Reconstruction-based Out-of-Distribution Detection for Short-Range FMCW Radar

Sabri Mustafa Kahya⋆   Muhammet Sami Yavuz⋆†   Eckehard Steinbach⋆

⋆Technical University of Munich, School of Computation, Information and Technology,
Department of Computer Engineering, Chair of Media Technology,
Munich Institute of Robotics and Machine Intelligence (MIRMI)
†Infineon Technologies AG

Abstract—Out-of-distribution (OOD) detection recently has drawn attention due to its critical role in the safe deployment of modern neural network architectures in real-world applications. The OOD detectors aim to distinguish samples that lie outside the training distribution in order to avoid the overconfident predictions of machine learning models on OOD data. Existing detectors, which mainly rely on the logit, intermediate feature space, softmax score, or reconstruction loss, manage to produce promising results. However, most of these methods are developed for the image domain. In this study, we propose a novel reconstruction-based OOD detector to operate on the radar domain. Our method exploits an autoencoder (AE) and its latent representation to detect the OOD samples. We propose two scores incorporating the patch-based reconstruction loss and the energy value calculated from the latent representations of each patch. We achieve an AUROC of 90.72% on our dataset collected by using 60 GHz short-range FMCW Radar. The experiments demonstrate that, in terms of AUROC and AUPR, our method outperforms the baseline (AE) and the other state-of-the-art methods. Also, thanks to its model size of 641 kB, our detector is suitable for embedded usage.

Index Terms—Out-of-distribution detection, 60 GHz FMCW radar, deep neural networks, autoencoders, reconstruction and energy scores

I. INTRODUCTION

In recent years, modern deep learning models have set the state-of-the-art (SOTA) in various applications. However, they tend to give high confidence to out-of-distribution (OOD) samples. For instance, during the inference of a classification task, an instance can be assigned to one of the classes with an overconfident prediction even though the sample does not belong to any class that appears in the training set. Since these models make closed-world assumptions, their predictions on real-world scenarios such as medical and autonomous driving applications may cause undesirable consequences. Hence, to mitigate the overconfident predictions on OOD data, various OOD detection studies [1]–[8] have been revealed. The same problem also exists for radar-based applications.

Radars, thanks to their insensitivity to illumination and environmental conditions like fog and rain, and their ability to preserve privacy, have attracted attention for several applications such as presence sensing [9], human activity classification [10], and gesture recognition [11]. However, none of the previous works consider how to act when a sample from a different distribution comes into evaluation. For instance, in [12], the authors aim to detect abnormal human activities like fainting and falling in a toilet cabin. However, they do not consider when a non-human moving object like a robot vacuum cleaner appears in the toilet. Typically, the system needs to say, "unknown object," but it says either normal or abnormal activity, which belongs only to human targets. This study aims to create an OOD detector that works with Range Doppler Images (RDIs) of 60 GHz L-shaped FMCW Radar. The OOD detector tries to detect moving objects other than a walking human.

An OOD detector works as follows; It first assigns a score to the test sample $\mathbf{x}$ by using a score function S. Then, using the validation data containing only in-distribution (ID) samples, a threshold τ is calculated. Finally, if $S(\mathbf{x}) < \tau$, the sample is treated as ID, otherwise as OOD. To determine the scores, some detectors rely on softmax probabilities [13], [14], some use the logits [15], and some exploit intermediate feature representations [8], [16]–[18]. In this work, we propose two scores: Patch-based reconstruction (PB-REC) and latent energy scores (PB-LSE). For PB-REC, we use a patch-based autoencoder architecture. For PB-LSE, on the other hand, we use the latent representation of the autoencoder by bringing it from a 1D vector ($\mathbb{R}^D$) to a scalar ($\mathbb{R}$). With the help of patch-based training (only with ID samples) and inference strategy, we focus on the local features of RDIs. Thus, the local features of an ID sample are better reconstructed, leading to better OOD detection during inference. Also, with the same idea, for a single RDI, we get as many latent codes as patches in inference time, which means that we get compressed but more detailed information from each patch. We combine the latent codes to get the final energy score (PB-LSE). This combination makes the assigned energy score more distinguishable for ID and OOD samples since we have more information about the data. The experiments in Section V also reflect the correctness of the proposed idea. Our key contributions are as follows:

- We propose a reconstruction-based OOD detector. Our detector uses two scoring functions that produce patch-based reconstruction and latent energy scores. We achieve an AUROC of 90.72% on our dataset, consisting of several ID and OOD samples.

- Compared to the baseline and the other SOTA methods, our method has superior results. Besides, when we only use the energy scores, we get an AUROC of 90.72%. Thus, we can differentiate between ID and OOD samples without needing the decoder, using only the encoder part of the network. The size of the encoder part is only 641 kB, making our detector very suitable for embedded usage.

II. RELATED WORK

The existing studies on OOD detection are mainly shaped by the image domain, especially for classification tasks. In this regard, as a baseline method, Hendrycks & Gimpel published their work [13], which is based on maximum softmax probability (MSP). Their main claim is that in the inference of deep neural networks, higher confidence values are given to ID samples compared to OOD samples. Therefore, with simple thresholding, the ID and OOD samples can be differentiated. The ODIN approach introduced in [14] improves [13] by introducing temperature scaling and input perturbations to increase the softmax scores of ID samples. This method is extended by introducing a model ensembling approach [19]. Mahalanobis distance based detector (MAHA) [16] aims to detect OOD samples using intermediate feature representations of DNNs. It creates a class conditional Gaussian distribution for each ID class and assigns an OOD score to the test sample using the Mahalanobis distance of the sample to each of the class conditional Gaussian distributions. It also uses layer ensembling and input perturbation like ODIN [14]. Similarly, [17], [18] utilize intermediate representations to detect OOD samples. Liu et al. [15] created an energy-based OOD detection method by applying the $LogSumExp(LSE)$ function on the logit layer and stated that ID samples have lower energy scores than OOD samples. The methods above work with any pre-trained model, but some of them require OOD samples to tune the hyper-parameters.

In the Outlier Exposure (OE) technique by Hendrycks et al. [20], a limited number of OOD samples are used for either fine-tuning or training from scratch with a new loss, which pushes the softmax probabilities of the OODs to the uniform distribution. In the inference step, they only use regular cross-entropy loss and propose a score-based OOD detector. Similar studies [21], [22] exploit OE while training their deep architectures with novel loss functions. In [23], a GAN architecture is used with a special loss function to generate artificial OOD samples to be used for OE. In the literature, many other studies [24], [25] utilize generative models like GANs and Normalizing Flows to detect OODs.

Some studies [14], [26] emphasize the importance of gradient information for the OOD detection task. For example, in GradNorm [26], the magnitudes of the gradients are utilized to distinguish between IDs and OODs. The gradients are calculated by backpropagating the KL divergence between the softmax output and uniform distribution. In this logic, the ID samples have higher magnitudes than OODs.

There are also some reconstruction-based OOD detection methods. For example, [27] uses AE, and as an OOD score, it combines the reconstruction loss with a Mahalanobis distance in the latent code. Some other studies [28], [29] benefit from Variational Autoencoder (VAE) and its latent representation to detect OODs.

Most of the works in this area are based on convolutional neural network (CNN) architectures. However, in recent years, some works [30]–[32] based on pre-trained transformers with attention mechanisms like VIT [33] and BERT [34] are revealed as OOD detectors, and they achieved SOTA results on some benchmark datasets.

In the radar domain, there are few OOD detection studies utilizing DNNs. One study [35] compares a few conventional and deep learning methods for low-resolution radar micro-Doppler signatures using a synthetically generated dataset. In [36], a hand gesture recognition system is proposed that leverages FMCW radar technology and includes the ability to detect OOD input data. [37] proposes a meta-reinforcement learning approach for robust radar tracking with OOD detection support.

III. SYSTEM DESIGN

A. Radar Configuration

Our sensor is based on Infineon's BGT60TR13C chipset with a 60 GHz L-Shaped FMCW radar. It consists of one transmit (Tx) and three receiver (Rx) antennas. The radar configuration is listed in Table I. The Tx antenna transmits N_c chirp signals, and the Rx antennas receive the reflected signals. The transmitted and reflected signals are mixed, and the mixture of the signals produces the intermediate frequency (IF). The final raw Analogue-to-Digital Converter (ADC) data is acquired by low-pass filtering and digitizing the IF signal. Each chirp has N_s samples, so with some rearranging, the dimensions of a frame are shaped as $N_{Rx} \times N_c \times N_s$, and the frame becomes ready for further digital signal processing.

B. Pre-processing

In our network, we use RDIs as inputs; therefore, we apply the following pre-processing steps to our raw data:
- **Range FFT**: We apply range FFT with Chebyshev window at 100 dB on fast time with mean removal to get one channel range data out of the three channels from three Rx antennas.
- **MTI**: Following, we apply simple frame-wise moving target identification (MTI) on range data to remove any static object in the field of view of the radar.
- **Doppler FFT**: We then perform doppler FFT with Chebyshev window at 100 dB on slow time for each sample of a chirp and have our final RDI with the dimension of 64×64.

IV. PROBLEM STATEMENT AND METHOD

In this work, we aim to have a system that can identify and appropriately respond to moving targets (OOD), except for a single walking human (ID) using the RDIs. Normally, this

TABLE I
FMCW RADAR CONFIGURATION PARAMETERS

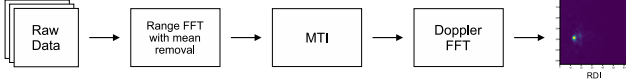| Configuration name	Symbol	Value
Number of Transmit Antennas	N_{Tx}	1
Number of Receive Antennas	N_{Rx}	3
Sampling frequency	f_s	2 MHz
Number of chirps per frame	N_c	64
Number of samples per chirp	N_s	128
Frame Period	T_f	50 ms
Chirp to chirp time	T_c	391.55 μs
Ramp start frequency	f_{min}	60.1 GHz
Ramp stop frequency	f_{max}	61.1 GHz
Bandwidth	B	1 GHz

Fig. 1. The diagram of the pre-processing steps to obtain RDIs

is a binary (0-1) classification task. The ID samples should be classified as one, while the OODs need to be labeled as zero. However, the OOD detection task differs from a typical binary classification application since there may be infinitely many OOD objects, which need to be correctly detected. That is why we test our pipeline with a large number of distinct OOD objects, which may appear with a high chance in indoor environments like offices and homes. We are also confident that our pipeline would perform well when tested with different OODs we do not use in our study.

A. Architecture and Training

For our reconstruction-based OOD detection pipeline, we use a patch-based AE architecture. The AE aims first to compress the input data to a latent space with smaller dimensions than the input sample, and from the latent space, it tries to reconstruct the input without losing information.

Our patch-based AE architecture consists of encoder and decoder parts. The encoder includes three 2D convolution layers with ReLu activation and a pooling layer after each convolution layer. Each convolution has a kernel size of 3×3. The first has 16 filters while the others have 32 and 64, respectively. With a flattening and a dense layer, the encoder brings the input to its latent space (128). The decoder, on the other hand, first takes the latent code and sequentially feeds it to dense and reshaping layers. Then it applies four transpose 2D convolutions with an up-sampling layer after the first three convolution layers. The first three convolution layers have ReLu activation, and a kernel size of 3×3. The first has 64 filters, while the others have 32 and 16. To reconstruct the input with the same dimensionality, the final convolution layer consists of a 3×3 kernel size with one filter, and sigmoid activation.

For the training, we only use ID samples. We divide each RDI (64,64,1) into four equally sized (32,32,1) patches. These patches are individually fed into our network while preserving

their original spatial arrangement within one compact, unified RDI representation. With this technique, the system focuses on the local features instead of learning the entire RDI once. Together with Adam optimizer, we use binary cross-entropy as the loss function and calculate it between the input RDI and the concatenated reconstructions.

B. OOD Detection

To detect the OOD data, we propose two scores consisting of reconstruction and latent energy scores. During inference, each instance in the RDI sequence is again divided into four patches, and each patch is fed to the pre-trained DNN model. The entire model gives the reconstruction results from each patch. We combine the reconstructed patches based on their position and calculate the mean of the reconstruction mean squared error (MSE). This score represents the reconstruction score (1). The encoder part of the model gives four latent codes coming from each patch. We first sum them and apply the LogSumExp (LSE) function to the summed latent code (1D vector) and calculate a latent energy score out of it (2). Then we simply use the scores to assess the performance of our method on different metrics. The overall pipeline is presented in Figure 2.

$$S_r(\mathbf{X}) = \frac{1}{q}\sum_{z=1}^{q}\frac{1}{hw}\sum_{i=1}^{h}\sum_{j=1}^{w}(D(E(\mathbf{P}_z))_{ij} - (\mathbf{P}_z)_{ij})^2 \quad (1)$$

$$S_e(\mathbf{X}) = \log\left[\sum_{j=1}^{k}\exp(\sum_{t=1}^{q}(\mathbf{z}_t(j)))\right] \quad (2)$$

where $\mathbf{X} \in \mathbb{R}^{64 \times 64}$ is the input RDI and consists of four patches $\{\mathbf{P}_1, \mathbf{P}_2, \mathbf{P}_3, \mathbf{P}_4\}$. E and D are encoder and decoder respectively. $\mathbf{z}_t(j)$ is the j-th element of $\mathbf{z} \in \mathbb{R}^k$ for patch $\mathbf{P}_t$. h and w are the dimensions of a patch (32,32), q is the number of patches (4), and k is the length of the latent vector (128).

Energy-based models (EBM) are used in the literature for OOD tasks. They mainly depend on a function that maps N_d dimensional data vectors to a scalar value called energy. [15] utilizes this idea by applying LSE on the logit layer, which involves class label information of a K-class classification task. We are inspired by [15] and apply LSE on the latent code that has all the information about the input in a compressed way. The main difference is that they have classification-based architectures and get the energy scores just before the softmax layer, which includes information from all classes, including the actual class the data belongs to. On the other hand, we use reconstruction-based architecture and have our energy scores from the latent representation of an AE. Meaning our energy scores come from the real data and only the actual class and are expected to be more informative. We expect lower energy scores for ID samples than for OOD samples, as well as reconstruction scores due to the fact that we train our network only with ID samples.

V. EXPERIMENTS

A. Dataset

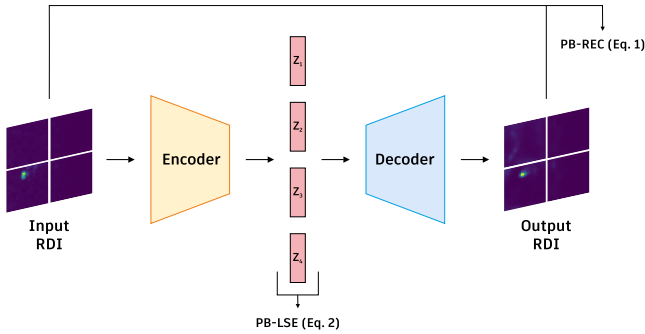The RF data is collected using Infineon's BGT60TR13C 60 GHz FMCW radar sensor with four individuals in 10 distinct

Fig. 2. The overall pipeline of the proposed method.

ID Activity	# of Training Samples	# of Test Samples
Walking	20000	10000
OOD Activity		
Table Fan	-	2000
RC Toy Car	-	2000
Swinging Laundry	-	2000
Robot Vacuum Cleaner	-	1500
Vacuum Cleaner	-	1500
Swinging Blinds	-	1000
Swinging Plants	-	1000
Swinging Paper Bag	-	1000
Total OOD	-	12000
Total	20000	22000

indoor places, including office and house rooms. The radar is placed at two meters height and tilted 15 degrees down from the front side. Our dataset consists of ID and OOD samples. ID samples comprise a human walking activity. For this scenario, the person in the radar field of view freely walks without supervision with a changing distance from 1 to 5m. The free walks contain any movement, such as forward-backward walks, from left to right and right to left, and circular and diagonal walks. On the other hand, OOD samples consist of various moving objects seen in indoor environments, including a table fan, a remote-controlled (RC) toy car, swinging plants, swinging blinds, swinging laundry, a swinging paper bag, a vacuum cleaner, and a robot vacuum cleaner. Each activity is recorded for around 1 to 2 minutes. After the pre-processing, we obtain a set of RDI data from the recordings of 10 separate rooms and use six rooms for the training and the remaining four for the test phase. For training, we use 20000 ID frames from six different rooms and four different individuals. These frames consist of a balanced number of individuals and rooms. For inference, we exploit many ID (10000) and OOD (12000) samples from four different rooms with a balanced number of individuals, other moving objects, and rooms. For similar radar-based applications, DL models tend to learn the room itself, so in some cases, an application may perform well in one room but poorly in different rooms. This means the system may overfit to the room. In our experiments, none of the rooms utilized during training appear in testing, and this reflects the robustness of our pipeline. The dataset details are provided in Table II.

B. Evaluation Metrics and Experimental Results

In OOD tasks, accuracy is not a reliable metric to be used, so researchers commonly evaluate their OOD detectors using the following metrics:

- **AUROC** is the area under the receiver operating characteristic (ROC) curve.
- **AUPR_IN** is defined as the area under the precision-recall curve when ID samples are assumed to be positives.
- **AUPR_OUT** is the area under the precision-recall curve when OOD samples are assumed to be positives.

Method	AUROC	AUPR_IN	AUPR_OUT
Baseline (AE)	75.44	78.07	76.80
MSP (LeNet5) [13]	88.66	81.03	93.26
Energy (LeNet5) [15]	88.85	82.49	93.31
ODIN (LeNet5) [14]	88.86	82.50	93.32
OE (LeNet5) [20]	50.34	32.31	67.80
MSP (ResNet-34) [13]	50.07	29.00	71.17
Energy (ResNet-34) [15]	68.09	64.90	74.29
ODIN (ResNet-34) [14]	68.68	65.35	74.67
OE (ResNet-34) [20]	17.38	20.61	49.69
MSP (ResNet-50) [13]	50.13	28.96	70.78
Energy (ResNet-50) [15]	89.82	79.48	94.36
ODIN (ResNet-50) [14]	89.86	79.51	**94.40**
OE (ResNet-50) [20]	7.99	18.89	47.21
PB-REC (Ours)	88.84	83.61	92.16
PB-LSE (Ours)	**90.72**	**87.67**	92.81

Table III demonstrates the superiority of the proposed method in terms of AUROC and AUPR_IN over the baseline and other SOTA methods. The baseline method (AE) uses the same architecture as our proposed methods, and instead of using patches, it directly uses the RDIs to train the AE. In the OOD detection phase, the baseline method only uses the reconstruction scores calculated from the mean of the reconstruction MSE in between input and reconstructed RDIs. In MSP, Energy, ODIN, and OE methods, LeNet5 [38], ResNet-34, and ResNet-50 [39] backbones are used. The architectures are trained in a one-class classification manner by only using ID samples, and the trained models are used to apply the SOTA methods.

VI. CONCLUSION

OOD detection has a crucial place in academia and industry since being able to detect unknown samples has great importance, especially for safety-critical applications like medical and autonomous driving. In the literature, there are several OOD detection methods, and they produce impressive results.

However, in the radar domain, this field is not studied broadly. Therefore, in this study, we addressed the OOD detection problem on top of a short-range FMCW radar. We propose two score values by using a reconstruction score from an autoencoder and a latent energy score from the latent representation of the same autoencoder to detect the OOD samples, which are all the moving objects in the field of view of the radar other than a walking person. With our approach, we reach an AUROC of 90.72%, which outperforms the baseline as well as other compared SOTA methods. Our experiments also demonstrate the effectiveness of patch-based strategy. Thanks to it, we get better reconstructions and more informative latent codes, resulting in more distinguishable reconstruction and energy scores for ID and OOD samples. Additionally, with its small size of 641 kB, our detector is appropriate for embedded devices.

REFERENCES

[1] J. Yang, K. Zhou, and Z. Liu, "Full-spectrum out-of-distribution detection," *arXiv*, 2022.

[2] H. Wang, Z. Li, L. Feng, and W. Zhang, "Vim: Out-of-distribution with virtual-logit matching," *arXiv*, 2022.

[3] X. Du, Z. Wang, M. Cai, and Y. Li, "Vos: Learning what you don't know by virtual outlier synthesis," *Proceedings of the International Conference on Learning Representations*, 2022.

[4] J. Yang, H. Wang, L. Feng, X. Yan, H. Zheng, W. Zhang, and Z. Liu, "Semantically coherent out-of-distribution detection," in *Proceedings of the IEEE International Conference on Computer Vision*, 2021.

[5] Y. Sun, C. Guo, and Y. Li, "React: Out-of-distribution detection with rectified activations," in *Advances in Neural Information Processing Systems*, 2021.

[6] R. Huang and Y. Li, "Mos: Towards scaling out-of-distribution detection for large semantic space," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2021.

[7] Y. C. Hsu, Y. Shen, H. Jin, and Z. Kira, "Generalized odin: Detecting out-of-distribution image without learning from out-of-distribution data," in *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 10948–10957, 2020.

[8] Y. Sun, Y. Ming, X. Zhu, and Y. Li, "Out-of-distribution detection with deep nearest neighbors," in *International Conference on Machine Learning (ICML)*, PMLR, 2022.

[9] N. Regev and D. Wulich, "Radar-based, simultaneous human presence detection and breathing rate estimation," *Sensors*, vol. 21, no. 10, 2021.

[10] P. Vaishnav and A. Santra, "Continuous human activity classification with unscented kalman filter tracking using fmcw radar," *IEEE Sensors Letters*, vol. 4, no. 5, pp. 1–4, 2020.

[11] S. Hazra and A. Santra, "Robust gesture recognition using millimetric-wave radar system," *IEEE Sensors Letters*, vol. 2, no. 4, pp. 1–4, 2018.

[12] W. Takabatake, K. Yamamoto, K. Toyoda, T. Ohtsuki, Y. Shibata, and A. Nagate, "Fmcw radar-based anomaly detection in toilet by supervised machine learning classifier," in *2019 IEEE Global Communications Conference (GLOBECOM)*, pp. 1–6, 2019.

[13] D. Hendrycks and K. Gimpel, "A baseline for detecting misclassified and out-of-distribution examples in neural networks," *Proceedings of International Conference on Learning Representations*, 2017.

[14] S. Liang, Y. Li, and R. Srikant, "Enhancing the reliability of out-of-distribution image detection in neural networks," in *International Conference on Learning Representations*, 2018.

[15] W. Liu, X. Wang, J. Owens, and Y. Li, "Energy-based out-of-distribution detection," in *Advances in Neural Information Processing Systems*, vol. 33, pp. 21464–21475, Curran Associates, Inc., 2020.

[16] K. Lee, K. Lee, H. Lee, and J. Shin, "A simple unified framework for detecting out-of-distribution samples and adversarial attacks," in *Proceedings of the 32nd International Conference on Neural Information Processing Systems*, NIPS'18, (Red Hook, NY, USA), p. 7167–7177, Curran Associates Inc., 2018.

[17] C. S. Sastry and S. Oore, "Detecting out-of-distribution examples with Gram matrices," in *Proceedings of the 37th International Conference on Machine Learning*, vol. 119 of *Proceedings of Machine Learning Research*, pp. 8491–8501, PMLR, 13–18 Jul 2020.

[18] H. Huang, Z. Li, L. Wang, S. Chen, B. Dong, and X. Zhou, "Feature space singularity for out-of-distribution detection," in *Proceedings of the Workshop on Artificial Intelligence Safety 2021 (SafeAI 2021)*, 2021.

[19] B. Lakshminarayanan, A. Pritzel, and C. Blundell, "Simple and scalable predictive uncertainty estimation using deep ensembles," in *Advances in Neural Information Processing Systems*, vol. 30, Curran Associates, Inc., 2017.

[20] D. Hendrycks, M. Mazeika, and T. Dietterich, "Deep anomaly detection with outlier exposure," in *International Conference on Learning Representations*, 2019.

[21] Q. Yu and K. Aizawa, "Unsupervised out-of-distribution detection by maximum classifier discrepancy," in *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, October 2019.

[22] A.-A. Papadopoulos, M. R. Rajati, N. Shaikh, and J. Wang, "Outlier exposure with confidence control for out-of-distribution detection," *Neurocomputing*, vol. 441, pp. 138–150, 2021.

[23] K. Lee, H. Lee, K. Lee, and J. Shin, "Training confidence-calibrated classifiers for detecting out-of-distribution samples," *arXiv*, 2017.

[24] X. Huang, Y. Li, O. Poursaeed, J. Hopcroft, and S. Belongie, "Stacked generative adversarial networks," in *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 1866–1875, 2017.

[25] P. Kirichenko, P. Izmailov, and A. G. Wilson, "Why normalizing flows fail to detect out-of-distribution data," in *Advances in Neural Information Processing Systems*, vol. 33, pp. 20578–20589, Curran Associates, Inc., 2020.

[26] R. Huang, A. Geng, and Y. Li, "On the importance of gradients for detecting distributional shifts in the wild," in *Advances in Neural Information Processing Systems*, 2021.

[27] T. Denouden, R. Salay, K. Czarnecki, V. Abdelzad, B. Phan, and S. Vernekar, "Improving reconstruction autoencoder out-of-distribution detection with mahalanobis distance," *arXiv*, 2018.

[28] Z. Xiao, Q. Yan, and Y. Amit, "Likelihood regret: An out-of-distribution detection score for variational auto-encoder," in *Advances in Neural Information Processing Systems*, vol. 33, pp. 20685–20696, Curran Associates, Inc., 2020.

[29] A. R. Venkatakrishnan, S. T. Kim, R. Eisawy, F. Pfister, and N. Navab, "Self-supervised out-of-distribution detection in brain ct scans," *arXiv*, 2020.

[30] D. Hendrycks, X. Liu, E. Wallace, A. Dziedzic, R. Krishnan, and D. Song, "Pretrained transformers improve out-of-distribution robustness," *arXiv*, 2020.

[31] S. Fort, J. Ren, and B. Lakshminarayanan, "Exploring the limits of out-of-distribution detection," *arXiv*, 2021.

[32] R. Koner, P. Sinhamahapatra, K. Roscher, S. Günnemann, and V. Tresp, "Oodformer: Out-of-distribution detection transformer," *arXiv*, 2021.

[33] A. Dosovitskiy, L. Beyer, A. Kolesnikov, D. Weissenborn, X. Zhai, T. Unterthiner, M. Dehghani, M. Minderer, G. Heigold, S. Gelly, J. Uszkoreit, and N. Houlsby, "An image is worth 16x16 words: Transformers for image recognition at scale," *arXiv*, 2020.

[34] M. Bauw, S. Velasco-Forero, J. Angulo, C. Adnet, and O. Airiau, "Near out-of-distribution detection for low-resolution radar micro-doppler signatures," 2022.

[35] M. Bauw, S. Velasco-Forero, J. Angulo, C. Adnet, and O. Airiau, "Near out-of-distribution detection for low-resolution radar micro-doppler signatures," *arXiv*, 2022.

[36] J.-W. Choi, C.-W. Park, and J.-H. Kim, "Fmcw radar-based real-time hand gesture recognition system capable of out-of-distribution detection," *IEEE Access*, vol. 10, pp. 87425–87434, 2022.

[37] J. Ott, L. Servadei, G. Mauro, T. Stadelmayer, A. Santra, and R. Wille, "Uncertainty-based meta-reinforcement learning for robust radar tracking," ICMLA, 2022.

[38] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," *Proceedings of the IEEE*, vol. 86, no. 11, pp. 2278–2324, 1998.

[39] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," *CoRR*, vol. abs/1512.03385, 2015.