Visual Question Answering on Multiple Remote Sensing Image Modalities

Supplementary Material

SAR data (Sentinel-1) pre-processing

The pre-processing pipeline to generate continuous SAR (Sentinel-1) patches consists of several steps including SAR and VHR data. We describe hereinafter each step in details:

1. We use the algorithm presented in [33] to find the position of the center of the VHR patch in the SAR image. To do so, we need the geographical position of the center of the VHR patch. This position must include altitude, due the geometrical distortions inherent in SAR, particularly layover;

2. The latitude and longitude of the VHR patches are extracted from the meta-data of the image. Using this latitude and longitude, the altitude is given by this https://geoservices.ign.fr/documentation/services/services-deprecies/calcul-altimetrique-rest and these information are stored in a JSON file, *latlonalt.json*;

3. The Sentinel-1 Single Look Complex (SLC) images are separated in three swath. To find the correct swath, the projection of the geographical point is applied, using the meta-data linked to each swath. The only swath for which the algorithm returns a valid position is selected;

4. In each swath, the S1 image is divided in different overlapping bursts, that are all stored in the same file separated by black lines. The S1 images need to be debursted (removing of the black line and of the overlap) to get a continuous image before to extract the S1 patch that is inputted in the model. This debursting is done using meta-data of the S1 image and by comparing the two consecutive burst. A result of the debursting process is shown in Figure 4. The debursted images are provided with the dataset. The projection algorithm is thus modified to give the position in this new deburst file. This position and the correct swath are stored in the JSON file containing all information on the VHR patches;

5. When a patch of size $L_{\text{SAR}} \times L_{\text{SAR}}$ (L_{SAR} specified by the user) is extracted, the two polarimetric channels are converted in dB, and the ratio is computed. A tail-value elimination procedure is performed on each channel separately using statistics information extracted over the whole dataset.

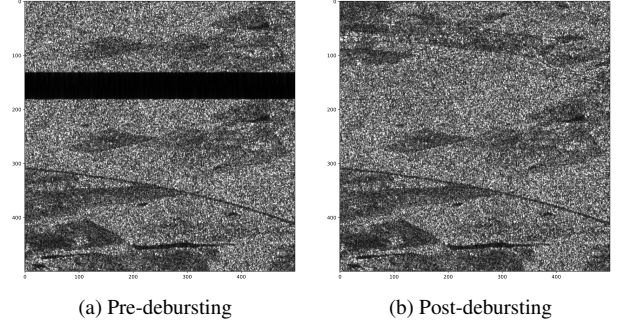(a) Pre-debursting   (b) Post-debursting

Figure 4. Sentinel-1 SLC images before (a) and after (b) the application of the proposed debursting method. The visualization is done using a threshold of 233.