WORKING WEEK 2025 AND Locate25 THE NATIONAL GEOSPATIAL CONFERENCE

FIG Geospatial Council of Australia

Collaboration, Innovation and Resilience: Championing a Digital Generation

Brisbane, Australia 6–10 April

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

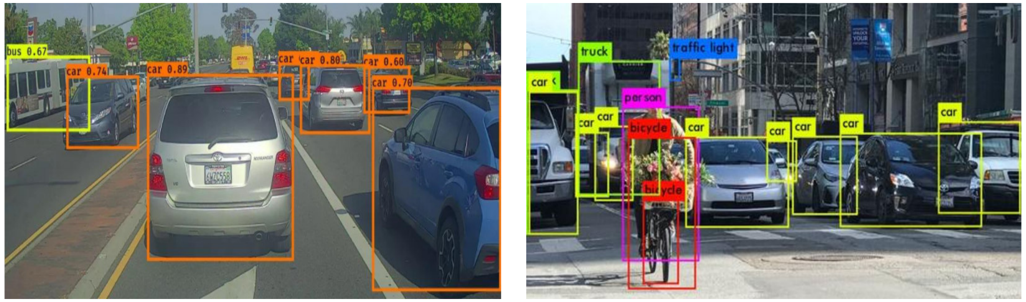
Authors: Ruiqian Zhang, Qin Yan, Hanchao Zhang, Xiaogang Ning

Affiliations: Institute of Photogrammetry and Remote Sensing, Chinese Academy of Surveying and Mapping, China

1. Background and Introduction
2. The Proposed Method
3. Experiments and Results
4. Conclusions

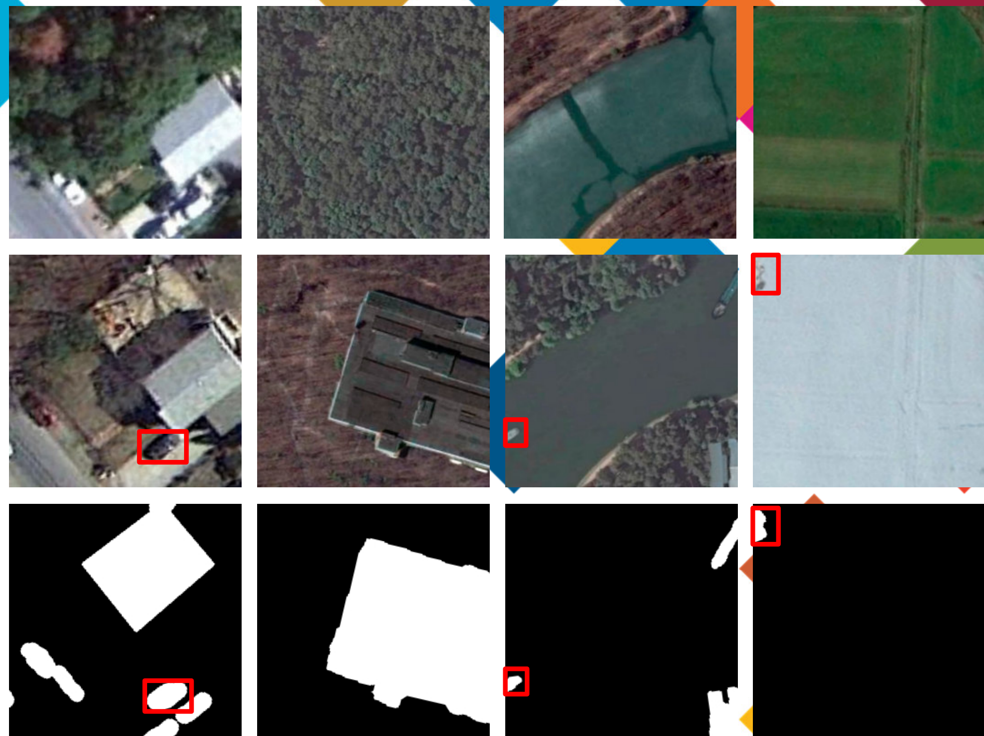
BACKGROUND AND INTRODUCTION

Object detection in natural images



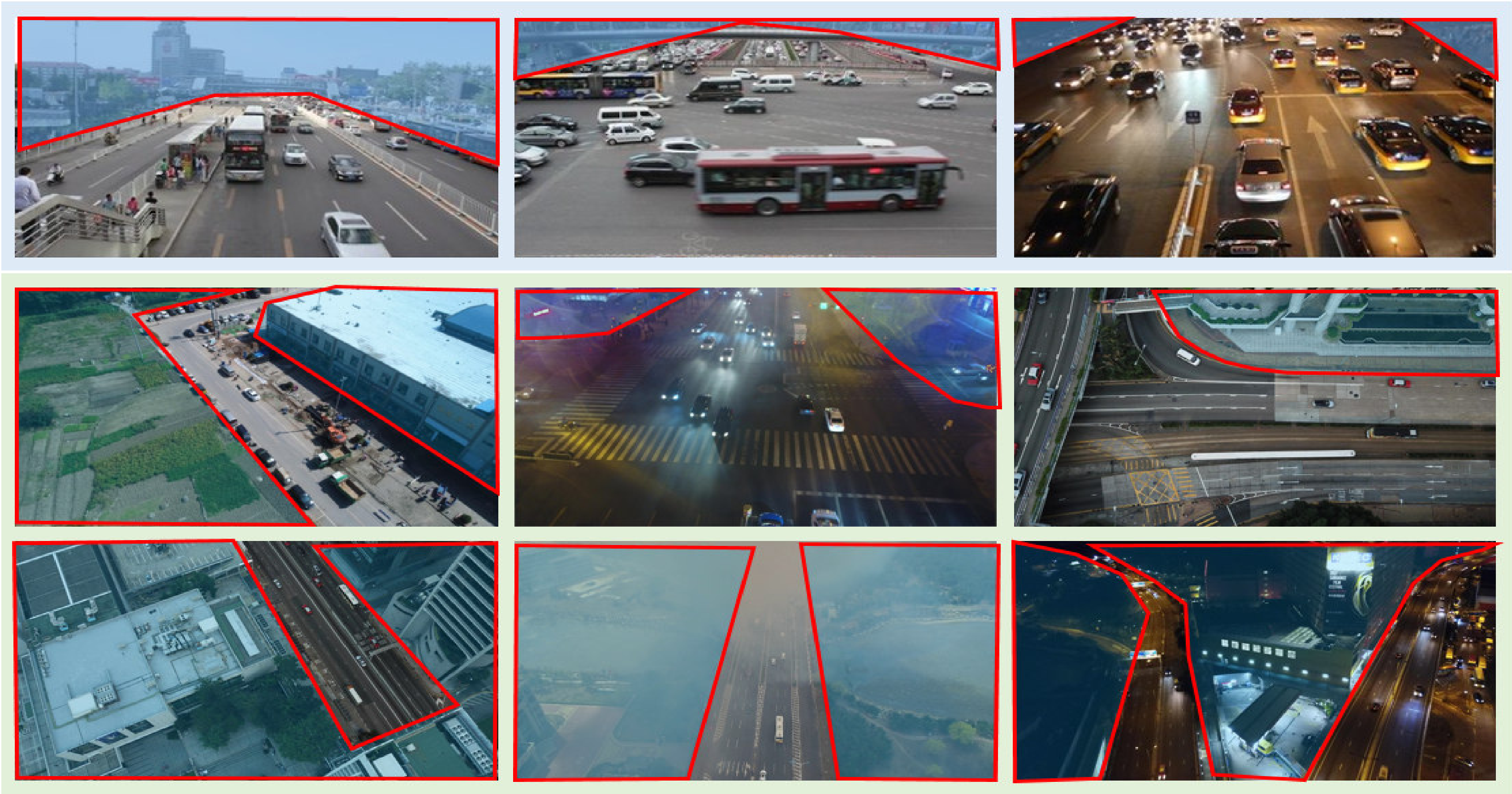
Object detection in RS images

Change detection in RS images



Background regions in Natural Images

Background regions in Remote Sensing Images



Research background

Current deep learning-based change detection methods can be categorized into **pixel-level** and **object-level** methods

Pixel-level methods → Can obtain **high detection accuracy**, but it is difficult to **distinguish each change object** when objects are densely distributed

Object-level methods → Can **distinguish changed objects**, but it is difficult to **obtain accurate boundary representation**

Motivation

Developing fine-grained object-level change detection with accurate boundary and distinguishing individual instances

Research background

Current deep learning-based change detection methods can be categorized into **pixel-level** and **object-level** methods

Therefore, we propose a Cascade Transformer-based Multi-Scale Framework

Motivation

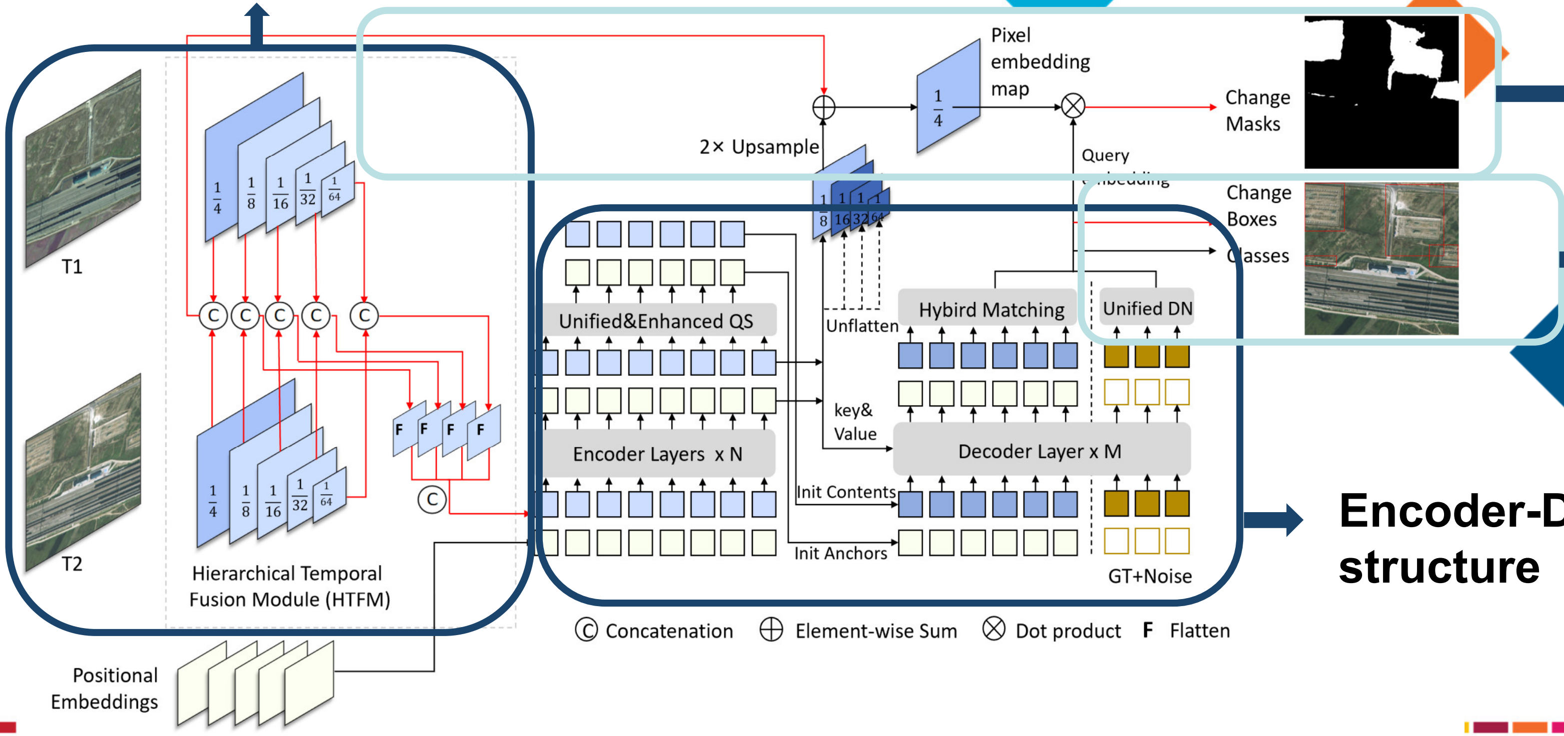
Developing fine-grained object-level change detection with accurate boundary and distinguishing individual instances

THE PROPOSED METHOD

ARCHITECTURE

This framework comprises a Hierarchical Temporal Fusion Module (HTFM), Transformer-based Encoder-Decoder, an object-level change detection head, and a segmentation branch

Hierarchical Temporal Fusion Module (HTFM)



segmentation branch

object-level change detection head

Encoder-Decoder structure

HTFM： Used to extract and fuse multi-scale features，the formula is expressed as

$$F_{bi} = Concatenate(F_{bi}^{1}, F_{bi}^{2}), (i = 1,2,3,4,5)$$
$$F_{hi} = Flatten(F_{bi}), (i = 2,3,4,5)$$
$$F_{e} = Concatenate(F_{hi}), (i = 2,3,4,5)$$

Loss Function

Includes a localization loss and a classification loss for object-level change detection, as well as a Mask loss for segmentation tasks

$$L_{hibird} = \lambda_{cls}L_{cls} + \lambda_{L1}L_{L1} + \lambda_{GIOU}L_{GIOU} + \lambda_{ce}L_{ce} + \lambda_{dice}L_{dice}$$

Encoder-Decoder structure based on Transformer

Used to get predictions for box and mask initialization contents and anchor box queries

Object-level change detection head and Segmentation branch

Obtain box representations of changed regions and fine-grained boundary representations

The proposed Cascade Transformer-based Multi-Scale Framework

01

The FIRST transformer-based object-level D&CD framework

Problems: Transformer-based CD methods are hard to train; existing methods lack precision.

Our Method: Inspired by the succeed models in CV field, effectively achieving transformer-based object-level change detection.

02

the FIRST unified object-level change detection and segmentation framework

Problems: Current methods output bboxes only, which are imprecise.

Our Method: Outputs results with the bbox and the fine boundary masks, and achieves better performance even better than pixel-level methods.

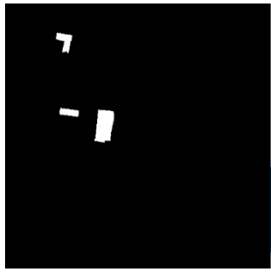
EXPERIMENTS AND RESULTS

Dataset https://github.com/xiaoxiangAQ/LIM-CD-dataset

LIM-CD: a large-scale high-resolution 2D change detection benchmark dataset, consists of 9,259 pairs of pre- and post-temporal high resolution images, divided into a training set with 6,547 pairs, a validation set with 1,776 pairs, and a test set with 936 pairs.

pre-temporal images

post-temporal images



label

Image sizes: 512x512 pixels

ground sampling distance ranging: from 0.5 to 2 meters

We compared our experimental results with the following SOTA methods:

(1) Transformed based pixel-level change detection methods include BIT-CD and ChangeFormer.

(2) Other CNN-based pixel-level change detection methods include FCEF, FC-Siam-diff, FC-Siam-conc, ISNet, SUNET_EP50 and SUNET.

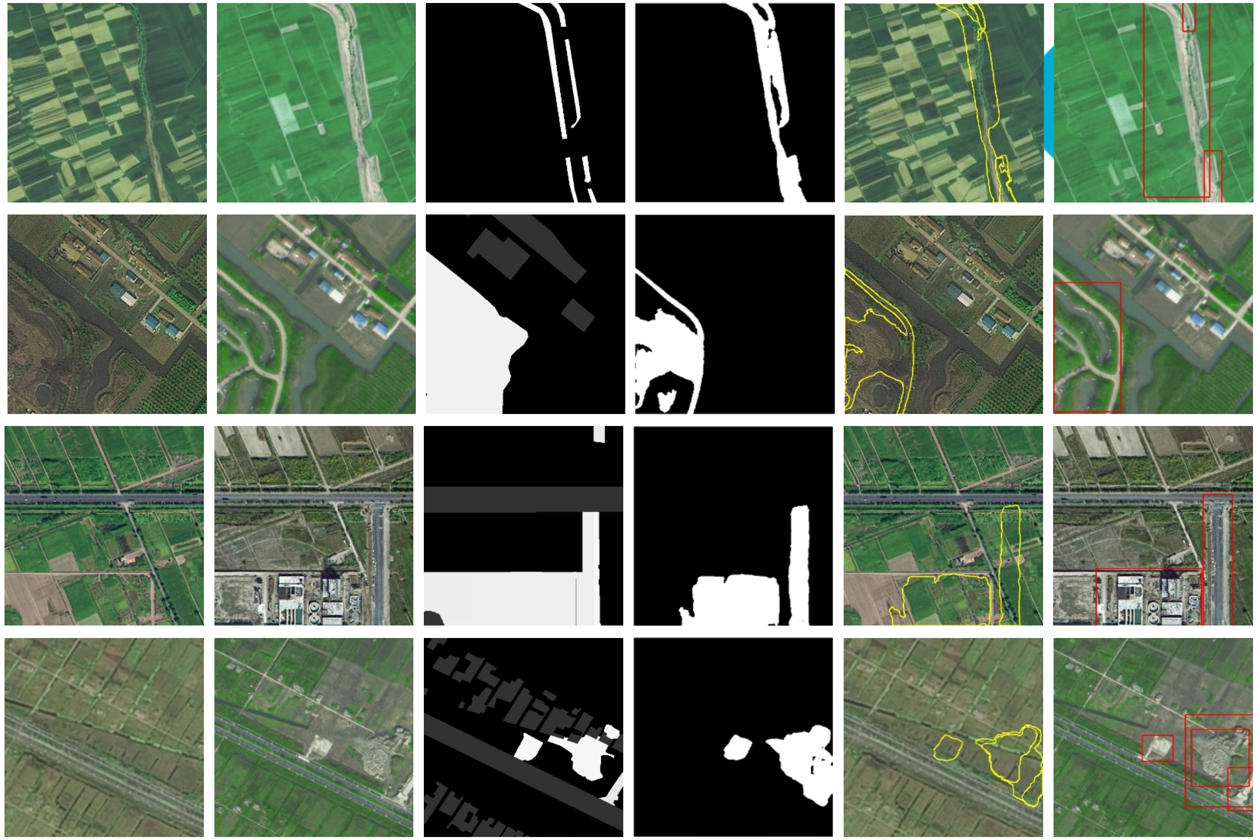
The dual output mode (box and mask) of the proposed framework addresses the challenge of comparing object-level and pixel-level change detection methods.

Experimental results of different methods on the LIM-CD dataset*

Method	Precision	Recall	IOU	F1
CNN-based pixel-level change detection methods				
FCEF	64.87	54.47	42.06	59.22
FC-Siam-diff	66.29	52.41	41.38	58.54
FC-Siam-conc	64.54	46.92	37.30	54.34
ISNet	66.41	54.63	42.80	59.95
SNUNET_EP50	72.01	55.99	45.98	63.00
SNUNET	73.27	57.19	47.31	64.24
Transformer-based pixel-level change detection methods				
BIT-CD	74.34	51.05	43.40	60.53
ChangeFormer	70.84	45.36	38.22	55.31
Our Method	67.30	64.01	48.83	65.62

*All values in the table are expressed as percentages (%)

Partial visualization of the proposed framework



(a) T1 image

(b) T2 image

(c) Ground Truth(GT)

(d) binary results from the change area Mask

(e) change area Mask (present on the former phase)

(f) change area box (present on the later phase)

Received third place in the competition supported by the National Natural Science Foundation of China, based on the enhanced proposed method with several additional strategies:

- **Two-stage Progressive Training:** Solves knowledge transfer in diverse scenes; 3 hours total training.
- **Rich Data Augmentation Techniques:** Significantly improve model generalization in challenging scenarios.
- **Efficient Inference:** Multi-process and multi-batch design boosts efficiency.


荣誉证书

CERTIFICATE OF HONOR

CASM_LIMCD 团队：

在 2023 年“国丰东方慧眼杯”遥感影像智能处理算法大赛对象级变化检测赛道决赛中表现突出，荣获

三 等 奖

特发此证，以资鼓励。

（团队成员：张瑞倩、谢予星、贺尤）

国家自然科学基金委员会信息科学部

“空间信息网络基础理论与关键技术”重大研究计划指导专家组

International Society for Photogrammetry and Remote Sensing

二〇二三年十月

NSFC isprs

ISPRS International Individual Tree Crown (ITC) Segmentation Contest, which attracted **over 40 teams** and around **200 participants** from **13 countries and regions**, including China, the United States, Canada, and France.

The competition ran from **January 29 to June 22, 2024**, and included two stages:
a **ranking stage** and a **final evaluation stage**.

- **Changed the dual-branch temporal input into a single-branch structure, using one-time remote sensing imagery.**
- **Introduced a lightweight Feature Pyramid Network (FPN) to better align multi-scale features across the network.**

INTERNATIONAL SOCIETY FOR PHOTOGRAMMETRY AND REMOTE SENSING
INTERNATIONALE GESELLSCHAFT FÜR PHOTOGRAMMETRIE UND FERNERKUNDUNG
SOCIÉTÉ INTERNATIONALE DE PHOTOGRAMMÉTRIE ET DE TÉLÉDÉTECTION

isprs NSFC

The ISPRS International Individual Tree Crown Segmentation, 2023

The Golden Prize

Presented to

Team casmli

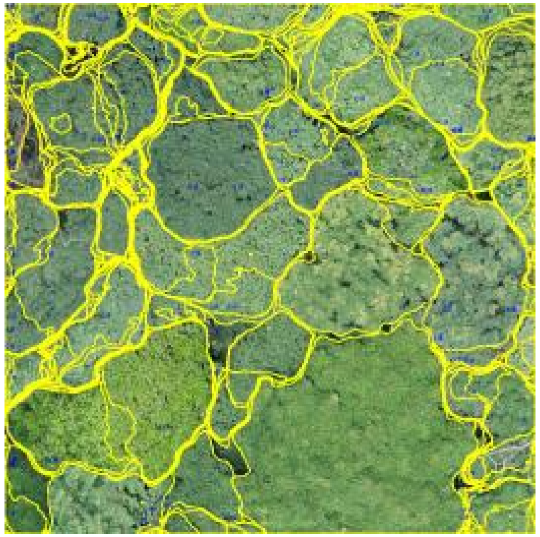
Ruiqian Zhang, Yuxing Xie, Xiaogang Ning, Hanchao Zhang

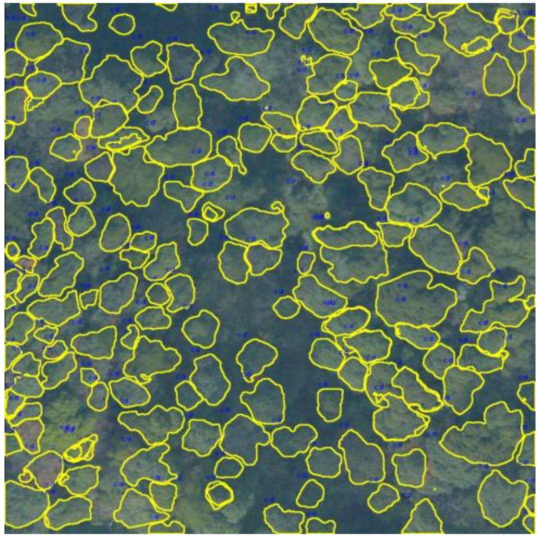
Laurent Polidori, ISPRS TC III, President
Jianya Gong, Advisor Committee, President
Mi Wang, ISPRS WG I/3, Co-Chair
Xinlian Liang, ISPRS WG III/1, Chair

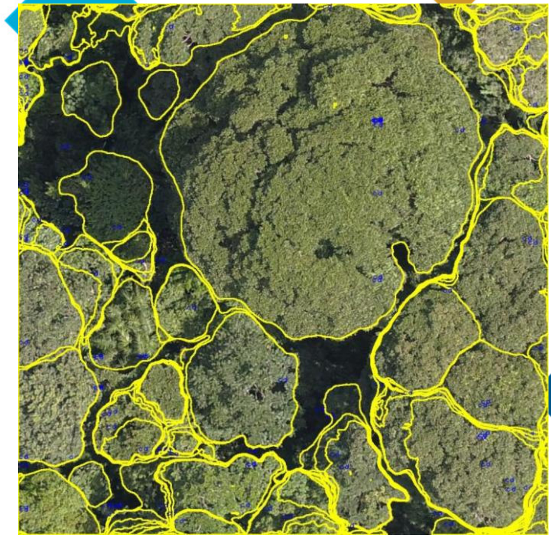
November 2024, Belém, Brazil

Received Gloden Prize (1st) in the ISPRS International Contest on Individual Tree Crown (ITC) Segmentation, based on the proposed framework with several additional strategies.

Experiment Results



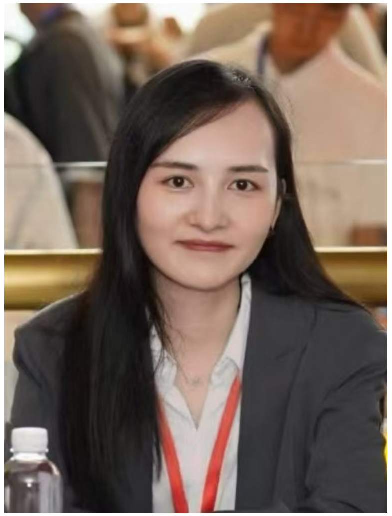




CONCLUSIONS

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

- We proposed a cascade Transformer-based multi-scale framework for object detection and instance segmentation in remote sensing imagery.
- The method integrates object-level detection and mask-level segmentation in a unified structure, and handles complex scenes with varying object scales.
- Originally designed for change detection, the framework was successfully adapted to single-image tasks, and achieved first place in the ISPRS ITC Segmentation Contest.



- **Name:** Ruiqian Zhang
- **Institution:** Chinese Academy of Surveying and Mapping
- **Academic Title:** Associate Research Professor
- **Degree:** PhD in Engineering
- **Research Interests:** Image processing, computer vision, remote sensing, deep learning
- **E-mail:** zhangrq@casm.ac.cn; zhangruiqian@whu.edu.cn

Homepage