International Symposium on Grids & Clouds 2018 (ISGC 2018) in conjunction with Frontiers in Computational Drug Discovery (FCDD)

Contribution ID: **21**   Type: **Oral Presentation**

Towards a crowdsourcing platform for labelling remote sensing images online

Wednesday, 21 March 2018 17:00 (30 minutes)

Vast quantities of remote sensing data are becoming available at an ever-accelerating rate, and it is transforming geosciences today. When investigating global-scale environmental phenomena, large scale of temporal and spatial remote sensing data need to be processed to extract useful information, especially the land use types. As most powerful remote sensing image processing methods need labeled training dataset, and there is no such open remote sensing image training dataset, researchers have to label the image all by themselves. Sometimes, researcher have to download and label large volume of remote sensing images to satisfy their machine learning methods, such as deep learning. This is not only time-consuming, but also a waste of researchers energy.

Based on our past work of having built a cloud-based platform for massive remote sensing data retrieving, downloading and processing, we have accumulated a large volume of remote sensing data, and more than 187 thousand users. Crowdsourcing is a way of utilizing the human intelligence of citizens. As it has been proved to be an effective way to tackle tasks that are difficult to be automated processed, such as Wikipedia, we decide to try to collect training datasets for remote sensing image processing through crowdsourcing. It not only can fully utilize the human intelligence of the large number of users, but also can build a large remote sensing labeled dataset of great value.

In this paper, we propose an on-line remote sensing image labeling platform. Through it, users neither have to install any software, nor download large remote sensing images. Remote sensing images are displayed on the web, and users can directly label the images by just drawing polygons and add tags. Tags are predefined. All the labeled dataset can be downloaded and shared with other people. Moreover, standards of the training dataset have been studied. And two components of the labeled data are stored, which are the true image values of the labeled region and their metadata, such as the bounding box, tag, and location information. In this way, the training data can be used directly for a variety of machine learning models. By implementing such an online remote sensing labelling platform, participants not only can visualize, label, download, and share training datasets only with a simple web-connection, but also will have a better understanding of the land types on our planet.

Primary author: Ms ZHAO, Jianghua (Computer Network Information Center, Chinese Academy of Sciences)

Co-authors: Dr LIN, Qinghui (Computer Network Information Center, Chinese Academy of Sciences); Dr WANG, Xuezhi (Computer Network Information Center, Chinese Academy of Sciences); Mr ZHOU, xiaohua (Computer Network Information Center, Chinese Academy of Sciences); Prof. ZHOU, yuanchuan (Computer Network Information Center, Chinese Academy of Sciences)

Presenter: Ms ZHAO, Jianghua (Computer Network Information Center, Chinese Academy of Sciences)

Session Classification: Earth, Environmental Science & Biodiversity Session

Track Classification: Earth & Environmental Sciences & Biodiversity Applications