A Novel Deep Learning Framework Based on Transfer Learning and Joint Time-Frequency Analysis

Z. Huang, C.O. Dumitru, Z. Pan, B. Lei, and M. Datcu

ABSTRACT:

We propose a novel SAR-specific deep learning framework Deep SAR-Net (DSN) for complex-valued SAR images based on transfer learning and joint time-frequency analysis. Conventional methods for deep convolutional neural networks usually take the amplitude information of single-polarization SAR images as input to learn hierarchical spatial features automatically, which may have difficulties in discriminating objects with similar texture but with discriminative scattering patterns. As a result, we analyzed complex-valued SAR images to learn both spatial texture information and the backscattering patterns of objects on the ground.

Firstly, we experimented on a large-scale SAR land cover dataset collected from TerraSAR-X images, with a hierarchical three-level annotation of 150 categories and comprising more than 100,000 image patches. With three main challenges of highly imbalanced classes, geographic diversity, and label noise, in automatically interpreting the dataset, a deep transfer learning method based on a similarly annotated optical land cover dataset (NWPU-RESISC45) was used to learn a deep Residual convolutional neural network, optimizing a combined top-2 smooth loss function with cost-sensitive parameters. Rather than applying the ImageNet pre-trained model of ResNet-18 to SAR images directly, the optical remote sensing land cover dataset narrows the gap between SAR and natural images which results in a significant improvement in feature transferability, and the proposed combined loss function is successful in accelerating the training process, and is reducing the model bias to noisy labels. The trained deep Residual CNN model shows a good generalization for other SAR image processing tasks, including MSTAR target recognition, land cover, and land use localization.

Based on this pre-trained model, we transferred the first two residual blocks to extract the mid-level representative spatial features from the intensity images of single-look complex (SLC) SAR data, which have a similar resolution and pixel spacing along range and azimuth directions to avoid large distortions. Then, a joint time-frequency analysis was applied to SLC data to obtain a 4-D representation with information in all sub-bands, where the radar spectrograms reveal the backscattering diversity versus range and azimuth frequencies of objects on the ground. A stacked convolutional auto-encoder was designed to learn the latent features from the radar spectrograms in the frequency domain, related to physical target properties. Later, the frequency features were spatially aligned corresponding to the spatial information in the 4-D representation to be fused with the transferred spatial features. A post-learning sub-net consisting of two bottleneck residual blocks was designed to make the final decisions.

This is the first time to exploit the full use of single-polarization SLC SAR data in deep learning. Compared with conventional CNNs which are based on intensity information only, the proposed DSN

shows a superior performance in SAR image land cover and land use classification, especially for man-made objects. In some cases, the shapes and textures are similar to intensity images which confuse CNNs to make a right decision, but the spectrogram amplitudes present prominently different characteristics, helping DSNs to reach a better understanding of the objects on the ground. On the other hand, for natural surfaces, the radar spectrograms present similar backscattering patterns without a specific mechanism for distinguishing the features in the frequency domain, so that they cannot provide enough extra information on natural surfaces to support the interpretation of SAR images. The experiments are conducted on Sentinel-1 Stripmap SAR images and we believe the proposed DSN can be also applied to TerraSAR-X SLC data.