Salient Region-of-Interest Detection in Multispectral Agricultural Imagery

Çağkan Gürsoy, Master Computer Vision & Data Science
Supervisor: Willem Dijkstra

Summer 2025

Introduction

- Precision agriculture enhances **yield** and **sustainability** through **field-level anomaly detection**.
- Traditional inspections are **labor-intensive** and **subjective**; drone-based AI offers a **scalable**, **objective** alternative.
- This study evaluates a **DETR-based framework** for **fine grained** anomaly detection.
- Assesses **NIR** vs. **RGB** value, addresses **fusion** challenges, and analyzes **resolution requirements** for reliable performance.

Abstract

Modern agriculture increasingly relies on remote sensing to detect crop stress and anomalies. We present a DETR[1]-based framework for detecting fine-grained anomalies in high-resolution RGB and NIR drone imagery, using spatially preserving tiling, RGB-NIR alignment, and both early and mid-level fusion. RGB input achieved the best precision (84.05%), with the green channel performing nearly as well, suggesting potential for standalone use. While mid-level fusion of NIR yielded competitive results, its added complexity offered limited benefit. Detection performance was highly sensitive to input resolution, emphasizing the dominance of visible-spectrum spatial cues for this task.

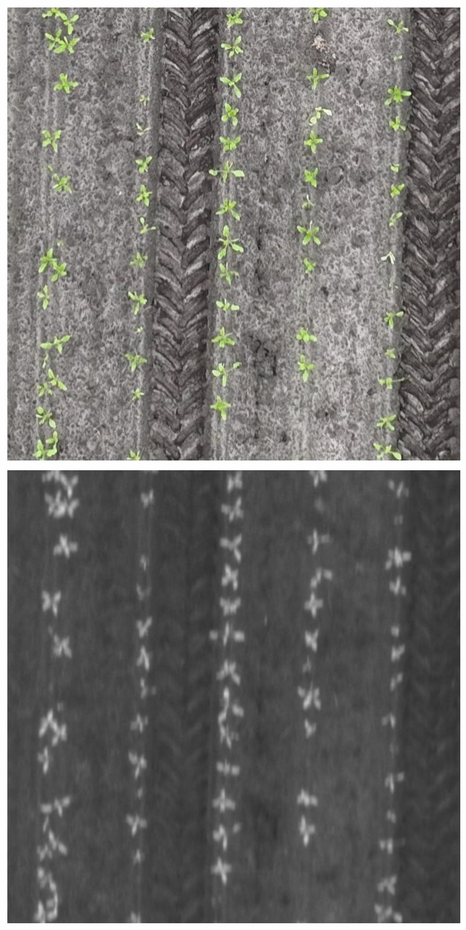


Figure 1. Example of spatially aligned and RGB (top) NIR (bottom) drone image tiles from an early crop growth stage.

Experiments and Results

- **RGB-only**, **early fusion** (RGB+NIR concatenation), **mid-level fusion** (parallel RGB/NIR backbones: addition, concatenation, attention)
- Single-channel ablation: **R**, **G**, **B**, **NIR-only**
- Resolution degradation: systematic **downsampling** to assess detection **performance** and spatial sensitivity

Table 1. Performance comparison of different input modalities evaluated in terms of precision, recall, F1 score, and mAP at IoU = 0.50.

Input Modality	Precision	Recall	F1 Score	mAP
RGB	0.8314	**0.8145**	0.8227	0.5067
Mid Fusion	**0.8405**	0.8087	**0.8241**	**0.5168**
Early Fusion	0.7959	0.7765	0.7857	0.4674
Green Channel	0.8255	0.8093	0.8169	0.5005

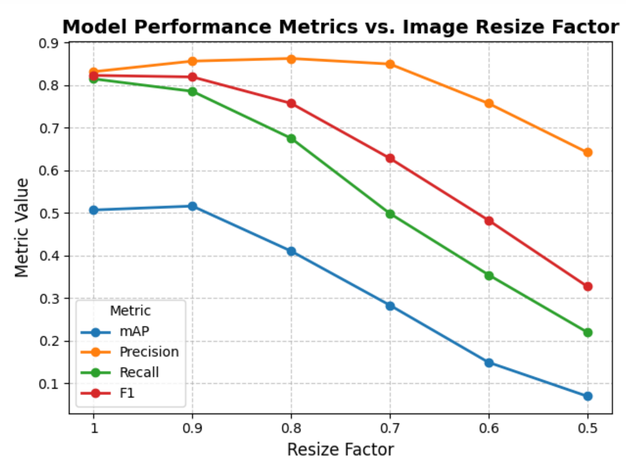


Figure 3. Performance metrics plotted against image resize factor.

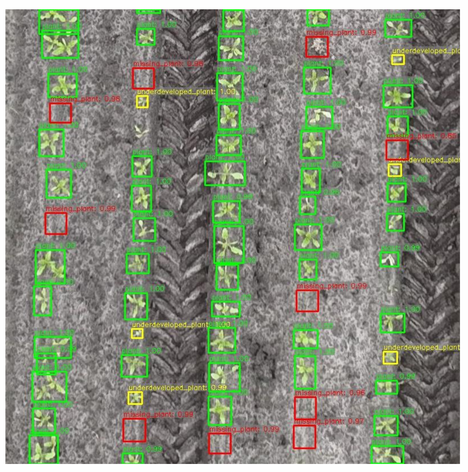


Figure 4. Inference result on a single tile. (Green: healthy, yellow: underdeveloped, red: missing plant)

- **Mid-level fusion** achieved **the highest detection performance**, slightly surpassing RGB.
- **Early fusion** resulted in **lower performance** in all metrics.
- **Green** channel showed performance **close** to **RGB**, with minor score reductions.
- **Detection accuracy** dropped sharply below **80%** resolution.

Materials and Methods

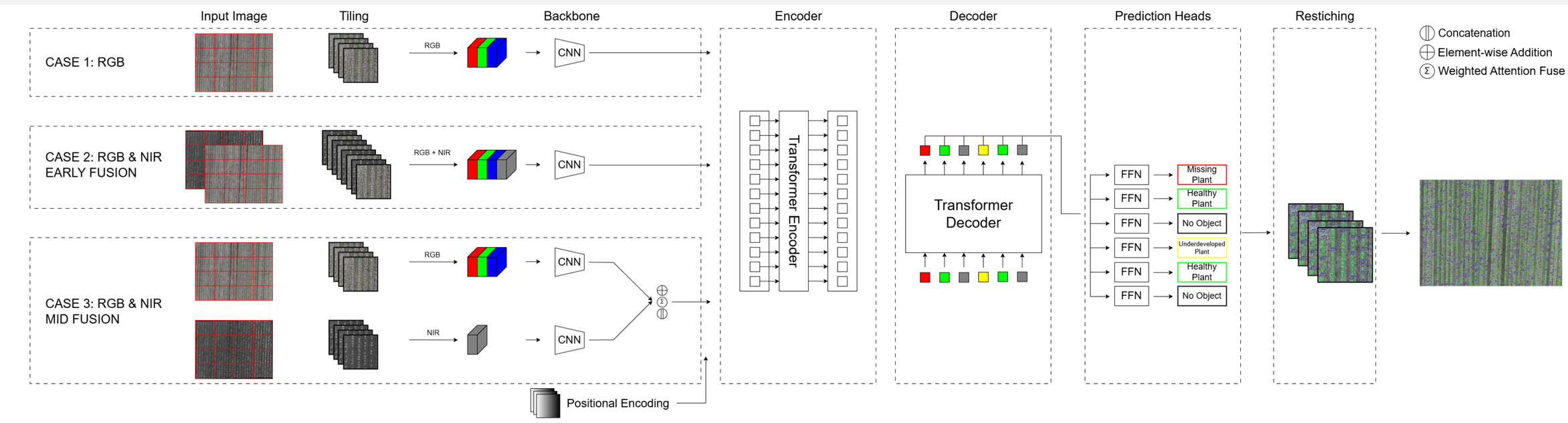


Figure 2. Overview of the proposed framework for fine-grained plant anomaly detection. Spatially tiled RGB and NIR images are processed through three configurations: RGB-only baseline, early fusion (channel concatenation of RGB and NIR), and mid-level fusion (parallel backbones with feature integration). Transformer-based detection identifies underdeveloped and missing plants, with outputs aggregated to reconstruct full field-level predictions.

- A **DETR-based** object detection framework with spatially aligned **RGB** and **NIR images** by homography transformations and feature-based refinement.
- **Image tiling** to preserve spatial detail; multispectral fusion via early fusion (channel concatenation) and mid-level fusion (parallel RGB/NIR backbones).
- Performance evaluated using **precision**, **recall, F1 score**, and **mAP** across multiple **IoU** thresholds.

Conclusions

- **Transformer-based** architectures enable reliable detection of **fine grained plant anomalies.**
- **RGB** spatial information, especially the **green channel**, provides provides the **strongest cues** for detecting plant patterns
- Existing **multispectral fusion** methods **underutilize** NIR contributions.
- Preserving **high spatial resolution** is essential for effective anomaly detection.

Acknowledgements

This project is financially supported by the European Union within the JTF-call project DigiAgro3.

NHL STENDEN
computer vision & data science

DIGI-AGRO ABDRONE Ministerie van Economische Zaken en Klimaat

Medegefinancierd door de Europese Unie

SNN SAMENWERKINGSVERBAND NOORD-NEDERLAND
• Stimuleert • Faciliteert • Verbindt

References

[1] Nicolas Carion, Francisco Massa, Gabriel Synnaeve, Nicolas Usunier, Alexander Kirillov, Sergey Zagoruyko. End-to-end Object Detection with Transformers. Computer Vision – ECCV, 2020.
[2] Hao Han, Zhiwei Liu, Jianfeng Li, Zhiwei Zeng. Challenges in Remote Sensing-based Climate and Crop Monitoring: Navigating the Complexities Using AI. Journal of Cloud Computing, 2024.