DSPANet: A deeply supervised pseudo-siamese attention-guided network for building change detection with intensity and coherence information of SAR images

Peng Chen[1,2,3], Qing Zhao[1,2,3,*]

[1]Key Laboratory of Geographical Information Science, Ministry of Education, East China Normal University, Shanghai 200241, China; [2]School of Geographic Sciences, East China Normal University, Shanghai 200241, China; [3]Key Laboratory of Spatial-Temporal Big Data Analysis and Application of Natural Resources in Megacities, Ministry of Natural Resources, Shanghai 200241, China

Abstract

Change detection (CD) is to quantitatively analyze and determine the characteristics and process of earth surface change based on remote sensing data in different periods. It is widely used in disaster dynamic detection, urban planning, and other fields. Compared with optical remote sensing images, Synthetic Aperture Radar (SAR) images have the unique advantage for CD. As active remote sensing systems, SAR has all-weather, day-night imaging capability, and permits synoptic views of large areas [1]. The backscatter intensity information of SAR images in urban areas is easily affected by comprehensive factors such as building layout, orientation, and surface materials, resulting in overall low radar echo intensity in some building areas. The introduction of coherence maps into change detection can improve the recognition of changed areas because urban buildings have high phase stability and coherence maps provide reliable information for building change detection (BCD). In this study, we mainly focus on two problems in practical application. First, speckle noise as an inherent characteristic of SAR images greatly influences the performance of BCD. Second, how to fuse intensity information and coherence information effectively is of significance for extracting the spatial features of changed areas. Thus, this study proposes a deeply supervised pseudo-siamese attention-guided network (DSPANet) for BCD, in which convolutional blocks have a strong ability in noise reduction owing to the large receptive fields [2], and the adopted pseudo-siamese structure does well in extracting intensity information and coherence information with the same network branch but different weights.

Keywords: Urban change detection; Data fusion; Deep learning; SAR;

1. Introduction

Deep learning is a very successful method in the field of machine learning, It can learn complex and deep features through large networks and network models can be optimized quickly through backpropagation. There are already some studies of change detection based on deep learning. In [3], a dual stream U-Net was proposed for the fusion of SAR and optical data. Sun et al.[4] substituted the convolutional layer of U-Net with Conv-LSTM to form a new architecture L-UNet. In this study, we use DSPA-Net model to learn more in-depth features. the convolutional block attention module (CBAM) is included to assign accurate labels to each pixel through the attention mechanism, thus enhance the feature learning of the network. The multi-level feature fusion (MFF) blocks with residual structure are used to avoid the gradient disappearance and gradient explosion caused by the deepening of the network. Amplitude information and coherent information are entered into the DSPA-Net model for fusing more features.

2. Methodology

Different from FCNs used for image semantic segmentation, DSPA-Net (Figure 1) is a dual-branch end-to-end network. Each branch with independent weights is a feature extraction network (FEN) that extracts individual features from intensity images and coherence maps, respectively. Then, the extracted features are transmitted to the Decoder layer to detect building changes. Notably, we use strided convolutions rather than poolings to achieve the downsampling of feature maps, aiming to facilitate the preservation of details.

In the Encoder layer, except for the first one, the other three stages contain a strided convolutional layer, two convolutional layer, and a CBAM block (Figure 3). CBAM block leads to increased attention on building changes. The output features of two branches are concatenated and fed into the Bottleneck layer for expanding the receptive fields. In the Decoder layer, first, a transpose convolutional layer is applied for upsampling purposes in each satge. Then, the output of CBAM block in the Encoder layer is concatenated with the upsampling features, and the concatenated features are fed into two sequential convolutional layers to extract deep features. Further, the MFF block (Figure 2) is applied to fuse multi-level features and remove the problem of vanishing and exploding gradients caused by the deepening of the network.

Besides the supervision on the output layer of the backbone network, DSPA-Net also introduces three deep supervision branches (DS1 and DS2) with the same structure. Note that change maps generated by deep supervision branches are only used for auxiliary training networks, and the final results of DSPA-Net are obtained at the stage3.

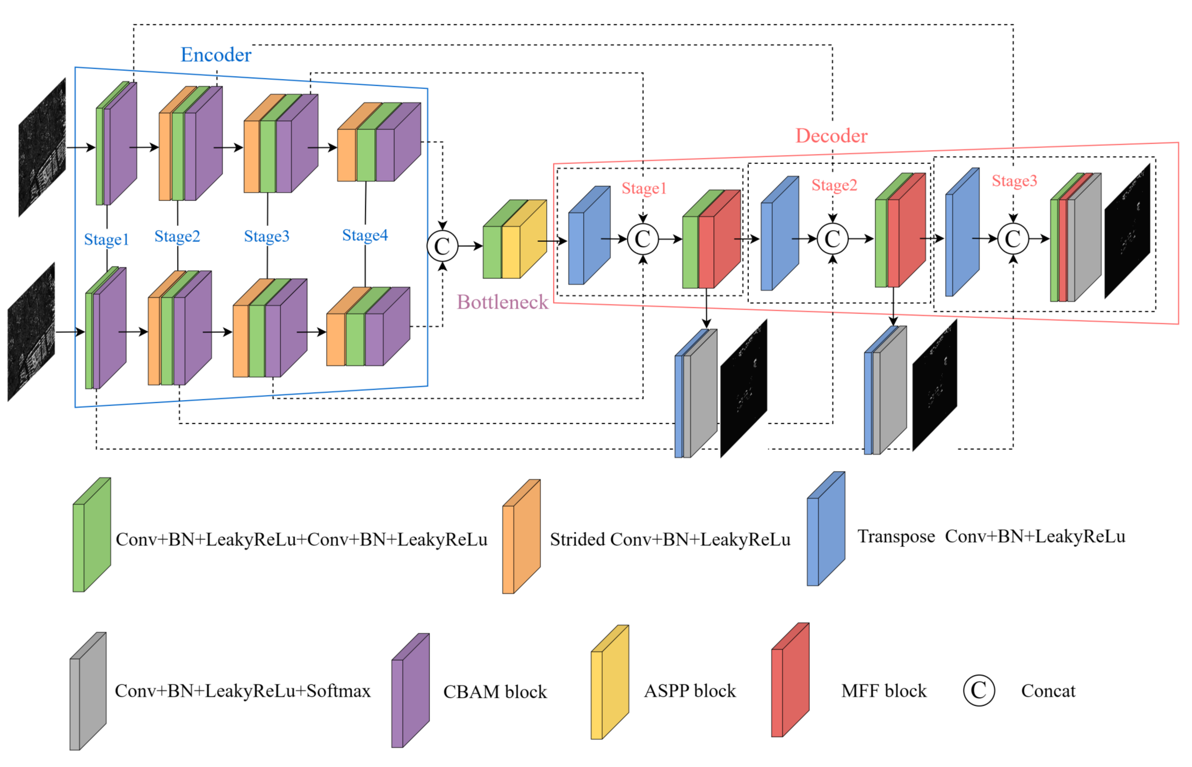

Figure 1. The structure of the DSPA-Net block

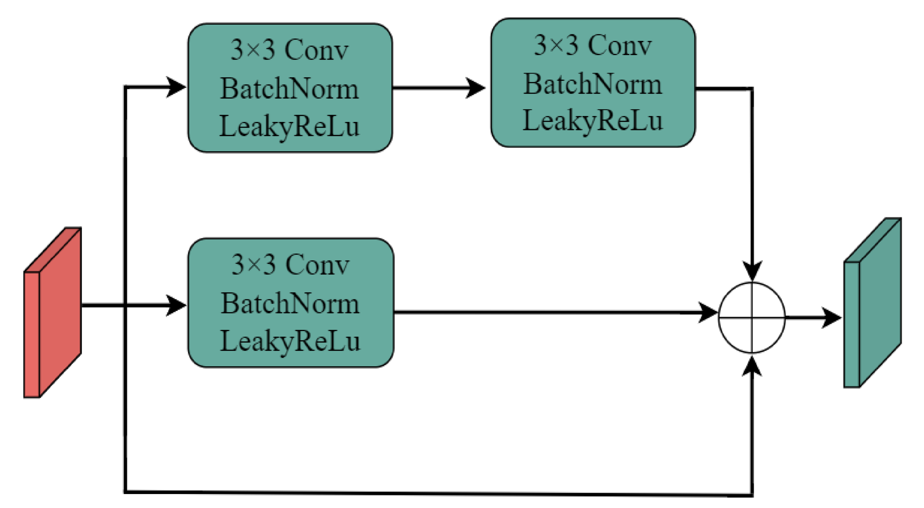

Figure 2. The structure of the MFF block

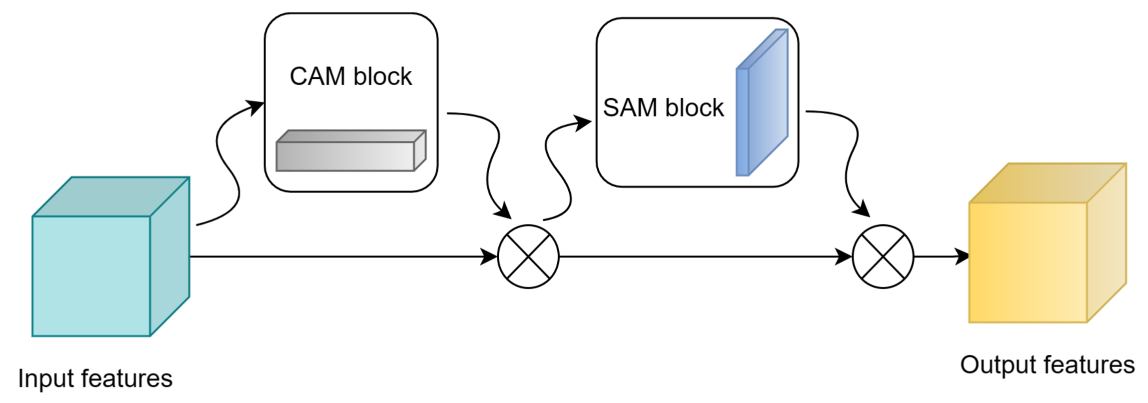

Figure 3. The structure of the CBAM block

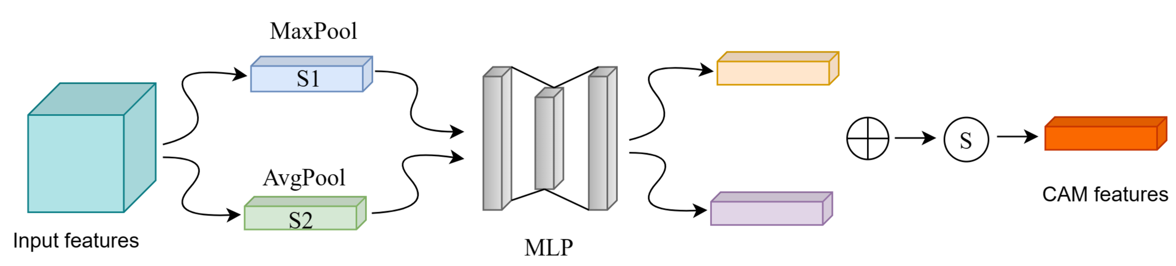

Figure 4. The structure of the CAM block

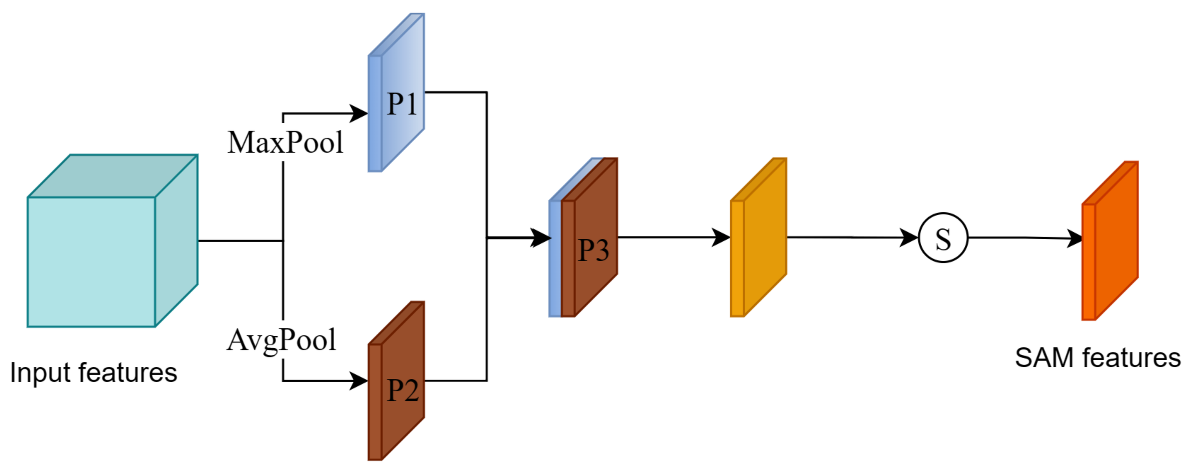

Figure 5. The structure of the SAM block

3. Experiments and results

A. Experimental Data

This study uses high-resolution TerraSAR-X (TSX) images covering Shanghai from a descending pass track. A set of four TSX images is acquired between 16 June 2019 and 10 September 2021. Before training the network, we pre-process TSX images to get intensity information and coherence information. Intensity information is obtained by pre-processing single-look complex (SLC) images, including radiometric correction, multi-looking processing, and geocoding. Coherence map (CM) is obtained by performing interferometry and computing coherence values of interferograms. PWR and CM are shown in Figure 6.

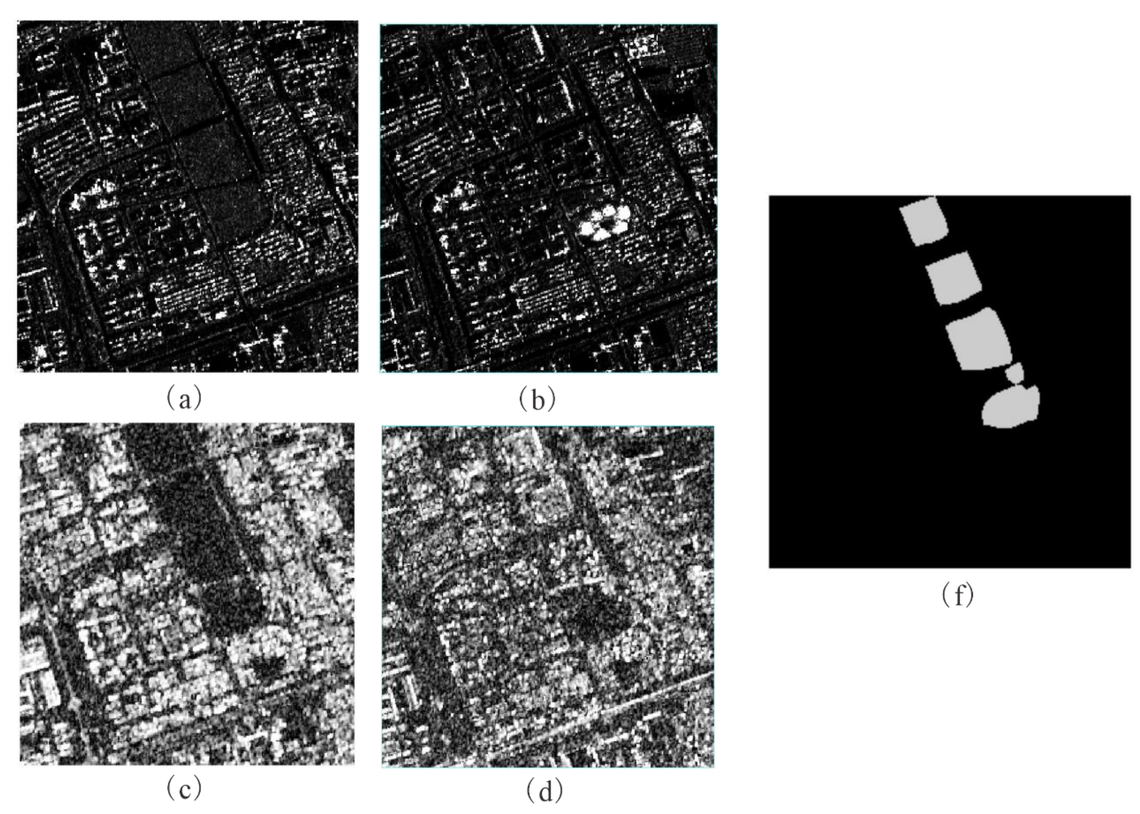

Figure 6. Data set 1. (a)Pre-change SAR intensity image. (b)Post-change SAR intensity image. (c)Pre-change coherence map. (d)Post-change coherence map. (e) Ground-truth image

B. Result and Accuracy

In this poster, the traditional U-Net method is compared with the proposed DSPA-Net. For a qualitative comparison, the visual interpretation of some BCD results derived from testing samples is displayed in Figure 7. There exist some missing detections in the results of U-Net, leading to some difficulties in detecting detailed changes, whereas the proposed DSPA-Net obtains the best visual performance. To quantitatively analyze the results, Figure 8 lists the evaluation values of these two methods and different input data. The U-Net model yields an improvement of 13.76%, 41.81%, 33.74%, 34.88%, and 2.78% for Precision, Recall, F1, Kappa, and Accuracy, respectively while adding the coherence maps into the U-Net model. In addition, the performance of the DSPA-Net model is superior to the U-Net model, with a gain of 1.4%, 6.17%, 3.97%, 4.19%, and 0.43% for Precision, Recall, F1, Kappa, and Accuracy, respectively.

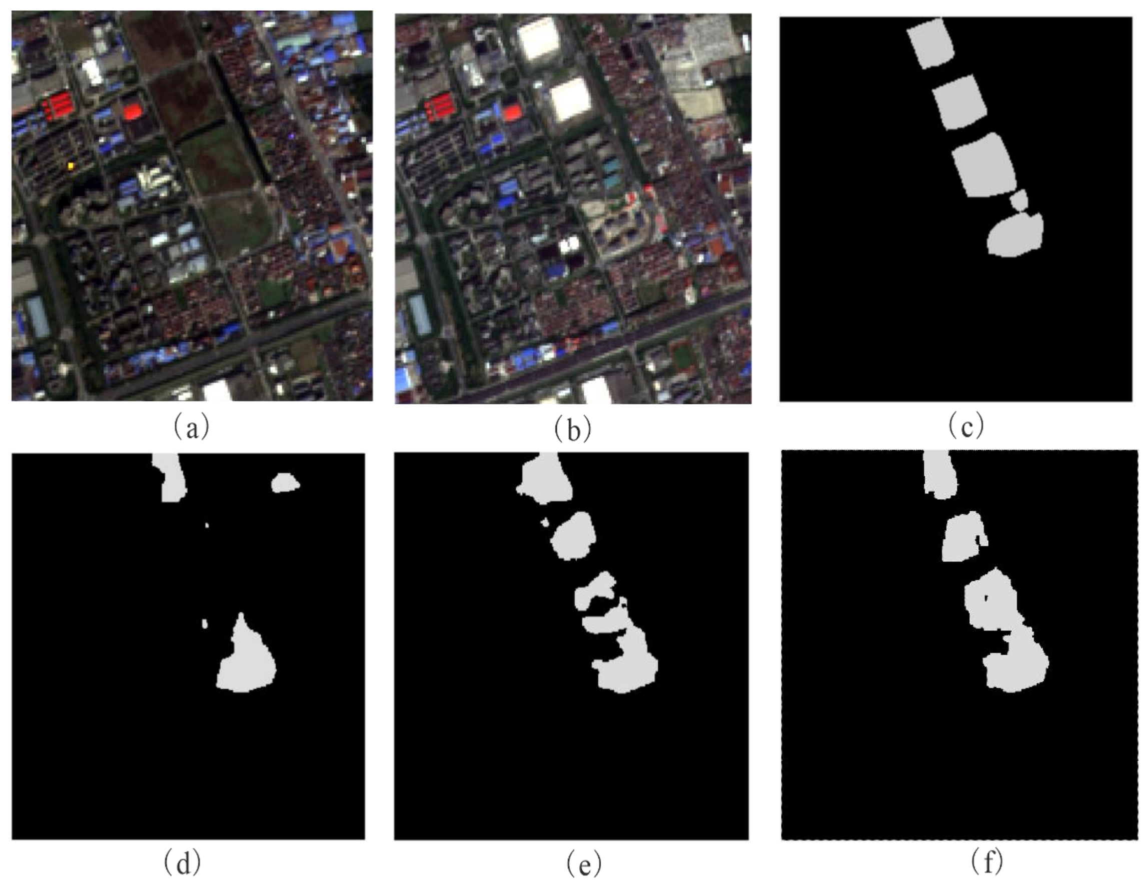

Figure 7. the change detection result of data set 1. (a) Pre-change optical image. (b) Post-change optical image. (c) Ground-truth image. (d) U-Net, PWR as input data. (e) U-Net, PWR and CM as input data (f) DSPA-Net, PWR and CM as input data

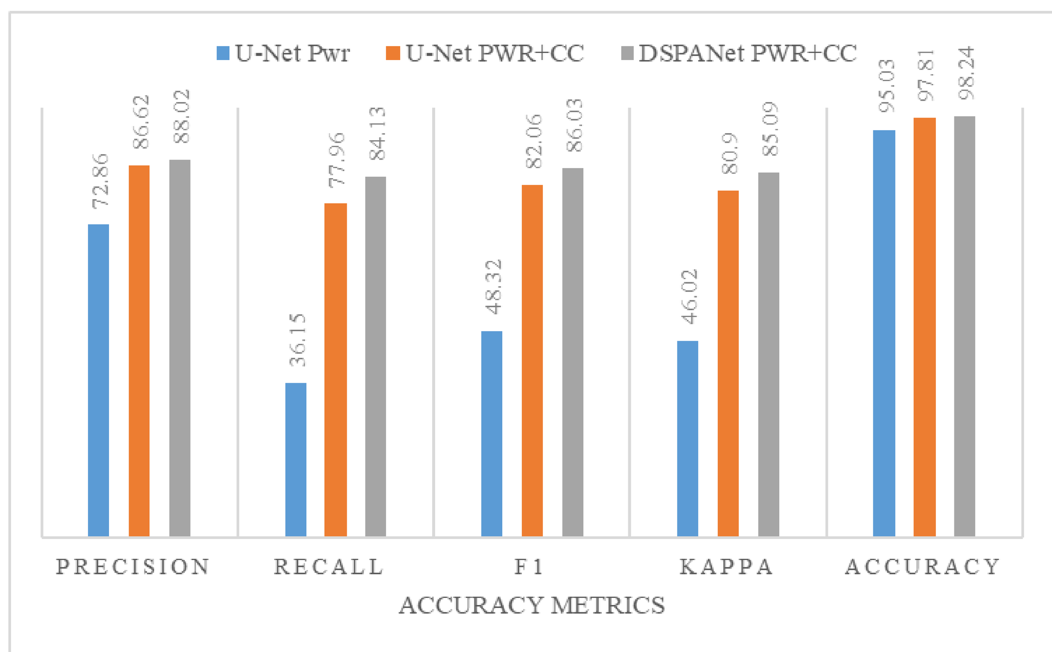

Figure 8. Comparisons of detection methods based on data set 1. PWR and PWR+CM represent the different sources of input data

4. Conclusion

In this poster, we presented an urban BCD approach that uses a network architecture for the joint use of PWR and CM data. Different from other BCD experiments, this study utilizes CM data that can judge whether the ground objects have changed. Compared with non-artificial objects, urban buildings have very high phase stability characteristics, so the interference coherence provides very reliable information for change detection. Besides, for coherence maps with long time intervals and short spatial vertical baseline conditions, urban buildings can still maintain high coherence coefficient. On the contrary, non-artificial objects may be completely decoherent due to the influence of temporal decoherence. Therefore, SAR coherence information can be introduced into urban BCD. The DSPA-Net model aggregates the information of both and extracts their features in order to train the network effectively. The results show that the method is reliable in urban change detection and performs better than other advanced change detection methods. Its Precision, Recall, F1, Kappa, and Accuracy reach 88.02%, 84.13%, 86.03%, 85.09%, and 98.24, respectively.

REFERENCES

[1] Zhuang H, Hao M, Deng K, et al. Change detection in SAR images via ratio-based Gaussian kernel and nonlocal theory[J]. IEEE Transactions on Geoscience and Remote Sensing, 2021, 60: 1-15.

[2] Qin T, Wu K, Xiu D. Data driven governing equations approximation using deep neural networks[J]. Journal of Computational Physics, 2019, 395: 620-635.

[3] Hafner S, Nascetti A, Azizpour H, et al. Sentinel-1 and Sentinel-2 data fusion for urban change detection using a dual stream u-net[J]. IEEE Geoscience and Remote Sensing Letters, 2021, 19: 1-5.

[4] Sun S, Mu L, Wang L, et al. L-UNet: An LSTM network for remote sensing image change detection[J]. IEEE Geoscience and Remote Sensing Letters, 2020.