IGARSS 2021 - 2021 IEEE International Geoscience and Remote Sensing Symposium | 978-1-6654-0369-6/21/$31.00 ©2021 IEEE | DOI: 10.1109/IGARSS47720.2021.9554626

CAN WE EVALUATE THE DISTINGUISHABILITY OF THE OPENSARURBAN DATASET ?

Ning Liao[1], Mihai Datcu[2], Zenghui Zhang[1], Weiwei Guo[3], Wenxian Yu[1]

1 Shanghai Key Lab of Intelligent Sensing and Recognition, Shanghai Jiao Tong University
2 German Aerospace Centre DLR
3 Center of Digital Innovation, Tongji University

ABSTRACT

In Synthetic Aperture Radar (SAR) image classification tasks, the performance depends on both the classifier and the dataset itself. However, in comparison with plenty of SAR classification methods, there is little work aimed at analyzing the distinguishability of the dataset. In the classification dataset, some classes are semantically different but their distinguishability is low, the classes are hard to be classified especially in some more practical cases that there are unknown classes without supervision exist. Referring to open set recognition (OSR), in this paper, we proposed the SAR Distinguishability Analysor (SAR-DA) to evaluate the distinguishability of the OpenSARUrban dataset. By modeling each class as a multivariate Gaussian distribution in latent space, SAR-DA can not only classify the classes having been seen in training phase, but also can recognize unknown samples if a test sample is out of each known distribution. Each class in OpenSARUrban is set unknown in turn, then we apply the SAR-DA on the split dataset in OSR and supervised setting. The distinguishability can be reflected by the unknown recognition recall rate. The experimental results show that the unknown recognition recall rate in OSR setting significantly decreased compared with those in supervised setting, indicating that even though the classes in OpenSARUrban are semantically different from each other, the latent distributions of some classes are quite similar and hard to be classified, thus these classes are of low distinguishability.

Index Terms— Synthetic Aperture Radar (SAR), distinguishability, open set recognition (OSR), SAR Distinguishability Analysor (SAR-DA), OpenSARUrban, multivariate Gaussian distribution.

1. INTRODUCTION

Synthetic Aperture Radar (SAR) is an active sensor, working well in all-weather and all-day conditions. It has been widely used in civilian and military fields over the past decades. As one of the most popular methods for SAR image classification, convolutional neural networks (CNNs) have been leading a new trend for this application [1]. Convolutional auto-encoder has been successfully used in high-resolution SAR image classification [2]. More practically, due to the lack of a large number of labeled data, it is an urgent problem to be addressed that the classifier can not only classify the classes seen in training phase, but also can recognize unknown classes. Some previous work in open set recognition (OSR) [3] carried out unknown recognition in natural images by modeling the data distribution of known classes [4][5], if the test sample is out of the distribution of each known class, the test sample will be recognized unknown.

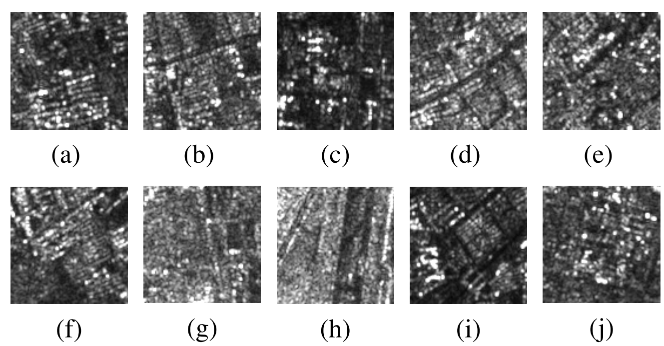

Fig. 1. [6] The patches in OpenSARUrban dataset, there are 10 different classes. (a) Denselow, (b) General-Residential, (c) Highbuildings, (d) SingleBuilding, (e) Skyscraper, (f) StorageArea, (g) Vegetation, (h) Airport, (i) Railway, (j) Highway.

Although many methods have been designed and applied to image classification tasks, they only focus on the algorithms. Fundamentally, the dataset itself is an important factor that affects the classification performance, particularly for SAR images. The patches of each class in OpenSARUrban [6] dataset are shown in Fig.1. In this figure, patches of Denselow, General Residential, Highbuildings and SingleBuilding are all composed of strong scattering points reflected from the building surface, it is even hard for experts to classify them correctly. If there are unknown classes without supervision exist, it will be intractable to classify the unknown classes if they are of low distinguishability. It can be seen that in SAR image classification tasks, the analysis of the dataset itself is of great importance and meaningful, but there is little work aimed at analyzing the dataset compared with plenty of classification algorithms, can we evaluate the distinguishability of the dataset?

Considering the complexity of the spatial distribution of the original images and the low dimensional nature of the semantic information of classes, we proposed to evaluate the distinguishability by modeling the latent multivariate Gaussian distribution of each class. In this paper, referring to the conditional Gaussian distribution learning (CGDL) [7], we proposed the SAR Distinguishability Analysor (SAR-DA) to investigate the OpenSARUrban dataset. We set each class in OpenSARUrban as unknown class in turn, and then apply SAR-DA on the split OpenSARUrban dataset for open set recognition (OSR) and supervised classification.

The rest of this paper is organized as follows. Section 2 presents the proposed SAR Distinguishability Analysor (SAR-DA). In section 3, the experimental setting is introduced and the results are discussed. Finally, conclusions are summarized in section 4.

2. METHODOLOGY

In this section, we describe the SAR Distinguishability Analysor (SAR-DA) method as shown in Fig.2. The architecture of SAR-DA is composed of encoder, decoder and classifier in supervised training phase.

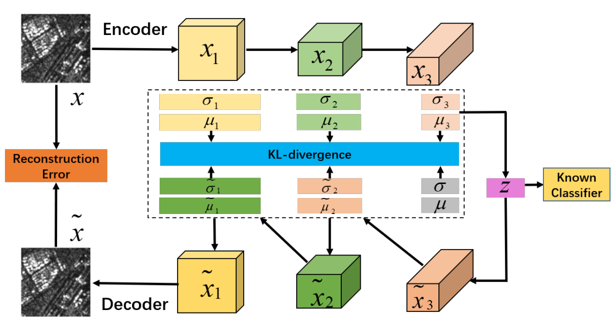

Fig. 2. SAR Distinguishability Analysor (SAR-DA)

Encoder. Each encoder layer adopts the probabilistic ladder [8] and the k-th encoder layer is denoted as follows:

$$\begin{aligned} \boldsymbol{x}_k &= \text{Conv}\left(\boldsymbol{x}_{k-1}\right) \\ \boldsymbol{\mu}_k &= \text{Linear}(\text{Flatten}\ (\boldsymbol{x}_k)) \\ \boldsymbol{\sigma}_k^2 &= \text{F}\ (\ \text{Linear}\ \ (\ \text{Flatten}\ \ (\boldsymbol{x}_k))) \end{aligned} \tag{1}$$

Flatten is adopted to flatten $\boldsymbol{x}_k$ into a 1-dimensional data. $\boldsymbol{\mu}_k$ and $\boldsymbol{\sigma}_k^2$ in each encoder layer can form the approximate multivariate Gaussian distribution and will be used for calculating KL-divergence. F is a non-linear function denoted as $log(1 + exp(\cdot))$. The latent representation $\boldsymbol{z}$ is obtained as $\boldsymbol{z} = \boldsymbol{\mu} + \varepsilon \cdot \boldsymbol{\sigma}, \varepsilon \sim N(\mathbf{0}, \mathbf{I})$ in the top encoder layer, in which $\mathbf{I}$ is an identity matrix.

Classifier. The classifier takes the latent representation $\boldsymbol{z}$ as input and the softmax layer ouputs the probability distribution over each known class.

Decoder. The k-th decoder layer is denoted as follows:

$$\begin{aligned} \boldsymbol{d}_{k+1} &= \text{Unflatten}\left(\text{Linear}(\tilde{\boldsymbol{z}}_{k+1})\right) \\ \tilde{\boldsymbol{x}}_{k+1} &= \text{TConv}\left(\boldsymbol{d}_{k+1}\right) \\ \tilde{\boldsymbol{\mu}}_k &= \text{Linear}\ (\ \text{Flatten}\ \ (\tilde{\boldsymbol{x}}_{k+1})) \\ \tilde{\boldsymbol{\sigma}}_k^2 &= \text{F}\ (\ \text{Linear}\ \ (\ \text{Flatten}\ \ (\tilde{\boldsymbol{x}}_{k+1}))) \\ \tilde{\boldsymbol{z}}_k &= \tilde{\boldsymbol{\mu}}_k + \varepsilon \cdot \tilde{\boldsymbol{\sigma}}_k^2, \varepsilon \sim N(\mathbf{0}, \mathbf{I}) \end{aligned} \tag{2}$$

TConv is a transposed Conv. Unflatten is inverse to Flatten that it converts the 1-dimensional data into 3-dimensional. In the final decoder layer, the input image is reconstructed.

The loss function in training phase is composed of 3 parts:

(1) The classification loss L_c is the softmax cross-entropy between the prediction and the ground-truth labels;

(2) The reconstruction loss L_r is the L_1 distance between the input image $\boldsymbol{x}$ and the final reconstructed image $\tilde{\boldsymbol{x}}$;

(3) The KL-divergence L_{KL} in the latent space and each paired encoder-decoder layer. Based on the latent representation $\boldsymbol{z} = \boldsymbol{\mu} + \varepsilon \cdot \boldsymbol{\sigma}, \varepsilon \sim N(\mathbf{0}, \mathbf{I})$, the conditional Gaussian posterior distribution is denoted as $\boldsymbol{q}_k = N\left(\boldsymbol{z}; \boldsymbol{\mu}, \boldsymbol{\sigma}^2\right)$. The posterior distribution is forced to approximate the multivariate Gaussian distribution $\boldsymbol{p}_k = N\left(\boldsymbol{z}; \boldsymbol{\mu}_k, \mathbf{I}\right)$ of each known class, in which $\boldsymbol{\mu}_k$ is the output of a fully-connected layer which takes the one-hot encoding of the input's label as input. In each paired encoder-decoder layer, the distributions are also multivariate Gaussian distributions. The KL-divergence forms the third part of the loss function as:

$$\begin{aligned} L_{KL} = -\frac{1}{K}\Big[& D_{KL}\left(q_{\boldsymbol{\phi}}(\boldsymbol{z} \mid \boldsymbol{x}) \| p_{\boldsymbol{\theta}}^{(l)}(\boldsymbol{z} \mid \boldsymbol{x})\right) \\ & + \sum_{k \neq K} D_{KL}\left(q_{\boldsymbol{\phi}}\left(\boldsymbol{x}_k \mid \boldsymbol{x}\right) \| p_{\boldsymbol{\theta}}\left(\tilde{\boldsymbol{x}}_k \mid \tilde{\boldsymbol{x}}_{k+1}\right)\right) \Big] \end{aligned} \tag{3}$$

Thus, the training loss function is summarized as:

$$L = -\left(L_r + L_c + L_{KL}\right) \tag{4}$$

After the training process is complete, we set the mean and variance of the latent representations of all the correctly classified training samples as $\boldsymbol{m}_l$ and $\boldsymbol{\sigma}_l^2$ in the l-th class. The multivariate Gaussian distribution of the l-th known class is modeled as $f_l(\boldsymbol{z}) = N\left(\boldsymbol{z}; \boldsymbol{m}_l, \boldsymbol{\sigma}_l^2\right)$.

In testing phase, if the reconstruction error is larger than that of known classes, or if the probability expressed in Eq.(5) is lower than the threshold we set in experiments. It means that the latent representation of a test sample locates in the low probability space of the d-dimension multivariate Gaussian distribution of each known class, the test sample will be recognized as belonging to unknown class.

$$P_l(\boldsymbol{z}) = 1 - \int_{m_1 - |z_1 - m_1|}^{m_1 + |z_1 - m_1|} \cdots \int_{m_d - |z_d - m_d|}^{m_d + |z_d - m_d|} f_l(\boldsymbol{t}) d\boldsymbol{t}. \tag{5}$$

3. EXPERIMENTS AND ANALYSIS

3.1. Experimental Setting

There are 5 functional classes and totally 10 classes in OpenSARUrban dataset, the structure of OpenSARUrban dataset is shown in Fig.3. In order to evaluate the distinguishability of OpenSARUrban dataset, we utilize SAR-DA to carry out the experiments on the split dataset.

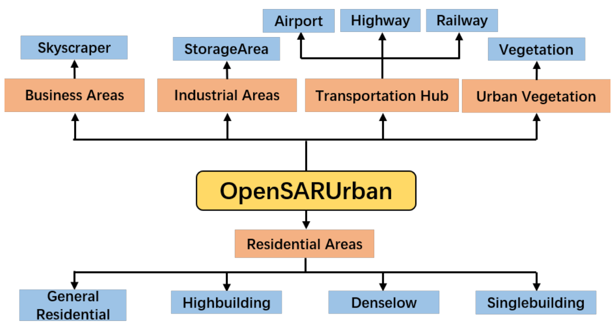

Fig. 3. [6] The structure of OpenSARUrban dataset. The 5 functional classes: Business Areas, Industrial Areas, Transportation Hub, Urban Vegetation, Residential Areas. Totally 10 classes: Skyscraper, StorageArea, Airport, Highway, Railway, Vegetation, General Residential, Highbuilding, Denselow, Singlebuilding.

For comparison, we conducted the experiments in OSR setting and supervised setting. We choose the recall rate as the measurement, which indicates how many samples are correctly classified or recognized in a class. Firstly, the comparison between the classification recall rate of the same class in supervised setting and OSR setting can reflect the impact caused by introducing unknown classes. On the other hand, the unknown recognition recall rate of each class can reflect the distinguishability when the class is set unknown.

In OSR setting, we set each of the 10 classes as unknown class, and the other 9 classes as known classes seen in training phase, 70% of the samples of the known classes are randomly selected as training set and the others as part of test set. In testing phase, the SAR-DA is used to classify the 9 classes seen in the training process and also recognize the classes not seen in training phase as belonging to another unknown class. The upper bound of the reconstruction error of known classes is set as the mean value plus 2 times standard deviation of the reconstruction error of the correctly classified samples, mathematically written by Eq.(6). The threshold of the probability corresponding to Eq.(5) is set as 0.5. If the probability of the latent representation of a test sample locating in each known distribution is lower than 0.5, the test sample is recognized unknown.

$$rec_{upb} = mean(\boldsymbol{rec}_{train}) + 2 * std(\boldsymbol{rec}_{train}) \qquad (6)$$

In supervised setting, we set each class in OpenSARUrban unknown in turn, the unknown class will not be used neither for training or testing. The other 9 classes are split into training set and test set. SAR-DA is adopted for supervised classification and will not calculate the probability in Eq.(5) to validate whether a test sample belongs to known or unknown classes. Besides, in supervised setting, we have conducted another experiment for supervised classification on the totally 10 classes.

3.2. Result Analysis

In OSR and supervised setting, each class is set as unknown class in turn. We conducted the experiments with 10 times training and testing in OSR setting and 11 times in supervised setting to investigate the distinguishability of the total 10 classes in OpenSARUrban dataset. The recall rate results of the experiments are shown in Table.1.

In Table.1, there are 2 columns of values for each class when they are set as unknown class. Values in the left column are the classification recall rates in OSR setting. Values in the right column are the classification recall rates in supervised setting. The bold numbers on the diagonal line indicate the recall rates of correctly recognizing the samples as belonging to the unknown class when the corresponding class is set unknown. Other values in each column indicate the recall rates of known classes. The last 3 rows indicate the m-precision, m-recall and m-fmeasure results.

As can be seen from Table.1, in supervised setting, according to the values in the right columns, the classification recall rate of each class under the 11 supervised conditions are all high. The m-recall results in supervised setting are all no less than 95.54%, which shows the excellent performance of SAR-DA in supervised classification on the OpenSARUrban dataset. In OSR setting, after introducing the test samples belonging to unknown class, in comparison with the classification recall rate of the known classes in supervised setting, most of the recall rate values reduced. For example, when the General Residential is set unknown, the recall rate of correctly classifying the Railway is greatly dropped from 100.0% to 10.71%. From the reduction of classification recall rate of the known classes, we can see the impact of introducing test samples belonging to unknown class. When the unknown class is introduced, due to the lack of strong supervision signals, the overlap between the distributions of the known classes and unknown class results in the wrongly classification and decrease in classification recall rate, especially when the unknown class is of low distinguishability.

In terms of the unknown recognition recall rate, compared with the recall rate in supervised setting of the 10 classes in the right most column, the recall rate results all significantly reduced, which indicates that in OpenSARUrban dataset, if the class is not under supervision, it is hard to be recognized correctly. Among all the unknown recognition results, when

Table 1. The recall rate(%) results of the classification under OSR setting and supervised setting. For each unknown class, the values in the left column are the classification results in OSR setting, values in the right column are those in supervised setting.

Classes	Unknown Classes																				All Known
	Denselow		Gen.Res		Highbuildings		SingleBuilding		Skyscraper		StorageArea		Vegetataion		Airport		Railway		Highway		
Denselow	**5.63**	—	81.81	100.0	97.16	99.82	96.18	99.91	84.65	99.38	94.50	99.82	91.13	99.65	92.37	99.91	46.85	99.82	76.40	98.94	98.23
Gen.Res	96.13	99.91	**40.95**	—	96.00	100.0	84.08	99.95	90.36	99.36	65.89	99.91	90.72	99.00	83.27	98.59	83.77	98.36	76.40	99.77	99.50
Highbuildings	94.80	99.81	97.03	99.94	**2.21**	—	92.31	99.94	86.69	99.68	71.36	99.90	90.39	98.85	93.26	99.87	87.10	99.36	81.86	99.94	99.90
SingleBuilding	93.41	99.06	90.89	99.53	90.11	100.0	**6.30**	—	90.89	99.06	84.93	99.69	91.37	99.53	77.71	100.0	92.46	91.99	84.30	100.0	99.37
Skyscraper	89.31	99.78	93.10	100.0	85.75	100.0	56.12	99.55	**31.07**	—	80.85	99.78	73.05	99.78	84.19	98.89	68.15	99.11	87.08	93.99	99.55
StorageArea	91.23	99.80	90.74	99.28	95.17	99.92	95.41	99.84	90.82	99.96	**29.71**	—	89.05	99.84	87.08	99.92	83.94	97.30	56.62	99.36	99.20
Vegetataion	79.06	100.0	73.62	99.83	67.67	99.67	81.99	99.92	76.38	99.16	71.44	100.0	**45.20**	—	72.19	99.41	69.93	93.63	74.12	99.83	98.83
Airport	72.31	100.0	56.92	100.0	61.54	98.46	74.63	98.51	72.31	100.0	73.85	100.0	56.92	92.31	**73.40**	—	67.69	84.62	66.15	98.46	93.85
Railway	85.71	100.0	10.71	100.0	92.86	85.71	82.14	85.71	75.00	96.43	85.71	100.0	92.86	96.43	100.0	96.43	**65.00**	—	92.86	98.18	100.0
Highway	87.05	100.0	89.21	99.28	79.14	95.68	85.61	98.56	88.49	100.0	81.30	98.56	79.86	100.0	85.61	99.28	84.17	95.68	**67.84**	—	98.56
m-precision	82.10	99.82	81.47	99.82	67.62	98.59	83.82	98.59	88.75	99.61	72.21	99.74	82.66	99.20	90.13	99.36	89.77	97.32	89.42	99.49	99.40
m-recall	79.46	99.82	72.50	99.76	76.76	97.70	75.48	97.99	78.67	99.23	73.95	99.74	80.05	98.38	84.91	99.14	74.91	95.54	76.36	98.52	98.70
m-fmeasure	80.36	99.82	73.55	99.79	69.98	98.07	78.91	98.25	82.69	99.41	72.80	99.74	80.96	98.76	83.83	99.24	77.18	96.31	78.56	98.99	99.04

Airport is set unknown, the probability calculated from Eq.(5) is lower than the threshold, most of the test samples can be correctly recognized as unknown, showing that most latent representations of Airport are out of the other known distributions, Airport is of the highest distinguishability. While when Highbuildings, Denselow or SingleBuilding is set unknown, the probability calculated from Eq.(5) is higher than the threshold, most test samples are hard to be distinguished, which reveals the high similarity and overlap between the known classes and Highbuildings, Denselow or SingleBuilding, even though they are semantically different from each other. According to the recall rate change, we infer that the most distinguishable class is Airport, followed by Highway and Railway, the Highbuildings, Denselow and SingleBuilding are the least distinguishable classes.

4. CONCLUSION

In comparison with plenty of classification methods, referring to open set recognition (OSR), we proposed the SAR-DA to evaluate the distinguishability of each class in OpenSARUrban dataset. The SAR-DA models the latent multivariate Gaussian distributions of the known classes seen in training phase. A test sample will be recognized as unknown if the extracted latent representation of the test sample locates in the low probability space of the distributions of each known class. We set each class in OpenSARUrban unknown alternately and carried out the experiments in OSR and supervised setting. The experimental results show the excellent performance of SAR-DA in supervised classification on OpenSARUrban dataset, while the classification recall rate significantly reduced after introducing unknown class. Most importantly, though the classes are semantically different from each other, some classes are similar and of low distinguishability. This may be the first work that adopts the OSR method to evaluate the distinguishability of SAR classification dataset.

Compared with natural image datasets, we still lack high-quality SAR image datasets. In future work, we will propose better methods to analyze the data and help constructing more SAR datasets.

5. ACKNOWLEDGMENTS

This work was supported in part by the National Natural Science Foundation of China under grant 62071333, U1830103 and ESA-MOST CHINA Dragon 5 programme ID.58190.

6. REFERENCES

[1] N. Ødegaard, A. O. Knapskog, C. Cochin, and J. Louvigne, "Classification of ships using real and simulated data in a convolutional neural network," in *2016 IEEE Radar Conference (RadarConf)*, 2016, pp. 1–6.

[2] J. Geng, J. Fan, H. Wang, X. Ma, B. Li, and F. Chen, "High-resolution sar image classification via deep convolutional autoencoders," *IEEE Geoscience and Remote Sensing Letters*, vol. 12, no. 11, pp. 2351–2355, 2015.

[3] W. J. Scheirer, A. de Rezende Rocha, A. Sapkota, and T. E. Boult, "Toward open set recognition," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 35, no. 7, pp. 1757–1772, 2013.

[4] Lukas Ruff, Robert Vandermeulen, Nico Goernitz, Lucas Deecke, Shoaib Ahmed Siddiqui, Alexander Binder, Emmanuel Müller, and Marius Kloft, "Deep one-class classification," in *International conference on machine learning*, 2018, pp. 4393–4402.

[5] Stanislav Pidhorskyi, Ranya Almohsen, and Gianfranco Doretto, "Generative probabilistic novelty detection with adversarial autoencoders," in *Advances in neural information processing systems*, 2018, pp. 6822–6833.

[6] J. Zhao, Z. Zhang, W. Yao, M. Datcu, H. Xiong, and W. Yu, "Opensarurban: A sentinel-1 sar image dataset for urban interpretation," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 13, pp. 187–203, 2020.

[7] X. Sun, Z. Yang, C. Zhang, K. V. Ling, and G. Peng, "Conditional gaussian distribution learning for open set recognition," in *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2020, pp. 13477–13486.

[8] Casper Kaae Sønderby, Tapani Raiko, Lars Maaløe, Søren Kaae Sønderby, and Ole Winther, "How to train deep variational autoencoders and probabilistic ladder networks," in *33rd International Conference on Machine Learning (ICML 2016)*, 2016.