# Multinomial Feature Matching for Out-of-Distribution Detection in Synthetic Aperture Radar

Christopher W. Pitts[a, b], Devin White[b], Trilce Estrada[a], and Gruia-Catalin Roman[a]

[a]University of New Mexico, 1155 University Blvd SE, Albuquerque, NM, USA
[b]Sandia National Laboratories, 1515 Eubank Blvd SE, Albuquerque, NM, USA

## ABSTRACT

Out-of-distribution (OOD) detection is an important part of automatic target recognition (ATR) systems. The capability to reject unknown classes improves reliability and trust in an ATR, and permits the use of otherwise closed-set classifiers where open-set recognition is necessary. In this paper we present multinomial feature matching (MFM), a method for detecting OOD data in the latent feature space of neural classifiers, and apply it to an EfficientNet-B7 model trained on the SAMPLE+ dataset. We show that MFM has efficient and low-overhead runtime characteristics, and that it exhibits a high level of performance when applied to OOD targets in SAMPLE+ and other SAR datasets, including both vehicle targets and clutter. MFM achieves a state-of-the-art area under the receiver operating characteristic (ROC) curve (AUROC) score and false positive rate (FPR) at a 95% true positive rate (TPR) (FPR@95) benchmark on the dataset, outperforming both mainstay benchmarks in OOD detection and contemporary work in the field.

**Keywords:** Machine learning, out-of-distribution detection, automatic target recognition, synthetic aperture radar

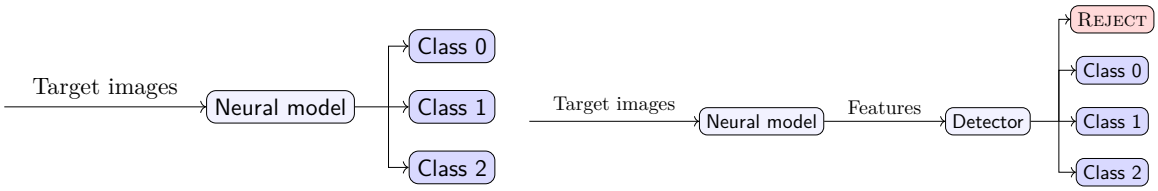
## 1. INTRODUCTION

Deep learning has dramatically expanded the field of automatic target recognition (ATR) research, with neural-based classifiers occupying an increasingly large share of publicly released algorithms.[1] However, while deep learning algorithms can achieve high performance on target classification, the majority of algorithms are *closed-set* (Figure 1a); that is, they will always classify a target as belonging to one of the training classes, even when the target is from a novel class not present in the training data. These unknown classes, known as *out-of-distribution (OOD)* classes, can achieve high confidence scores in deep learning models despite not being part of the training data.[2] Deployment of closed-set classifiers into an open-set environment can lead to undesired behavior in autonomous systems; the proliferation of computer vision and autonomous systems into areas as diverse as medical imaging,[3,4] satellite or airborne image processing,[5–9] and autonomous navigation[10–15] underscores the need for open-set recognition. ATR systems are particularly vulnerable to encountering unknown classes due to their use in a wide variety of application areas, necessitating open-set algorithms for target recognition.

To adapt closed-set classifiers to operate in an open-set context, and thus become suitable for use in ATR systems, OOD detection techniques are used to identify OOD data and reject it instead of allowing a label to be assigned to the unknown-class target (Figure 1b). While natively open-set classifiers exist,[15–20] the development of effective OOD detectors is a critical part of continued research into ATR that can take advantage of modern computing hardware and large-scale data processing. A significant body of research has been done on closed-set classification techniques, resulting in highly effective classifiers that can be applied in open-set contexts with the addition of an OOD detector. Enabling the use of a large set of industry-standard classifiers that can run on commodity and edge hardware is a strong motivator for the application of OOD techniques to ATR problems.

The remainder of this paper is structured as follows: Section 2 gives an overview of the field of OOD detection, with particular emphasis on methods that have been applied to synthetic aperture radar (SAR) data. Section 3 describes the design of the multinomial feature matching (MFM) detector, and how it can be applied to the features extracted by deep neural networks (DNNs). Section 4 introduces the neural model, dataset, and

Corresponding author: Christopher Pitts – cwpitts@unm.edu/cwpitts@sandia.gov




(a) Closed-set ATR operation (b) Open-set ATR operation

Figure 1: OOD detection converts closed-set classifiers into open-set with the addition of a detector module than can produce a rejection label in lieu of a class label from the training set.

evaluation of the proposed detector, and Section 5 analyzes the results and shows that MFM is a high-performing OOD detector, achieving higher performance than a suite of prior work on OOD detection, selected to include both common benchmark algorithms such as maximum softmax probability (MSP)[2] and recent work in the field such as Neural Collapse Inspired (NCI),[21] giving broad coverage of prior work in the field. Section 6 summarizes the research, experiments, and results, while laying out directions for future research.

## 2. RELATED WORK

OOD detection research is a broad field, and typically broken up into taxonomies to help categorize research. In this paper we adopt the taxonomy proposed by Lu *et al.*,[22] which breaks the OOD problem space into three main parts: training-driven, training-agnostic, and large pre-trained model-based. Training-driven approaches modify the training of a DNNs in order to improve a downstream OOD detector in the image processing pipeline. The pre-trained model approaches focus on either using a visual language model (VLM) for multimodal detection (e.g. by using a large language model (LLM) to produce feature embeddings for comparison to the classifier's embeddings, as proposed by Sun *et al.*[23]), or by taking a pre-trained classifier and fine-tuning it on the in-distribution (ID) dataset in question, often with the use of OOD data to improve downstream detection. The third branch of the taxonomy is of most interest to us, as our method falls under this category. Training-agnostic approaches to OOD detection assume the existence of a trained classification model, and focus on exploiting the structure of that model to perform OOD detection. Approaches in this area are distinguished by which part of the classifier the detector uses. Briefly, they are:

**Output-based** methods, which use the final output of the classifier. MSP[2] is prominent example of this type of detector, along with Generalized ENtropy (GEN)[24] and MaxLogit.[25]

**Gradient-based** methods, which use model gradients to make predictions. Grad[26] and GradNorm[27] are examples of this type of detector.

**Density-based** methods, which estimate the true density of the ID distribution. Examples of this type of detector include Gaussian mixture based Energy Measurement (GEM)[28] and ConjNorm.[29]

**Feature-based** methods, which model ID classes using the activation vectors of features extraction layers. Out of DIstribution detector for Neural networks (ODIN)[30,31] and OpenMax[19] are examples of this type of detector, along with, Activation SHaping (ASH),[32] rectified activations (ReAct),[33] and various neural collapse-inspired approaches.[21,34,35]

**Distance-based** methods, which use feature embeddings in the latent space and employ distance metrics to compute OOD scores. Many variations on Mahalanobis distance have been used as OOD detectors.[4,36–38] Various k-nearest neighbors (KNN)-based approaches have also been used.[39,40]

Open-set techniques from all parts of the taxonomy have been applied to the ATR problem. Our work is firmly post-hoc, so our analysis will focus on these methods. Conceptually, our method (MFM) computes statistics over feature embeddings, although instead of a distance metric (such as Mahalanobis distance), we compute a penalty score based on the position of the embedding vector relative to the ID classes. It is most similar to feature- and distance-based methods, and does not use gradients or model outputs for detection. We will focus on feature- and distance-based detectors for comparing MFM to prior work, but will include examples

from other post-hoc detectors for a broad comparison. Mahalanobis-based methods have been used for SAR OOD detection,[41] as well as OpenMax,[42] MSP,[43] and ODIN.[41,43]

## 3. ALGORITHM DESIGN

Our approach to open-set recognition applies MFM to the feature embeddings extracted by a neural network trained to identify targets of interest. The embeddings are vectors of length $d$ that describe a position in the model's latent space. Because neural networks optimize to project different classes into distinct regions of the latent space,[44–46] there exists a class-specific characteristic set of features which will tend to have higher activations for the targets of that class than the other features in an embedding. While the exact numeric value of these class-characteristic features may vary, the general pattern will be for those features to have higher activations relative to the other features. This relative intensity can be captured by using a quantization function to assign each feature in the embedding to a quantile bin, resulting in a categorical representation of the embedding. Given the $N$ samples in a training set, each of which is represented as a $d$-dimensional feature embedding $z$ (so $|z| = d$), the probability of a particular feature appearing in each of $q$ quantiles can be empirically estimated. The likelihood of a given feature $z_i$ obtaining a specific quantile $q$, assuming the overall probability vector $\vec{p}$ is known can be written as in Equation 1.

$$Pr(z_i = q'|\vec{p}) = p_{q'} \tag{1}$$

Because the true value of $\vec{p}$ is *not* known, we approximate it over the $N$ images in the training set by computing a vector $\vec{g}$ with the number of times each quantile was obtained and normalizing it to obtain a probability distribution, as shown in Equation 2

$$\hat{p} = \frac{\sum \vec{g}}{N} \approx \vec{p} \tag{2}$$

The conjugate prior of a categorical distribution is the Dirichlet distribution, and the quantile count vector $\vec{g}$ can be used as the $\alpha$ parameter for a empirically estimated Dirichlet-Multinomial distribution. Modeling the problem with the assumption that it is distributed Dirichlet allows for a closed-form solution to the expected value and the variance of the distribution, permitting a closed-form computation of normalization parameters that allow for a one-sided $t$-test at inference. While this is done in multinomial pattern matching (MPM),[16–18] that algorithm has different goals, and is designed to operate on a per-pixel basis over raw SAR images, without any neural component. For our algorithm, computing $\hat{p}$ suffices, and the assumption of being distributed Dirichlet allows us to compute online updates as the dataset is processed. The computation of $\hat{p}$ for each feature in an embedding $\vec{z}$ results in a set of probability distributions that, taken together, form a statistical model of the relative intensity of a particular target's features in latent space. We refer to this collection of distributions as a *class template*.

The process for computing class-wise templates given a particular feature layer in the network and $N$ training images is given in Algorithm 1.[47] Features are first extracted from some layer $l_j$ in the model and reduced to an embedding by taking the mean of each feature map. These embeddings are then quantized, and for each feature in the embedding the count of the quantile assignment is incremented. Once all target instances have been processed, the counts are divided by the number of target images to produce a quantile distribution over the $q$ quantiles.

In this algorithm each entry in the $T_c$ data structure is a $\vec{\hat{p}}$ corresponding to a particular feature. This operation is repeated for each ID class $c$ in a training set $C$, yielding a set of $|C|$ templates. The templates will each have the shape $d \times q$, where $d$ is the number of features output by the layer chosen for feature extraction. The full structure will thus have the shape $|C| \times d \times q$.

The selection of the reduction operation $f_r$ and quantization operation $f_q$ are left as design choice to the implementer, with the constraints that $f_r$ must take features output by $l_j$ and produce feature embeddings of shape $1 \times d$ and $f_q$ must take $1 \times d$ embeddings and return categorical representations of shape $1 \times d$, where each entry has been assigned a quantile. We set the reduction function $f_r$ to be the mean value across the feature

**Algorithm 1** Computation of an MFM class template

**Require:** $l_j : \mathbb{R}^{d_{j-1} \times h_{j-1} \times w_{j-1}} \mapsto \mathbb{R}^{d_j \times h_j \times w_j}$, the $j^{\text{th}}$ DNN layer
**Require:** $M_{j-1} \in \mathbb{R}^{N_c \times d_{j-1} \times h_{j-1} \times w_{j-1}}$, the dataset features from $l_{j-1}$ that correspond to the $N_c$ class images
**Require:** $q$, the number of quantiles to use
**Require:** $f_r : \mathbb{R}^{d \times h_j \times w_j} \mapsto \mathbb{R}^{1 \times d_j}$, a reduction function
**Require:** $f_q : \mathbb{R}^{1 \times d_j} \mapsto \mathbb{N}^{1 \times d_j}$, a quantization function
**Ensure:** $T_c$, a $d_j \times q$ class-conditional template

**procedure** ComputeTemplate($l_j$, $M_{j-1}$, $q$, $f_r$, $f_q$ )
  Initialize $T_c \leftarrow 0^{d \times q}$
  **for** $m_{j-1} \in M_{j-1}$ **do**
    $m_j \leftarrow l_j(m_{j-1})$
    $\vec{z} \leftarrow f_r(m_j)$
    $\vec{z_q} \leftarrow f_q(\vec{z})$
    **for** $i \leftarrow 0..d$ **do**
      $T_c[i][\vec{z_q}[i]] \leftarrow T_c[i][\vec{z_q}[i]] + 1$
    **end for**
  **end for**
  **for** $i \leftarrow 0..d$ **do**
    $T_c[i] = \frac{T_c[i]}{N_c}$
  **end for**
  **return** $T_c$
**end procedure**

maps and $f_q$ to be the linear quantization method as described by Hyndman *et al.*[48] Experimentally, we have found that the exact method of quantization does not produce significant results in the assignment of quantile values, only small differences in the thresholds.

At inference time, features $m_j$ are extracted from the target image by layer $l_j$, reduced to an embedding $\vec{z}$, and quantized as $\vec{z_q}$. For each of the entries in $z_q$, a penalty score is computed by applying a penalty function to the quantile assignment. These per-feature penalties are then summed to produce a per-embedding score as given in Equation 3, where $z_{q_i}$ is the $i^{\text{th}}$ quantized feature and is a scalar value that can be used to index $\hat{p}$.

$$s(z_q) = \sum_{i=0}^{d} P(\vec{\hat{p}}, z_{q_i}) \tag{3}$$

The selection of the scoring function $P$ is the final piece to complete the basic MFM detector. In this paper we consider two penalty functions: negative log-likelihood (NLL) and unlikelihood. The NLL score is computed as the NLL of the probability score obtained by indexing the probability vector $\vec{\hat{p}}$ (Equation 4), with a small numerical constant $\epsilon$ added for numerical stability.

$$P(\vec{\hat{p}}, z_{q_i}) = -\log(\vec{\hat{p}}[z_{q_i}] + \epsilon) \tag{4}$$

The unlikelihood score is computed as the complement of the probability score obtained from $\vec{\hat{p}}$ (Equation 5). This score indicates the unlikelihood of a true instance of the target obtaining a specific quantile.

$$P(\vec{\hat{p}}, z_{q_i}) = 1 - \vec{\hat{p}}[z_{q_i}] \tag{5}$$

In addition to the basic single-layer detector, it is possible to combine scores from different layers to form a multilayer ensemble detector. It has been observed in OOD detection research that different OOD classes may be more readily detectable at different layers in a network.[4,36] Drawing on this insight, an MFM detector can be computed for each layer in a network and combined into a meta-detector. In this detector, the scores from

each layer are combined as weighted sum, using weights found by logistic regression. Using a multilayer scoring method imposes a constraint on targets; a given instance must conform to the ID statistics at every layer in order to be classified as OOD. If an OOD target does in fact conform to the statistics at every layer, detection is rendered functionally impossible. This method requires a sample of OOD data to use, but such data is generally available for a specific mission or minable from the background of images.[49] The multilayer implementations of MFM can be referred to as MFM-LU and MFM-LN, indicating the unlikelihood and NLL penalty scores, respectively. In this paper we will evaluate both MFM-LU and MFM-LN.

For all variants of MFM, the penalty scores are computed using a fixed number of quantiles. Because of this, there is a finite number of possible penalty scores, and a penalty lookup table can be precomputed for a given probability vector, allowing for fast lookups of scores at inference time. We present the results of a timing test in Section 5 using precomputed penalty tables, and observe that this allows the multilayer variants to achieve competitive inference speed in the combined classifier/detector pipeline.

## 4. EXPERIMENTS

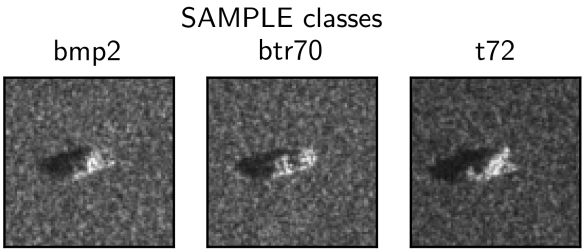
### 4.1 Datasets

The Synthetic and Measured Paired Labeled Experiment (SAMPLE)+ dataset[8] is a combination of the original SAMPLE[7] and Moving and Stationary TARgets (MSTAR)[50] datasets. Both datasets contain SAR data for a set of targets, with some overlap (Table 1). SAMPLE+ also includes clutter images that contain no targets, both as full scenes and a tiled set of chips. Such clutter images are useful for assessing the false alarm rate (FAR) of an ATR, and we do so in this work.

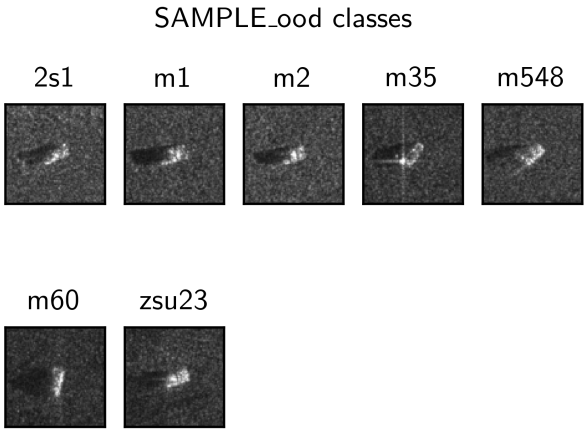
| Dataset | Targets |
|---|---|
| SAMPLE only | M1, M2, M35, M548, M60 |
| MSTAR only | BRDM2, BTR60, D7, T62, ZIL131, SLICY |
| Shared | 2S1, BTR70, ZSU23-4, BMP2, T72 |

Table 1: Available targets classes in the SAMPLE and MSTAR datasets

We use three SAMPLE classes for the ID set: T72, BMP2, and BTR70 (see Figure 2a). The other seven SAMPLE classes are used as an OOD set, which we refer to as SAMPLE-OOD (Figure 2b). Note here that the SLICY target is a non-vehicle target constructed to exhibit particular features in SAR images, and is not used.




(a) In-distribution targets




(b) Out-of-distribution targets

Figure 2: The ID/OOD target split for SAMPLE

In addition to SAMPLE-OOD, we use the following SAR datasets as OOD:

- Fudan University SAR (FUSAR)-Ship,[51] a maritime dataset that includes maritime target such as ships and port infrastructure
- SAR Aircraft Classification Dataset (SAR-ACD),[52] a SAR dataset with aviation targets
- SAR Detection (SARDet)-100k,[53] a large-scale SAR dataset with maritime targets, including ships, harbors, and ground vehicles such as tanks
- Random chips extracted from the SAMPLE clutter scenes
- The MSTAR-only targets

A summary of the OOD targets is given in Table 2.

| **Dataset** | **OOD Targets** |
|---|---|
| SAMPLE-OOD | M1, M2, M35, M548, M60, 2S1, ZSU23-4 |
| FUSAR-ship | BulkCarrier, CargoShip, ContainerShip, Dredger, Fishing, GeneralCargo, Tanker |
| SAR-ACD | A220, A320321, A330, ARJ21, Boeing737, Boeing787 |
| SARDet-100k | Ship, Aircraft, Car, Tank, Bridge, Harbor |
| SAMPLE-clutter | N/A |
| MSTAR | BRDM2, BTR60, D7, T62, ZIL131 |

Table 2: OOD targets in the datasets

Of these datasets, the SAMPLE-OOD is the "nearest"-OOD, given the sourcing of the targets from the same collection as SAMPLE. While there is no perfect method for measuring the "OOD-ness" of a dataset, work has been done on quantifying the distance between image distributions. In,[54] Heusel *et al.* introduced the Fréchet inception distance (FID), a method which computes the Fréchet distance[55] between datasets using the feature embeddings produced by a trained Inception-v3[56] classifier. We computed the mean Fréchet and cosine distances between datasets using the feature embeddings of the final layer of the EfficientNet-B7 classifier for the ID and OOD datasets, shown in Table 3.

| | SAMPLE | SAMPLE-OOD | SAMPLE-clutter | SAMPLE (test) | MSTAR-OOD | SARDet-100k | FUSAR-Ship | SAR-ACD |
|---|---|---|---|---|---|---|---|---|
| Fréchet | 0 | 2.19E5 | 2.94E5 | 1.6E5 | 2.93E8 | 1.77E11 | 3.23E11 | 1.57E5 |
| Cosine | 0 | 0.77 | 0.94 | 0.72 | 0.66 | 0.80 | 0.67 | 0.91 |

Table 3: The SAMPLE-derived ID/OOD sets are generally significantly closer to each other than the non-SAMPLE datasets.

In this table several interesting things can be seen. First, recall that the training set was defined as all of the synthetic data and 20% of the measured data for all ID targets. The ID test set, the remaining 80% of measured data, can be observed to have a similar average Fréchet distance from the training set as the SAMPLE-OOD dataset, which consists of all measured data for the OOD SAMPLE targets. This suggests that the detection of SAMPLE-OOD data will be difficult for a detector downstream from a classifier, with the caveat that a single distance metric represents an incomplete picture of the similarity between datasets. Second, the SAMPLE-clutter data, which consists of random clutter chips taken from the clutter scenes in SAMPLE+, is significantly further from the ID training set than either SAMPLE-OOD or the ID training set. Conversely, this suggests that an OOD detector should be readily able to reject clutter data as non-target. Finally, we observe that the non-SAMPLE datasets are generally significantly further away from the ID data that the OOD SAMPLE data (with the exception of the clutter chips), confirming the intuition that the SAMPLE-OOD detection task will be the most difficult for a detector.

### 4.2 Classifiers

For our classifier, we trained an EfficientNet-B7[57] model on the ID SAMPLE targets. We applied three data augmentations: color jittering, random erasing, and Gaussian noise. These three augmentations were found to improve accuracy on the SAMPLE dataset by Baffour *et al.*,[58] although we note that applying the transforms as described resulted in a model which converged on the training data but failed to converge on the measured testing data. We included 20% of the measured data with the synthetic images and found that they had a regularizing effect on the training, resulting in a model that performed well on both the mixed training dataset (94.29% accuracy) and the measured-only testing dataset (93.55% accuracy). We trained the model for 200 epochs using stochastic gradient descent (SGD) as the optimizer with an initial learning rate of 0.001 and cosine annealing as the learning rate scheduler, setting the $T_{max}$ parameter to 200.

### 4.3 Evaluation

To evaluate the performance of the MFM detector, we will compute three standard metrics: area under the receiver operating characteristic (ROC) curve (AUROC), area under the precision-recall curve (AUPR), and false positive rate (FPR) at a 95% true positive rate (TPR) (FPR@95). These metrics are standard evaluations in OOD research,[59] and are useful indicators of the performance of a detector. To compare our detector to other standard detectors, we will use PyTorch-OOD[60] implementations of the following detectors:

| **Post-hoc Family** | **Detectors** |
|---|---|
| Output-based | MSP,[2] MaxLogit,[25] Energy[61] |
| Gradient-based | GradNorm[27] |
| Feature-based | ODIN,[30] OpenMax,[19] ASH,[32] NCI[21] |
| Distance-based | Mahalanobis,[36] Relative Mahalanobis distance (RMD),[38] KNN[39] |

Table 4: Comparing against a broad subset of the extant literature enables a deep comparison

## 5. RESULTS

MFM achieved high performance on the SAMPLE/SAMPLE-OOD and SAMPLE/SAMPLE-clutter detection tasks. The unlikelihood penalty variant achieved an AUROC of 0.8834, with an FPR@95 of 0.5429 and an AUPR of 0.9887 on the SAMPLE-OOD task. The corresponding values for the clutter rejection task were 0.9929 (AUROC), 0.0204 (FPR@95), and 0.9843 (AUPR).

The NLL detector variant achieved an AUROC of 0.886, an FPR@95 of 0.1373, and an AUPR of 0.9781 on the SAMPLE/SAMPLE-OOD task. For clutter rejection the corresponding values were 0.9929 (AUROC), 0.0204 (FPR@95), and 0.9739 (AUPR).

Full results for all of the benchmark methods and the two MFM detectors across all datasets used in the experiments are given in Table 5.

Speed is an important consideration in OOD detection, particularly for ATRs. To assess the inference speed of a combined EfficientNet-B7 + MFM-LU pipeline, we timed each of the OOD detectors in processing the SAMPLE dataset, with a batch size of 100. As shown in Figure 5, MFM-LU is faster than many other detectors, falling behind only those which are essentially $\mathcal{O}(|C|)$, where $C$ is the set of ID classes. However, MFM-LU significantly outperforms the faster detectors in terms of raw AUROC score, making it the best-performing detector overall.

## 6. CONCLUSION

In this paper we applied our proposed MFM detector to the problem of OOD detection in neural classifiers to SAR data. We showed that MFM has low computational overhead, fast runtime, and achieves high-level performance on the SAMPLE/SAMPLE-OOD detection task, as well as achieving a high clutter rejection rate. MFM's strengths lie in the probability-based scoring and fast runtime, allowing for the provisioning of a downstream

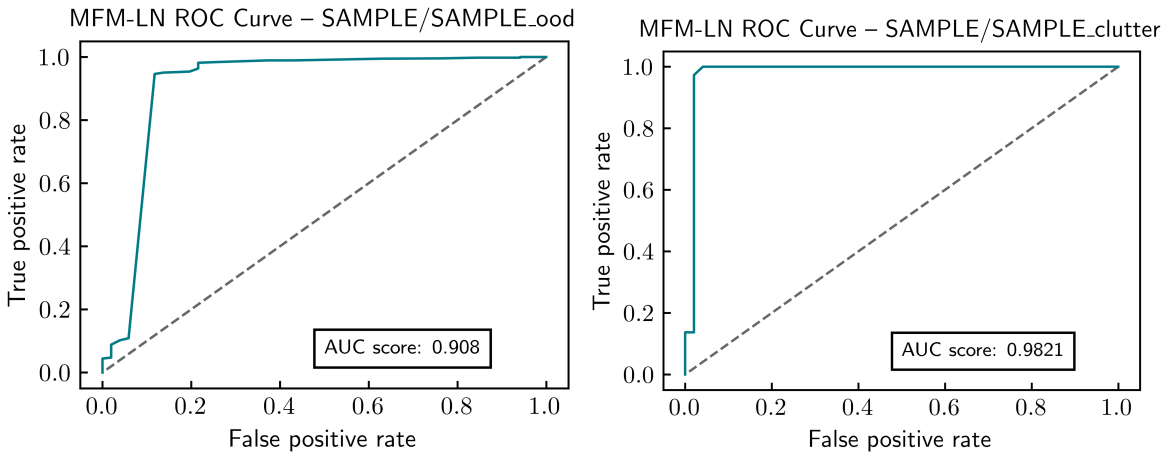



Figure 3: MFM-LN achieves a high level of performance on the SAMPLE-OOD and SAMPLE-clutter detection tasks

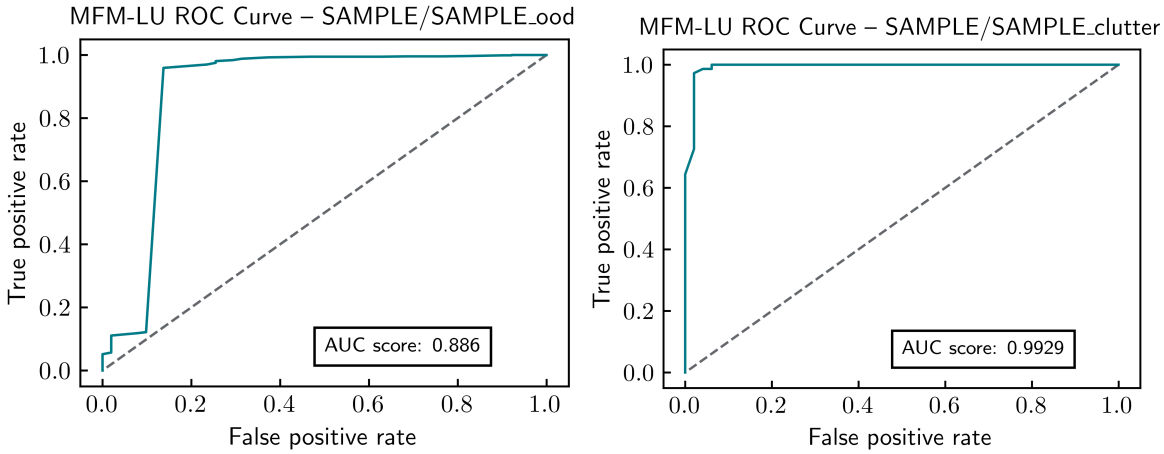



Figure 4: MFM-LU also achieves a high level of performance on the SAMPLE-OOD and SAMPLE-clutter detection tasks

consumer of detector decisions to have a probabilistic assessment of the correctness of the decision, which is valuable for consumers of ATR outputs.

Future work in the application of MFM to SAR will include the application of confidence-tuning methods to the detector scores, the optimization of quantile distributions to improve performance, and investigation into training-based methods to more strongly separate feature embeddings in the latent space. Applications to multimodal sensing will also be of interest, as both combined and modality-specific scores could be combined in an ensemble detector.

## 7. ACKNOWLEDGEMENTS

### 7.1 Funding

Part of this work was funded by Sandia National Laboratories through the University Part-Time Program.

### 7.2 NTESS External Communications Statement

Sandia National Laboratories is a multimission laboratory managed and operated by National Technology & Engineering Solutions of Sandia, LLC, a wholly owned subsidiary of Honeywell International Inc., for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-NA0003525.

| Detector | AUROC/FPR@95TPR/AUPR | | | | | |
|---|---|---|---|---|---|---|
| | SAMPLE-OOD | SAMPLE-clutter | MSTAR-OOD | FUSAR-Ship | SAR-ACD | SARDet-100k |
| ASH | 0.6506/0.8902/0.958 | 0.7084/0.561/0.9823 | 0.2878/1.0/0.7545 | 0.2351/1.0/0.7988 | 0.5447/0.9715/0.9991 | 0.456/1.0/0.9925 |
| Energy | 0.8129/0.8943/0.9681 | 0.939/0.0894/0.9858 | 0.2884/1.0/0.7732 | 0.2356/1.0/0.8026 | 0.5487/0.9715/0.999 | 0.4562/1.0/0.9927 |
| GradNorm | 0.4914/1.0/0.8182 | 0.4309/1.0/0.1667 | 0.5689/1.0/0.7987 | 0.5005/1.0/0.6822 | 0.4619/1.0/0.7584 | 0.5004/1.0/0.9575 |
| KNN | 0.6467/1.0/0.9005 | 0.9112/0.1016/0.6585 | 0.6889/1.0/0.9763 | 0.7604/1.0/0.9907 | 0.4916/1.0/0.968 | 0.5474/1.0/0.9976 |
| Mahalanobis | 0.4953/1.0/0.9036 | 0.6367/1.0/0.6952 | 0.6033/1.0/0.9712 | 0.6104/1.0/0.9885 | 0.5031/1.0/0.9698 | 0.5054/1.0/0.9963 |
| MaxLogit | 0.7809/0.8537/0.9685 | 0.9757/0.0163/0.9866 | 0.2878/1.0/0.7746 | 0.2356/1.0/0.8028 | 0.5729/0.9715/0.999 | 0.4562/1.0/0.9927 |
| MSP | 0.8608/0.626/0.9664 | 0.9885/0.0366/0.9869 | 0.4731/1.0/0.7968 | 0.4286/1.0/0.7921 | 0.9881/0.0488/0.9986 | 0.7685/1.0/0.9931 |
| NCI | 0.5334/0.9106/0.9529 | 0.5733/0.8333/0.9327 | 0.292/1.0/0.7289 | 0.2324/1.0/0.7876 | 0.5265/0.9756/0.9975 | 0.4551/1.0/0.9917 |
| ODIN | 0.5599/1.0/0.8816 | 0.9451/0.0976/0.8527 | 0.4934/1.0/0.6402 | 0.5275/1.0/0.7469 | 0.6894/1.0/0.9697 | 0.5682/1.0/0.9742 |
| OpenMax | 0.7028/0.6748/0.8752 | 0.7315/0.3049/0.388 | 0.8545/0.6382/0.9585 | 0.8659/0.313/0.9633 | 0.7396/0.3211/0.9134 | 0.7965/0.3211/0.9925 |
| RMD | 0.7421/0.7073/0.9142 | 0.9454/0.0732/0.7535 | 0.5852/1.0/0.9295 | 0.5896/1.0/0.9811 | 0.5031/1.0/0.9804 | 0.5077/1.0/0.9962 |
| **MFM-LN** | 0.908/0.1373/0.9763 | 0.9821/0.0204/0.9739 | 0.9115/0.4583/0.9929 | 0.999/0.0/1.0 | 1.0/0.0/1.0 | 0.9998/0.0/1.0 |
| **MFM-LU** | 0.886/0.1373/0.9781 | 0.9929/0.0204/0.9946 | 0.8902/0.4375/0.9907 | 0.9998/0.0/1.0 | 0.999/0.0/1.0 | 0.999/0.0/1.0 |
| **Avg. MFM** | 0.897/0.1373/0.9772 | 0.9875/0.0204/0.9843 | 0.9008/0.4479/0.9918 | 0.9994/0.0/1.0 | 0.9995/0.0./1.0 | 0.9994/0.0/1.0 |

Table 5: MFM outperforms numerous other detectors across a variety of datasets, including near-OOD like SAMPLE-OOD and far-OOD like SARDet-100k. Best values for each metric are highlighted in blue.

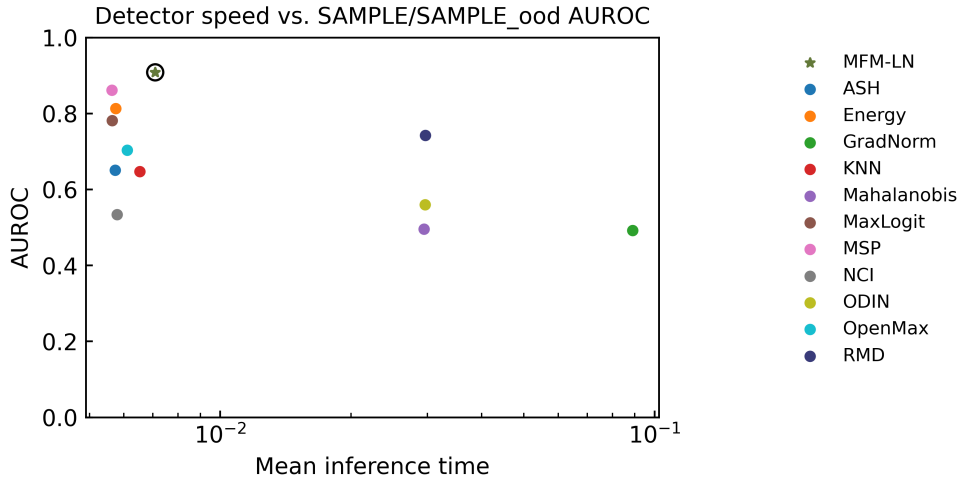



Figure 5: Multilayer MFM maintains a competitive speed while achieving a significantly higher AUROC score than all other detectors

### 7.3 NTESS Objectivity Statement

This paper describes objective technical results and analysis. Any subjective views or opinions that might be expressed in the paper do not necessarily represent the views of the U.S. Department of Energy or the United States Government.

## REFERENCES


[1] Li, J., Yu, Z., Yu, L., Cheng, P., Chen, J., and Chi, C., "A comprehensive survey on SAR ATR in deep-learning era," *Remote Sensing* **15**(5) (2023).

[2] Hendrycks, D. and Gimpel, K., "A baseline for detecting misclassified and out-of-distribution examples

in neural networks," in [*5th International Conference on Learning Representations, ICLR 2017, Toulon, France, April 24-26, 2017, Conference Track Proceedings*], OpenReview.net (2017).

[3] Cao, A., Klabjan, D., and Luo, Y., "Open-set recognition of breast cancer treatments," *Artificial Intelligence in Medicine* **135**, 102451 (January 2023).

[4] Anthony, H. and Kamnitsas, K., "On the use of mahalanobis distance for out-of-distribution detection with neural networks for medical imaging," in [*Uncertainty for Safe Utilization of Machine Learning in Medical Imaging*], Sudre, C. H., Baumgartner, C. F., Dalca, A., Mehta, R., Qin, C., and Wells, W. M., eds., 136–146, Springer Nature Switzerland, Cham (2023).

[5] Jiang, C., Zhang, H., Zhan, R., Shu, W., and Zhang, J., "Open-set recognition model for SAR target based on capsule network with the KLD," *Remote Sensing* **16**(17) (2024).

[6] Yang, J., Gu, J., Xin, J., Cong, Z., and Ding, D., "Liosr-sar: Lightweight open-set recognizer for sar imageries," *Remote Sensing* **16**(19) (2024).

[7] Lewis, B., Scarnati, T., Sudkamp, E., Nehrbass, J., Rosencrantz, S., and Zelnio, E., "A SAR dataset for ATR development: the synthetic and measured paired labeled experiment (sample)," in [*Algorithms for Synthetic Aperture Radar Imagery XXVI*], Zelnio, E. and Garber, F. D., eds., **10987**, 109870H, International Society for Optics and Photonics, SPIE (2019).

[8] Lewis, B., Ashby, M., and Zelnio, E., "SAMPLE with a side of MSTAR: extending SAMPLE with outliers and target variants from MSTAR," in [*Algorithms for Synthetic Aperture Radar Imagery XXX*], Zelnio, E. and Garber, F. D., eds., **12520**, 1252007, International Society for Optics and Photonics, SPIE (2023).

[9] Mao, S., Lu, S., Du, Z., Jiao, L., Gou, S., Mou, L., Lu, X., Xiong, L., and Zhang, Y., "Cross-rejective open-set sar image registration," in [*2025 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*], 23027–23036 (June 2025).

[10] Ren, X., Yang, T., Li, L. E., Alahi, A., and Chen, Q., "Safety-aware motion prediction with unseen vehicles for autonomous driving," in [*Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*], 15731–15740 (October 2021).

[11] Chitta, K., Prakash, A., and Geiger, A., "Neat: Neural attention fields for end-to-end autonomous driving," in [*Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*], 15793–15803 (October 2021).

[12] Azak, S., Bozkaya, F., Tı̆glıoğlu, S., Yusefi, A., and Durdu, A., "A unified monocular vision-based driving model for autonomous vehicles with multi-task capabilities," *IEEE Transactions on Intelligent Vehicles* **10**, 4397–4408 (Sep. 2025).

[13] Chen, M., Liu, Y., Zhang, Z., and Guo, W., "Rcrfnet: Enhancing object detection with self-supervised radar–camera fusion and open-set recognition," *Sensors* **24**(15) (2024).

[14] Zhao, F., Lou, W., Sun, Y., Zhang, Z., Ma, W., and Li, C., "Open set vehicle detection for uav-based images using an out-of-distribution detector," *Drones* **7**(7) (2023).

[15] Akhtarshenas, A. and Toosi, R., "An open-set framework for underwater image classification using autoencoders," *SN Applied Sciences* **4**, 229 (Jul 2022).

[16] Horvath, M. S. and Rigling, B. D., "Multinomial pattern matching revisited," in [*Algorithms for Synthetic Aperture Radar Imagery XXII*], Zelnio, E. and Garber, F. D., eds., **9475**, 94750H, International Society for Optics and Photonics, SPIE (2015).

[17] Koudelka, M. L., Richards, J. A., and Koch, M. W., "Multinomial pattern matching for high range resolution radar profiles," in [*Algorithms for Synthetic Aperture Radar Imagery XIV*], Zelnio, E. G. and Garber, F. D., eds., **6568**, 65680V, International Society for Optics and Photonics, SPIE (2007).

[18] Koch, M. and Malone, K., "A sequential vehicle classifier for infrared video using multinomial pattern matching," in [*2006 Conference on Computer Vision and Pattern Recognition Workshop (CVPRW'06)*], 127–127 (June 2006).

[19] Bendale, A. and Boult, T. E., "Towards open set deep networks," in [*2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*], 1563–1572 (June 2016).

[20] Lu, Y., Pei, J., Zhang, Y., Huo, W., Huang, Y., Wang, W., and Yang, J., "A sar open-set recognition method aided by hierarchically reconstructive latent representation learning," in [*IGARSS 2024 - 2024 IEEE International Geoscience and Remote Sensing Symposium*], 8966–8970 (July 2024).

[21] Liu, L. and Qin, Y., "Detecting out-of-distribution through the lens of neural collapse," in [*Proceedings of the Computer Vision and Pattern Recognition Conference (CVPR)*], 15424–15433 (June 2025).

[22] Lu, S., Wang, Y., Sheng, L., He, L., Zheng, A., and Liang, J., "Out-of-distribution detection: A task-oriented survey of recent advances," *ACM Comput. Surv.* **58** (Sept. 2025).

[23] Sun, H., He, R., Han, Z., Lin, Z., Gong, Y., and Yin, Y., "Clip-driven outliers synthesis for few-shot ood detection," (2024).

[24] Liu, X., Lochman, Y., and Zach, C., "Gen: Pushing the limits of softmax-based out-of-distribution detection," in [*2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*], 23946–23955 (June 2023).

[25] Hendrycks, D., Basart, S., Mazeika, M., Zou, A., Kwon, J., Mostajabi, M., Steinhardt, J., and Song, D., "Scaling out-of-distribution detection for real-world settings," in [*Proceedings of the 39th International Conference on Machine Learning*], Chaudhuri, K., Jegelka, S., Song, L., Szepesvari, C., Niu, G., and Sabato, S., eds., *Proceedings of Machine Learning Research* **162**, 8759–8773, PMLR (17–23 Jul 2022).

[26] Lee, J. and AlRegib, G., "Gradients as a measure of uncertainty in neural networks," in [*2020 IEEE International Conference on Image Processing (ICIP)*], 2416–2420 (Oct 2020).

[27] Huang, R., Geng, A., and Li, Y., "On the importance of gradients for detecting distributional shifts in the wild," in [*Advances in Neural Information Processing Systems*], Ranzato, M., Beygelzimer, A., Dauphin, Y., Liang, P., and Vaughan, J. W., eds., **34**, 677–689, Curran Associates, Inc. (2021).

[28] Morteza, P. and Li, Y., "Provable guarantees for understanding out-of-distribution detection," *Proceedings of the AAAI Conference on Artificial Intelligence* **36**, 7831–7840 (Jun. 2022).

[29] Peng, B., Luo, Y., Zhang, Y., Li, Y., and Fang, Z., "Conjnorm: Tractable density estimation for out-of-distribution detection," in [*The Twelfth International Conference on Learning Representations, ICLR 2024, Vienna, Austria, May 7-11, 2024*], OpenReview.net (2024).

[30] Lee, K., Lee, K., Lee, H., and Shin, J., "A simple unified framework for detecting out-of-distribution samples and adversarial attacks," (2018).

[31] Liang, S., Li, Y., and Srikant, R., "Enhancing the reliability of out-of-distribution image detection in neural networks," (2020).

[32] Djurisic, A., Bozanic, N., Ashok, A., and Liu, R., "Extremely simple activation shaping for out-of-distribution detection," in [*The Eleventh International Conference on Learning Representations*], (2023).

[33] Sun, Y., Guo, C., and Li, Y., "React: Out-of-distribution detection with rectified activations," in [*Advances in Neural Information Processing Systems*], Ranzato, M., Beygelzimer, A., Dauphin, Y., Liang, P., and Vaughan, J. W., eds., **34**, 144–157, Curran Associates, Inc. (2021).

[34] Ammar, M. B., Belkhir, N., Popescu, S., Manzanera, A., and Franchi, G., "NECO: NEural collapse based out-of-distribution detection," in [*The Twelfth International Conference on Learning Representations*], (2024).

[35] Zhang, J., Chen, Y., Jin, C., Zhu, L., and Gu, Y., "Epa: Neural collapse inspired robust out-of-distribution detector," in [*ICASSP 2024 - 2024 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*], 6515–6519 (April 2024).

[36] Lee, K., Lee, K., Lee, H., and Shin, J., "A simple unified framework for detecting out-of-distribution samples and adversarial attacks," (2018).

[37] Müller, M. and Hein, M., "Mahalanobis++: Improving OOD detection via feature normalization," in [*Forty-second International Conference on Machine Learning*], (2025).

[38] Ren, J., Fort, S., Liu, J., Roy, A. G., Padhy, S., and Lakshminarayanan, B., "A simple fix to mahalanobis distance for improving near-ood detection," (2021).

[39] Sun, Y., Ming, Y., Zhu, X., and Li, Y., "Out-of-distribution detection with deep nearest neighbors," in [*Proceedings of the 39th International Conference on Machine Learning*], Chaudhuri, K., Jegelka, S., Song, L., Szepesvari, C., Niu, G., and Sabato, S., eds., *Proceedings of Machine Learning Research* **162**, 20827–20840, PMLR (17–23 Jul 2022).

[40] Park, J., Chai, J. C. L., Yoon, J., and Teoh, A. B. J., "Understanding the feature norm for out-of-distribution detection," (2023).

[41] Geng, Z., Xu, Y., Wang, B.-N., Yu, X., Zhu, D.-Y., and Zhang, G., "Target recognition in sar images by deep learning with training data augmentation," *Sensors* **23**(2) (2023).

[42] Oveis, A. H., Giusti, E., Ghio, S., and Martorella, M., "Open set recognition in sar images using the openmax approach: Challenges and extension to boost the accuracy and robustness," in [*EUSAR 2022; 14th European Conference on Synthetic Aperture Radar*], 1–4 (July 2022).

[43] Pitts, C. and Hummel, M., "Application of ODIN to SAR Data, and Mitigation of an Unexpected Bimodality Effect," in [*2024 ATR Working Group (ATRWG) Conference*], (2024).

[44] Cohen, U., Chung, S., Lee, D. D., and Sompolinsky, H., "Separability and geometry of object manifolds in deep neural networks," *Nature Communications* **11**(746) (2020).

[45] Montúfar, G. F., Pascanu, R., Cho, K., and Bengio, Y., "On the number of linear regions of deep neural networks," *Advances in Neural Information Processing Systems (NeurIPS)* **27** (2014).

[46] Schilling, M., Willsch, D., Becker, S., Lück, A., and Nagler, J., "Quantifying the separability of data classes in neural networks," *Neural Networks* **139**, 278–293 (2021).

[47] Pitts, C., Estrada, T., and Roman, G.-C., "Multinomial feature matching for out-of-distribution detection in deep neural networks," *Submitted to Journal of Computer Vision and Image Understanding* **SomeVolume**(SomeNumber), Some–Pages (2026).

[48] Hyndman, R. J. and Fan, Y., "Sample quantiles in statistical packages," *The American Statistician* **50**(4), 361–365 (1996).

[49] Pei, S., "Image background serves as good proxy for out-of-distribution data," in [*The Twelfth International Conference on Learning Representations, ICLR 2024, Vienna, Austria, May 7-11, 2024*], OpenReview.net (2024).

[50] Keydel, E. R., Lee, S. W., and Moore, J. T., "Mstar extended operating conditions: A tutorial," in [*Algorithms for Synthetic Aperture Radar Imagery III*], *Proceedings of SPIE* **2757**, 228–242, SPIE (1996).

[51] Hou, X., Ao, W., Song, Q., Lai, J., Wang, H., and Xu, F., "Fusar-ship: building a high-resolution sar-ais matchup dataset of gaofen-3 for ship detection and recognition," *Science China Information Sciences* **63**, 140303 (mar 2020).

[52] Sun, X., Lv, Y., Wang, Z., and Fu, K., "Scan: Scattering characteristics analysis network for few-shot aircraft classification in high-resolution sar images," *IEEE Transactions on Geoscience and Remote Sensing* **60**, 1–17 (2022).

[53] Li, Y., Li, X., Li, W., Hou, Q., Liu, L., Cheng, M.-M., and Yang, J., "Sardet-100k: Towards open-source benchmark and toolkit for large-scale sar object detection," in [*The Thirty-eighth Annual Conference on Neural Information Processing Systems (NeurIPS)*], (2024).

[54] Heusel, M., Ramsauer, H., Unterthiner, T., Nessler, B., and Hochreiter, S., "Gans trained by a two time-scale update rule converge to a local nash equilibrium," in [*Advances in Neural Information Processing Systems*], Guyon, I., Luxburg, U. V., Bengio, S., Wallach, H., Fergus, R., Vishwanathan, S., and Garnett, R., eds., **30**, Curran Associates, Inc. (2017).

[55] Fréchet, M. M., "Sur quelques points du calcul fonctionnel," *Rendiconti del Circolo Matematico di Palermo (1884-1940)* **22**, 1–72 (Dec 1906).

[56] Szegedy, C., Vanhoucke, V., Ioffe, S., Shlens, J., and Wojna, Z., "Rethinking the inception architecture for computer vision," in [*2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*], 2818–2826 (June 2016).

[57] Tan, M. and Le, Q., "EfficientNet: Rethinking model scaling for convolutional neural networks," in [*Proceedings of the 36th International Conference on Machine Learning*], Chaudhuri, K. and Salakhutdinov, R., eds., *Proceedings of Machine Learning Research* **97**, 6105–6114, PMLR (09–15 Jun 2019).

[58] Baffour, A. A., Osei Agyemang, I., Adjei-Mensah, I., and Nuhoho, R. E., "Towards fully synthetic training: Exploring data augmentations for synthetic-to-measured sar in automatic target recognition," in [*2025 IEEE World AI IoT Congress (AIIoT)*], 0008–0017 (May 2025).

[59] Zhang, J., Yang, J., Wang, P., Wang, H., Lin, Y., Zhang, H., Sun, Y., Du, X., Li, Y., Liu, Z., Chen, Y., and Li, H., "Openood v1.5: Enhanced benchmark for out-of-distribution detection," *arXiv preprint arXiv:2306.09301* (2023).

[60] Kirchheim, K., Filax, M., and Ortmeier, F., "Pytorch-ood: A library for out-of-distribution detection based on pytorch," in [*Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*], 4351–4360 (2022). Library available via PyPI: https://pypi.org/project/pytorch-ood/.

[61] Liu, W., Wang, X., Owens, J., and Li, Y., "Energy-based out-of-distribution detection," in [*Advances in Neural Information Processing Systems*], Larochelle, H., Ranzato, M., Hadsell, R., Balcan, M., and Lin, H., eds., **33**, 21464–21475, Curran Associates, Inc. (2020).