Abstract

Our study delves into harnessing the spatial awareness and performance capabilities of pre-trained Vision Transformers (ViTs) for remote sensing-based Earth observation, with a specific focus on coral reef extent mapping—a crucial task in assessing climate change impacts on vulnerable ecosystems. Utilizing the pre-trained DinoV2 ViT as our frozen encoder, we constructed three semantic segmentation heads for reef extent mapping: a linear head, a U-Net head, and a UNETR head. Through rigorous evaluation against a baseline U-Net model, our experiments, conducted using reflectance data from the Sentinel-2 satellite and ground truth data from the Allen Coral Atlas, revealed competitive performance of transformer-based models. Notably, the UNETR head showcased superior performance, achieving a significant improvement in both the F1 Score and the IoU metric. Addressing the limitation of ViTs in processing multi-spectral data, we enhanced the UNETR architecture to incorporate bathymetric information, resulting in further performance gains. Our findings suggest that the UNETR with a DinoV2 backbone could pave the way for a new era of transformer-based segmentation models in Earth observation, offering promising avenues for addressing critical environmental challenges such as coral reef monitoring and conservation.