

An End-to-end Large-area Tree Detection Approach on High Resolution Imagery Using Detection Transformer

Tao Jiang

Georg-August-Universität Göttingen
Forest Inventory and remote sensing department
June 17, 2024 Umeå



Object detection models in Deep Learning

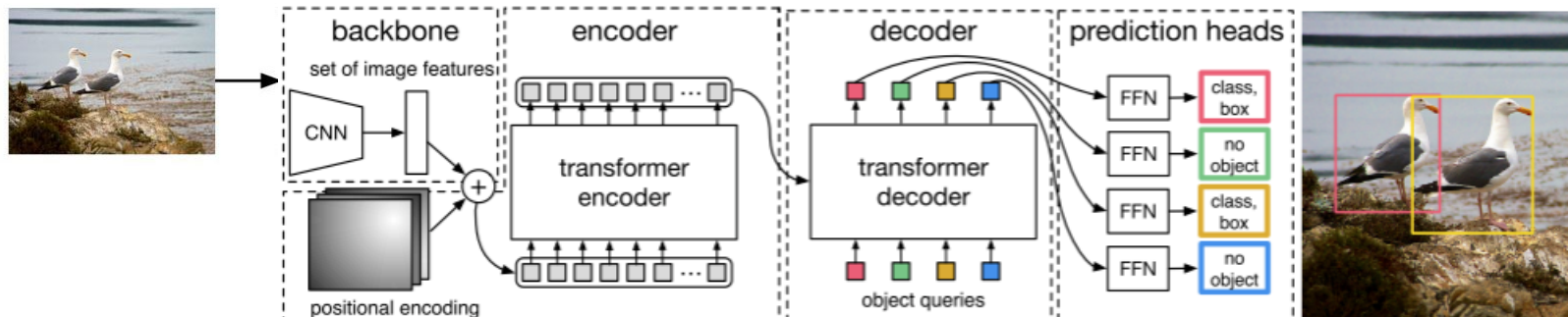
Convolutional Neural Networks

- RCNN, Fast RCNN, Faster RCNN, Mask RCNN
- YOLO and RetinaNet

Transformer

- Detection Transformer (DETR) (Carion 2020)
- Deformable DETR (Zhu 2020)
- DETR with Improved deNoising anchor boxes (DINO) (Zhang 2022)

Overview of Detection Transformer architecture

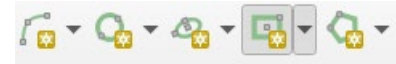


Carion et al. 2020

Tree crown detection dataset and model training

Dataset labelling

- 330 WorldView-3 image patches (320×320)
- 23,643 tree crown bounding boxes



Tree crown detection dataset and model training

Dataset compiling

- COCO format
- Dataset splitting (300, 15, 15)

```
"images": [{"id": 1, "width": 320, "height": 320,
            "file_name": "tile_WV3_2016-
            11_north_106_0.tif"}],
"annotations": [{
"segmentation":
[[5.0, 158.5, 1.0, 157.5, 0, 155.0, 1.0, 148.5, 7.5,
150.0, 8.5, 156.0, 5.0, 158.5]],
"area": 65,
"iscrowd": 0,
"image_id": 1,
"bbox": [0.0, 149.0, 9.0, 10.0],
"category_id": 1,
"id": 1}]
```

Tree crown detection dataset and model training

Model training

- OpenMMLab
- Hyperparameters



The screenshot shows the GitHub repository page for MMDetection. At the top, there are links for README, Code of conduct, and Apache-2.0 license. The main heading is "MMDetection" with a stylized logo. Below the heading, there are links for "OpenMMLab website" (HOT) and "OpenMMLab platform" (TRY IT OUT). A progress bar shows various metrics: pypi v3.3.0, docs latest, build failing, codecov 64%, license Apache-2.0, open issues 12%, and issue resolution 4 d. There is also a button for "OpenXLab Demo". At the bottom, there are links for Documentation, Installation, Model Zoo, Update News, Ongoing Projects, and Reporting Issues. The language is set to English, with a link for 简体中文. Social media icons for GitHub, Discord, Twitter, YouTube, and WeChat are also present.

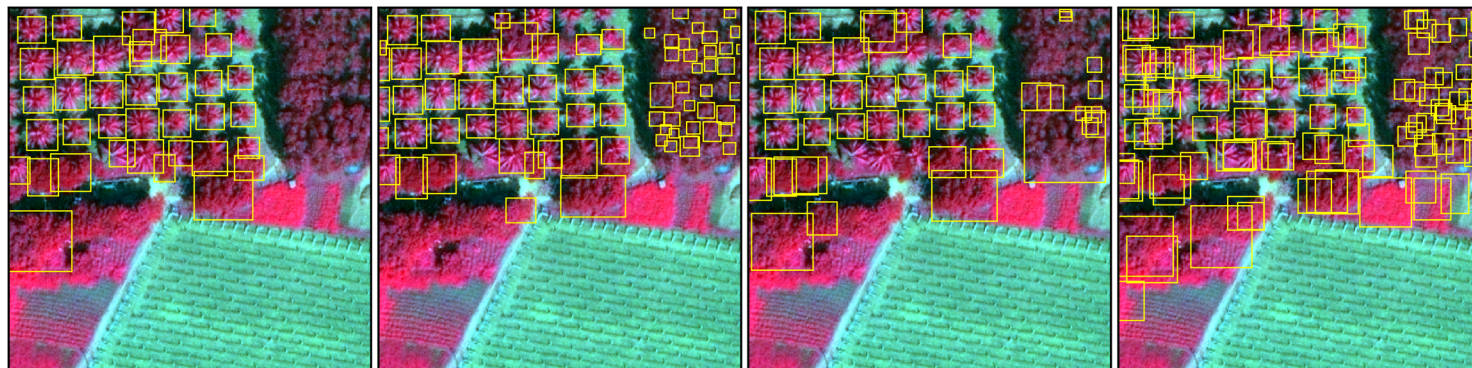
A better performance for DINO-SWIN compared to other DL object detection methods

Table 3: Summary of detection models' performance computed based on the test dataset.

Detection models	DT	GT	TP	FP	DR(%)	Precision(%)	Recall(%)	F1(%)	AP(%)
Faster RCNN	318	916	243	75	35	76	27	39	28
YOLO	649	916	462	187	71	71	50	59	56
RetinaNet	562	916	191	371	61	34	21	26	19
DETR	1284	916	182	1102	140	14	20	17	5
Deformable-DETR	594	916	381	213	65	64	42	50	41
DINO-Res50	810	916	563	247	88	70	61	65	60
DINO-SWIN	871	916	663	208	95	76	72	74	76

Note: DT (detections), GT (ground truths), TP (true positives), FP (false positives), DR (detection rate) and AP (average precision).

A better performance for DINO-SWIN compared to other DL object detection methods

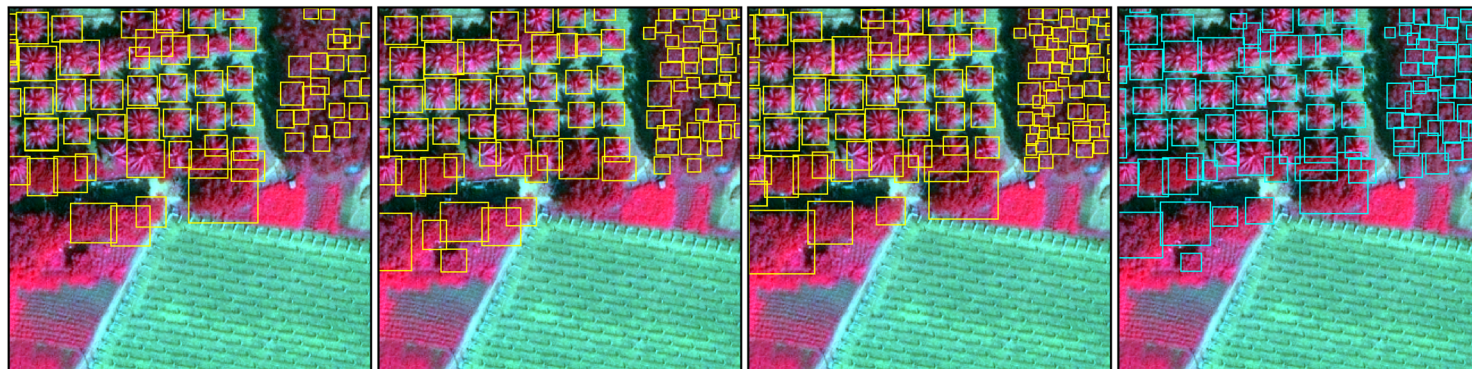


Faster_RCNN

YOLO

Retinanet

DETR



Deformable_DETR

DINO_Res50

DINO_SWIN

GT

A better performance for DINO-SWIN compared to other DL object detection methods

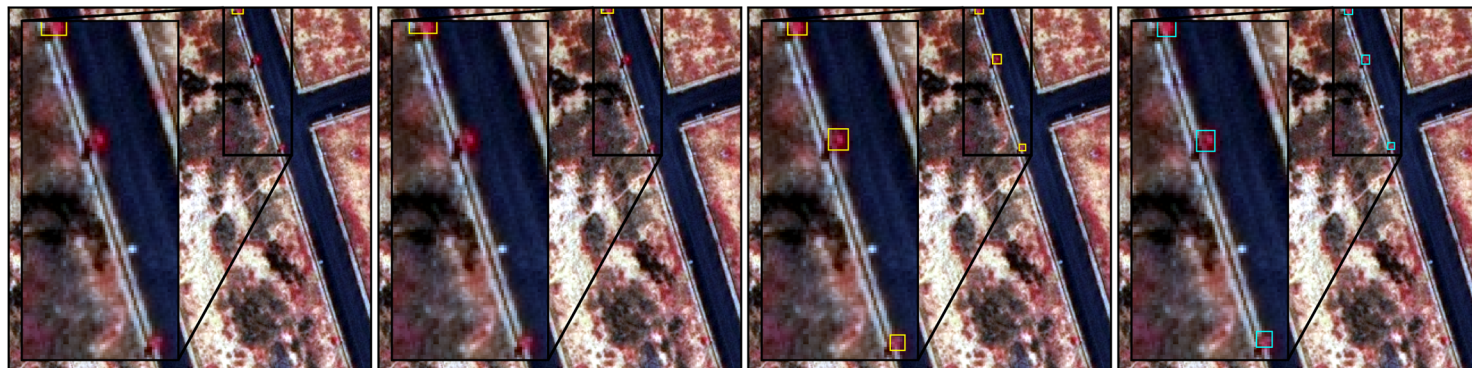


Faster_RCNN

YOLO

Retinanet

DETR



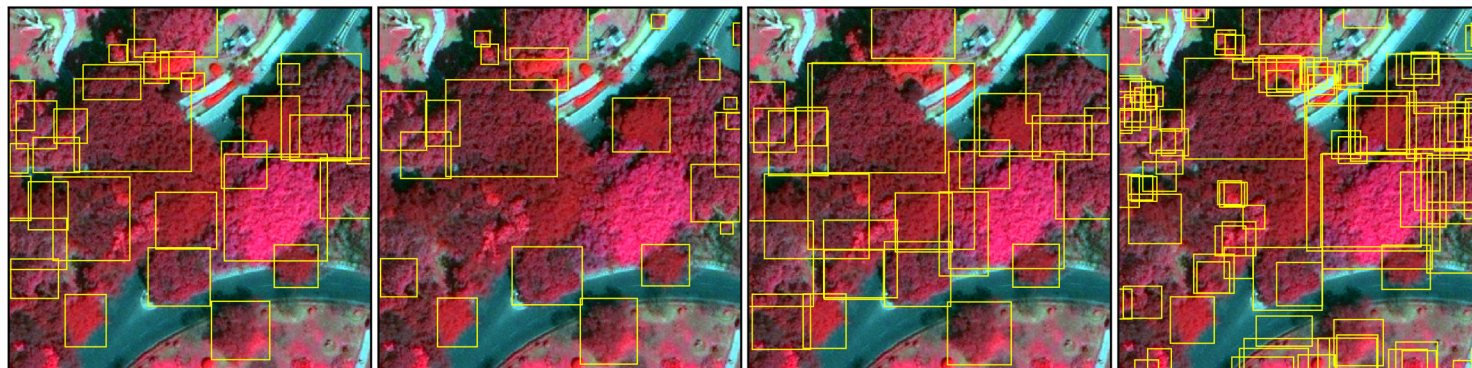
Deformable_DETR

DINO_Res50

DINO_SWIN

GT

A better performance for DINO-SWIN compared to other DL object detection methods

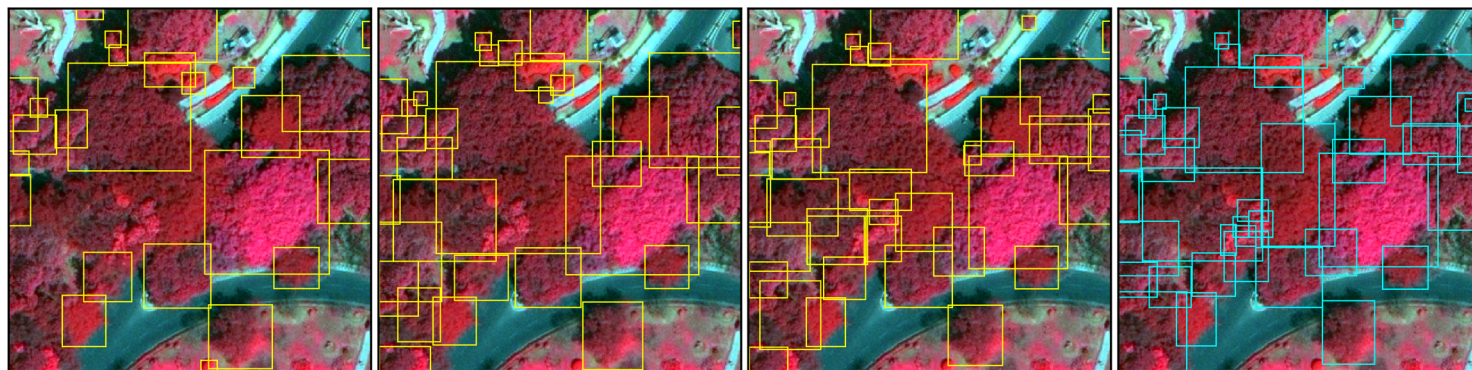


Faster_RCNN

YOLO

Retinanet

DETR



Deformable_DETR

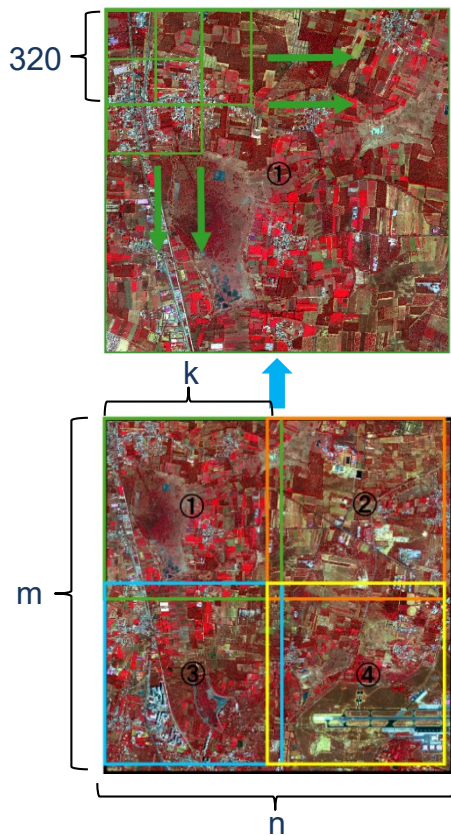
DINO_Res50

DINO_SWIN

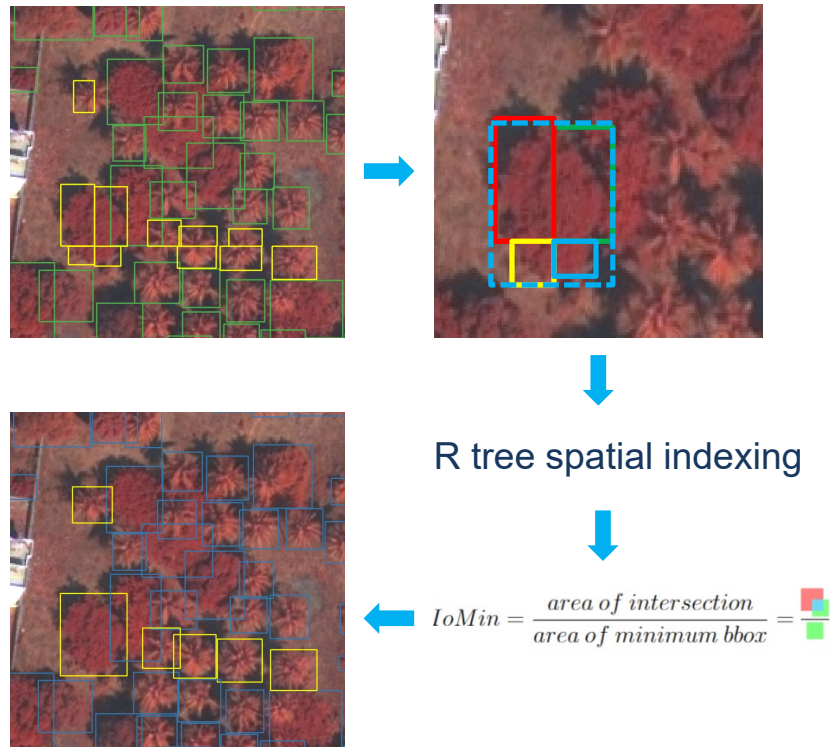
GT

Large-area tree detection approach consists of two steps

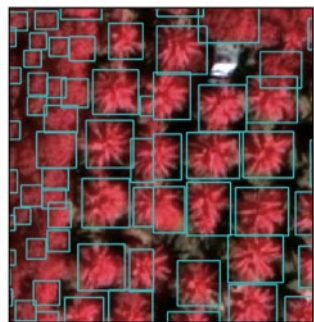
Step 1: two-level partitioning scheme



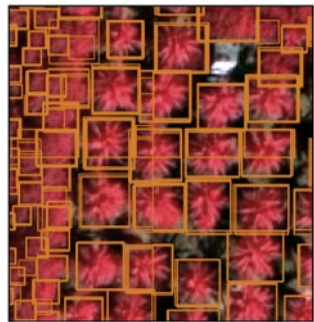
Step 2: R-tree based Box Merging



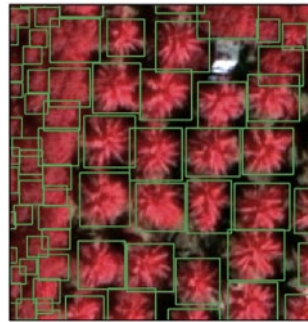
More examples of detection results using proposed approach



No overlapping



Overlapping without Merging



Overlapping and Merging



Bengaluru (250 km²)

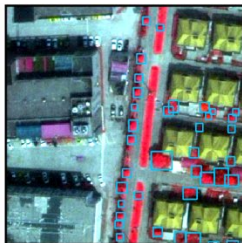
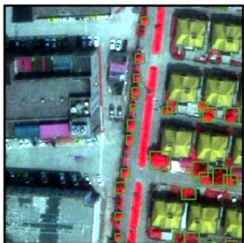
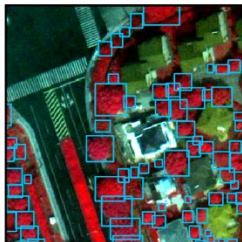
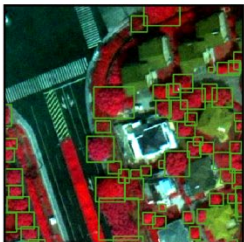
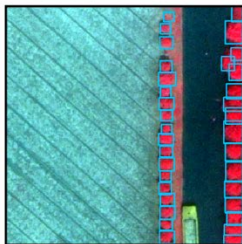
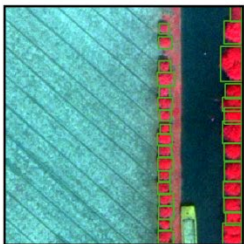
~1.3 million trees

Precision: 76%

Recall: 72%

F1 score: 74%

Proposed approach tested on two additional locations



0 1 2 km

Shanghai (26 km²)

~90k trees

Precision: 76%

Recall: 70%

F1 score: 73%

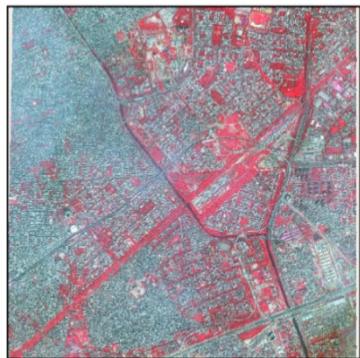
 Detections

 Ground truth

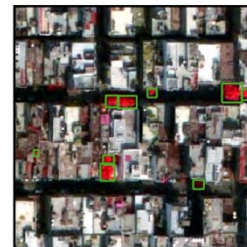
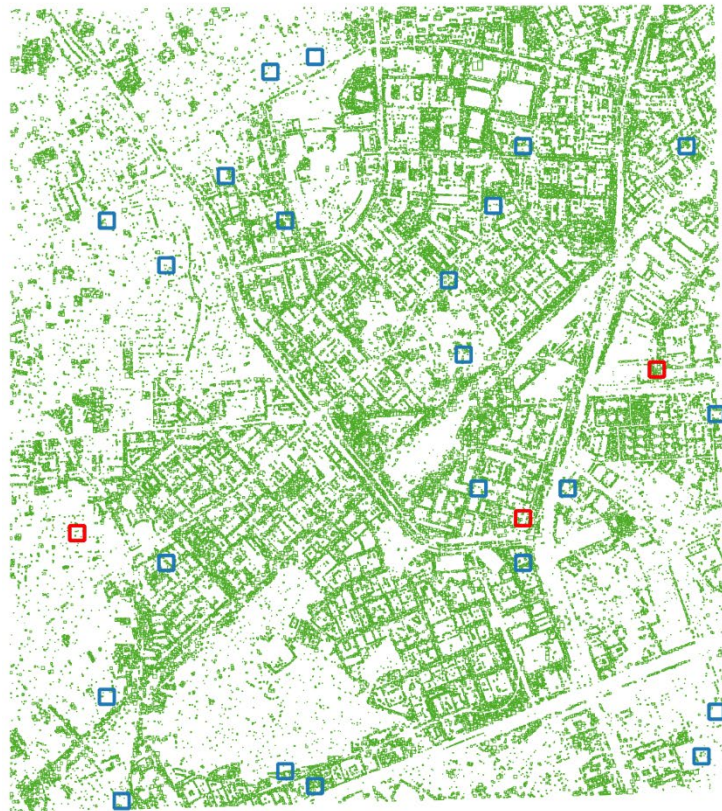
 Test plots

 Test plots displayed

Proposed approach tested on two additional locations



Delhi (26 km²)
~70k trees
Precision: 86%
Recall: 87%
F1 score: 87%



Test plots



Test plots displayed



Detections



Ground truth

An End-to-end Large-area Tree Detection Approach on High Resolution Imagery Using Detection Transformer

- a) We explored the potential of detection transformer (DETR) models on identifying individual tree from high resolution satellite imagery.
- b) We devised an end-to-end tree detection approach utilizing DINO as classifier, incorporating two-level partitioning scheme, and employing a R-tree based Box Merging technique.
- c) The proposed tree detection approach has been further tested on two additional locations with different image quality, showing its strong applicability.