**Multimodal model-based water body extraction algorithm on SAR imagery**

Lin Y. 1, He L.1\*, Song Y.F.1 and Zhong D.Q.1

1College of Surveying and Geo-informatics, Tongji University, Shanghai 200092, China

[\*helin19950830@gmail.com](mailto:*helin19950830@gmail.com) (\*Corresponding author’s email only)

***ABSTRACT***

The application of remote sensing technology for the extraction of inland water bodies has proven to be an effective method for large-scale water body investigations. Synthetic Aperture Radar (SAR) images offer the advantage of all-weather observation, unlike multispectral remote sensing images, which suffer from data quality degradation due to cloud cover. Whereas, the efficient and accurate extraction of water bodies from SAR data, which is characterized by limited spectral information, remains a significant challenge. This article proposes an image segmentation algorithm based on a multimodal deep neural network model for extracting inland water bodies from SAR images. Initially, the method achieves feature alignment of SAR images and text based on the Transformer-based multimodal network. Subsequently, the image features of SAR images are encoded through a linear classifier applied to the scene classification task. Finally, an image decoder, integrating Transformer and Convolutional Neural Network (CNN) structures, generates the results for the inland water body segmentation task. Experimental results using Sentinel-1 satellite data demonstrate that the multimodal feature encoder can effectively align SAR images with text. More than 90% precision and recall for the defined scene classification task can be achieved by the SAR image-text model. The proposed segmentation algorithm has also been validated for the extraction of inland water bodies.

**Keywords:** SAR image, multimodal model, segmentation, inland water extraction