EUSAR 2021 . ISBN 978-3-8007-5457-1 / ISSN 2197-4403 . ©VDE VERLAG GMBH . Berlin . Offenbach

Urban Area Analysis in Single-polarized SAR Images Based On Unsupervised Deep Learning

Zhongling Huang[a] and Mihai Datcu[b]
[a]School of Automation, Northwestern Polytechnical University, Xi'an, China
[b]Remote Sensing Technology Institute (IMF), German Aerospace Center (DLR), Wessling, Germany

Abstract

Urban mapping from remote sensing images is important for monitoring urbanization. In this paper, we propose an unsupervised learning algorithm for high-resolution single-polarized synthetic aperture radar (SAR) image to extract man-made targets for urban area analysis. The proposed method mainly focuses on the special physical characteristics of man-made targets that are different from natural areas. Without polarimetric information, we propose the sub-band scattering pattern based on time-frequency analysis to describe the physical properties of targets, and then design an end-to-end neural network to learn the latent features and potential clusters. The proposed method is evaluated on three different urban areas acquired at C-band by Sentinel-1 and Gaofen-3, and X-band by TerraSAR-X, respectively. The experiments present the visualized result of man-made targets extraction and analyze some specific targets to show the effectiveness of our proposed method.

1 Introduction

Analyzing urban area with Synthetic Aperture Radar (SAR) images has attracted much attention in recent years, due to the all-day and all-weather observation capability of SAR. The backscattering is complicated in urban area where man-made targets are densely arranged, resulting in difficulties for visual interpretation. The backscattering behaviors of targets can be various under different observation conditions, such as polarization, incidence angle, etc. Also, the intrinsic physical properties of targets, like material, shape, and orientation, have influences on SAR scattering. Consequently, knowing the physical properties of targets is essential for urban understanding.

Many studies are working with polarimetric SAR (PolSAR) to reveal the inherent scattering characteristics of man-made objects, which is very different from that of natural areas [1]. As more advanced SAR satellites have been launched, the very high resolution (VHR) SAR images are available, providing detailed texture information of objects on the ground. Many deep learning based algorithms were proposed for SAR image urban built-up extraction, that learn the hierarchical spatial features of SAR images automatically only with intensity information [2], [3]. The SAR-specific physical scattering model is ignored in these methods, since many VHR SAR data is single- or dual-polarized without enough polarimetric information for physical property extraction.

In this paper, we explore the urban man-made targets extraction with deep neural networks in a new perspective, that is, learning to distinguish the different physical scattering characteristics. The main idea is inspired by studies about time-frequency analysis on SAR data [4], [5], revealing that the man-made targets in urban areas are with

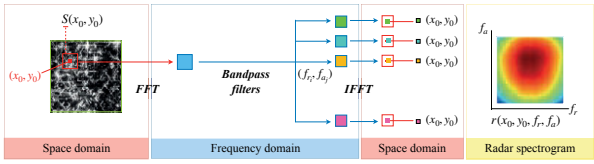

Figure 1 Short-time Fourier transform based 2-D continuous time-frequency analysis.

specific characteristics in different azimuth angles of observation and frequencies of illumination other than natural objects, especially for VHR SAR data. The complex single-looked SAR image contains the full backscattering information of both amplitude and phase values, embedded with physics messages. We propose to apply the short-time Fourier transform to extract the backscattering behavior variations of targets along range and azimuth directions as a physical characteristic representation. Then, a deep neural network is designed for learning the latent backscattering patterns automatically and exploring the distinguishing physical properties of man-made targets. Our experiments are conducted on three urban SAR images acquired by different sensors, with various bandwidths, resolutions, and incident angles.

The organization of this paper is as follows. In Section II, the proposed method is presented. The experiments and some results are demonstrated in Section III. Section IV gives the conclusion.

2 Methodology

As shown in **Figure 1**, the defined "sub-band scattering pattern" $|r|$ is derived from 2-D continuous sub-

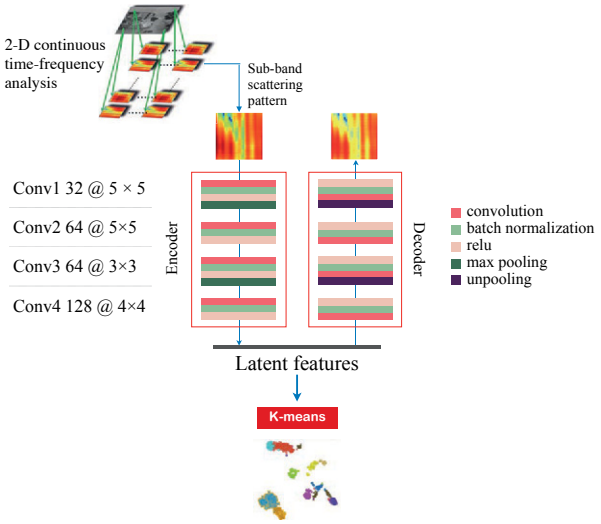

Figure 2 The proposed unsupervised deep embedding learning network.

band decomposition based on short-time Fourier transform, along range and azimuth directions. The common used sub-aperture decomposition [6] processes a set of low-resolution images which depict the phenomenon of different looks of a synthesized antenna. High resolution SAR transmits chirp signals with a large bandwidth spectrum and the sub-band decomposition in range direction is capable to obtain a series of scene reflectivity characteristics at different observation frequencies. The 2-D continuous sub-band decomposition produces a more compact signal in frequency domain that reveal the backscattering diversity along range and azimuth directions. The processing is formulated as:

$$r(x_0, y_0, f_r, f_a) = \mathrm{FFT}^{-1}\{\mathrm{FFT}(S) \cdot w(f_r, f_a)\}(x_0, y_0). \tag{1}$$

where S denotes the small area centered in position (x_0, y_0), and $w(f_r, f_a)$ denotes a series of bandwidth filters centered in frequency pair (f_r, f_a).

Consequently, a large amount of data is generated for a SAR image which makes it possible for deep learning. The sub-band scattering patterns are regarded as physical characteristics of targets. Our purpose is to discover the inherent features of sub-band scattering patterns and distinguish the man-made targets from natural areas according to their different physical properties. A deep convolutional autoencoder network, as shown in **Figure 2**, is designed for pre-training with sub-band scattering patterns in order to initiate the embedded features. The encoder part $\mathcal{G}_F$ contains four convolutional block, each with a convolutional layer, a batch normalization layer, an activation layer, and with or without a max-pooling layer. The architecture of decoder $\widetilde{\mathcal{G}}_F$ is symmetric to project the features back to input size. The optimization is conducted by minimizing the mean square error between the output of the decoder $|\hat{r}|$ and the input of the encoder $|r|$ that

$$Loss_{mse} = \sum ||\widetilde{\mathcal{G}}_F(\mathcal{G}_F(|r|)) - |r|||_2^2. \tag{2}$$

Then, k initial cluster centers are obtained with k-means algorithm. The deep embedding learning network is fine-tuned to optimize the feature extractor (encoder layers) and cluster centers simultaneously in an unsupervised manner. Specifically, for an input sub-band scattering pattern $|r_i|$, to measure the similarity between latent feature ϕ_i and initial cluster center μ_j, a soft assignment is given based on the Student's t-distribution that

$$q_{ij} = \frac{(1 + ||\phi_i - \mu_j||^2)^{-1}}{\sum_j (1 + ||\phi_i - \mu_j||^2)^{-1}}. \tag{3}$$

The soft assignment q_{ij} can be considered as the probability of assigning data $|r_i|$ to cluster j. In order to refine the cluster centroids iteratively, a Kullback-Leibler (KL) divergence distance between the soft assignment Q and the target distribution P is given for optimization:

$$Loss_{kl} = KL(P||Q) = \sum_i \sum_j p_{ij} \log \frac{p_{ij}}{q_{ij}}, \tag{4}$$

where p_{ij} is defined as

$$p_{ij} = \frac{q_{ij}^2 / \sum_i q_{ij}}{\sum_{j'} (q_{ij'}^2 / \sum_i q_{ij'})}. \tag{5}$$

The cluster center μ and the parameters in encoder layers are simultaneously updated with the gradients of $Loss_{kl}$ with respect to cluster center and encoder parameters, respectively. The implementation can be referred to [7].

In order to obtain the man-made target cluster, we borrow the idea of hierarchical clustering top-down approach [8]. The sub-band scattering patterns are firstly split into two clusters, and the significant one can be manually selected to continue clustering until the natural areas are mainly excluded.

3 Experiments

In the experiments, we explore some single-polarized (or single-polarized channel) SAR images from different sensors: Gaofen-3, Sentinel-1, and TerraSAR-X, including urban areas in Paris and Houston. **Table 1** describes the experimented data and some important parameters.

Table 1 Data Description

Urban City	Paris	Houston	Houston
Satellite	GaoFen-3	Sentinel-1	TerraSAR-X
Band	C-band	C-band	X-band
Imaging Mode	Quad Stripmap	Stripmap	Spotlight
Polarization	HH/HV/VH/VV	HH/HV	HH
Applied Channel	HH	HH	HH
Selected Area (pixel)	2854×3120	2048×2048	3072×6144
Resolution (m)	8	5	1
Incident Angle (°)	35∼37	31.2	30

The experimental setup is as follows. The TFA analysis window size of complex-valued SAR image is set to 32×32 and the continuous bandpass filters group is with a bandwidth of half full-aperture bandwidths both in range and azimuth. The encoder-decoder architecture pre-training is implemented by stochastic gradient descent

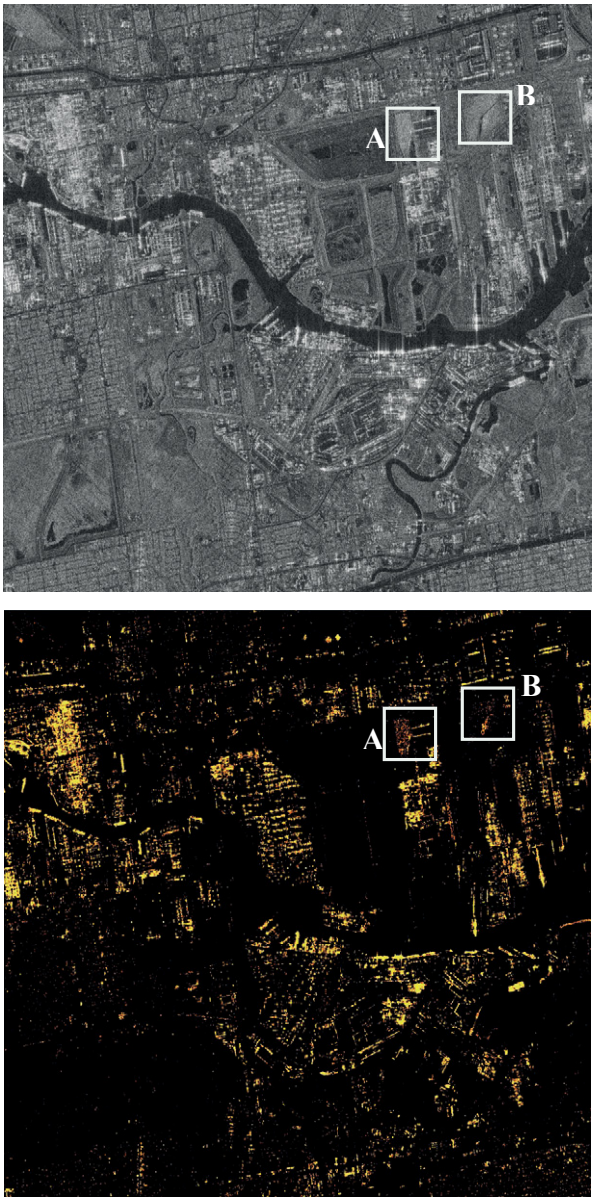

Figure 3 The urban area in Houston, USA, acquired from Sentinel-1 satellite in Stripmap mode.

(SGD) optimizer, with a weight decay value 0.0005 and momentum 0.9, and a fixed learning rate of 0.01. The learning rate of fine-tuning the feature extractor and cluster centers is set to 10^{-4}. The hierarchical clustering iteration number is set to 4.

Figure 3, **Figure 4**, and **Figure 5** show the experimented single-polarized amplitude image and the generated man-made targets map of Sentinel-1, Gaofen-3, and TerraSAR-X SAR data, respectively. In each figure, the final iteration result of clustering is shown and the man-made targets are marked with red and yellow colors. Gaofen-3 SAR image has the lowest resolution of 8 meters among the three, so that there could be multiple man-made targets in one resolution cell. The sub-band scattering pattern characteristics could be less representative. With a highest resolution of 1 meter of TerraSAR-X Spotlight data, **Figure 5** show a very clear extraction result of man-made targets. For comparison, more natural targets could be mixed in the clustering result in Sentinel-1 and Gaofen-3 cases.

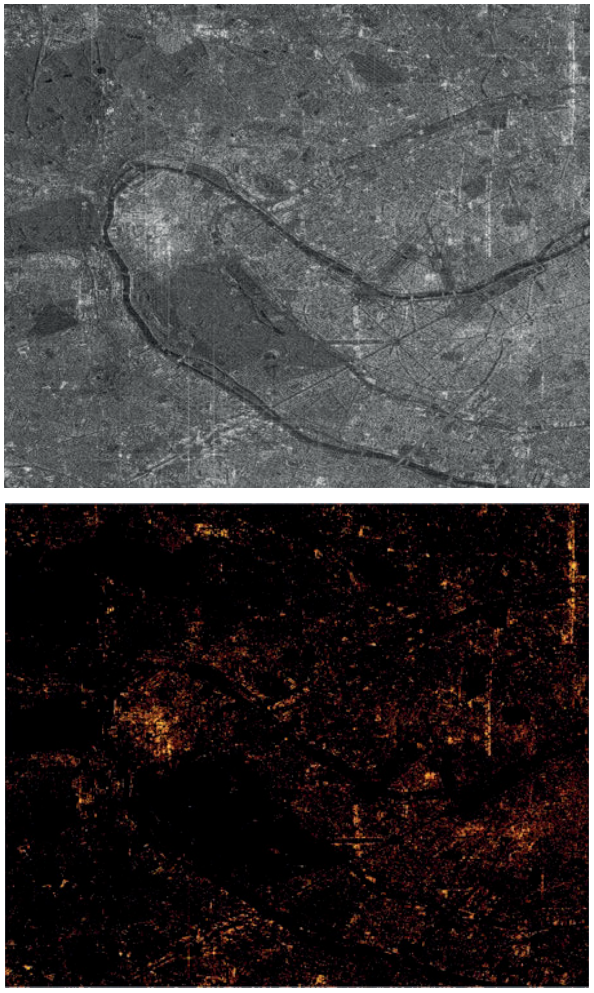

Figure 4 The urban area in Paris, France, acquired from Gaofen-3 satellite in Quad-stripmap mode.

Also, we select some specific man-made targets in urban areas to demonstrate what our proposed method could learn from single-polarimetric SAR data without any texture information. As shown in **Figure 6**, the selected patch A and B from Figure 3 are both with bright pixels but different targets. In patch B, the trains on railways can be hardly distinguished from the surrounding natural vegetation visually. However, our proposed method has detected the "hidden" trains, as well as trains in patch A. **Figure 7** shows a metallic dock with severe side-lobes which are as bright as the target. Nevertheless, the proposed method has extracted the target and removed the interference.

Besides, the proposed method can tell the multi-path scattering characteristics as shown in **Figure 8**. In this figure, the bridge represents a main scattering, double-bounce, and two other multiple bounce scattering. The bright pixels may fool the image-based machine learning methods to decide the multi-path scattering as true targets. However, the proposed method can extract the main scattering marked as yellow and most double-bouncing scattering as red. The other multiple scattering characteristics (third and fourth) are more different from the previous ones, so that they are

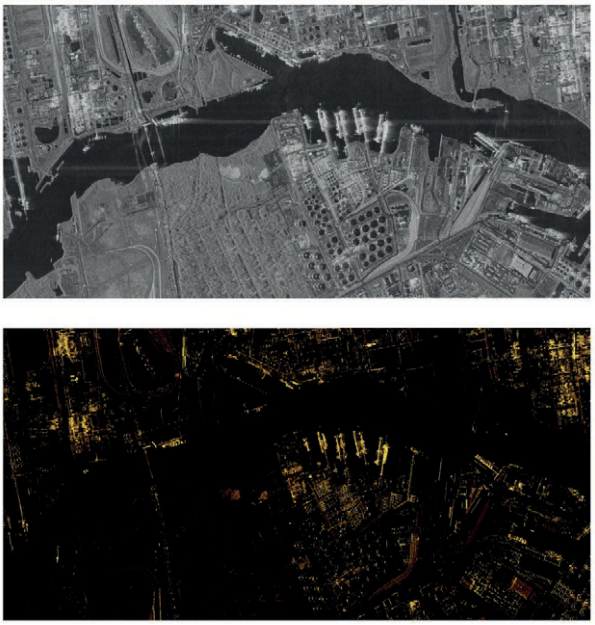

Figure 5 The urban area in Houston, USA, acquired from TerraSAR-X satellite in Spotlight mode.

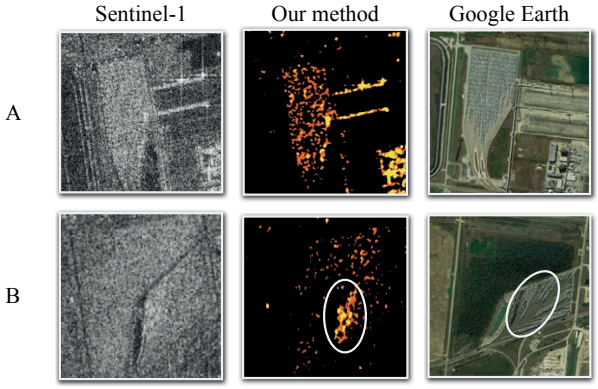

Figure 6 A and B from Figure 3 are visually similar with bright pixels. The trains/railways in B could not be easily distinguished from the surrounding environments, but can be extracted by our proposed method.

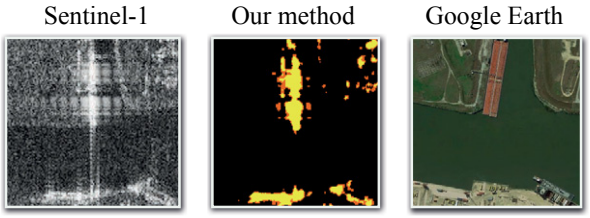

Figure 7 Our method can extract the man-made targets while ignoring the severe side-lobes with high intensity in SAR images.

not detected finally.

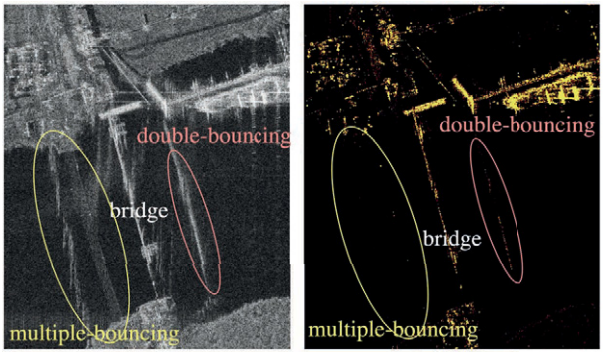

Figure 8 The man-made target bridge and its multi-path scattering characteristics.

4 Conclusion

The proposed unsupervised deep learning method focuses on the physical scattering property extraction of targets in complex-valued single-polarized SAR image. Based on 2-D continuous time-frequency analysis, the sub-band scattering pattern is generated to represent the physical characteristics of targets. The neural network is designed for automatically learning the features of scattering patterns and distinguish the different natural areas and man-made targets. As three different satellite SAR images are experimented, the proposed method is proved to be effective in extracting meaningful characteristics of man-made targets without any texture information. This paper is a previous study of our work in [9], where more detailed information and further exploration of the method and analysis are presented. In the future study, we will improve the urban mapping method and further explore the physics interpretability of the neural network based method.

5 Literature

[1] Ratha, D.: Novel Techniques for Built-Up Area Extraction From Polarimetric SAR Images. IEEE Geoscience and Remote Sensing Letters, Vol. 17, No. 1, 2020, pp 177181. https://doi.org/10.1109/LGRS.2019.2914913

[2] Zhen Xu: Building extraction from high-resolution SAR imagery based on deep neural networks, Remote Sensing Letters. Vol. 8, No. 9, 2017, pp. 888-896

[3] Li, J: Multiscale convolutional neural network for the detection of built-up areas in high-resolution SAR images, 2016 IEEE International Geoscience and Remote Sensing Symposium (IGARSS). IEEE, 2016, pp. 910-913.

[4] Tison, C.: Point target behaviour in high resolution sar images: time-frequency versus polarimetric analysis, EUSAR, 2006

[5] Spigai, M.: Time-frequency analysis in high-resolution SAR imagery, IEEE Trans. Geosci. Remote Sens. Vol. 49, No. 7, 2011, pp. 2699-2711

[6] Tupin, F.: Sub-aperture decomposition for SAR urban area analysis, 2004 EUSAR, pp. 431434, 2004.

[7] Xie, J.: Unsupervised Deep Embedding for Clustering Analysis, Proceedings of The 33rd International Conference on Machine Learning, Vol. 48, 2016, pp. 478-487, New York

[8] Liao, K.: A sample-based hierarchical adaptive K-means clustering method for large-scale video retrieval. Knowledge-Based Systems, 2013, 49: 123-133.

[9] Huang, Z.: HDEC-TFA: An Unsupervised Learning Approach for Discovering Physical Scattering Properties of Single-Polarized SAR Image, IEEE Trans. Geosci. Remote Sens., pp. 118, 2020, doi: 10.1109/TGRS.2020.3014335.