A Cascade Transformer-Based Multi-Scale Framework for Object Detection and Instance Segmentation in Remote Sensing Imagery

Ruiqian Zhang, Qin Yan, Xiaogang Ning and Hanchao Zhang (China, PR)

Key words: Remote sensing; Young surveyor

SUMMARY

Object-level land-cover detection and segmentation in remote sensing imagery are crucial for accurately identifying the location and distribution of land-cover types on the Earth's surface. This technology has wide applications in fields such as land-use surveys, urban planning, environmental monitoring, and disaster assessment. However, remote sensing imagery often presents complex backgrounds and objects of varying scales, which pose significant challenges for achieving accurate object detection and instance segmentation. To address these challenges, we propose a unified Cascade Transformer-Based Multi-Scale Framework for object detection and instance segmentation in remote sensing imagery. The framework employs a cascade encoder-decoder structure, integrating an object-level detection head and an instance segmentation branch. Specifically, the encoder-decoder utilizes a Swin Transformer backbone combined with a MaskDINO head, while a lightweight Feature Pyramid Network (FPN) is used to align multi-scale features effectively. Furthermore, we design a semantic hybrid loss function, which combines localization and classification losses for object detection with mask losses for segmentation tasks, effectively addressing class imbalance and the challenges of multi-task learning. We validate the proposed framework on publicly available datasets for remote sensing change detection and conduct comparative experiments against state-of-the-art methods. The results demonstrate that our framework achieves superior performance in complex remote sensing scenarios, significantly improving the accuracy and robustness of object-level detection and segmentation. Notably, our approach secured first place in the ISPRS International Contest on Individual Tree Crown (ITC) Segmentation, highlighting its effectiveness and potential for practical applications.
