its limited working volume and large size, it does show promise as a tool for generating large datasets of annotated images with limited worker interaction to enhance the performance of current machine learning models.

7.1 Future Work

One path forward with this work could be to further constrain the use cases of anomaly map preprocessing. While we demonstrated mild performance increases for dent and residue segmentation with the addition of a fourth anomaly channel to the dataset, PaDiM was originally created for images of the same camera perspective. This dataset used mostly the same perspective with somewhat different angles, which may have contributed to error. This could potentially affect the initial feature extraction process PaDiM uses to compute the embedding vectors (meaning, the representation of what is typical). Using a dataset entirely from one perspective could alleviate the issue of high anomaly scores clustering around table top corners, for example. Integrating this capability into our in-house ML tool could also be of use for pre-processing applications beyond Slip Lid containers.

Another promising path forward may be to use the segmentation and masking capabilities from PaDiM for detection. Because PaDiM computes anomaly maps, a threshold can be set as a metric for contouring and masking. Should the threshold be optimized, the model could perform segmentation tasks similarly to supervised CNN architectures such as the Detectron2 Mask R-CNN implementation. This means that it potentially it could not only perform segmentation faster, but segment instances of damage that are novel to a supervised model. This could be applicable for hybrid or semi-supervised approaches to segmenting all damage on containers. A supervised approach could, for example, make inferences on all types of damage known to it, while an unsupervised approach could segment anomalous instances of damage, deferring to the supervised model if there is too much overlap.

Finally, one of the most promising paths forward from this work is the development of a more closely integrated robotic system for damage detection and data generation. The test setup utilized a UR5e robotic manipulator with an inspection table designed for the smaller UR3e model, but the software tools and sensor package developed for the bench-top proof-of-concept will transfer quickly to the other robot. Other sensors could be integrated to expand the imaging capabilities to include the inside of containers for annual inspection activities.

References

[1] J. Shaw. *DOE M 441.1-1 Chg 1 (Admin Chg), Nuclear Material Packaging*. U.S. Department of Energy, 2016.

[2] K.P. Reeves et al. *Surveillance Report on SAVY-4000 and Hagan Nuclear Material Storage Containers for FY 2017 (LA-UR-17-31263)*. Los Alamos National Laboratory, 2017.

[3] T. Karns et al. *Surveillance Report on SAVY 4000 and Hagan Nuclear Material Storage Containers for Fiscal Year 2018 (LA-UR-18-31604)*. Los Alamos National Laboratory, 2018.

[4] T. Karns et al. *Surveillance Report on SAVY 4000 and Hagan Nuclear Material Storage Containers Update for Fiscal Year 2019 (LA-UR-19-32444)*. Los Alamos National Laboratory, 2019.

[5] T. Karns et al. *Surveillance Report on SAVY 4000 and Hagan Nuclear Material Storage Containers Update for Fiscal Year 2020 (LA-UR-21-20040)*. Los Alamos National Laboratory, 2021.

[6] Steven Douglas Lukow et al. "Advancing Vision-based Feedback and Convolutional Neural Networks for Visual Outlier Detection". In: (Sept. 2022). DOI: 10.2172/1889960. URL: https://www.osti.gov/biblio/1889960.

[7] Steven Lukow et al. *A Demonstration of Intelligent Container Surveillance Using Stationary and Mobile Camera Platforms*. Accessed 19-Aug-2022. 2022.

[8] Thomas Defard et al. "PaDiM: a Patch Distribution Modeling Framework for Anomaly Detection and Localization". In: *CoRR* abs/2011.08785 (2020). arXiv: 2011.08785. URL: https://arxiv.org/abs/2011.08785.