

Mitigating SAR Out-of-distribution Overconfidence based on Evidential Uncertainty

Xiaoyan Zhou, Tao Tang, Zhongzhen Sun, Gangyao Kuang, Janne Heikkilä, and Li Liu

Abstract—Synthetic Aperture Radar (SAR) Automatic Target Recognition (ATR) is extensively applied in both military and civilian sectors. Nevertheless, test and training data distribution may differ in the open world. Therefore, SAR Out-of-Distribution (OOD) detection is important because it enhances the reliability and adaptability of SAR systems. However, most OOD detection models are based on maximum likelihood estimation and overlook the impact of data uncertainty, leading to overconfidence output for both in-distribution (ID) and OOD data. To address this issue, we consider the effect of data uncertainty on prediction probabilities, treating these probabilities as random variables and modeling them using Dirichlet distribution. Building on this, we propose an Evidential Uncertainty aware Mean Squared Error (UMSE) loss function to guide the model in learning highly distinguishable output between ID and OOD data. Furthermore, to comprehensively evaluate OOD detection performance, we have compiled and organized some publicly available data and constructed a new SAR OOD detection dataset named SAR-OOD. Experimental results on SAR-OOD demonstrate that the UMSE approach achieves state-of-the-art performance, reducing the average FPR95 by up to 64.8%. The code and data are available at: <https://github.com/Xiaoyan-Zhou/UMSE-SAR-OOD-Detection>.

Index Terms—SAR ATR, out-of-distribution detection, uncertainty estimation

I. INTRODUCTION

SYNTHETIC Aperture Radar (SAR) has emerged as a critical tool for earth observation, which can work independently of daylight and atmospheric conditions. This advantage renders SAR invaluable in both military and civilian applications. The evolution of SAR imagery resolution, combined with advancements in machine learning, has propelled deep learning-based SAR Automatic Target Recognition (ATR) to the forefront. SAR ATR plays a pivotal role in port surveillance, airport management, and military reconnaissance applications. However, the assumption that testing and training data share the same distribution may not always hold in the open world. When this assumption falters, the reliability of model prediction is compromised. Consequently, there is a pressing need to explore out-of-distribution (OOD) detection within SAR ATR.

This work was partially supported by The National Key Research and Development Program of China No. 2021YFB3100800, and the National Natural Science Foundation of China under Grant 62376283, 62022091, 62201588, Natural Science Foundation of Hunan Province No. 2021JJ30780. (Corresponding Author: Li Liu, email: liuli_nudt@nudt.edu.cn).

Xiaoyan Zhou, Tao Tang, Zhongzhen Sun, Gangyao Kuang, and Li Liu are with the College of Electronic Science and Technology, National University of Defense Technology, Changsha, Changsha, 410073, China.

Janne Heikkilä is with the Faculty of Information Technology and Electrical Engineering, University of Oulu, Oulu, Finland.

OOD detection aims to identify “in-distribution” (ID) target classes and reject OOD target categories, which is a sub-topic of trustworthy classification. Outlier exposure [1], a branch of OOD detection, involves training models with OOD data to refine ID/OOD differentiation. Despite its effectiveness, obtaining a comprehensive OOD dataset is challenging, especially within SAR imagery. Therefore, developing OOD score functions to identify OOD data through a deeper understanding of ID data distribution becomes paramount. Noteworthy approaches in this realm include Maximum Softmax Probability (MSP [2]), Maximum Logit (Maxlogit [3]), and Decoupled MaxLogit (DML [4]), each contributing uniquely to OOD detection efficacy. Subsequently, Inkawich *et al.* combined MSP with the Mahalanobis-distance detector [5] for enhanced OOD detection [6].

However, deep learning models suffer from an overconfidence problem, which produces predictions that are incorrect but are assigned high probabilities or confidence scores. To address this problem, Wei *et al.* propose LogitNorm loss, which uses an input-dependent temperature to mitigate the overconfidence of the deep model [7]. However, most existing OOD detection methods overlook high data uncertainty within SAR images, such as low signal-to-noise ratios, which compromise model output reliability. Consequently, overconfident outputs and high data uncertainty in SAR contexts affect the performance of OOD data detection. To counteract this, we advocate for introducing uncertainty estimation in SAR OOD detection by modeling the prediction probabilities with Dirichlet distribution, aiming to reduce overconfidence and improve detection performance across various OOD score functions. As a result, a deep learning model with uncertainty estimation will enhance feature-based, logit-based, and probability-based OOD score functions. Our contributions are summarized as follows:

- Utilizing some existing public datasets include MSTAR [8], SAMPLE [9], SAR-ACD [10], FUSAR-ship [11], we have constructed a new dataset for SAR OOD detection called SAR-OOD.
- We propose an Evidential Uncertainty aware Mean Squared Error (UMSE) loss function, which integrates evidential uncertainty with the MSE loss function. The model supervised by UMSE can generate discriminative and stable features and logit between ID and OOD data. UMSE becomes a state-of-the-art (SOTA) in SAR OOD detection compared with the most popular Cross-Entropy (CE) and LogitNorm loss functions.
- We highlight the critical role of uncertainty estimation in

SAR OOD detection.

II. METHODOLOGY

A. Problem Setting and Motivation

1) *Problem Setting*: OOD detection is a subclass of classification instead of detection that aims to discern test samples from a distribution distinct from the training set. We consider the training data distribution as $p_{\text{train}}(x, y)$ and the test data distribution as $p_{\text{test}}(x, y)$, where x presents the image and y presents the label of x . In OOD detection, $p_{\text{test}}(x, y) \neq p_{\text{train}}(x, y)$. Grounded by Bayes' theorem, where $p(x, y) = p(x)p(y|x) = p(y)p(x|y)$, OOD detection can be categorized into two situations. (1) Covariate shift: $p(x)$ changes, and $p(y|x)$ is fixed. (2): Label shift: $p(y)$ changes, and $p(x|y)$ is fixed. In this paper, we pay attention to the situation of label shift, which means that OOD samples should not have overlapping labels with training data. It is crucial to note that the training dataset often comprises multiple classes, and the process of OOD detection should maintain the capability to classify ID samples accurately.

2) *Motivation*: The motivation for integrating the UMSE approach into OOD detection stems from a crucial challenge faced by many SAR ATR methods: their loss functions driven by MLE, result in models suffering from overconfidence [7]. Such a scenario leads to a degradation in the performance of many OOD score functions. Additionally, SAR images suffer from high data uncertainty. For example, there is speckle noise in SAR images, which is irreducible because it is an inherent characteristic of SAR systems. This noise adversely impacts the target's logit and feature representation, leading to unreliable predictions.

Therefore, uncertainty estimation becomes paramount in the realm of SAR OOD detection. It offers a solution to mitigate the model's overconfidence and achieves more reliable predictions, especially when there is high data uncertainty. In the classification problem, the predicted probability satisfies the constraint that the sum is 1, and the probability obeys a multinomial distribution. If the probability of each category is regarded as a random variable, then their distribution is the Dirichlet distribution, which is also the conjugate prior of the multinomial distribution. Therefore, if the Dirichlet distribution is considered when designing the maximum a posteriori probability (MAP) based loss function, it will mitigate the overconfidence of the deep learning model from a Bayesian statistic perspective. Our motivation together with an example of a three-class classification task is illustrated in Fig. 1.

B. Evidential Uncertainty aware Mean Squared Error Loss

In Section II-A, we recognized the importance of considering uncertainty in SAR OOD detection tasks. However, most deep learning loss functions are based on maximum likelihood estimation (MLE) and do not have uncertainty estimation capabilities. Reviewing uncertainty estimation in classification problems, Sensoy *et al.*'s work on Evidential Deep Learning employs Dirichlet distribution for prediction modeling [12].

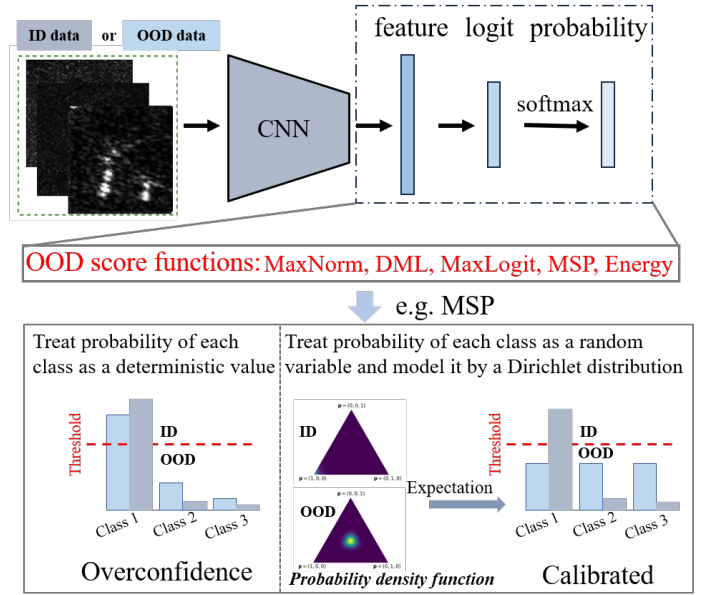


Fig. 1. Motivation of UMSE in SAR OOD detection. Overconfidence hampers the performance of OOD detection, while calibrating by uncertainty estimation can mitigate this problem. The expectation of the probability density function is the calibrated probability, which can be obtained through the Dirichlet parameter with a single pass.

This approach merges objective logic with Dirichlet distribution and calibrates probability through uncertainty estimation. Drawing inspiration from this, we introduce evidential uncertainty in SAR OOD detection and propose UMSE to alleviate the adverse effects of data uncertainty on feature space, logit space, and probability. Compared with EDL [12], UMSE removes the Kullback-Leibler (KL) divergence term. The KL divergence term causes the model to predict a uniform distribution in the face of uncertainty, which could lead to blurred decision boundaries between classes. This ambiguity might interfere with the model's ability to recognize OOD samples, as it produces the same output for label shift OOD and low-confidence ID samples.

For a K ($K \geq 2$) classification problem, the predicted probabilities of input x_i can be expressed as $\mathbf{p}_i = [p_{i1}, p_{i2}, \dots, p_{iK}]$, which adhere to the following constraint:

$$\sum_{k=1}^K p_{ik} = 1, p_{ik} \in [0, 1]. \quad (1)$$

We treat \mathbf{p}_i as a random variable following a Dirichlet distribution with parameters $[\alpha_{i1}, \dots, \alpha_{iK}]$. The probability density function is given by:

$$f(p_{i1}, \dots, p_{iK}; \alpha_{i1}, \dots, \alpha_{iK}) = \frac{1}{B(\boldsymbol{\alpha}_i)} \prod_{j=1}^K p_{ij}^{\alpha_{ij}-1}, \quad (2)$$

where $B(\boldsymbol{\alpha}_i)$ is the multivariate beta function, which can be expressed in terms of the gamma function:

$$B(\boldsymbol{\alpha}_i) = \frac{\prod_{j=1}^K \Gamma(\alpha_{ij})}{\Gamma(S_i)}, S_i = \sum_{j=1}^K \alpha_{ij}, \quad (3)$$

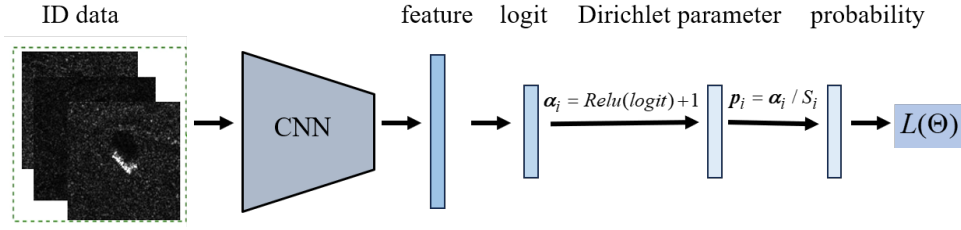


Fig. 2. Framework based on UMSE loss for model training. The model training process is equivalent to the Dirichlet parameter estimation.

where S_i denotes the Dirichlet strength. We treat the probability of each class as a random variable to address the impact of data uncertainty on prediction probability. This is achieved by constructing the UMSE loss function, which integrates the Dirichlet distribution into the traditional MSE loss function. This uncertainty-aware loss function assists in obtaining a calibrated logit and probability, enhancing the model's performance on the OOD detection task. The UMSE loss function can be expressed as:

$$\begin{aligned} L(\Theta) &= \int \|\mathbf{y}_i - \mathbf{p}_i\|_2^2 \frac{1}{B(\alpha_i)} \prod_{j=1}^K p_{ij}^{\alpha_{ij}-1} d\mathbf{p}_i \\ &= \sum_{j=1}^K \underbrace{(y_{ij} - \alpha_{ij}/S_i)^2}_{\mathcal{L}_{ij}^{\text{err}}} + \underbrace{\frac{\alpha_{ij}(S_i - \alpha_{ij})}{S_i^2(S_i + 1)}}_{\mathcal{L}_{ij}^{\text{var}}}. \end{aligned} \quad (4)$$

In equation (4), \mathbf{y}_i is a one-hot vector representing the ground truth label, and \mathbf{p}_i is an expectation of Dirichlet distribution, also called calibrated probability, predicted by the network for input \mathbf{x}_i :

$$\mathbf{p}_i = \alpha_i / S_i. \quad (5)$$

In deep learning, the Dirichlet parameter α_i is a vector and can be calculated from the model output logit:

$$\alpha_i = \text{Relu}(\text{logit}) + 1. \quad (6)$$

C. Model Framework based on UMSE Loss

The UMSE-based model maps the logit to the Dirichlet parameter α . Therefore, the model trained with UMSE is similar to that trained with CE. For the CE-based model, the probability can be obtained by softmax after logit, while the UMSE-based model calculates probability using the equation (5). It is a calibrated probability by uncertainty estimation. Specifically, the proposed framework is shown in Fig. 2.

III. EXPERIMENTS

A. Dataset and Evaluation Metrics

Currently, there is no comprehensive dataset to validate the performance of SAR OOD detection methods. To address this gap, we have curated some existing publicly available datasets, including vehicle targets (MSTAR [8], SAMPLE [13]), airplane targets (SAR-ACD [10]), and ship targets (FUSAR-ship [11]), and constructed a new dataset for SAR OOD detection. This new data set is called SAR-OOD, and its details are provided in Table I.

In the SAR OOD detection task, we only use a train set of MSTAR (17°) in Table I for training, and a test set of MSTAR (15°) to evaluate the performance of ID classification. The other OOD data presented in Table I is utilized to assess the performance of OOD detection. We use the following metrics to validate our methods: area under the receiver operating characteristic curve (AUROC) and the false positive rate of OOD examples when the true positive rate of ID data is 95% (FPR95).

B. Implementation Details

We implement experiments using the Pytorch framework. In the training process, the batch size is set to 8, and Adam is applied as an optimizer. The learning rate is set to 1e-5. We conduct experiments on a single NVIDIA A100 and train for 100 epochs to make the model converge. To comprehensively verify the performance of different losses in SAR OOD detection, resnet18 [14], wrn-50-2 [15], and densenet121 [16] are used as feature extraction models.

C. Results

1) *Performance of ID*: We compared CE, LogitNorm [7], and UMSE loss function to obtain the recognition accuracy on the MSTAR test, as shown in Fig. 3. The temperature parameter of LogitNorm $\tau = 0.1$. Results on MSTAR test data show that the UMSE loss function with resnet18 achieves the best performance, and the LogitNorm has the lowest recognition accuracy in all models. The reasons for this result are as follows: (1) Compared with resnet18, densenet121 increases the dense connection between layers, and wrn-50-2 widens the residual network to improve recognition performance on the ImageNet dataset. However, MSTAR has a smaller amount of data than ImageNet, and using a more complex network can easily lead to overfitting and reduce recognition performance. (2) UMSE is based on MAP, while CE and LogitNorm are based on MLE. It's well known that MAP-based loss function has a stronger generalization performance on small-scale datasets than MLE-based loss functions.

2) *Performance of OOD Detection*: We conducted OOD detection performance evaluations using different OOD score functions as follows: (1) **Maxlogit** [3]: logit-based OOD score function. (2) **MaxNorm**: feature-based OOD score function. (3) **MSP** [2]: probability-based method, where the sample belongs to ID data if the maximum probability is bigger than a threshold. (4) **DML** [4]: OOD score function based on feature and logit. (5) **Energy** [17]: calculate the energy of

TABLE I
DATASET DETAILS OF SAR-OOD

Dataset Name	Target	ID / OOD	Category
MSTAR	Vehicle	ID (train: 17°, test: 15°)	2S1, BMP2, BRDM_2, BTR_60,BTR70, D7, ZIL131, T62, T72, ZSU_23_4
SAMPLE	Vehicle	OOD	m1, m2, m35, m60, m548
SAR-ACD	Airplane	OOD	A220, A330, A320321, ARJ21, Boeing737, Boeing787
FUSAR-ship	Ship	OOD	BulkCarrier, CargoShip, ContainerShip, Dredger, Tanker, Fishing, GeneralCargo

TABLE II
OOD DETECTION PERFORMANCE COMPARISON USING CE LOSS AND UMSE LOSS. ↑INDICATES LARGER VALUES ARE BETTER, AND ↓ INDICATES SMALLER VALUES ARE BETTER. **BOLD** NUMBERS ARE SUPERIOR RESULTS.

Dataset	Evaluation	SAMPLE		SAR-ACD		FUSAR-ship		Average	
		AUROC ↑	FPR95 ↓	AUROC ↑	FPR95 ↓	AUROC ↑	FPR95 ↓	AUROC ↑	FPR95 ↓
CE / UMSE (our proposed)									
densenet121	MaxLogit	0.642 / 0.931	1.000 / 0.441	0.916 / 0.982	0.378 / 0.084	0.762 / 0.833	0.623 / 0.52	0.773 / 0.915	0.667 / 0.348
	MaxNorm	0.057 / 0.928	1.000 / 0.494	0.302 / 0.978	0.98 / 0.065	0.474 / 0.901	0.958 / 0.438	0.277 / 0.935	0.979 / 0.332
	MSP	0.878 / 0.963	0.819 / 0.154	0.954 / 0.981	0.256 / 0.083	0.829 / 0.833	0.579 / 0.506	0.887 / 0.926	0.551 / 0.248
	DML	0.091 / 0.937	1.000 / 0.419	0.456 / 0.987	0.911 / 0.044	0.58 / 0.875	0.859 / 0.485	0.376 / 0.933	0.924 / 0.316
	Energy	0.624 / 0.93	1.000 / 0.441	0.907 / 0.982	0.428 / 0.084	0.759 / 0.833	0.632 / 0.52	0.763 / 0.915	0.686 / 0.348
resnet18	MaxLogit	0.878 / 0.963	0.951 / 0.186	0.96 / 0.992	0.227 / 0.007	0.821 / 0.939	0.541 / 0.295	0.886 / 0.965	0.573 / 0.163
	MaxNorm	0.315 / 0.948	1.000 / 0.417	0.714 / 0.995	0.573 / 0.004	0.737 / 0.947	0.601 / 0.275	0.588 / 0.963	0.725 / 0.232
	MSP	0.955 / 0.977	0.338 / 0.049	0.971 / 0.989	0.166 / 0.018	0.851 / 0.944	0.563 / 0.286	0.926 / 0.97	0.356 / 0.118
	DML	0.546 / 0.959	1.000 / 0.209	0.814 / 0.993	0.45 / 0.005	0.775 / 0.944	0.579 / 0.278	0.712 / 0.965	0.676 / 0.164
	Energy	0.862 / 0.963	0.974 / 0.186	0.957 / 0.992	0.233 / 0.007	0.82 / 0.939	0.53 / 0.295	0.88 / 0.965	0.579 / 0.163
wrn-50-2	MaxLogit	0.719 / 0.955	0.977 / 0.291	0.942 / 0.991	0.329 / 0.027	0.869 / 0.901	0.609 / 0.492	0.843 / 0.949	0.638 / 0.27
	MaxNorm	0.818 / 0.959	0.984 / 0.254	0.863 / 0.995	0.367 / 0.017	0.844 / 0.904	0.464 / 0.451	0.842 / 0.953	0.605 / 0.241
	MSP	0.786 / 0.945	0.883 / 0.406	0.915 / 0.979	0.518 / 0.133	0.882 / 0.893	0.686 / 0.615	0.861 / 0.939	0.695 / 0.385
	DML	0.804 / 0.957	0.981 / 0.27	0.896 / 0.994	0.331 / 0.02	0.863 / 0.903	0.489 / 0.468	0.854 / 0.951	0.6 / 0.253
	Energy	0.713 / 0.955	0.988 / 0.291	0.943 / 0.991	0.324 / 0.027	0.867 / 0.901	0.624 / 0.492	0.841 / 0.949	0.646 / 0.27

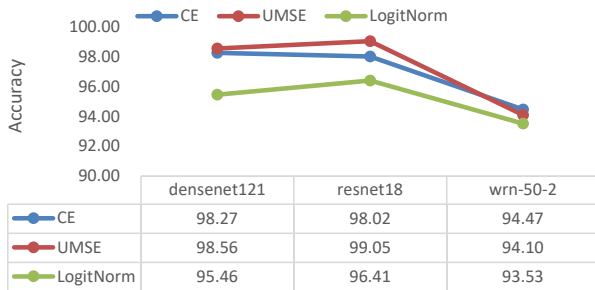


Fig. 3. Recognition comparison between CE, LogitNorm, and UMSE using different models.

prediction as OOD score function; the parameter T in energy score function is equal to 1 in our experiments. From Fig. 3, the recognition performance of LogitNorm is far away from CE and UMSE. Therefore, LogitNorm is not considered in the following OOD performance comparison. The OOD detection results on SAR-OOD are presented in Table II.

The results show that UMSE achieves the SOTA under various conditions. On the one hand, **a wide range of downstream OOD scoring functions benefit from UMSE**. Specifically, for the Maxlogit OOD score function based on resnet18, the FPR is reduced from 0.593 to 0.162, and the AUROC is increased from 0.882 to 0.965. This means the model with UMSE loss, which considers data uncertainty, can get highly distinguished features, logit, and probability between ID and OOD data. We visualize the density of the

maximum feature, maximum logit, and maximum probability on ID and OOD. The results are shown in Fig. 4, and the distribution of ID and OOD data related to UMSE overlaps less than CE. That’s why the OOD score functions based on feature, logit, and probability can be improved. On the other hand, **regardless of the model used, UMSE-based models achieve SOTA in SAR OOD detection**. This indicates that UMSE is closer to the essence of the SAR OOD detection problem in form and can more effectively promote the model to learn useful information. Overall, it is important to consider uncertainty in loss function design, especially when dealing with images with high data uncertainty, such as SAR images.

IV. CONCLUSION

SAR images are characterized by high data uncertainty, which can impact the model’s ability to discern image features and logit representations, consequently affecting OOD detection performance. In this paper, we account for the influence of data uncertainty on prediction outcomes using Dirichlet distribution, thereby endowing the model with capabilities for uncertainty estimation and calibration. Specifically, we propose a loss function named UMSE, guiding the model in learning highly distinguishing features, logit, and probability for OOD data detection. To validate the performance of the proposed algorithm, we constructed the SAR-OOD dataset based on some existing public data. UMSE achieved SOTA performance on this dataset. These findings underscore the importance of uncertainty estimation for mitigating overconfidence in SAR OOD detection.

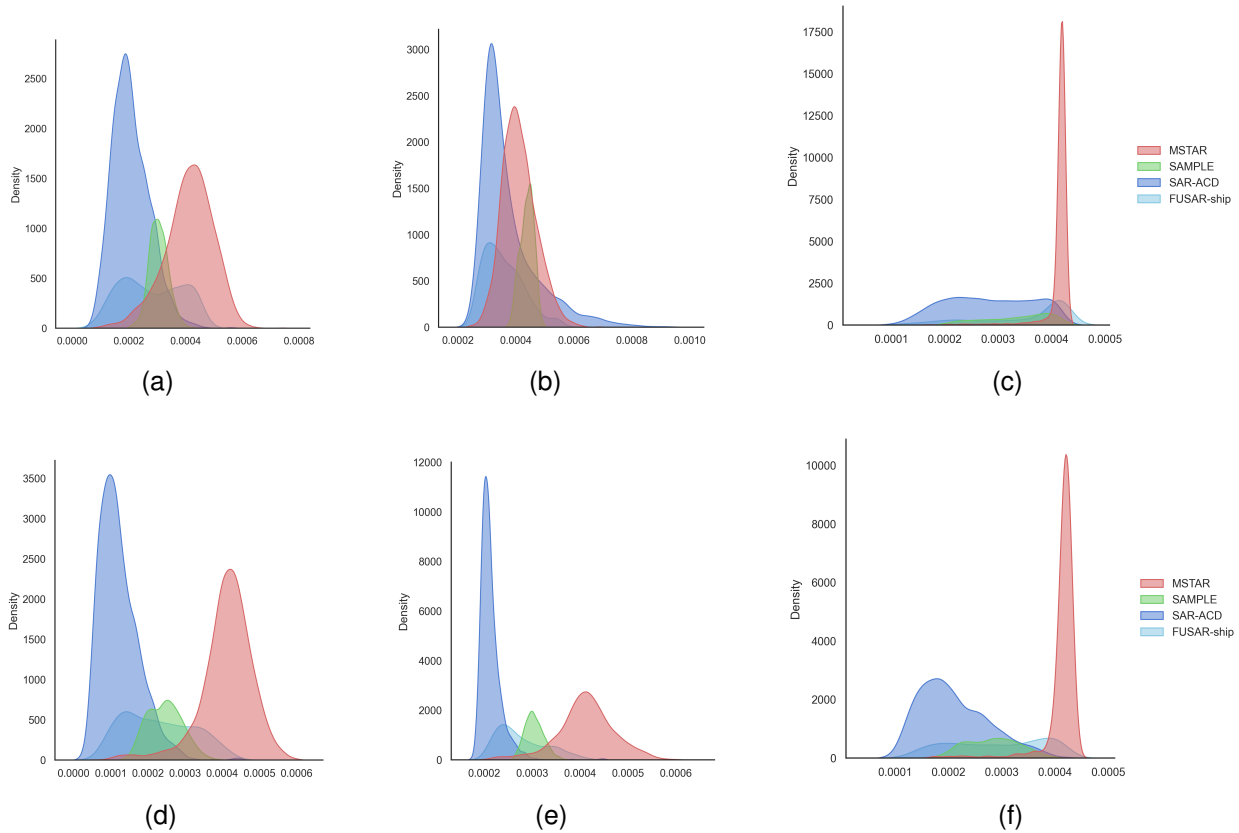


Fig. 4. Distribution of ID and OOD based on resnet18 with CE loss and UMSE loss. (a) Maximum logit with CE loss. (b) Maximum feature norm with CE loss. (c) Maximum probability with CE loss. (d) Maximum logit with UMSE loss. (e) Maximum feature norm with UMSE loss. (f) Maximum probability with UMSE loss. Based on the features, logit, and probability obtained from the model supervised by UMSE loss shows better separability between ID (MSTAR) and OOD data (SAMPLE, SAR-ACD, FUSAR-ship) because of the less distribution overlap.

REFERENCES

- [1] N. Inkawhich, J. Zhang, E. K. Davis, R. Luley, and Y. Chen, “Improving out-of-distribution detection by learning from the deployment environment,” *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 15, pp. 2070–2086, 2022. **1**
- [2] D. Hendrycks and K. Gimpel, “A baseline for detecting misclassified and out-of-distribution examples in neural networks,” *arXiv preprint arXiv:1610.02136*, 2016. **1, 3**
- [3] D. Hendrycks, S. Basart, M. Mazeika, A. Zou, J. Kwon, M. Mostajabi, J. Steinhardt, and D. Song, “Scaling out-of-distribution detection for real-world settings,” *arXiv preprint arXiv:1911.11132*, 2019. **1, 3**
- [4] Z. Zhang and X. Xiang, “Decoupling maxlogit for out-of-distribution detection,” in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2023, pp. 3388–3397. **1, 3**
- [5] K. Lee, K. Lee, H. Lee, and J. Shin, “A simple unified framework for detecting out-of-distribution samples and adversarial attacks,” *Advances in neural information processing systems*, vol. 31, 2018. **1**
- [6] N. Inkawhich, M. J. Inkawhich, E. K. Davis, U. K. Majumder, E. Tripp, C. Capraro, and Y. Chen, “Bridging a gap in sar-atr: Training on fully synthetic and testing on measured data,” *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 14, pp. 2942–2955, 2021. **1**
- [7] H. Wei, R. Xie, H. Cheng, L. Feng, B. An, and Y. Li, “Mitigating neural network overconfidence with logit normalization,” in *International conference on machine learning*. PMLR, 2022, pp. 23 631–23 644. **1, 2, 3**
- [8] J. R. Diemunsch and J. Wissinger, “Moving and stationary target acquisition and recognition (mstar) model-based automatic target recognition: Search technology for a robust atr,” in *Algorithms for synthetic aperture radar Imagery V*, vol. 3370. SPIE, 1998, pp. 481–492. **1, 3**
- [9] B. Lewis, T. Scarnati, E. Sudkamp, J. Nehrbass, S. Rosencrantz, and E. Zelnio, “A sar dataset for atr development: the synthetic and measured paired labeled experiment (sample),” in *Algorithms for Synthetic Aperture Radar Imagery XXVI*, vol. 10987. SPIE, 2019, pp. 39–54. **1**
- [10] X. Sun, Y. Lv, Z. Wang, and K. Fu, “Scan: Scattering characteristics analysis network for few-shot aircraft classification in high-resolution sar images,” *IEEE Transactions on Geoscience and Remote Sensing*, vol. 60, pp. 1–17, 2022. **1, 3**
- [11] X. Hou, W. Ao, Q. Song, J. Lai, H. Wang, and F. Xu, “Fusar-ship: Building a high-resolution sar-ais matchup dataset of gaofen-3 for ship detection and recognition,” *Science China Information Sciences*, vol. 63, pp. 1–19, 2020. **1, 3**
- [12] M. Sensoy, L. M. Kaplan, and M. Kandemir, “Evidential deep learning to quantify classification uncertainty,” in *Advances in neural information processing systems*, 2018, pp. 3183–3193. **2**
- [13] B. Lewis, T. Scarnati, E. Sudkamp, J. Nehrbass, S. Rosencrantz, and E. Zelnio, “A sar dataset for atr development: the synthetic and measured paired labeled experiment (sample),” in *Algorithms for Synthetic Aperture Radar Imagery XXVI*, vol. 10987. SPIE, 2019, pp. 39–54. **3**
- [14] K. He, X. Zhang, S. Ren, and J. Sun, “Deep residual learning for image recognition,” in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 770–778. **3**
- [15] S. Zagoruyko and N. Komodakis, “Wide residual networks,” *arXiv preprint arXiv:1605.07146*, 2016. **3**
- [16] G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger, “Densely connected convolutional networks,” in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 4700–4708. **3**
- [17] W. Liu, X. Wang, J. Owens, and Y. Li, “Energy-based out-of-distribution detection,” *Advances in neural information processing systems*, vol. 33, pp. 21 464–21 475, 2020. **3**