Radiation Anomaly Detection Using an Adversarial Autoencoder

Charles Sayre, Eric C. Larson
Department of Computer Science
Southern Methodist University
Dallas, Texas
{sayrec, eclarson}@smu.edu

Gabs DiLiegro, Joseph Camp
Department of Electrical and Computer Engineering
Southern Methodist University
Dallas, Texas
{gdiliegro, camp}@smu.edu

Bruce Gnade
Department of Materials Science and Engineering
Univertsity of Texas
Dallas, Texas
gnade@utdallas.edu

***Abstract*— Scintillators are the primary devices used for radiation detection, especially at national borders and other ports of entry. Reducing the size of these detectors for placement on mobile devices such as drones can allow for better detection and localization of radiation sources. Detection of radiation with supervised machine learning can be a challenge when looking for previously unobserved radiation sources. Therefore, anomaly detection methods are investigated. In this work we employ adversarial autoencoders trained to classify spectra from radioactive sources as either background or anomalous. This allows the model to detect anomalies regardless of radiation source and outperform supervised methods on newly encountered sources.**

I. INTRODUCTION AND MOTIVATION

Radiation detection in high traffic areas such as ports and border crossings is a key area of national security according to the Department of Homeland Security (DHS). At present scintillators exist in large static forms at ports-of-entry [1] or as bulky handheld devices [2]. These scintillators are capable of providing the energy spectra of radiation in an area, which is important for characterizing sources of harmful radiation as different sources emit different energies. While there has been some effort in radiation detection using machine learning methods, these have generally only used supervised methods such as multilayer perceptrons [3], linear models, and support vector machines [4]. However, supervised learning methods require training examples of the radiation types they are expected to classify. This makes them inappropriate for applications that do not have calibration examples for training or in applications where exact types of radiation are not known. It is therefore desirable to create a radiation detection system that is small enough to mount to a mobile system and that utilizes the latest developments in anomaly detection to accurately detect radiation sources without specific calibration.

Some work has been carried out using unsupervised methods including the use of Autoencoders [5][6]. In this work, we build from and improve upon these previous works by proposing the use of an Adversarial Autoencoder (AAE)[7] in the detection of radioactive material. AAEs are state-of-the-art anomaly detection models across many fields ranging from machine fault detection [8] to video anomaly detection [9] to medical image detection [10].

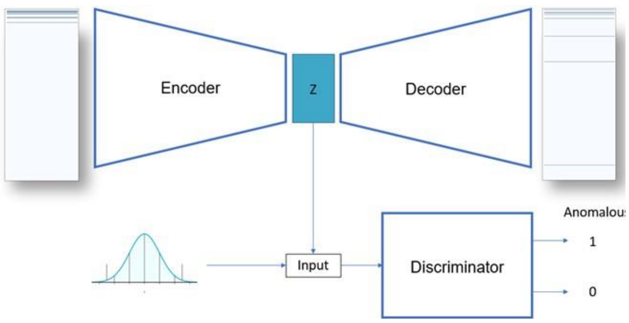
AAEs are well matched for the radiation detection problem as anomalies can come from a wide variety of radiation sources. While traditional supervised classes require positive and negative examples for training, AAEs are generative models that only require baseline observations. This is advantageous in radiation detection as it allows the model to be agnostic to any particular radiation source. In this work, we show that AAEs are excellent models for detecting anomalous radiation spectra using custom scintillators. These scintillators are based on CsI(TI), and therefore lower cost than many other currently available sensors. Scintillator detectors has lower resolution compared to solid state detectors, therefore this provides a challenging use case for our AAE models.

II. DATA

We collect data for use in the design and evaluation of our AAE models. Using our custom detectors, radiation spectra were collected from background and from three radioactive sources: Co^{60}, Mn^{54}, and Cs^{137}. Each source was profiled at three distances, 0 cm, 15 cm, and 30 cm for one hour per distance, with 3 hours of background radiation recorded. This data was further divided into 5-second time windows, yielding 2049, 2043, 2086, and 2064 windows collected for Co^{60}, Mn^{54}, Cs^{137}, and background, respectfully. We note that our AAE model does not require samples of radiation sources for training. However, we compare the AAE model to supervised methods that do require calibration examples. Hence, we split our data as follows: Each of the sources and the background were divided into 80-20 training-testing splits and were constructed into files that contained half radiative data (stratified across distances) and half background data. Training and testing data come from contiguous time segments in the data collection. Thus, a training set consisted of all radiation sources at varying intensity and background examples. During the data collection process, the facility was relatively uncontrolled, with personnel entering and leaving the area periodically. This ingress and egress did not appear to influence the collected data in any impactful way.

III. METHODOLOGY

FIGURE I. THE OVERALL ADVERSARIAL AUTOENCODER MODEL



A. *AAE Architecture*

The AAE model consists of three neural networks: Encoder, Decoder, and Discriminator, as shown in Figure I. The encoder acts to compress the input radiation spectra into a reduced dimensionality latent space, which the decoder uses to recover the original input. Like in other autoencoders the encoder-decoder pair train together to create a latent space of reduced dimensionality, that is also performant for reconstructing samples. The model is trained only on what is meant to be regular behavior, in our case background radiation. Mean Squared Error is used as the loss function for measuring the fidelity of reconstruction. Binary Cross Entropy is used to measure the correctness of the discriminator's prediction. Where AAEs differ from other autoencoders is in the use of a discriminator which is trained on the latent space representation of the input data and data that are generated from a normal distribution. The encoder is trained to "fool" the discriminator while the discriminator trains to differentiate the differences between encoded and generated data. In this way, the adversarial losses of the encoder and discriminator ensure that the latent space is sufficiently normally distributed. Therefore, points that are encoded into the latent space that depart from a normal distribution are considered anomalies.

B. *Testing Methods*

Testing falls into two main categories, comparative and real-time. Comparative results will focus on showcasing the viability of the AAE model at detecting radiation versus some baseline classifiers. Real-time results show the model working on a distributed system making predictions to illustrate the potential of this model working in an environment representative of the real world. With the goal of this project being to mount this onto drones it is vital to understand how the model performs detecting radiation in real time.

For our comparative tests the AAE is compared to six supervised training models: K-Nearest Neighbors (KNN), Support Vector Classifier (SVC), Random Forest Classifier (RFC), XGBoost Classifier (XGB) [11], CatBoost Classifier (CBC) [12], and Multilayer Perceptron (MLP). The baseline models were trained using only the Co^{60} training set which consists of 3291 samples (1640 with radiation present and 1651 background). This was done to better illustrate one of the advantages of using an unsupervised method like the AAE to accomplish radiation detection. Since the baseline models are only trained on one of the radiation sources, they may not generalize to the detection of different radiation samples.

Real-time results take a multitude of forms to try to recreate the applications that this system will have in the real-world. Firstly, we need to create a software system that can take readings directly from the detector, process it into the format required for the model, make a prediction, and then return that prediction. Secondly, since this system is going to be mounted to a series of drones, it needs to generate output that will be useful to the piloting of those drones. Thirdly, the system should be responsive enough for the done network to make decisions quickly. Finally, the system needs to have outputs that are stable and reproducible so that for a given output of the model we know that the environment it is in is similar to other environments for the same output.

IV. RESULTS

A. *Comparative Results*

We report the F-beta score, which is the weighted geometric mean of the precision and recall in Table I. A beta value of 2 was chosen (emphasizing recall over precision) as identifying possible threats is more important than correctly identifying non-threats in a defensive context. The results illustrate that a Supervised model trained only on Co^{60} can outperform the AAE on that same source. In this case MLP, SVC, and RFC are superior to AAE. However, the supervised models perform worse when tested on Mn^{54} and Cs^{137}, two sources that the models have never seen before.

TABLE I. F-BETA SCORE BY RADIOACTIVE SOURCE

Model	Cobalt-60	Manganese-54	Cesium-137
KNN	0.839637	0.646417	0.720604
SVC	0.97263	0.752661	0.810398
RFC	0.960784	0.680412	0.67596
XGB	0.871743	0.612565	0.632306
CBC	0.904573	0.649014	0.666842
MLP	**0.976563**	0.706605	0.775951
AAE	0.931085	**0.798507**	**0.83958**

Table II shows the accuracy of the models by the strength of the radioactive source, in terms of gamma rays per centimeter squared, which is the half-life adjusted kilo Becquerel count divided by the distance from the source squared. These show another advantage of the AAE model—it can more reliably detect a radioactive source when the source is very weak. This advantage is most prominent for radioactive sources upon which the supervised classifiers were not trained upon. The ultimate goal is to place these detectors into a mobile swarm that can work together to pinpoint the location of a potential radioactive source. Having a model that can more reliably detect the presence of radiation from a weaker source would be advantageous to the success of the process. One of the main drawbacks to using an AAE is that it requires large amounts of background data to reduce false positives. It has the least

TABLE II. ACCURACY BY RADIATIVE SOURCE AND STRENGTH, Γ-RAYS/CM^2

Model	**Cobalt-60**			**Manganese-54**			**Cesium-137**			**Background**
γ-rays ***/cm^2***	***6.56 x 10^4***	***23.2***	***5.8***	***1.66 x 10^4***	***5.9***	***1.5***	***3.62 x 10^4***	***12.8***	***3.2***	***0***
KNN	1.000	0.993	0.439	1.000	0.727	0.050	1.000	0.893	0.144	0.990
SVC	1.000	0.993	0.906	1.000	**0.986**	0.137	1.000	**0.943**	0.388	1.000
RFC	1.000	0.993	0.835	1.000	0.835	0.072	0.985	0.879	0.043	1.000
XGB	0.993	0.993	0.554	1.000	0.604	0.065	0.985	0.700	0.072	1.000
CBC	0.993	0.993	0.669	1.000	0.755	0.029	0.985	0.857	0.022	1.000
MLP	1.000	0.993	**0.921**	1.000	0.842	0.151	1.000	0.936	0.281	0.998
AAE	1.000	0.993	0.784	1.000	0.899	**0.396**	0.985	0.893	**0.583**	0.877

accuracy on background radiation classification. This is likely because the model is trained to create a latent space of the training data so when the testing data has even subtle differences from baseline, the model might classify it incorrectly. The solution to this problem is simply more examples of background spectra for the AAE to train on. At present, it only has roughly half of the training data that the other models have, which was done for the robustness of the experiments, but more varied samples of background radiation will only improve the model's classification. Moreover, we note that averaging the baseline samples over time is effective at reducing false positives because the detections tend to be spurious, not sustained over time.

B. *Real-time Results*

Ultimately, we intend to create a fully automated system that can detect and locate a radioactive source by a network of mobile agents in a reliable manner. As such, it then becomes a requirement to build out and test the basic infrastructure of that network. For the case of this paper the model was deployed onto an Intel UP Xtreme Series computer where it was loaded in a single instance on a locally hosted API. API calls were made to it via another instance on the same board that was handling the processing of data from the detector. The decision to put the model into an API instance comes down to two factors. First, if the model needs to run on the board located on the mobile device, we wanted to ensure that the model is only loaded once even if the system has access to multiple scintillators on the device. Secondly, it may become necessary to run the API on a basestation that is also controlling the agents remotely.

The first real-time test was done to show that there is a link between the distance from the source to the detector and the error value produced by the reconstruction of the spectra. Figure II shows an experiment where a source of Cs^{137} was moved 2.5 centimeters away from the detector every 10 minutes. As expected, the reconstruction error decreased as the source moved farther away from the detector. This is intuitive from a physics perspective and a modeling perspective. A source that is farther away should result in fewer gamma rays contacting the detector and because of that the closer the resulting spectrum should be to a background spectrum, where there is a minimal amount of gamma rays contacting the detector. Additionally, we see in the probability output of the discriminator that there is a decrease in accuracy as the source moves away. While the output of the discriminator becomes progressively worse, it is worth noting that when presented with a background environment with no radioactive sources nearby it rarely identifies a false positive. This in and of itself could be

FIGURE II. ERROR OF RECONSTRUCTION AND PROBABILTY OF RADIATION OF SOURCE MOVING 2.5CM EVERY 10 MINUTES

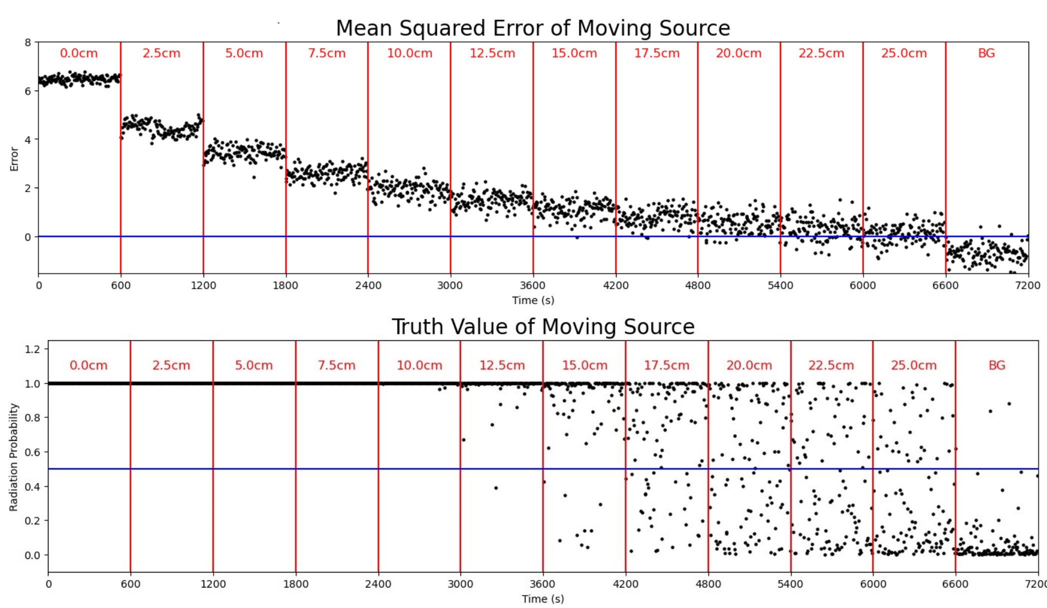
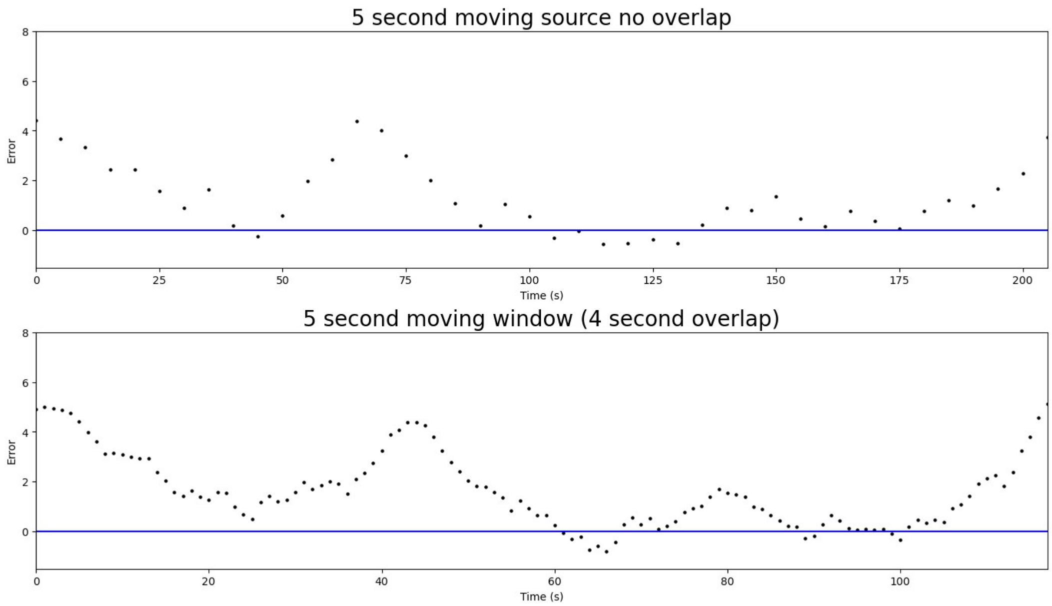


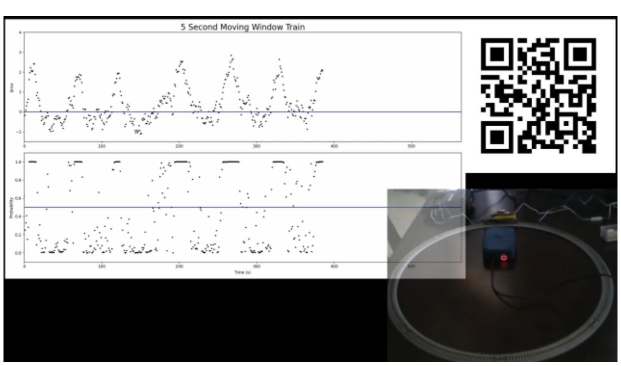
FIGURE III. EFFECTS OF DATA OVERLAP IN RADIATION DETECTION



useful in an operational environment, as if the system ever identifies radiation, it is likely that there is a radioactive source in the vicinity. Therefore, the outputs of both the discriminator and the reconstruction can be used in tandem to help identify the presence of radiation as well as the relative position of the source to the detector.

The second real-time test was determining a sufficient time resolution to make predictions on. All of the previous comparative and real-time tests were conducted on 5-second aggregated channel counts. However, because the use case of this model requires it to be mounted on a mobile agent that may be moving at a significant speed, we needed to create a way to make predictions in a significantly shorter amount of time. The method proposed here is to use an overlapping window such that a new prediction is made every second. To accommodate this, after creating a 5-second buffer and making a prediction, the oldest second of data is removed and the newest is added to make the next prediction. Figure III shows the results of this process. For both tests a source was moved along a straight line with the closest distance being 0 cm, representative of the highest values of reconstruction error, and 30 cm being the furthest, corresponding to the lowest reconstruction error. While for each test the sources were moved in a slightly different speed, the general pattern of the movement was maintained to the best of our abilities. As is shown, using a 4 second overlap allows us to have a better resolution on the movement of the source as it moves near the detector. Being able to make predictions using this overlapping methodology may also allow for the system to make predictions in an environment that is rapidly changing.

FIGURE IV. FRAME FROM LIVE DEMONSTRATION ON A MODEL TRAIN



The final real-time test that we wanted to create is an experiment to show the model predicting on a consistent series of values. We utilized a small locomotive, attached a radioactive source to it, and then placed the detector offset from the center of a circular track so that the source would move closer to and farther away from the detector in a regular pattern. The track was approximately 50 cm in diameter and the detector was offset such that the closest point between it and the track was around 10 cm. In Figure IV a sample still frame from the recorded video of the live test is shown. In this frame you can clearly see the setup described above as well as the consistent output we expected to see as a result of the source moving in a regular pattern around the track. As the source moves closer to the detector there are clear peaks with error values around 2.5 occurring every 50 to 70 seconds. Also included in Figure IV is a link to the original video of the live demonstration. This video is of the locomotive running for 10 minutes around the track and is sped up to four times real speed for the sake of brevity. The AAE model was not trained with background from that specific geographic location, not did it have access to training samples of the radiation type. Thus, the results from this test help show the efficacy of our model and the repeatability of the model when given similar inputs. This gives us confidence that the model will be able to work in real-world conditions.

V. CONCLUSIONS

Unsupervised anomaly detection is a well-matched technique for radiation detection as there are many varied sources of radiation with numerous different signatures. The AAE trained in this project leverages this by learning to create a normally distributed latent space of background radiation and marking radiation as anomalous because it lies outside of that distribution. Through this process we have shown that the AAE is a viable alternative to supervised models, especially in situations where a supervised model is tested on radiation sources that that are previously unseen. Additionally, training the AAE is simpler than training a supervised model because obtaining radioactive sources can be challenging. Because the end goal of this project is to mount this on a mobile agent swarm, we also needed to show that the system can quickly and accurately make detections while a source is moving relative to a detector. We have shown that the reconstruction and discriminator can be used in tandem to detect a source and its relative position. Also, we can augment the input data to increase the temporal resolution of the output without sacrificing that accuracy. Finally, the model, when presented with similar environments, does produce consistent results as illustrated in the test with the model locomotive. Ultimately, this work shows that AAEs are well matched to the problem of radiation anomaly detection.

VI. FUTURE WORK

In the immediate future we hope to be able to work with new and varied sources for detection. At present we have only been able to work with only the small single microcurie sources of Cobalt, Manganese, and Cesium which are not necessarily indicative of the types of radiation that may be encountered in the field. Various radioactive sources will need to be detected and can be more intense. Additionally, because the sources we have been working with are so small we are unable to quantify the model output in response to a strong source. Also, we want to test our AAE against other unsupervised classifiers. Finally, we need to create the software infrastructure to make a fully autonomous agent system. For now, we have only been able to run off of a single Intel UP Xtreme board which runs the API locally, however in the end project we will need to have multiple boards communicating to a base station that is running both the software to control the agents as well as the API for making radiation predictions.

ACKNOWLEDGMENT

The funding for this project is provided through the Department of Homeland Security via grant 18DNARI000029-05-00. The detectors were created and provided to us by Dr. Manuel Quevedo's research group at the University of Texas at Dallas as a part of this project.

REFERENCES

[1] Kouzes, R. T. (2010). Neutron and gamma ray detection for border security applications. *2010 1st International Nuclear & Renewable Energy Conference (INREC)*.

[2] Milbrath, B. D., Choate, B. J., Fast, J. E., Hensley, W. K., Kouzes, R. T., & Schweppe, J. E. (2006). Comparison of labr3:CE and nai(tl) scintillators for Radio-Isotope Identification Devices.

[3] Kangas, L. J., Keller, P. E., Siciliano, E. R., Kouzes, R. T., & Ely, J. H. (2008). The use of Artificial Neural Networks in PVT-based Radiation Portal Monitors. Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment, 587(2-3), 398–412.

[4] Morgan, D., Pilania, G., Couet, A., Uberuaga, B. P., Sun, C., & Li, J. (2022). Machine learning in nuclear materials research. Current Opinion in Solid State and Materials Science, 26(2), 100975.

[5] Ghawaly, James M. Jr, "A Datacentric Algorithm for Gamma-ray Radiation Anomaly Detection in Unknown Background Environments. " PhD dissertation, University of Tennessee, 2020. https://trace.tennessee.edu/utk_graddiss/6943

[6] Ghawaly, J. M., Nicholson, A. D., Archer, D. E., Willis, M. J., Garishvili, I., Longmire, B., Rowe, A. J., Stewart, I. R., & Cook, M. T. (2022). "Characterization of the autoencoder radiation anomaly detection (ARAD) model." Engineering Applications of Artificial Intelligence, 111, 104761. https://doi.org/10.1016/j.engappai.2022.104761

[7] Makhzani, Alireza, Jonathon Shlens, Navdeep Jaitly, Ian Goodfellow, and Brendan Frey. "Adversarial autoencoders." arXiv preprint arXiv:1511.05644 (2015).

[8] Wu, J., Zhao, Z., Sun, C., Yan, R., & Chen, X. (2020). Fault-attention generative probabilistic adversarial Autoencoder for Machine Anomaly Detection. IEEE Transactions on Industrial Informatics, 16(12), 7479–7488.

[9] Li, N., & Chang, F. (2019). Video anomaly detection and localization via multivariate gaussian fully convolution adversarial autoencoder. *Neurocomputing*, *369*, 92–105.

[10] Zhang, H., Guo, W., Zhang, S., Lu, H., & Zhao, X. (2022). Unsupervised deep anomaly detection for medical images using an improved adversarial autoencoder. *Journal of Digital Imaging*, *35*(2), 153–161.

[11] "Xgboost: A scalable tree boosting system." In Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining, pp. 785-794. 2016.

[12] Prokhorenkova, Liudmila, Gleb Gusev, Aleksandr Vorobev, Anna Veronika Dorogush, and Andrey Gulin. "CatBoost: unbiased boosting with categorical features." Advances in neural information processing systems 31 (2018).